

ZURECHT: NEUE RECHERCHEMÖGLICHKEITEN IN KORPORA GESPROCHENER SPRACHE FÜR GESPRÄCHSANALYSE UND DEUTSCH ALS FREMD- UND ZWEITSPRACHE

Elena Frick, Leibniz-Institut für Deutsche Sprache Mannheim
Henrike Helmer, Leibniz-Institut für Deutsche Sprache Mannheim
Franziska Wallner, Herder-Institut der Universität Leipzig

Abstract

ZuRecht steht für *Zugang zur Recherche in Transkripten*. Es handelt sich um eine prototypische Implementierung einer webbasierten grafischen Benutzeroberfläche, welche Zugriff auf Transkripte gesprochener Sprache aus dem Archiv für Gesprochenes Deutsch (AGD) des Leibniz-Instituts für Deutsche Sprache (IDS) bietet. Der Zugriff erfolgt über die neue, im Projekt „*ZuMult*“ entwickelte Schnittstelle zur Suche in mündlichen Korpora. *ZuRecht* dient einerseits der Demonstration der Möglichkeiten der neuen Schnittstelle, indem es komplexe Suchanfragen mit der speziell für die Korpusrecherche entwickelten Anfragesprache CQP auf Transkriptionen gesprochener Sprache erlaubt. Andererseits kommt *ZuRecht* als Erweiterung der Datenbank für Gesprochenes Deutsch (DGD) zum Einsatz und eröffnet den DGD-Nutzer:innen viele neue Forschungsmöglichkeiten, insbesondere auf den Gebieten der Gesprächsanalyse und der DaF/DaZ-bezogenen Forschung. Im Beitrag werden die Funktionalitäten von *ZuRecht* ausführlich vorgestellt und ihre Einsatzmöglichkeiten in den genannten Disziplinen exemplarisch vorgeführt.

Keywords: gesprochene Sprache; mündliche Korpora; Korpusrecherche; Suchanfragesprache; korpusbasierte Gesprächsforschung; Korpora in DaF/DaZ

Abstract

ZuRecht is a German abbreviation for *Zugang zur Recherche in Transkripten* (Engl. „search access to transcripts“). It is a prototype implementation of a web-based application providing query access to spoken language transcripts from the Archive for Spoken German (AGD) at the Leibniz-Institute for German Language (IDS). The query access happens via the new search query interface developed in the „*ZuMult*“ project. *ZuRecht* is, on the one hand, implemented for demonstration purposes — specifically to illustrate complex search queries on transcriptions of spoken language by using CQP – the query language specially developed for corpus research. On the other hand, *ZuRecht* is currently used to extend the Database for Spoken German (DGD) by means of several new search possibilities, which can be beneficial for conversation analysis and research related to German as a Foreign and Second Language (GFL/GSL). In this paper, we describe the new search functionalities and give examples of how *ZuRecht* can be applied in the research disciplines mentioned above.

Keywords: spoken language; oral corpora; corpus research; search query language; corpus-based conversation research; corpora in GFL/GSL

1. Einführung

Die Erforschung gesprochener Sprache hat in der linguistischen Forschungslandschaft seit vielen Jahren eine derart hohe Relevanz, dass Forscher(gruppen) nicht nur eigene Korpora erheben und aufbereiten, sondern sich zentralisierte Angebote etabliert haben, die auf spezifische Bedarfe der Forschungscommunity abzielen und eben jene große Schritte weitergebracht haben. So wird etwa am Leibniz-Institut für Deutsche Sprache seit 2008 das Forschungs- und Lehrkorpus Gesprochenes Deutsch (*FOLK*) aufgebaut, eine Sammlung von aufgenommenen Interaktionen, die, orientiert an

Bedarfen von Gesprächsforscher:innen und Interaktionalen Linguist:innen, methodisch fundiert erhoben, aufbereitet und nutzbar gemacht werden (vgl. Reineke / Deppermann / Schmidt 2023; Schmidt 2023 in dieser Themenausgabe). Das Vergleichskorpus zur gesprochenen Wissenschaftssprache (*GeWiss*; vgl. Schmidt et al. 2023 in dieser Themenausgabe sowie Wallner 2023 in dieser Themenausgabe und Fandrych / Wallner 2023) ist ein weiteres umfangreiches Angebot, welches eine systematische empirische Erforschung der mündlichen Hochschulkommunikation ermöglicht und zudem auch als Grundlage für die Vermittlung der gesprochenen Wissenschaftssprache genutzt werden kann. Mit der Datenbank für Gesprochenes Deutsch¹ (DGD, vgl. Schmidt 2017) hat sich eine breit genutzte Korpusplattform etabliert, die für verschiedene Forschungsbereiche ein zentral zugängliches Angebot darstellt und Korpora des Archivs für Gesprochenes Deutsch² (AGD, vgl. Stift / Schmidt 2014), wie *FOLK* und *GeWiss*, weitere Gesprächskorpora oder vielfältige Variationskorpora, zur Durchsichtung und Nutzung anbietet. Die DGD richtet sich primär u.a. an Forschende aus den Bereichen der Gesprächsforschung, Interaktionalen Linguistik, Variationslinguistik und Korpuslinguistik und bietet diesen Möglichkeiten, unterschiedlichste Forschungsfragen systematisch zu untersuchen. Daneben ist sie besonders in den letzten Jahren zunehmend für Forscher:innen sowie für Lehrende und Lernende aus dem Kontext Deutsch als Fremd- und Zweitsprache attraktiv geworden (vgl. Fandrych et al. 2016), die einerseits auf der Grundlage von Korpora gesprochener Sprache didaktisch relevante Vermittlungsgegenstände identifizieren und andererseits Korpusdaten für die Sprachvermittlung bzw. -aneignung nutzen (vgl. Fandrych / Meißner / Wallner 2021).

Gerade aufgrund dieser vielfältigen Nutzungsinteressen und Nutzungsmöglichkeiten zeigen sich jedoch auch Einschränkungen von existierenden Korpusplattformen wie der DGD. So gibt es selbstverständlich immer Verbesserungsbedarf bezüglich der Effizienz und Benutzerfreundlichkeit der Suchmaschine, aber auch ein Ausbau nutzergruppenspezifischer Suchfunktionalitäten wird zurzeit, wie eine Nutzerstudie (vgl. Fandrych et al. 2016) gezeigt hat, dringend benötigt. An diesem Desideratum orientiert sich die neue Korpusrechercheplattform *ZuRecht*³. Diese bietet einen im Vergleich zur DGD erweiterten Zugriff auf Korpora gesprochener Sprache und richtet sich insbesondere an Gesprächsforscher:innen und Interaktionale Linguist:innen sowie an Forscher:innen und Praktiker:innen aus dem Kontext Deutsch als Fremd- und Zweitsprache. Sie bedient demnach mehrere Nutzer:inneninteressen und -bedürfnisse gleichzeitig und geht somit über das Angebotspektrum anderer Korpusoberflächen und Plattformen deutlich hinaus (vgl. Batinić / Frick / Schmidt 2021).

In unserem Beitrag beschreiben wir zunächst die grundlegenden Suchfunktionen der DGD und deren technische Grenzen (Abschnitt 2). Hiernach stellen wir die Korpusplattform *ZuRecht* vor und geben einen ersten Überblick über deren technische Besonderheiten (Abschnitt 3). In den folgenden beiden Abschnitten werden sie exemplarisch anhand von verschiedenen Fragestellungen und Beispielen vertiefend erklärt und illustriert. Wir demonstrieren dabei, wie *ZuRecht* für Forscher:innen im Bereich der Gesprächsanalyse/Interaktionalen Linguistik (Abschnitt 4) sowie im Kontext von Deutsch als Fremd- und Zweitsprache genutzt werden kann (Abschnitt 5) und wie die nutzerspezifischen Bedarfe dabei adressiert sind. Abschnitt 6 fasst den Beitrag zusammen und gibt einen kurzen Ausblick.

2. DGD

Die Datenbank für Gesprochenes Deutsch (DGD) ist eine am Leibniz-Institut für Deutsche Sprache entwickelte Online-Plattform zur computergestützten Analyse von Korpora gesprochener Sprache.

¹ <https://dgd-ids.mannheim.de> (07.10.2022).

² <https://agd.ids-mannheim.de> (07.10.2022).

³ <http://zumult.ids-mannheim.de/ProtoZumult/jsp/zuRecht.jsp> (07.10.2022).

EREIGNISSE		SPRECHER	TRANSKRIPTE	
ausgewählte Transkripte: aligniert: 796 Suchausdruck: zurecht%_				
zurecht%_		max. 10	Treffer anzeigen	Suchen
Der Suchausdruck wurde gefunden in 20 Dokument(en). Gesamtscore: 22				
#	Transkript-ID	KWIC-Liste	Score	Hörprobe
1	FOLK_E_00193_SE_01_T_02 ▶	zurechtrückt des isch...kro ne zurechtrücke weiderlaufe	2	▶
2	FOLK_E_00271_SE_01_T_02 ▶	dann selber zurechtschneiden wenn se...zurechtschneiden sondern	2	▶
3	FOLK_E_00432_SE_01_T_01 ▶	sich immer zurechtmachen für online	1	▶
4	FOLK_E_00021_SE_01_T_09 ▶	be kommen zurechtgelegt für n	1	▶
5	FOLK_E_00026_SE_01_T_01 ▶	äh zurechtzubiege also wir	1	▶
6	FOLK_E_00036_SE_01_T_01 ▶	zurechtkommen	1	▶
7	FOLK_E_00038_SE_01_T_01 ▶	gesellschaft auch zurechtkomme	1	▶
8	FOLK_E_00042_SE_01_T_01 ▶	äh zurechtgelegt haben aber	1	▶
9	FOLK_E_00329_SE_01_T_05 ▶	dich zurechtfin den	1	▶
10	FOLK_E_00203_SE_01_T_01 ▶	mir des zurechtzulegen zum interpretiern	1	▶

Abbildung 1
Volltextrecherche der DGD mit Oracle Text CONTAINS Query

Sie bietet Zugriff auf Korpora aus dem Archiv für Gesprochenes Deutsch (AGD) und stellt zahlreiche Funktionen zum Browsen und Durchsuchen großer Datensammlungen von Audio- und Videoaufnahmen, den dazugehörigen Transkriptionen und Annotationen zur Verfügung.

Die Suchkomponenten in der DGD sind mit Hilfe einer objekt-relationalen Oracle-Datenbank umgesetzt, die native XML-Daten speichern und verarbeiten kann. Den DGD-Nutzer:innen stehen zwei Suchmöglichkeiten zur Verfügung. Zum einen kann eine dokumentbasierte Volltextrecherche in Transkriptionstexten (ohne Annotationen) mit Hilfe der datenbankeigenen Abfragesprache Oracle Text CONTAINS Query⁴ vorgenommen werden. Die Darstellung der Suchergebnisse erfolgt in Form einer Liste von Dokumenten, die den gesuchten Ausdruck enthalten. Abbildung 1 zeigt ein Beispiel für eine mögliche Suchanfrage dieser Art. Sie findet Dokumente, die mindestens eine transkribierte Form enthalten, die mit „zurecht“ anfängt ("_" steht für ein beliebiges Zeichen, "%" für eine optionale Folge beliebiger Zeichen).

Zum anderen kann in der DGD eine Wortindex-basierte Suche auf Transkriptionen und Wort-Token-Annotationen ausgeführt werden, wobei die Anfrage durch eine Spezifikation der transkribierten und normalisierten Form eines Tokens sowie seines Lemmas und seiner Wortart (POS = „Part of Speech“) formuliert werden kann. Abbildung 2 zeigt die vier Sucheingabefelder (Transkribiert, Normalisiert, Lemma und POS), die einzeln oder in Kombination verwendet werden können. Das Suchergebnis wird als KWIC (Keyword-in-Context)-Konkordanz angezeigt und kann durch Metadaten und einzelne Token im benachbarten Kontext schrittweise gefiltert werden.⁵

⁴ https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.help_a (07.10.2022)

⁵ Konkrete Beispiele für Suchanfragen und weitere Informationen zu den Suchmöglichkeiten finden sich auf der DGD-Webseite https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.help_b (07.10.2022).

The screenshot shows the DGD search interface. At the top, there are tabs for POSITION, TOKEN (selected), KONTEXT, METADATEN, and ANZEIGE. Below the tabs, there are input fields for 'Transkribiert:', 'Lemma:', and 'Normalisiert: zurecht'. There is also a 'POS:' field and a checkbox for 'Reguläre Ausdrücke'. A 'Suche starten' button and a 'CQP' button are visible. Below the search bar, there is a 'Recherche' section with a 'Tokens' tab. A status bar indicates 'KWIC wird angezeigt'. The main area shows search results for 'zurecht', with 15 results displayed. Each result includes a checkbox, a speaker ID (e.g., FOLK_00013_01), a speaker name (e.g., CJ), and a snippet of text with the word 'zurecht' highlighted in red. The results are paginated, showing 'Seite 1 von 2'.

	Sprechereignis	Sprecher	Treffer
<input checked="" type="checkbox"/>	FOLK_00013_01	CJ	die libelle rückt ihre brille zurecht mh sieht lecker aus
<input checked="" type="checkbox"/>	FOLK_00016_01	CJ	die libelle rückt ihr brille zurecht mh sieht lecker aus
<input checked="" type="checkbox"/>	FOLK_00016_01	CJ	die libelle rückt ihre brille zurecht
<input checked="" type="checkbox"/>	FOLK_00021_01	MT	das fragst du zurecht
<input checked="" type="checkbox"/>	FOLK_00055_01	US	ehrllich gsagt auch wenn ma wenn ma zurecht irgendwas m es gibt ja manchma au wirklich sachen sie
<input checked="" type="checkbox"/>	FOLK_00069_01	HG	der herr inghofen ist auch nicht damit zurecht gekommen
<input checked="" type="checkbox"/>	FOLK_00069_01	BH	äh wächst ja wohl der optimismus ob zurecht hm is sehr dahingestellt des projekt kippen zu können
<input checked="" type="checkbox"/>	FOLK_00118_01	ME	halt nim mehr mit sie z sich zurecht wusste net was er wie s weitergehen soll war damals
<input checked="" type="checkbox"/>	FOLK_00147_01	EH	kommen die mit den tiern zurecht und äh wie sind die so drauf
<input checked="" type="checkbox"/>	FOLK_00147_01	RK	wie kommen die mit den tiern so zurecht ich mein die machen krach stinkt
<input checked="" type="checkbox"/>	FOLK_00152_01	PB_ga	ja ich glaub wir kommen zurecht
<input checked="" type="checkbox"/>	FOLK_00174_01	TN	mit mit dem wesemann komm ich nich zurecht oder wenn herr wesemann sacht nee also des passt nich
<input checked="" type="checkbox"/>	FOLK_00201_01	GB	gehabt aber der kam da gut mit zurecht weil der f wohl für opern war und lorenzo der
<input checked="" type="checkbox"/>	FOLK_00202_01	ME	glaub der is nich mehr so gut zurecht ne
<input checked="" type="checkbox"/>	FOLK_00210_01	PS	was sind die perspektiven zurecht

Abbildung 2
Wortindex-basierte Suche der DGD

Über die Jahre ist die DGD zu einem mächtigen Korpusanalyse- und -managementsystem gewachsen und stellt heutzutage ein unverzichtbares korpuslinguistisches Werkzeug für Untersuchungen in Korpora des gesprochenen Deutsch dar. Sie hat aktuell etwa 16.500 registrierte Nutzer:innen (Stand: Mai 2023) und wird in verschiedenen wissenschaftlichen Forschungsbereichen, in der Hochschullehre und im Studium verwendet. Nichtsdestotrotz ist die Nutzung der DGD mit einigen Einschränkungen verbunden, die im Folgenden kurz beschrieben werden sollen:

1. Die KWIC-Ansicht ist aus technischen Effizienzgründen der aktuellen Implementierung der DGD auf 10.000 Treffer eingeschränkt. Dies erweist sich bei Analysen hochfrequenter linguistischer Formen wie z.B. Artikel oder Negationspartikel bzw. grammatischer Strukturen, welche die genannten Wortarten enthalten, oft als hinderlich, da so z.B. quantitative Auswertungen nur auf der Basis der gezogenen Stichprobe vorgenommen werden können.
2. Die Suche auf Annotationen ist in der Datenbank auf eine feste Menge von tokenbasierten Eigenschaften (Normalisierung, Lemma, POS) beschränkt. Andere, z.B. tokenübergreifende Annotationen (engl. *span annotations*) können über die Datenbank nicht durchsucht werden. So sind z.B. im *FOLK*-Korpus Handlungssequenzen annotiert (Kaiser / Schedl 2021), die über die DGD den Nutzer:innen nicht zugänglich gemacht werden können.
3. Für die Erstellung der KWIC-Konkordanz fehlen der DGD noch einige grundlegende Suchmöglichkeiten wie z.B. Negationsoperatoren. Diese wären nützlich, um die Suche nach grammatischen Strukturen nach weiteren syntaktischen Kriterien einzuschränken. Beispielsweise könnten bei der Suche nach *ich weiß nicht* (vgl. Helmer / Deppermann / Reineke 2017) Belege ausgeschlossen werden, auf die ein subordinierter Komplementsatz mit *ob* folgt. Auch eine Einschränkung in Bezug auf Metadaten könnte durch Negationsoperatoren erfolgen – z.B. könnten damit Belege gesucht werden, in denen *ich weiß nicht* von Sprecher:innen geäußert werden, die nicht Deutsch als Erstsprache sprechen.

4. Komplexe Suchen werden in der DGD über ein schrittweises Ausfiltern der Suchergebnisse in der KWIC-Ansicht umgesetzt. Während dies ein adäquates Mittel ist, um die Auswirkung einzelner Komponenten komplexerer Suchanfragen transparent nachvollziehbar zu machen, entspricht es nicht der für die Korpuslinguistik kanonischen Herangehensweise. In Ergänzung dazu wäre es daher wünschenswert, eine komplexe Suchanfrage mit Hilfe einer Suchanfragesprache effizient stellen zu können (ohne sich durch mehrere Filter durchklicken zu müssen). Beispielsweise sind in der DGD für die Suche nach längeren Wortverbindungen, wie *ich weiß nicht*, mehrere Filteranwendungen erforderlich. Eine speziell für korpuslinguistische Zwecke entwickelte Suchanfragesprache kann eine solche mehrschrittige Suche in einem einzigen Vorgang zusammenfassen (mehr dazu in Kapitel 4.1).
5. Da die DGD in erster Linie für die Arbeit mit Gesprächs- und Variationskorpora ausgerichtet ist, sind ihre Benutzeroberfläche und Recherchemöglichkeiten auch vornehmlich an den Bedarfen von Gesprächsforscher:innen und Dialektolog:innen orientiert. Andere Nutzergruppen (wie z.B. aus den Bereichen der Lexikographie, DaF/DaZ-Forschung und -Lehre) sind weniger berücksichtigt.

Im Rahmen des *ZuMult*-Projekts wurden die hier genannten Einschränkungen aufgegriffen: Es wurde eine komplett neue technische Umgebung für die Suche entwickelt, die nun durch eine prototypisch implementierte grafische Benutzeroberfläche unter dem Namen *ZuRecht* in die DGD integriert wurde.

3. *ZuRecht*

Die Konzeption eines neuen Tools für die Recherche in den AGD-Korpora begann mit einer Sichtung von zwei Dutzend moderner internationaler Online-Plattformen, die verschiedene Arten von Zugängen zu mündlichen Korpora anbieten. Die Ergebnisse dieser Korpusplattformanalyse (vgl. dazu Batinic / Frick / Schmidt 2021) zeigten, dass Portale, die systematisches Durchsuchen von Transkripten mit Hilfe einer Suchanfragesprache (meistens CQP) erlauben, ursprünglich für geschriebene Sprache entwickelt wurden (z.B. CQPWeb⁶, KonText⁷, OpenSoNaR⁸). Die dahinterstehenden Suchmaschinen erlauben zwar komplexe Suchanfragen auf Daten mit tokenübergreifenden und Mehr-Ebenen-Annotationen, jedoch bieten sie keine Möglichkeiten zur Berücksichtigung der besonderen Merkmale gesprochener Sprache wie z.B. Sprecherüberlappungen oder zeitbasierte Annotationen. Dagegen steht bei den Anwendungen, die speziell für gesprochensprachliche Korpora entwickelt wurden, eher das Browsing von Transkripten im Vordergrund. Die Suchfunktionalität ist oft eingeschränkt und erfolgt über spezielle GUI-Elemente (Formulare, Filter) ohne die Möglichkeit, eine korpuslinguistische Suchanfragesprache zu nutzen (z.B. CLAPI⁹, ESLO¹⁰, TGDA¹¹). Des Weiteren existieren unter diesen Plattformen bislang nur wenige, die gezielt sprachdidaktisch relevante Bedürfnisse aufgreifen (CLAPI-FLE¹² und Plattform Gesprochenes Deutsch¹³). Genau diese Problematiken hatte die Entwicklung des neuen Tools im Blick. Mit *ZuRecht* ist eine innovative Plattform entstanden, die eine

⁶ Zugang zu Spoken BNC2014, <https://cqpweb.lancs.ac.uk> (07.10.2022)

⁷ Zugang zu mündlichen Korpora aus CNC (Czech National Corpus), <https://www.korpus.cz> (07.10.2022)

⁸ Zugang zum Spoken Dutch Corpus (Corpus Gesprochen Nederlands, CGN), https://portal.clarin.inl.nl/opensonar_frontend/opensonar/search (07.10.2022)

⁹ Corpus de Langue Parlée en Interaction, <http://clapi.icar.cnrs.fr> (07.10.2022)

¹⁰ Corpus linguistique composé d'enregistrements sonores et de leurs transcriptions réalisés à Orléans, <http://eslo.humanum.fr> (07.10.2022)

¹¹ Texas German Dialect Archive, <https://tgdp.org/dialect-archive> (07.10.2022)

¹² <http://clapi.icar.cnrs.fr/FLE/> (07.10.2022)

¹³ <https://dafdaz.sprache-interaktion.de> (07.10.2022)

mächtige korpuslinguistische Suchanfragesprache zum Durchsuchen von authentischen Interaktionsdaten nutzt und zugleich Recherchemöglichkeiten anbietet, die Spezifika gesprochener Sprache und für die Sprachvermittlung wichtige Aspekte fokussieren (mehr dazu in den folgenden Kapiteln).

Auf *ZuRecht* kann entweder per Link aus der DGD (unter dem Reiter „Recherche“) oder direkt auf der Webseite des *ZuMult*-Projekts¹⁴ zugegriffen werden. Zum Einloggen benötigt man eine Registrierung bei der DGD¹⁵. Als Erweiterung der DGD bietet *ZuRecht* eine Vielzahl von neuen Funktionalitäten, die im Folgenden anhand von sechs unterschiedlichen Aspekten (Search Engine, Transkriptformat, Anfragesprache, Suchoptionen, Ergebnisdarstellung und Hilfestellung) kurz dargestellt werden.

3.1 MTAS-basierte Search Engine

Die in der DGD verwendete Oracle Datenbank bietet eine mächtige Query-Syntax an. Wenn es um größere Datenmengen geht, sind Lucene¹⁶-basierte Suchmaschinen jedoch in Geschwindigkeit und Skalierbarkeit relationalen Datenbanken gegenüber überlegen. Zwar hat man im Fall von mündlichen Korpora noch nicht mit vergleichbar großen Datenmengen wie bei Korpora geschriebener Sprache zu tun (mehr dazu bei Kupietz / Schmidt 2015). Es lohnt sich aber trotzdem auf mittlere und längere Sicht, in Bezug auf die DGD den Arbeitsaufwand nicht nur in die Weiterentwicklung der gegenwärtigen Software zu investieren, sondern schon heutzutage mit Aussicht auf die Zukunft neue alternative Architekturlösungen zu testen, welche einen effizienten Umgang mit sehr großen Mengen mündlicher Daten ermöglichen und diese innerhalb einer den Endnutzer:innen zumutbaren Zeit durchsuchen können. So ist im *ZuMult*-Projekt eine vollkommen neue Suchmaschine für die Recherche in den AGD-Korpora entstanden, die mit Hilfe des speziell für Korpussuchen entwickelten Lucene-basierten Frameworks MTAS (Multi-Tier Annotation Search, vgl. Brouwer et al. 2016) implementiert wurde. Im Vergleich zu den anderen Lucene-basierten und ebenso für Korpusrecherchen entwickelten *open source* Frameworks wie z.B. KorAP¹⁷ und BlackLab¹⁸ zeichnet sich MTAS durch eine große Flexibilität bei der Erstellung der Suchindizes aus. Diese können direkt aus den in unserem Projekt verwendeten TEI-basierten und durch die ISO standardisierten Transkriptformaten erstellt und mittels einer Korpussuchanfragesprache abgefragt werden.

Die Dienste der neu implementierten *ZuMult*-Suchmaschine können über RESTful Web Services¹⁹ genutzt werden. Sie enthalten keine Einschränkungen für die Anzahl und Darstellung der Suchergebnisse: Alle Treffer können vom Client (hier *ZuRecht*) in kleinen Teilen angefragt und dem/der Nutzer:in angezeigt werden.

¹⁴ <http://zumult.ids-mannheim.de/ProtoZumult/index.jsp> (07.10.2022)

¹⁵ Der Zugang zur DGD kann nur erteilt werden, wenn es sich um eine wissenschaftliche und nicht-kommerzielle Nutzung für Forschung, Lehre und Studium an einer Hochschule oder vergleichbaren wissenschaftlichen Institution handelt (vgl. dazu DGD-Nutzungsbedingungen unter https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.sys_use (07.10.2022)).

¹⁶ <https://lucene.apache.org/> (07.10.2022)

¹⁷ <https://github.com/KorAP> (07.10.2022)

¹⁸ <https://inl.github.io/BlackLab> (07.10.2022)

¹⁹ Hierbei handelt es sich um eine spezielle Art von Online-Diensten, die auf den Prinzipien des Representational State Transfer (REST) Architekturstils mit seinem Ressourcen-Konzept aufbauen und über die HTTP-Anfragen Zugriff auf Daten gewährleisten.

3.2 ISO-TEI-Standard für Transkripte

Wie oben bereits erwähnt, arbeitet die neue Suchmaschine mit den Transkripten im TEI-basierten und dem von der ISO standardisierten Dateiformat (ISO 24624:2016). Im Gegensatz zu *FOLKER*-Transkripten²⁰, die aktuell von der DGD verwendet werden, bietet der ISO/TEI-Standard Mittel zur Integration der token- und sprecherübergreifenden sowie zeitbasierten Annotationen, die nun über *ZuRecht* durchsucht werden können. Als Beispiel können die manuell erstellten Annotationen von sprachlichen Handlungen (mehr dazu in Kapitel 5.2) und automatisch hinzugefügten *speech rate*-Annotationen (siehe Kapitel 4.3) im *FOLK* und *GeWiss*-Korpora genannt werden.

3.3 Suchanfragesprache (CQP)

Für die Suchanfragen in *ZuRecht* wird die MTAS eigene Variante der CQP-Syntax verwendet. CQP ist ursprünglich die Abfragesprache des Corpus Query Processors – einer linguistischen Suchmaschine, die an der Universität Stuttgart als Teil der IMS Open Corpus Workbench (CWB)²¹ entwickelt wurde. CQP kennzeichnet sich durch eine besondere Syntax aus. Diese besteht aus einer Reihenfolge von Tokenspezifikationen, welche aus einer oder mehreren Attribut-Wert-Paaren zusammengesetzt und durch eckige Klammer dargestellt werden. (1) ist ein Beispiel für eine mögliche CQP-Suchanfrage. Diese findet alle orthographisch als *auf* normalisierten Formen, gefolgt von einem Element der Wortart „Artikel“ (ART) und dem Lemma *Seite*. Zwischen dem Artikel und dem Nomen kann optional (dies spezifiziert der reguläre Ausdruck `?`) ein Element der Wortart „attribuierendes Indefinitpronomen“ (PIDAT) vorkommen. Die Suchanfrage findet damit Sequenzen wie *auf die seite, uf de seit, auf der andern seite, uff der annere seite* etc.

(1) `[norm="auf"] [pos="ART"] [pos="PIDAT"]? [lemma="Seite"]`

Da heutzutage viele Suchmaschinen (wie z.B. Poliqarp²², SketchEngine²³, BlackLab²⁴) eine ähnliche Syntax für ihre Anfragesprachen verwenden, wird mittlerweile oft eine ganze Familie von Suchanfragesprachen, die eine CQP-ähnliche Syntax aufweisen, mit dem Begriff *CQP* bezeichnet, im Gegensatz zu den anderen z.B. graphenbasierten Anfragesprachen mit einer ganz anderen Syntaxform (vgl. z.B. ANNIS Query Language²⁵). Eine Studie zur Nutzung mündlicher Korpora (vgl. Fandrych et al. 2016) ergab, dass CQP auch diejenige Sprache ist, die den meisten (wenn auch immer noch wenigen) DGD-Nutzer:innen geläufig ist. Nun kann diese Sprache verwendet werden, um AGD-Korpora via *ZuRecht* zu durchsuchen. Es handelt sich um die MTAS eigene Variante der CQP, die eine modifizierte Form der Suchanfragesprache der CWB darstellt, erweitert durch einige Operatoren der Corpus Query Language von *SketchEngine*. Abbildung 3 zeigt die oben erwähnte CQP-Suchanfrage eingegeben in das Suchanfragefeld von *ZuRecht*²⁶.

²⁰ <https://agd.ids-mannheim.de/folker.shtml> (07.10.2022)

²¹ <https://cwb.sourceforge.io/> (07.10.2022)

²² <http://nkjp.pl/poliqarp/> (07.10.2022)

²³ <https://www.sketchengine.eu/documentation/corpus-querying/> (07.10.2022)

²⁴ <https://inl.github.io/BlackLab> (07.10.2022)

²⁵ <http://korpling.github.io/ANNIS/4.3/user-guide/aql/> (07.10.2022)

²⁶ Eine genauere Beschreibung der Suchanfragesprache findet sich in *ZuRecht* unter dem Fragezeichen-Button neben dem CQP-Suchanfragefeld.

Suche ▾ [norm="auf"] [pos="ART"] [pos="PIDAT"]? [lemma="Seite"]



Abbildung 3
Beispiel für eine mögliche CQP-Suchanfrage in *ZuRecht*

3.4 Suchoptionen

Die MTAS-basierte Search Engine von *ZuRecht* bietet mehrere Vorteile gegenüber den aktuellen Suchmöglichkeiten der DGD. Zum einen konnten mit der neuen Search Engine neue Elemente aus den Transkriptionen indiziert und für die Suche verfügbar gemacht werden, z.B. können jetzt nicht nur Word-Token und tokenbasierte Annotationen, sondern auch zeitbasierte, token- und sprecherübergreifende Annotationen, Sprecherüberlappungen, transkribierte Pausen (inkl. Dauer), Ein- und Ausatmen sowie andere verbale und nonverbale Phänomene recherchiert werden. Zusätzlich wurden unterschiedliche Typen von Suchindizes erstellt, um die Suche mit oder ohne Satzzeichen und die sprecherbasierte Suche (mehr dazu in Kapitel 4.2 und in Frick / Helmer / Schmidt 2022) zu ermöglichen.

Zum anderen besteht nun die Möglichkeit, die AGD-Korpora mit Hilfe einer extra für Korpusrecherchen entwickelten Suchanfragesprache durchsuchen zu können. Das bedeutet erstens, dass eine komplexe Suchanfrage als ein geschlossener Ausdruck formuliert und in einem einzigen Schritt (ohne mehrstufiges Filtern) abgeschickt werden kann. Zweitens heißt es, dass die Suchanfragen für die AGD-Korpora komplexer als bis jetzt gestaltet werden können. Es kann z.B. direkt in der Suchanfrage eine Negation für die Ausschließung bestimmter Token, Annotationen und Metadaten formuliert werden. Zudem können optionale Query-Elemente (mit Hilfe regulärer Ausdrücke) und Vergleichsoperatoren (z.B. für die Suche nach Pausen mit einer bestimmten Länge) verwendet werden. Suchen nach unbestimmten Token, nach Strukturen und die Positionssuche (z.B. mit Hilfe der Präzedenzoperatoren) sind ebenfalls möglich. Hier sind einige Beispiele:

- (2) `[norm="wir" & word.type=".*assimilated.*"]`
findet alle assimilierten Formen von „wir“, z.B. [da ham] a [n problem], [kö] mer [auch]
- (3) `[pause.dur.ceil>6]`
findet alle Pausen länger als 6 Sekunden
- (4) `[word="die"] [pos="ADJA"] {0,2} [word="woche"]`
findet in der Transkription die Formen „die“ und „woche“ direkt aufeinander folgend oder mit maximal zwei Adjektiven dazwischen
- (5) `[norm="ich"] [norm="verstehe"] [] {2} [norm="nicht"]`
*findet alle orthographisch auf „ich verstehe“ und „nicht“ normalisierten Formen mit zwei beliebigen Token dazwischen, z.B. **ich verstehe** das jetzt **nicht**, **ich versteh** s gar **nicht**, **ich verstehe** den streit **nicht***
- (6) `<word/> followedby [lemma="schön"]`
*findet alle Wort-Token vor dem lemma „schön“ wie z.B. **sehr** [schön], **ganz** [schön], **wie** [schön]*
- (7) `[norm="ja"] !precededby <annotationBlock>`
findet alle orthographisch als „ja“ normalisierten Formen, die nicht am Anfang eines Sprecherbeitrags vorkommen
- (8) `[norm="und"] </annotationBlock>`
findet alle orthographisch als „und“ normalisierte Formen am Ende eines Sprecherbeitrags

Darüber hinaus bietet die *ZuRecht*-Nutzeroberfläche neben der CQP-basierten Suche noch einen weiteren Lemma- und Suchanfragelisten-basierten Zugang zur Recherche in Transkripten. Es besteht die Möglichkeit, eine benutzerdefinierte Liste der Lemmata als Text-Datei zu erstellen, sie in *ZuRecht* hochzuladen und sich die Transkripte anzeigen zu lassen, die die meisten Lemmata aus dieser Liste enthalten. Die Ergebnisse können anschließend nach absoluten oder relativen Häufigkeiten der Treffer sowie auch nach den gefundenen Wortformen sortiert werden. Mit der Entwicklung dieser Funktionalität sollten Bedürfnisse bestimmter Nutzergruppen Berücksichtigung finden wie etwa die von Lexikograf:innen/Lexikolog:innen bei der Suche nach Neologismen-Beispielen, oder die von DaF/DaZ-Lehrenden bei der Bearbeitung einer für das Thema des Unterrichts relevanten Wortschatzliste (mehr dazu in Kapitel 5.1).

3.5 Arbeiten mit Suchergebnissen

Die Ergebnisse der CQP-Suche werden in *ZuRecht* als KWIC-Konkordanz angezeigt (Abbildung 4). Zu jedem Treffer können Sprecher- und Gesprächsmetadaten abgerufen werden. Zusätzlich kann jeder Treffer im Transkriptbrowser *ZuViel* (Zugang zu Visualisierungselementen für Transkripte) im größeren Kontext geöffnet, angehört und nach Bedarf heruntergeladen werden (für weitere Informationen siehe Schmidt / Schwendemann / Wallner 2023 in dieser Themenausgabe).

The screenshot shows a search results page titled 'Ergebnisse' for the query '[norm="zurecht"] (in FOLK)'. It includes buttons for 'Treffer gruppieren', 'Metadatenansicht', and 'Download der KWIC'. Below the search bar, it indicates 'Insgesamt: 28' results and provides navigation links for 'Erste Seite', 'Vorherige Seite', '1', '2', '3', 'Nächste Seite', and 'Letzte Seite'. The main content is a table of search results with columns for ID, metadata, snippet, and actions.

ID	Metadata	Snippet	Actions
1	FOLK_E_00055_SE_01_T_05 US	... ma wenn ma zurecht irgendwas (.) m es ...	DGD ZuViel
2	FOLK_E_00069_SE_01_T_07 BH	... der optimismus (.) ob zurecht (.) hm is sehr ...	DGD ZuViel
3	FOLK_E_00013_SE_01_T_02 CJ	... rückt ihre brille zurecht (.) mh (.) sieht lecker ...	DGD ZuViel
4	FOLK_E_00016_SE_01_T_01 CJ	... rückt ihr brille zurecht (.) mh (.) sieht lecker ...	DGD ZuViel
5	FOLK_E_00016_SE_01_T_01 CJ	... rückt ihre brille zurecht (0.72) °h (.) im sturzflug (0.21) ((summt)) schwirrt ...	DGD ZuViel
6	FOLK_E_00069_SE_01_T_01 HG	... auch nicht damit zurecht gekommen das freut ...	DGD ZuViel

Abbildung 4
KWIC-Konkordanz in *ZuRecht*

Die Suchergebnisse lassen sich nach Metadaten, Annotationen und der Anzahl der Token im Treffer gruppieren. Eine durch Nutzer:innen definierte Anzahl der Treffer inklusive der benötigten Metadaten und dem spezifizierten Kontext kann in der KWIC-Form im XML-Format heruntergeladen werden, um z.B. im Excel weiterverarbeitet werden zu können (ein Beispiel dazu findet sich in Kapitel 4.1).

3.6 Hilfe zum Erstellen von Suchanfragen

Die Suche in *ZuRecht* bietet viele Möglichkeiten, erfordert aber auch gewisse Vorkenntnisse der CQP-Syntax und der in Korpora verfügbaren Metadaten und Annotationen. Um den Nutzer:innen

den Einstieg zu erleichtern, wurde ein dem Query Builder²⁷ ähnliches Konstrukt in das Suchanfragefeld integriert, das automatisch – nach dem Mausklick in das Suchanfragefeld – erscheint und mit dessen Hilfe eine Auswahl der für die Suche vorhandenen Annotationen und Metadatenkategorien sowie deren möglichen Spezifikationswerte (z.B. POS-Annotationstags) getroffen werden kann (Abbildungen 5 und 6). Zusätzlich stehen eine umfangreiche Dokumentation der in *ZuRecht* verwendeten Suchanfragesprache und zahlreiche Suchanfragebeispiele direkt im User Interface (über den Hilfe-Button mit dem Fragezeichen) zur Verfügung.

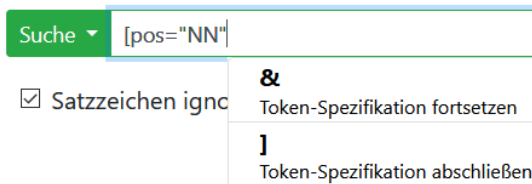


Abbildung 5
Hilfe zur CQP-Syntax

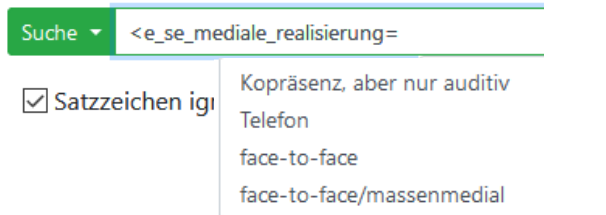


Abbildung 6
Auswahl der verfügbaren Spezifikationswerte
für die Metadaten-Kategorie „mediale Realisierung“

Derzeit befindet sich *ZuRecht* noch in einem Prototyp-Status. Auch wenn es noch auf die Schnelligkeit und Skalierbarkeit für größere Datenmengen getestet werden muss, eröffnet es den DGD-Nutzer:innen bereits jetzt vielfältige neue Forschungs- und Einsatzmöglichkeiten. Diese sollen in den nächsten Kapiteln exemplarisch skizziert werden.

4. Nutzungsmöglichkeiten in der Gesprächsanalyse

Formbasierte Korpusrecherchen haben sich als nützliches methodisches Element gesprächsanalytischer Studien erwiesen (vgl. Deppermann / Schmidt 2014), insbesondere für die aus der Gesprächsanalyse hervorgegangene Interaktionale Linguistik, die sich unter anderem für Fragen der Grammatik des gesprochenen Deutsch interessiert (vgl. Selting / Couper-Kuhlen 2000)²⁸. Formbasierte Recherchen in Korpora sind darüber hinaus auch im Kontext von Deutsch als Fremd- und Zweitsprache relevant: Forschende können über sie gezielt Formen und Strukturen finden, die sie interessieren. Zudem können korpusbasierte, gesprächsanalytische und interaktionslinguistische Untersuchungen zu gesprochenem Deutsch als relevante Grundlagenforschung für die Vermittlung im DaF/DaZ-Kontext herangezogen werden (vgl. Fandrych / Meißner / Wallner 2021).

FOLK kann nun auch mithilfe der neuen Korpusrechercheplattform *ZuRecht* durchsucht werden, welche den Forschungsprozess für Forschende effizienter macht und die Bearbeitung bestimmter Fragestellungen überhaupt erst ermöglicht. Anhand von einigen exemplarischen Fallbeispielen wird der Rechercheprozess in *ZuRecht* dargestellt. Zunächst geht es dabei um die effiziente Suche nach Mehrworteinheiten, danach um das Suchen von Tokens im Kontext von Sprecherwechseln und schließlich um die Suche nach Tokens mit einer gewünschten Sprechgeschwindigkeit und im Kontext von Sprecherüberlappungen auf der Basis tokenübergreifender Annotationen.

²⁷ Der Query Builder ist eine visuelle Komponente zur schrittweisen Erstellung der Suchanfragen über die grafische Benutzeroberfläche, die von Nutzer:innen ohne Kenntnisse der Syntax der jeweiligen Suchanfragesprache bedient werden kann.

²⁸ Dies zeigt nicht zuletzt eine große Anzahl von Publikationen, die mit dem *FOLK*-Korpus arbeiten, vgl. eine (stetig erweiterte) Publikationsliste hier: <https://www.ids-mannheim.de/prag/muendlichekorpora/bibliographie-folk/> (07.10.2022).

4.1 Suche nach Mehrworteinheiten am Beispiel von *ich weiß nicht*

In Untersuchungen zu grammatischen Strukturen des gesprochenen Deutsch wurden regelmäßig Mehrworteinheiten untersucht, etwa Reduplizierungen wie *ja ja* (vgl. Barth-Weingarten 2011), Kommentarphrasen wie *ohne Scheiß* (vgl. Torres Cajo 2017), Modalverbkonstruktionen wie *darf ich/kann ich* (vgl. Deppermann / Gubina 2021), Quotativkonstruktionen wie *nach dem Motto* und *von wegen* (vgl. Moroni 2021) oder aus mehreren Wörtern bestehende Diskursmarker wie *ich sag mal so* (vgl. Auer / Günthner 2004), *keine Ahnung* (vgl. Bergmann 2017) oder *ich weiß nicht* (vgl. Helmer / Deppermann / Reineke 2017).

Entsprechende Suchen lassen sich bereits in der DGD vornehmen. Mit *ZuRecht* eröffnet sich den Forschenden allerdings die Möglichkeit, die Suche nach Mehrworteinheiten noch effizienter und kompakter, vor allem ohne die Notwendigkeit mehrerer Filteranwendungen und zusätzlich mit neuen Suchoptionen (z.B. Negation), durchzuführen. Dies lässt sich am Beispiel von *ich weiß nicht* gut illustrieren.

In der DGD müssen Forschende zunächst eines der Wörter auswählen, das sie im Recherche-reiter für die Tokensuche (siehe Wortindex-basierte Recherche in Kapitel 2) als „Basistoken“ eingeben. Bei dem Format *ich weiß nicht* muss dies das Worttoken *weiß* sein, weil die Token *ich* und *nicht* zu häufig sind und ihre Belegstellen wegen der Treffereinschränkung der Datenbank nur unvollständig angezeigt werden können²⁹. Der Suche nach diesem ersten Token folgen in der DGD mehrere Schritte, in denen der linke und rechte Kontext des Basistokens bestimmt werden. So wird in dem hier beschriebenen Beispiel zuerst nach *ich* links des Basistokens *weiß* gesucht. Die Ergebnisse werden ausgefiltert und anschließend das Token *nicht* im rechten Kontext des Basistokens gesucht. Da nicht mehrere Token gleichzeitig im Kontext gesucht werden können, muss zwischen den Suchen immer wieder der Schritt des Ausfilterns der nicht passenden Belege aus der KWIC-Liste erfolgen. Interessieren Forschende ausschließlich Fälle von *ich weiß nicht*, denen kein Komplementsatz folgt (z.B. *ich weiß nicht, ob* oder Fälle mit Relativ- und Interrogativpronomen wie *ich weiß nicht, wer/wann/welche*)³⁰, müssen über die Kontextsuche in mehreren weiteren Schritten diverse Wortformen und POS-Tags ausgeschlossen werden, etwa *ob* in der normalisierten Form sowie die POS-Tags *PWS*, *PWAV*, und *PWAT*, mit denen die oben genannten Relativ- und Interrogativpronomen getaggt sind.

Durch die Möglichkeit, die CQP-Suchanfragesprache zu verwenden, lassen sich solche komplexen Suchanfragen in *ZuRecht* deutlich effizienter und kompakter gestalten. In einer einzigen Zeile kann die gesamte Suchanfrage mit dem folgenden Suchausdruck (Beispiel 9) eingegeben und an den Server abgeschickt werden (vgl. auch Abbildung 7):

(9) `[norm="ich"][norm="weiß"][]{0,3}[norm="nicht"] !followedby
[pos="PW.*"|norm="ob"]`

²⁹ Da *FOLK* stetig wächst, wird dies zukünftig auch Suchen nach anderen häufig vorkommenden Wortformen betreffen, die aktuell die Frequenz von 10.000 Belegen noch nicht überschreiten.

³⁰ Dies ist der Fall in der Studie von Helmer / Deppermann / Reineke (2017), in der die Autor:innen alleinstehende Fälle der grammatisch vollständig realisierten Formen *ich weiß es/das nicht* mit Fällen der grammatisch reduzierten Formen (*ich*) *weiß nicht* vergleichen. In beiden Fällen interessieren nur Belege ohne folgenden Komplementsatz.

The screenshot shows the ZuRecht search interface. At the top, there's a navigation bar with 'ZuRecht', 'CQP-Suchanfragen in den Korpora des AGD', 'Startseite', 'Projekt', and a language dropdown set to 'Deutsch'. On the left, a sidebar lists corpora: DH, FOLK (checked), GWSS, MEND, DNAM, BETV, and HMAT. The main area is titled 'Suche mittels der Corpus Query Language (CQP)' and prompts the user to enter a CQP query and select a search mode. The search bar contains the query: `[norm="ich"][norm="weiß"][]{0,3}[norm="nicht"]!followedby [pos="PW.*"][norm="ob"]`. Below the search bar, there are options for 'Satzzeichen ignorieren' (checked) and links for 'Beispiele' and 'XML'. The results section shows 'Ergebnisse für die Suchanfrage [norm="ich"][norm="weiß"][]{0,3}[norm="nicht"]!followedby [pos="PW.*"][norm="ob"] (in FOLK)'. There are buttons for 'Treffer gruppieren', 'Metadatenansicht' (highlighted with a red box), and 'Download der KWC'. A pagination bar shows 'Insgesamt: 1213' and page numbers 1 through 5. The first four results are listed with their transcript IDs and speaker initials (SZ, HM, MS).

Abbildung 7
Die einzeilige Suchanfrage in ZuRecht

Durch `[norm="ich"][norm="weiß"][]{0,3}[norm="nicht"]` im Beispiel (9) kann direkt zusammenhängend nach der Folge *ich weiß nicht* gesucht werden (wobei auch dialektale Varianten wie *ick weeiß nich* eingeschlossen sind), ohne dabei ein Basistoken auswählen zu müssen. In der Suchanfragesyntax kann außerdem spezifiziert werden, in welchem Abstand die Token aufeinander folgen können. Im Beispiel oben sollen *ich* und *weiß* direkt aufeinander folgen; dagegen gibt `[]{0,3}` an, dass zwischen *weiß* und *nicht* drei Token liegen dürfen, was Belege wie *ich weiß es/das echt gar nicht* mit einschließt. Die Abstandsangabe kann beliebig variiert werden, ohne dass bei einer Änderung wieder ein mehrschrittiger Suchprozess von Anfang an begonnen werden muss, wie es in der DGD der Fall wäre.

Durch den Negationsoperator **!** in Kombination mit dem Präzedenzoperator **followedby** kann außerdem direkt bei der Suchanfrage spezifiziert werden, welche Wortformen oder Wortarten nicht auf *ich weiß nicht* folgen sollen. Wie bereits in der oben beschriebenen DGD-Suche können dies verschiedene Pronomen sein sowie die Konjunktion *ob*. Durch eine Kombination mit dem regulären Ausdruck `.*` werden durch `!followedby [pos="PW.*"][norm="ob"]` sowohl alle Pronomen, deren POS-Tags mit „PW“ beginnen³¹, aussortiert als auch alle normalisierten Formen von *ob*. Die so formulierte komplette Suchanfrage liefert insgesamt 1213 Treffer (Stand: September 2022, vgl. Abbildung 7).

Durch Anklicken der Transkript-ID (gestrichelt umrandet in Abbildung 7) wird man zur Datenbank für Gesprochenes Deutsch weitergeleitet und bekommt dort die Metadaten des entsprechenden Gesprächs angezeigt. Auch die Metadaten der Sprecher:innen lassen sich auf diese Weise bei einem Klick auf die Sprechersiglen (gepunktet umrandet in Abbildung 7) anzeigen.

Die Trefferliste lässt sich außerdem über den Button „Metadatenansicht“ (mit einer durchgezogenen Linie umrandet in Abbildung 7) nach verschiedenen Metadaten-Kriterien gruppieren. Es

³¹ Dies sind die oben erwähnten Relativ- und Interrogativpronomen. Alternativ lassen diese sich auch explizit auflisten: `[pos="(PWS|PWAV|PWAT)"]`.

könnte Forschende etwa interessieren, in welcher Art Interaktionstyp in *FOLK* *ich weiß nicht* besonders häufig zu finden ist. Zu diesem Zweck wäre als Deskriptor „SE: Kurzbezeichnung („Art“)“ auszuwählen und die Ergebnisse nach „Treffer (abs.) absteigend“ zu sortieren. Im oben beschriebenen Fall zeigt sich, dass *ich weiß [es/das] nicht* ohne Komplementsatz in *FOLK* aktuell am häufigsten in Telefongesprächen geäußert wird (vgl. Abbildung 8)³².

Ergebnisse			
für die Suchanfrage <code>[norm="ich"][norm="weiß"][]{0,3}[norm="nicht"] !followedby [pos="PW.*" norm="ob"]</code> (in FOLK)			
Metadaten Deskriptor (Wählen Sie): <input type="text" value="SE: Kurzbezeichnung („Art“)"/>			
Sortiert nach (Wählen Sie): <input type="text" value="Treffer (abs.) absteigend"/>			
Insgesamt: 91 SE: Kurzbezeichnung („Art“) Treffer, Insgesamt: 1213			
1	Telefongespräch	231	KWIC öffnen
2	Tischgespräch	94	KWIC öffnen
3	Sprachbiografisches Interview	77	KWIC öffnen
4	Spielinteraktion zwischen Erwachsenen	66	KWIC öffnen
5	Meeting in einer sozialen Einrichtung	53	KWIC öffnen

Abbildung 8
Ergebnisse der Metadatenansicht

Möchten Forschende schon im ersten Schritt nur Belege aus Interaktionen finden, die bestimmten Kriterien entsprechen, können Angaben zu Metadaten auch direkt in den Suchausdruck integriert werden. Im Beispiel könnten Forschende die KWIC-Ergebnisliste direkt auf Belege aus Telefongesprächen einschränken wollen, z.B., weil dort die meisten Belege zu finden sind, oder mit der Motivation, nur solche Interaktionen zu berücksichtigen, in denen Mimik und Gestik keine Rolle spielen (können). Die entsprechende Einschränkung lässt sich mit `within <e_se_mediale_realisierung="Telefon"/>` vornehmen, wobei auch hier die Hilfe bei der Generierung der Suchsyntax schon mögliche Metadatenoptionen auflistet (vgl. hierzu Abbildung 6 in Kapitel 3). Der gesamte vorherige Suchausdruck muss bei der neuen Anfrage in runde Klammern gesetzt werden wie das in Beispiel (10) gezeigt ist. Die so weiter eingeschränkte Suchanfrage liefert insgesamt 231 Treffer (Stand: September 2022) in der KWIC-Ergebnisliste.

(10) `([norm="ich"][norm="weiß"][]{0,3}[norm="nicht"] !followedby [pos="PW.*"|norm="ob"]) within <e_se_mediale_realisierung="Telefon"/>`

Was auf den ersten Blick zunächst komplex und voraussetzungsreich wirkt, ist lediglich eine kompakte Verschriftlichung des oben beschriebenen Suchprozesses und umgeht das Durchlaufen mehrerer Schritte bei der Kontextsuche und dem Filtern. Die in Kapitel 3 beschriebene nutzerfreundliche

³² Bei der Trefferangabe handelt es sich um absolute Häufigkeiten, d.h. das Ergebnis liegt nicht etwa (notwendigerweise) daran, dass Sprecher:innen relativ gesehen *ich weiß nicht* besonders häufig am Telefon äußern würden, sondern es gibt lediglich viele Aufnahmen (mit entsprechend vielen Token) dieser Art des Interaktionstyps in *FOLK*. Vor einer Interpretation gefundener Ergebnisse ist es deshalb immer maßgeblich, dass Forschende sich mit dem Korpus, den verschiedenen Interaktionstypen sowie der Anzahl und Länge von berücksichtigten Interaktionen auseinandersetzen.

automatische Vervollständigung der Suchanfragesyntax erleichtert das Vorgehen zudem sehr. Insbesondere, wenn Forschende bei der Eingabe der Suchsyntax Routine gewinnen sowie vorab schon eine Vorstellung davon haben, welche Bedingungen für ihre Suche gelten sollen (z.B. ausgeschlossene oder spezifizierte Wortformen und POS-Tags oder bestimmte Kriterien, die in Metadaten erfasst sind), gestaltet sich die Suche in *ZuRecht* deutlich effizienter und schneller als eine vergleichbare Suche in der DGD. Sie hat zudem eine höhere Fehlertoleranz in dem Sinne, dass Fehler in der Suchanfrageleiste selbst korrigiert werden können und nicht ein unter Umständen komplexer, mehrschrittiger Suchprozess von vorne begonnen werden muss.

Forschende können die Transkripte zu den Treffern in der KWIC-Ergebnisliste in der DGD oder im Tool *ZuViel* (vgl. Schmidt / Schwendemann / Wallner 2023 in dieser Themenausgabe) ansehen und die dazugehörigen Audio- (und ggf. Video-)Dateien abspielen. Typischerweise möchten Forschende aus der Gesprächsforschung die gefundenen Belege außerdem weiterbearbeiten, um erstens ihre Kollektion weiter zu schärfen (etwa weil sich noch falsche Positive darin befinden³³) und zweitens Belege nach bestimmten Kriterien weiter zu kodieren, d.h. kategorisieren (z.B. nach bestimmten grammatischen Aspekten oder bestimmten Sprechhandlungen). Zu diesem Zweck können Forschende die Treffer der KWIC-Ergebnisliste herunterladen. Beim Klick auf den Button „Download KWIC“ öffnet sich ein Fenster (vgl. Abbildung 9), in dem bestimmte Angaben wie die Größe des gewünschten Kontextes der KWIC-Belege spezifiziert werden können. Im Beispiel wurde der rechte Kontext größer gewählt, da sich so auf den ersten Blick schnell Belege identifizieren lassen, in denen *ich weiß nicht* potenziell als epistemischer oder pragmatischer Diskursmarker über den Folgekontext operiert. Zudem lassen sich auch in diesem Fenster weitere Metadaten auswählen – deren Ausprägungen werden zusätzlich in die herunterladbare Datei aufgenommen. Im Beispiel könnte es sinnvoll sein, als Zusatzinformation die Erstsprache der Sprecher:innen zu erfassen, um bei der weiteren Arbeit zwecks Ausschlusses potenziell beobachtbarer Effekte Personen auszuschließen, die nicht Deutsch als Erstsprache sprechen.

Die herunterladbare xml-Datei können Forschende in Excel öffnen. Dort finden sich in den Spalten die Namen der Transkripte, die Sprechersiglen, der linke und der rechte Kontext, die zuvor im Auswahlfenster markierten Metadaten-Angaben sowie je ein Link zum Transkriptausschnitt in der DGD und in *ZuViel* (vgl. Abbildung 10).

Im weiteren Forschungsprozess können Forschende in der Excel-Datei ihre weiteren Kodierungen in hierfür angelegten Spalten vornehmen. In der Studie von Helmer / Deppermann / Reineke (2017) etwa wurden Angaben zum Vorhandensein vs. Fehlen grammatischer Konstituenten (wie z.B. das Objekt *das/es* in *ich weiß [es/das] nicht*) sowie Kategorisierungen der nach *ich weiß nicht* folgenden Sprechhandlung vorgenommen (Vermutung, Vorschlag, Bekräftigung von Nichtwissen usw.).³⁴

³³ Dies ist auch der Fall bei der hier beschriebenen Suche. In Abbildung 7 findet sich in Zeile 2 z.B. noch ein Beleg mit *ob* nach *ich weiß nicht*, der aufgrund einer längeren Latenz durch Pausen und Atmen nicht aussortiert wurde.

³⁴ Ein entsprechender Forschungsprozess, wie er in Gesprächsforschung und Interaktionalen Linguistik prototypisch abläuft, wird auf der Basis des *FOLK*-Korpus und der DGD ausführlich in Reineke / Deppermann / Schmidt (2023) skizziert.

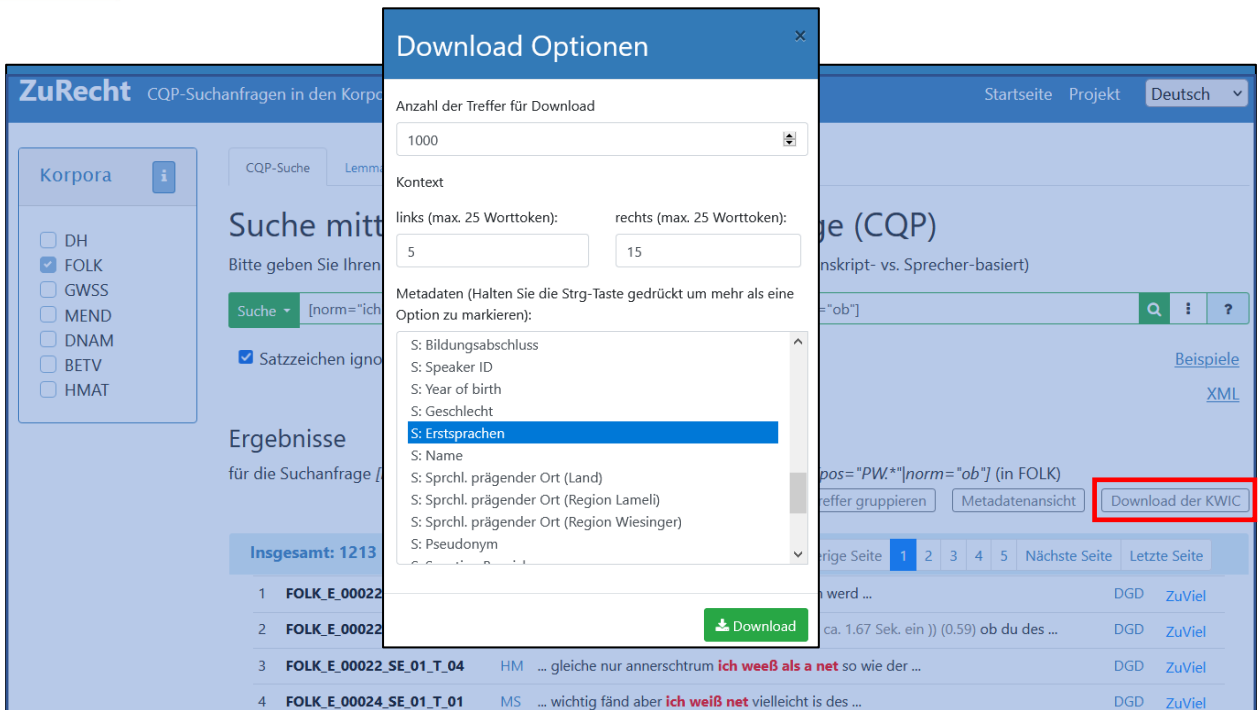


Abbildung 9
Auswahlfenster für den Download der KWIC-Belege als xml-Datei

	A	B	C	D	E	F	G	H
	transcript-id	spea	left-context	match	right-context	dgd-link	zumult-link	s_muttersprache
1	FOLK_E_00084_EG	EG	halt nur im krankenhaus äh	ich weiß auch nicht	das hat wahrscheinlich irgendwelche	http://zumult.ids-m	http://zumult.ids	Deutsch
2	FOLK_E_00084_EG	FR	er ja nich (0.46) nein ((Lachansatz)) h*	ich weiß is er auch nicht	(.) nein nein nein hab nich was gesagt	http://zumult.ids-m	http://zumult.ids	Deutsch
3	FOLK_E_00403_MT	MT	dann dann dann haben das (.) *h	ich weiß net	das ist (.) das natürlich ((Sprechansatz))	http://zumult.ids-m	http://zumult.ids	Deutsch
4	FOLK_E_00322_CE	CE	positiv ((Lachansatz)) ich weiß es net	ich weiß es echt net	*h mann die arbeit die war echt nich s	http://zumult.ids-m	http://zumult.ids	Deutsch
5	FOLK_E_00084_FR	FR	folge in der dritten staffel	ich weiß nicht	welche und äh verrat mir keine spoile	http://zumult.ids-m	http://zumult.ids	Deutsch
6	FOLK_E_00084_FR	FR	freitag mit ja (0.84) ja (.) ja	ich weiß ja au noch nicht	vielleicht bin ich auch am ersten märz	http://zumult.ids-m	http://zumult.ids	Deutsch
7	FOLK_E_00322_MP	MP	langsam voll ge (1.36) ja (.) weil (.)	ich weiß nicht	oah (0.78) ich will endlich geld verdie	http://zumult.ids-m	http://zumult.ids	Deutsch
8	FOLK_E_00084_EG	EG	halt erst (.) so (.) vermarkten ja	ich weiß es nicht	weil ich ha äh eigentlich (0.32) ich w	http://zumult.ids-m	http://zumult.ids	Deutsch
9	FOLK_E_00084_EG	EG	dinge wegen taurin und so *h (0.26)	ich weiß nicht	und halt ah gut ich mein wie das das j	http://zumult.ids-m	http://zumult.ids	Deutsch
10	FOLK_E_00403_MT	MT	ob ich s (.) ((lacht)) (0.24) ((lacht, 1.8s))	ich weiß net	((Sprechansatz)) du mochtest dann eh	http://zumult.ids-m	http://zumult.ids	Deutsch
11	FOLK_E_00322_MP	MP	du hast mich ja ermutigt	ich weiß nicht	mehr wo wir warn ah ja (0.83) des war	http://zumult.ids-m	http://zumult.ids	Deutsch
12	FOLK_E_00084_FR	FR	nich genug nährstoffe kriegen und *h	ich weiß es nicht	*h aber grade dieses ja (.) h* ja (0.47)	http://zumult.ids-m	http://zumult.ids	Deutsch
13	FOLK_E_00084_FR	FR	ja kommt drauf an also	ich weiß nicht	die letzten wochenenden war immer	http://zumult.ids-m	http://zumult.ids	Deutsch
14	FOLK_E_00322_MP	MP	gemeint ja ich liebe ähm (.)	ich weiß nicht	mehr wie der typ hieß nennen wir ihr	http://zumult.ids-m	http://zumult.ids	Deutsch
15	FOLK_E_00084_FR	FR	für level zwei sind hardcore (.)	ich weiß nicht	(0.74) wusst auch gar nicht dass es das	http://zumult.ids-m	http://zumult.ids	Deutsch
16	FOLK_E_00403_NH	NH	uni oder h* mh nee ich *h (0.31)	ich weiß aber nicht	so genau was sie hat (.) politik studier	http://zumult.ids-m	http://zumult.ids	Russisch

Abbildung 10
In Excel geöffnete xml-Datei

4.2 Suche im sprecherbasierten Modus

ZuRecht bietet neben der Suche in gesamten Transkripten auch einen sprecherbasierten Suchmodus an. Die Beiträge aller Sprecher:innen werden in diesem Modus ähnlich wie Annotationen auf die Äußerungen anderer Sprecher:innen projiziert. Außerdem werden die Token-Distanz und -Reihenfolge anders als bei der gewöhnlichen transkriptbasierten Suche berechnet, und zwar nur in Bezug auf eine/n einzige/n Sprecher:in (ohne Berücksichtigung von Beiträgen anderer Sprecher:in-

nen). Diese Art der Suche ermöglicht präzisere Recherchen mit Berücksichtigung des Sprecherwechsels bei überlappenden Sprecherbeiträgen³⁵. Möchten Forschende eine bestimmte Wortfolge finden, die von einer einzigen Person geäußert wird (z.B. die Kombination der Responsive *ja genau*), ist der entsprechende Suchmodus in der Suchzeile auszuwählen (vgl. Abbildung 11). Mit der Suchanfrage (11) im sprecherbasierten Modus werden Belege ausgeschlossen, die durch den transkriptbasierten Suchmodus enthalten wären, in denen *ja* und *genau* von zwei unterschiedlichen Personen geäußert werden.

(11) `[norm="ja"][norm="genau"]`

The screenshot shows the 'ZuRecht' search interface. At the top, it says 'CQP-Suchanfragen in den Korpora des AGD'. On the left, there is a 'Korpora' sidebar with a list of corpora: DH, FOLK (checked), GWSS, MEND, DNAM, BETV, and HMAT. The main search area is titled 'Suche mittels der Corpus Query Language (CQP)'. Below the title, it says 'Bitte geben Sie Ihren CQP-Suchausdruck ein und wählen Sie den Suchmodus (Transkript- vs. Sprecher-basiert)'. The search bar contains the query `norm="ja"[norm="genau"]` and a dropdown menu is open, showing 'Sprecher-basierte Suche' selected. Below the search bar, there is a checkbox for 'Satzzeichen ignorieren' which is checked. To the right of the search bar are buttons for 'Beispiele' and 'XML'. Below the search bar, the results section is titled 'Ergebnisse' and shows 'für die Suchanfrage `[norm="ja"][norm="genau"]` (in FOLK)'. There are buttons for 'Treffer gruppieren', 'Metadatenansicht', and 'Download der KWIC'. The results table shows 4 results out of a total of 1282. The first result is '1 FOLK_E_00357_SE_01_T_01 CP irgendwo hin ah ja genau (...) des is (...) des ...' with 'DGD' and 'ZuViel' links. The second result is '2 FOLK_E_00357_SE_01_T_02 CP spielen ja ja genau was also des ...' with 'DGD' and 'ZuViel' links. The third result is '3 FOLK_E_00366_SE_01_T_01 CR gürtel war dazu ja genau das s der ...' with 'DGD' and 'ZuViel' links. The fourth result is '4 FOLK_E_00438_SE_01_T_01 CR ja hmhm (0.35) ja (0.9) *h (...) ja genau und gestern hatten ...' with 'DGD' and 'ZuViel' links.

Abbildung 11
Der sprecherbasierte Suchmodus in *ZuRecht*

ZuRecht bietet im sprecherbasierten Suchmodus Forschenden weiterhin die Möglichkeit, nach Sprecherwechseln sowie nach Token vor oder nach einem Sprecherwechsel zu suchen. Dies erleichtert die Suche nach hochfrequenten Wortformen wie *ja*, *genau* oder *eben*, die in verschiedenen Positionen und Funktionen vorkommen (z.B. auch als Modalpartikel innerhalb eines Sprecherbeitrags), für die aber ganz spezifische Funktionen gefunden werden sollen, etwa die Vorkommnisse als Responsivpartikel. Eine entsprechende Suche kann durch die Angabe des Präzedenzoperators **precededby** und die Angabe, dass zuvor **<another-speaker/>** geredet haben soll, eingegrenzt werden. So lassen sich durch die Suchanfrage (12) gezielt Belege des hochfrequenten *ja* als Responsivpartikel finden. In der transkriptbasierten Suche würde demgegenüber die Suchanfrage nach normalisiertem *ja*, selbst mit der Einschränkung, dass nur Fälle gefunden werden sollen, in denen *ja* als Interjektion oder Responsivpartikel getaggt ist (`[norm="ja" & pos="NGIRR"]`), sehr viele falsche Positive liefern, z.B. Fälle

³⁵ Im transkriptbasierten Suchmodus werden Sprecherbeiträge immer nacheinander indexiert, was dazu führt, dass Sprecherbeiträge auch bei Überlappung nur sequenziell dargestellt werden können (vgl. hierzu und zum sprecherbasierten Modus Frick / Helmer / Schmidt 2022).

von selbstresponsivem oder auto-reflexivem *ja* innerhalb von Sprecherbeiträgen (z.B. „er hat dann irgendwie dann (2.3) ja keinen blick dann“ (FOLK_E_00022_T04_c588)).

(12) `[norm="ja"] precededby <another-speaker/>`

Durch weitere Angaben, die die CQP-Suchsyntax in *ZuRecht* ebenfalls erlaubt, lassen sich zudem weitere Spezifikationen angeben; etwa dahingehend, ob für die gewünschten Belege Überlappungen ausgeschlossen oder im Gegenteil eine Bedingung sein sollen oder/und ob Pausen und andere, paraverbale Token im Kontext zu finden sein sollen. So könnten für eine Untersuchung von *oder?* als *question tag* bestimmte Spezifikationen interessant sein, um gezielt und effizient reaktionseinerfordernde Funktionen zu finden, d.h. Fälle, in denen die Reaktion eines Folgesprechers im Sinne eines *response pursuits* (vgl. Pomerantz 1984) durch das *oder?* elizitiert wurde, nachdem zunächst keine Reaktion erfolgt³⁶. Diese Belege sollen zum einen nach einer Pause (d.h. potenziell nach dem hörbaren Ende einer Äußerungseinheit einer Person), zum anderen nicht in Überlappung mit einem potenziellen Folgesprecher und zudem vor einer Folgereaktion einer anderen Person zu finden sein (vgl. Abbildung 12).

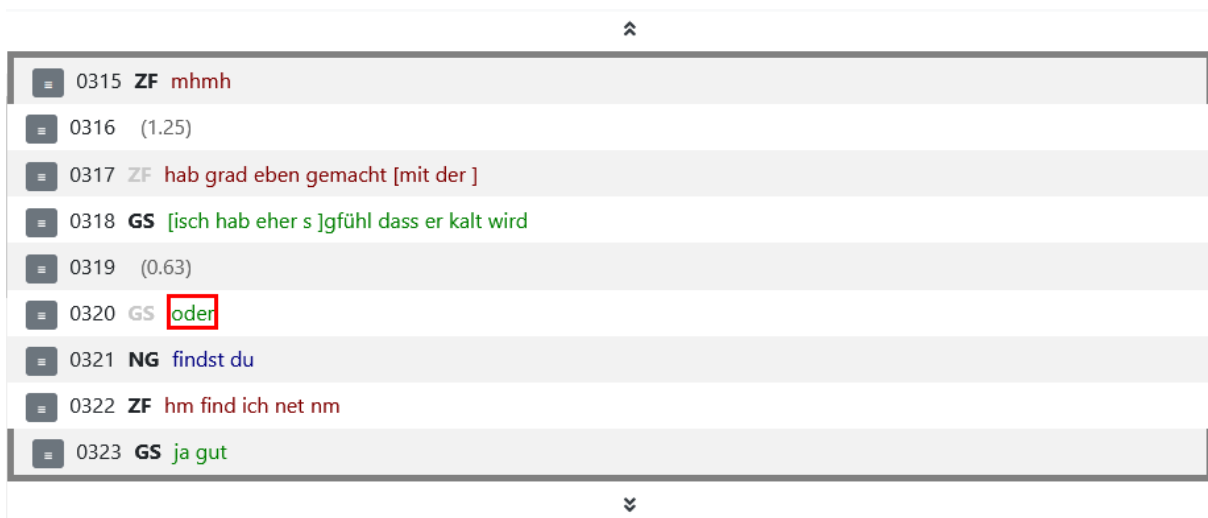


Abbildung 12
Mittels der beschriebenen Suchanfrage im sprecherbasierten Modus gefundener Beleg
(FOLK_E_00293_SE_01_T_02_c320) eines *oder* als *response pursuit*,
hier in der Transkriptdarstellung durch das Tool *ZuViel*

Eine solche Suche könnte im sprecherbasierten Modus durch die Anfrage (13) erfolgen. Diese Anfrage sucht nach Fällen, in denen alle transkribierten Formen von *oder* (`[norm="oder"]`) gefunden werden, die nicht in Überlappung (`!within <speaker-overlap/>`) stehen, denen eine Pause vorausgeht (`precededby <pause.dur/>`)³⁷, und denen ein Beitrag einer anderen Person folgt (`followedby <another-speaker/>`).

³⁶ Vgl. auch König (2017) zur wiederholten Herstellung einer Reaktionsrelevanz von *question tags* wie *ne?*

³⁷ Mit `<pause.dur/>` werden Pausen gesucht, die eine Angabe zur Dauer der Pause enthalten, d.h. keine Mikropausen (die nach GAT2 bzw. cGAT mittels eines Punkts gekennzeichnet sind: (,)). Nutzer:innen finden in der Query-Hilfe in *ZuRecht* weitere Informationen dazu, wie man welche Arten von Pausen suchen kann.

- (13) `(([norm="oder"] !within <speaker-overlap/>) precededby <pause.dur/>) followedby <another-speaker/>`

4.3 Zugriff auf Sprechgeschwindigkeit und Sprecherüberlappungen

Eine weitere neue Suchmöglichkeit in *ZuRecht* bietet sich durch den Zugang zu tokenübergreifenden Annotationen (engl. *span annotations*), auf die vorher über die DGD nicht zugegriffen werden konnte. Die in *ZuRecht* durchsuchbaren Korpora wie *FOLK* enthalten etwa Informationen über die Sprechgeschwindigkeit – genauer gesagt über die *speech rate*, die sich aus der Anzahl der Silben pro Dauer eines Sprecherbeitrags errechnet. So beschreibt Schegloff (2000) u.a. die Varianz von Sprechgeschwindigkeit als ein relevantes Phänomen im Fall von Sprecherwechseln zur Sicherung oder Beibehaltung des Rederechts, insbesondere bei Überlappungen mit anderen Sprechern. Ein Einbezug dieser Informationen ermöglicht Forschenden gezielt nach besonders schnellen (oder besonders langsamen) Beiträgen in Überlappungen zu suchen, vgl. dazu Beispiel (14). Der reguläre Ausdruck `([6-9][1-9][0-9])\.[0-9]+` in diesem Beispiel definiert, dass die Sprechgeschwindigkeit schneller als oder gleich 6.0 sein soll, und findet alle *speechrate*-Annotationen, die die Angaben zwischen 6.0 und 99.9 aufweisen, also Stellen, die ziemlich schnell gesprochen werden. **within <speaker-overlap>** schränkt die Suche ein, indem nur Beiträge innerhalb einer Überlappung gefunden werden.

- (14) `<speech-rate ="([6-9][1-9][0-9])\.[0-9]+"/> within <speaker-overlap/>`

Anschließend könnten die Treffer danach qualitativ ausgewertet werden, ob die potenziell kompetitiven Sprecherwechsel erfolgreich verlaufen oder nicht und ob die Sprechgeschwindigkeit dabei ein relevanter Faktor sein könnte. Auch zu diesem Zweck kann eine weitere Bearbeitung der gefundenen Belege durch das Herunterladen der KWIC-Belege, Aussortieren und Kodieren nach bestimmten Kriterien in Excel erfolgen sowie über eine detaillierte Sequenzanalyse des Transkriptausschnitts mit größerem Kontext mithilfe von *ZuViel*. *ZuViel* bietet zudem die Möglichkeit, die Sprechgeschwindigkeit zu visualisieren (vgl. Schmidt / Schwendemann / Wallner 2023 in dieser Themenausgabe).

Weiterhin lässt sich in *ZuRecht* durch die Funktion „Treffer gruppieren“ herausfinden, welche (normalisierten) Wortformen und Wortverbindungen besonders häufig entsprechend schnell artikuliert werden – im Falle der oben gezeigten Suchanfrage etwa Responsive wie *hmhm*, *ja*, *okay* oder *genau* (vgl. Abbildung 13).

ZuRecht CQP-Suchanfragen in den Korpora des AGD

Korpora: DH, FOLK, GWSS, MEND, DNAM, BETV, HMAT

Suche mittels der CQP-Suche

Suche: `<speech-rate = "[[6-9]][1-9][0-9]]\[0-9]+"/>`

Satzzeichen ignorieren

Ergebnisse
für die Suchanfrage `<speech-rate = "[[6-9]][1-9][0-9]]\[0-9]+"/>` within `<speaker-overlap>` (in FOLK)

Gruppiert nach (Wählen Sie): **Normalisierte Form**

	Insgesamt: 6424	Normalisierte Form	Treffer, Insgesamt: 10236	
1	hmhm &	792	KWIC öffnen	
2	ja	622	KWIC öffnen	
3	hm	236	KWIC öffnen	
4	okay	224	KWIC öffnen	
5	genau	177	KWIC öffnen	

Ergebnisse
für die Suchanfrage `<speech-rate = "[[6-9]][1-9][0-9]]\[0-9]+"/>` within `<speaker-overlap>` (in FOLK)

[Treffer gruppieren](#) [Metadatenansicht](#) [Download der KWIC](#)

Insgesamt: 10236

	Erste Seite	Vorherige Seite	1	2	3	4	5	Nächste Seite	Letzte Seite
1	FOLK_E_00022_SE_01_T_04	SZ	was anderes hmhm also ich mein (0.33) er isch jetzt ...	DGD	ZuViel				
2	FOLK_E_00022_SE_01_T_04	HM	äh reagier oder nee is okay *hh nee des mit ...	DGD	ZuViel				
3	FOLK_E_00022_SE_01_T_04	HM	eins hier weg auf jeden fall wieder da (1.03) ((schnalzt)) (2.55) eh was is ...	DGD	ZuViel				
4	FOLK_E_00022_SE_01_T_04	BS	ja eh eben okay (4.52) okay (0.55) na gut ...	DGD	ZuViel				
5	FOLK_E_00024_SE_01_T_01	AW	nachher einfach her dann bringen s en hh* (0.25) ich hab aufgelegt ...	DGD	ZuViel				

Abbildung 13
Nach normalisierter Form gruppierte Treffer der Suchanfrage

5. Nutzungsmöglichkeiten im Kontext von Deutsch als Fremd- und Zweitsprache

Auch für den Kontext der Sprachvermittlung eröffnet *ZuRecht* vielfältige Optionen zur Nutzung mündlicher Korpora. Diese sind sowohl für die Unterrichtspraxis als auch aus Forschungsperspektive von großem Nutzen. So bietet *ZuRecht* verschiedene Möglichkeiten zur Ermittlung und Analyse vermittlungsrelevanter Phänomene sowie für die Beschreibung kommunikativer Praktiken. Im Folgenden soll dies anhand von Beispielen zur Arbeit mit Wortschatzlisten und zum Zugriff auf sprachliche Handlungen genauer dargestellt werden.

5.1 Arbeit mit Wortschatzlisten

Ein häufiger Einwand von Lehrenden in Bezug auf die Nutzung von Korpusdaten für den Unterricht ist, dass die Suche nach geeigneten Daten viel Zeit kostet. Neben einer schnellen Filterung der einzelnen Interaktionen (im Folgenden auch Sprechereignisse genannt) nach ausgewählten Metadaten (bspw. Gesprächstyp, Region, Dauer), nach schwierigkeitsbezogenen Parametern (bspw. Sprechgeschwindigkeit, Niveaustufenzugehörigkeit des enthaltenen Wortschatzes) sowie nach einem hohen oder auch niedrigen Anteil an Mündlichkeitsphänomenen (wie dies bereits in *ZuMal* möglich ist, siehe Fandrych et al. 2023 in dieser Themenausgabe) ist es für Lehrende ebenfalls wichtig, dass die im Unterricht eingesetzten Sprachdaten möglichst auch thematisch zur Unterrichtseinheit passen. Das Auffinden von Sprechereignissen, die durch einen bestimmten für das Lektionsthema relevanten Wortschatz geprägt sind, erforderte bis jetzt allerdings langwierige Recherchen. Mit der in *ZuRecht* neu geschaffenen Anwendung „Suche mit Lemma- und Suchanfragen“ wird diese Recherche deutlich

vereinfacht. Sie ermöglicht es, gezielt nach Sprechereignissen zu suchen, die einen hohen Anteil des gewünschten Wortschatzes aufweisen. Hierzu stehen einerseits drei fertig vorbereitete Wortschatzlisten zur Verfügung, welche die Lexik der Wortschatzbereiche „Essen“, „Haus und Wohnung“ und „Schule und Ausbildung“ umfassen und auf dem Übungswortschatz „Sage und Schreibe“ (vgl. Fandrych / Tallowitz 2019) basieren. Andererseits können aber auch eigene Listen genutzt werden, die den gewünschten Wortschatz enthalten. Diese Listen können von den Lehrenden selbst erstellt werden, wobei bspw. der Wortschatz eines beliebigen Themenfeldes oder auch einer bestimmten Lektion im Lehrwerk in die Liste aufgenommen werden kann. Für die Liste kann eine einfache Textdatei (.txt) genutzt werden. Es besteht zum einen die Möglichkeit, eine reine Lemmaliste zu erstellen, in der die Grundformen der gewünschten Wörter untereinander aufgeführt werden (vgl. Abbildung 14). Zum anderen ist es aber auch möglich, eine Liste von CQP-Suchanfragen zu erstellen. Auf diese Weise können in die Suche nach Sprechereignissen neben Wortschatz in der Grundform auch aussprachenahne Realisierungen von Wortschatzeinheiten (wie bspw. *nee* statt *nein*), Flexionsformen und Wortsequenzen einbezogen werden. Zudem ist es möglich, annotierte Informationen wie z.B. die Wortart, sprachliche Handlungen (bspw. Begrüßungen und Verabschiedungen) sowie Metadaten zu berücksichtigen.

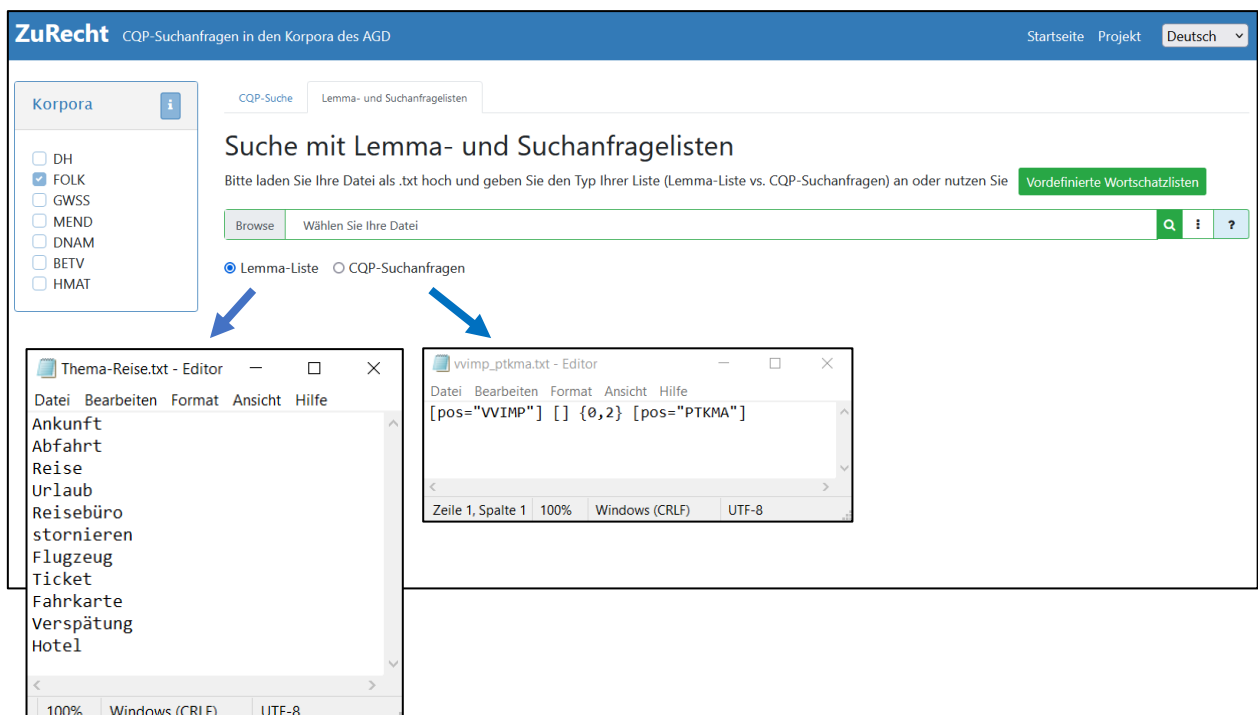


Abbildung 14
Suche mit Lemma- und Suchanfragelisten in *ZuRecht*

Hierzu sei nun ein Beispiel dargestellt: Bereits im Anfängerunterricht werden naturgemäß Aufforderungen thematisiert – nicht zuletzt auch um die Unterrichtskommunikation zu gewährleisten (*Öffnet doch bitte mal das Arbeitsbuch! Nutzt ruhig das Wörterbuch!*). Um Aufforderungen höflicher zu gestalten, werden im Deutschen häufig Modalpartikeln genutzt (vgl. Fandrych / Thurmair 2021: 294). Um diese Praktik anhand mündlicher Kommunikation zu veranschaulichen, kann es für Lehrende nützlich sein, gezielt Sprechereignisse mit einem hohen Anteil an solchen Beispielen zu finden. Mit der Suchanfrage (15) werden Vorkommen von Verben im Imperativ gefunden, auf die im Abstand von maximal zwei Token eine Modalpartikel folgt. Wird diese Suchanfrage in einer txt-Datei abgespeichert, kann sie unter dem Reiter „Suche mit Lemma- und Suchanfragelisten“ hochgeladen und

zur Ermittlung von Sprechereignissen genutzt werden, die möglichst viele solcher Sequenzen enthalten. Abbildung 14 veranschaulicht die Suche mit eigenen Wortschatzlisten am Beispiel einer Lemmaliste zum Thema „Reise“ und der hier erläuterten CQP-Suchanfrage.

(15) `[pos="VVIMP"] [] {0,2} [pos="PTKMA"]`

Die Suche mit Hilfe der Listen ergibt eine Übersicht über Sprechereignisse, die den gesuchten Wortschatz bzw. die mit der Suchanfrage vorgegebenen Annotationen und Eigenschaften enthalten. Die Ergebnisse lassen sich dabei nach absoluter und relativer Anzahl der Treffer sowie nach der Anzahl der Types sortieren. Abbildung 15 zeigt einen Ausschnitt aus der Ergebnisansicht zur oben dargestellten Suchanfrage der Form „Verb im Imperativ + Modalpartikel“. Besonders viele Vorkommen dieser Art enthält ein „Mädelsabend“ (FOLK_E_00287_SE_01) mit insgesamt 63 Treffern, was einem Anteil von 0,61% aller Token dieses Sprechereignisses entspricht (Stand: September 2022). Ausgehend von der Ergebnisübersicht ist es auch möglich, über den Link „Lemma-Ansicht öffnen“ (rot umrandet in Abbildung 15) diejenigen Lemmata aus der Liste in einem gesonderten Tab aufzurufen, die im jeweiligen Sprechereignis vorkommen. Für das oben genannte Beispiel wären dies die im Imperativ geäußerten Verben in Kombination mit den danach geäußerten Modalpartikeln. Wahlweise ist hier auch die Anzeige der transkribierten oder auch der normalisierten Formen möglich (vgl. Abbildung 15, Fenster „Ergebnisse“, Auswahlelement „Gruppiert nach“). Lehrende können sich damit schnell einen Überblick über die angezeigten Sprechereignisse verschaffen und prüfen, welche davon für ihren Unterricht geeignet sind.

ZuRecht CQP-Suchanfragen in den Korpora des AGD

Korpora: DH, FOLK, GWSS, MEND, DNAM, BETV, HMAT

CQP-Suche: Lemma- und Suchanfragelisten

Suche mit Lemma- und Suchanfragelisten

Bitte laden Sie Ihre Datei als .txt hoch und geben Sie den Typ an

Browse: C:\fakepath\vwimp_ptkma.txt

Lemma-Liste CQP-Suchanfragen

Ergebnisse

für die Suchanfrage `((pos="VVIMP") [] {0,2} [pos="PTKMA"]) within <t_dgd_kennung="FOLK_E_00287_SE_01_T_03"/>` (in FOLK)

Gruppiert nach (Wählen Sie): **Normalisierte Form**

Sortiert nach (Wählen Sie): **Treffer (abs.) absteigend**

Insgesamt: 26 Normalisierte Form Treffer, Insgesamt: 63

Id	Normalisierte Form	Treffer	Aktionen
1	guck mal	18	KWIC öffnen
2	zeig mal	13	KWIC öffnen

Ergebnisse

für die Wortschatzsuche aus der Datei C:\fakepath\vwimp_ptkma.txt (in FOLK)

Insgesamt: 649	Transkript ID	Token (Insgesamt)	Kurzbezeichnung ("Art")	Treffer	Types	Aktionen
1	FOLK_E_00287_SE_01_T_03	10311	Mädelsabend	63 (0.61%)	25	Lemma-Ansicht öffnen Ergebnisse als KWIC öffnen Transkript öffnen
2	FOLK_E_00217_SE_01_T_01	5869	Gespräch beim Renovieren	60 (1.02%)	27	Lemma-Ansicht öffnen Ergebnisse als KWIC öffnen Transkript öffnen
3	FOLK_E_00218_SE_01_T_01	5287	Gespräch beim Renovieren	45 (0.85%)	23	Lemma-Ansicht öffnen Ergebnisse als KWIC öffnen Transkript öffnen
4	FOLK_E_00217_SE_01_T_02	6101	Gespräch beim Renovieren	43 (0.70%)	30	Lemma-Ansicht öffnen Ergebnisse als KWIC öffnen Transkript öffnen

Abbildung 15

Ergebnisansicht nach der Suche von ‚Verb im Imperativ + Modalpartikel‘ über die Lemma- und Suchanfragelisten

Darüber hinaus besteht die Möglichkeit, ausgehend von der Ergebnisübersicht die Konkordanzen sowie die Transkripte im Transkriptbrowser *ZuViel* aufzurufen. Anhand der Transkripte lässt sich dann noch besser die Eignung des Sprechereignisses für den eigenen Unterricht beurteilen. Die Lemmata

aus der Liste werden in *ZuViel* rot umrandet. Für die Nutzung im Unterricht können im Transkriptbrowser noch weitere Markierungen (darunter Wortarten und Niveaustufenzugehörigkeit der einzelnen Token) vorgenommen werden (vgl. Schmidt / Schwendemann / Wallner 2023 in dieser Themenausgabe).

Die Option „Suche mit Lemma- und Suchanfragelisten“ ist darüber hinaus auch aus Perspektive der Forschung im Bereich DaF/DaZ sowie für lexikologische und lexikographische Untersuchungen interessant, da sie umfassende Abgleiche von Lemma- und Suchanfragelisten mit den Transkripten ermöglicht. Auf diese Weise ist es etwa möglich zu prüfen, ob bzw. inwieweit ein bestimmter Wortschatz oder die jeweils ausgewählten Phänomene besonders häufig in bestimmten Interaktionsdomänen oder auch Gesprächsarten vorkommen. So kann bspw. untersucht werden, in welchen Domänen und Gesprächsarten die Verben der fachübergreifenden Lexik der Geisteswissenschaften (GeSiG-Inventar³⁸) im *GeWiss*-Korpus besonders häufig sind und um welche Verben es sich dabei im Einzelnen handelt.

5.2 Zugriff auf annotierte sprachliche Handlungen

Sprachliche Handlungen sind ein wichtiger Gegenstand der Sprachvermittlung. Sie ermöglichen es den Lernenden, ihre kommunikativen Absichten erfolgreich umzusetzen. Lehrmaterialien bieten zu zentralen sprachlichen Handlungen sogenannte Redemittelsammlungen. Diese basieren jedoch häufig nicht auf empirischen Untersuchungen zum tatsächlichen Sprachgebrauch, sondern beruhen vielmehr auf dem individuellen Sprachgefühl der Lehrmittelautor:innen. Die Auswahl und das bereitgestellte Repertoire sind entsprechend begrenzt. Korpora eröffnen die Möglichkeit, die typischen Realisierungen von sprachlichen Handlungen im realen Kontext zu identifizieren und durch sogenannte pragmatische Annotationen für Vermittlungszwecke zugänglich zu machen (vgl. hierzu auch Kaiser 2023 in dieser Themenausgabe). Die Annotation von sprachlichen Handlungen ist jedoch ein sehr zeitaufwendiger Prozess. Das erklärt, warum derartige Aufbereitungen in Korpora nur in begrenztem Umfang zur Verfügung stehen. Für mündliche Korpusdaten beschränkt sich das Angebot bislang auf die in *FOLK* annotierten Handlungssequenzen³⁹ (vgl. Kaiser / Schedl 2021 und Kaiser 2023 in dieser Themenausgabe) und die im *GeWiss*-Korpus vorgenommenen Annotationen von Diskurskommentierungen⁴⁰ (vgl. Fandrych 2014; Meißner 2017) sowie von Verweisen und Zitaten⁴¹ (vgl. Sadowski 2017). Ein weiteres Problem war bislang auch die Zugänglichkeit dieser Annotationen. So können – wie eingangs erwähnt – wegen technischer Einschränkungen weder die in *FOLK* annotierten Handlungssequenzen noch die für *GeWiss* vorliegenden Annotationen von Diskurskommentierungen und

³⁸ Das GeSIG-Inventar ist eine nach Wortarten filterbare Wortschatzliste, die diejenige Lexik umfasst, die in den Geisteswissenschaften fachübergreifend verwendet wird. Die Liste ist unter <https://www.esv.info/t/gesig/aktualisierung.html> (07.10.2022) frei verfügbar. Ausführliche Informationen zum GeSIG-Inventar finden sich in Meißner / Wallner (2019).

³⁹ Darunter Begrüßungs- und Verabschiedungsformen in Eröffnungs- und Beendigungssequenzen sowie Modalverbformate zur Realisierung bestimmter sozialer Handlungen wie u.a. Aufforderungen, Bitten und Vorschläge (vgl. Kaiser 2023 in dieser Themenausgabe).

⁴⁰ Bei Diskurskommentierungen handelt es sich um einen der Gliederung und Rezipient:innenorientierung dienenden Handlungstyp, der für wissenschaftliche Vorträge beschrieben (vgl. Fandrych 2014) und in den deutschsprachigen L1-Expertenvorträgen des *GeWiss*-Korpus annotiert wurde (vgl. Baur et al. 2014).

⁴¹ Zitate und Verweise umfassen mündliche Bezugnahmen auf andere Forschungsarbeiten und wurden in den deutschsprachigen L1-Expertenvorträgen und in ausgewählten studentischen Vorträgen des *GeWiss*-Korpus annotiert (vgl. Maier / Sadowski / Schmidt 2015; Sadowski 2017).

Verweisen und Zitaten über die DGD abgefragt werden⁴². Mit *ZuRecht* wurde nunmehr eine Möglichkeit geschaffen, pragmatische Annotationen mit Hilfe der CQP-Suchanfragesprache in den Korpusdaten abzurufen. Hierzu ein Beispiel: Um etwa im Kontext von studienvorbereitenden und -begleitenden Deutschkursen im *GeWiss*-Korpus nach Beispielen für mündlich realisierte Verweise und Zitate zu suchen, können Lehrende und Lernende die Suchanfrage (16) eingeben. Die Suche ergibt 273 Treffer (Stand: September 2022). Die Suchanfrage lässt sich aber auch noch spezifizieren. So ist es bspw. möglich, mit der Suchanfrage (17) die Suche nach wörtlichen Zitaten einzugrenzen. Zudem besteht die Möglichkeit, Metadaten bei der Suche zu berücksichtigen. So kann z.B. mit der Suchanfrage (18) die Suche auf studentische Vorträge eingegrenzt werden. Für internationale Studierende, die einen Vortrag in der Fremdsprache Deutsch vorbereiten, können die so gewonnenen Belege eine sinnvolle Orientierung für die Versprachlichung eigener Bezugnahmen auf Forschungsarbeiten in Vorträgen darstellen. Darüber hinaus ist es möglich, die Suche nach annotierten sprachlichen Handlungen auch mit der Suche nach bestimmten Wortartkategorien zu kombinieren. So lassen sich mit der Suchanfrage (19) alle Verweise und Zitate ermitteln, die einen Eigennamen enthalten. Verweise auf Studien oder Publikationen, bei denen keine Autorin bzw. kein Autor genannt werden, würden hier dann nicht in der Treffermenge erscheinen.

(16) `<VZ/>`

(17) `<VZ="Z_woert1"/>`

(18) `<VZ/> within <e_se_art="Studentischer Vortrag"/>`

(19) `<VZ/> containing [pos="NE"]`

Eine wichtige Vorarbeit für die Vermittlung der Zitat- und Verweispraxis in der gesprochenen Wissenschaftssprache ist deren systematische Erforschung. Hierzu gehört etwa die Erfassung wiederkehrender sprachlicher Mittel innerhalb von mündlichen Zitaten und Verweisen. *ZuRecht* bietet die Möglichkeit, diese wiederkehrenden sprachlichen Mittel gezielt zu erfassen. Um bspw. zu ermitteln, welche Vollverben innerhalb von Verweisen und Zitaten üblicherweise vorkommen, eignet sich Suchanfrage (20).

(20) `[pos="V.+"] within <VZ/>`

Während in der Konkordanz der Gebrauch der Verben in den Verweisen und Zitaten veranschaulicht wird, führt der Button „Treffer gruppieren“ zu einer quantitativen Auswertung der Belege. Mittels der Sortierung nach der Kategorie „Lemma“ lässt sich schließlich anzeigen, welche Verben wie häufig in den Belegen vorkommen. Deutlich wird dabei, dass *sagen*, *geben* und *schreiben* am häufigsten innerhalb der im *GeWiss*-Korpus annotierten Verweise und Zitate vorkommen (vgl. Abbildung 16).

⁴² Die *GeWiss*-Annotationen konnten bislang lediglich über das ursprünglich im Projekt entwickelte *GeWiss*-Portal (<https://gewiss.uni-leipzig.de> (07.10.2022)), das allerdings nicht mehr weiterentwickelt wird und in seinen Suchmöglichkeiten begrenzt ist, abgerufen werden, nicht aber über die Oberfläche der DGD, die das Korpus auch langfristig der Öffentlichkeit zugänglich macht.

ZuRecht CQP-Suchanfragen in den Korpora des AGD

Korpora: DH, FOLK, GWSS, MEND, DNAM, BETV, HMAT

Suche mittels der Corpus Query

Bitte geben Sie Ihren CQP-Suchausdruck ein und wählen Sie die Korpora aus

Suche: [pos="VV.+"] within <VZ/>

Satzzeichen ignorieren

Ergebnisse für die Suchanfrage [pos="VV.+"] within <VZ/> (in GWSS)

Treffer gruppieren Metadatenansicht Download der KWIC

Insgesamt: 129	Lemma	Treffer, Insgesamt: 266	
1	sagen	20	KWIC öffnen
2	geben	17	KWIC öffnen
3	schreiben	10	KWIC öffnen
4	nennen	9	KWIC öffnen
5	kommen	9	KWIC öffnen

Beispiele [XML](#)

Insgesamt: 266

Erste Seite Vorherige Seite 7 8 9 10 11 Nächste Seite Letzte Seite

81	GWSS_E_00028_SE_01_T_01	IT_0235 ... argumentiert kommunikative kompetenz gibt (0.33) und dazu auch ...	DGD	ZuViel
82	GWSS_E_00028_SE_01_T_01	IT_0235 ... auf n punkt bringt von 'h aus der ...	DGD	ZuViel
83	GWSS_E_00028_SE_01_T_01	IT_0235 ... spielstruktur des textes schreibt sie beziehungsweise des ...	DGD	ZuViel
84	GWSS_E_00028_SE_01_T_01	IT_0235 ... in sandigs textstilistik findet sich das ähm ...	DGD	ZuViel
85	GWSS_E_00028_SE_01_T_01	IT_0235 ... eben barbara sandig fordert stilanalysen (.) dass stilanalysen ...	DGD	ZuViel
86	GWSS_E_00028_SE_01_T_01	IT_0235 ... wie coupland 'h das nennt verdeutlichen sehr schön ...	DGD	ZuViel

Abbildung 16

Beispiel für die Suche nach Vollverben innerhalb von Verweisen und Zitaten im *GeWiss*-Korpus

6. Fazit und Ausblick

Mit *ZuRecht* wurde eine Korpusrechercheplattform geschaffen, welche einen umfassenden Zugriff auf mündliche Korpusdaten ermöglicht und dabei einen Großteil der mit der DGD verbundenen Einschränkungen überwindet. *ZuRecht* zeigt sich somit als eine hilfreiche Erweiterung und teilweise unverzichtbare Weiterentwicklung der Möglichkeiten, die die DGD bietet. Dies wurde hier anhand ausgewählter Suchanfragen aus der Gesprächsforschung und der Interaktionalen Linguistik sowie aus dem Kontext Deutsch als Fremd- und Zweitsprache demonstriert. Dabei ist deutlich geworden, dass mithilfe von *ZuRecht* sowohl linguistisch als auch didaktisch motivierte Fragestellungen bearbeitet werden können. Darüber hinaus ist auch eine Nutzung im Kontext von Sprachvermittlung und -aneignung möglich. Die in *ZuRecht* genutzte Suchanfragesprache gestattet sowohl einfache als auch hoch komplexe Suchanfragen und ist damit von Nutzer:innen mit unterschiedlicher korpuslinguistischer Expertise gleichermaßen nutzbar. Zudem erlaubt die Suchanfragesprache für routinierte Nutzer:innen deutlich effizientere Recherchen, als bisherige Korpusrecherchertools zur Analyse gesprochener deutscher Sprache, bei denen komplexe Suchanfragen in der Regel mehrschrittige Prozesse erfordern. Mit dem Query Builder und einem umfassenden Hilfedokument werden für die Nutzer:innen vielfältige Unterstützungsmöglichkeiten für die Generierung von Suchanfragen bereitgestellt.

Aktuell handelt es sich bei *ZuRecht* um einen Prototyp, mit dem wir demonstrieren konnten, welche Möglichkeiten es gibt, die mit der DGD verbundenen Einschränkungen zu kompensieren. Nach einer erfolgreichen Erprobung der implementierten Funktionalitäten durch Nutzer:innen ist perspektivisch eine dauerhafte Ausstattung der DGD mit diesen Funktionalitäten denkbar.

Literatur und Ressourcen

- Auer, Peter / Günthner, Susanne (2004): Die Entstehung von Diskursmarkern im Deutschen – ein Fall von Grammatikalisierung? In: Leuschner, Torsten / Mortelsmans, Tanja (Hrsg.) *Grammatikalisierung im Deutschen*. Berlin: de Gruyter, 335-362.
- Barth-Weingarten, Dagmar (2011): Double sayings of German JA – more observations on their prosodic-phonetic make-up and alignment function. In: *Research on Language and Social Interaction* 44: 2, 157-185.
- Batinić, Josip / Frick, Elena / Schmidt, Thomas (2021): Accessing spoken language corpora: an overview of current approaches. In: *Corpora* 16: 3. <https://www.eupublishing.com/doi/10.3366/cor.2021.0229> (07.10.2022).
- Baur, Benedikt / Gräfe, Karen / Lange, Daisy / Schmidt, Julia (2014): Dokumentation zur Annotation der Diskurskommentierungen. https://gewiss.uni-leipzig.de/fileadmin/documents/Annotationsdokumentation_GeWiss.pdf (07.10.2022).
- Bergmann, Pia (2017): Gebrauchsprofile von *weiß nich* und *keine Ahnung* im Gespräch – Ein Blick auf nicht-responsive Vorkommen. In: Blühdorn, Hardarik / Deppermann, Arnulf / Helmer, Henrike / Spranz-Fogasy, Thomas (Hrsg.): *Diskursmarker im Deutschen. Reflexionen und Analysen*. Göttingen: Verlag für Gesprächsforschung, 157-182.
- Brouwer, Matthijs / Brugman, Hennie / Kemps-Snijders, Marc (2016): MTAS: A Solr/Lucene based Multi-Tier Annotation Search solution. In: *Selected papers from the CLARIN Annual Conference 2016, Aix-en-Provence* 136: 2, 19-37. <https://ep.liu.se/ecp/136/002/ecp17136002.pdf> (07.10.2022).
- Deppermann, Arnulf / Schmidt, Thomas (2014): Gesprächsdatenbanken als methodisches Instrument der Interaktionalen Linguistik - Eine exemplarische Untersuchung auf Basis des Korpus FOLK in der Datenbank für Gesprochenes Deutsch (DGD2). In: Domke, Christine / Gansel, Christa (Hrsg.): *Korpora in der Linguistik - Perspektiven und Positionen zu Daten und Datenerhebung*. Mitteilungen des Deutschen Germanistenverbandes 61: 1, 4-17.
- Deppermann, Arnulf / Gubina, Alexandra (2021): Positionally-sensitive action-ascription. Uses of *Kannst du X?* 'can you X?' in their sequential and multimodal context. In: *Interactional Linguistics* 1: 2, 183-215.
- Fandrych, Christian (2014): Metakomentierungen in wissenschaftlichen Vorträgen. In: Fandrych, Christian / Meißner, Cordula / Slavcheva, Adriana (Hrsg.): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron, 95-111.
- Fandrych, Christian / Frick, Elena / Hedeland, Hanna / Iliash, Anna / Jettka, Daniel / Meißner, Cordula / Schmidt, Thomas / Wallner, Franziska / Weigert, Kathrin / Westpfahl, Swantje (2016): User, who art thou? User Profiling for Oral Corpus Platforms. In: Calzolari, Nicoletta / Choukri, Khalid / Declerck, Thierry / Goggi, Sara / Grobelnik, Marko (Hrsg.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia. Paris: European Language Resources Association (ELRA), 280-287.
- Fandrych, Christian / Meißner, Cordula / Schwendemann, Matthias / Wallner, Franziska (2023): ZuMal: Zielgruppenspezifische Gesprächsauswahl aus Korpora gesprochener Sprache. In: *Korpora Deutsch als Fremdsprache* 3: 1, 13-43.
- Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (2021): Korpora gesprochener Sprache und Deutsch als Fremd- und Zweitsprache: Eine chancenreiche Beziehung. In: *Korpora Deutsch als Fremdsprache* 1: 2, 5-30.
- Fandrych, Christian / Tallowitz, Ulrike (2019): *Sage und Schreibe: Übungswortschatz Grundstufe A1–B1 mit Lösungen. Neubearbeitung mit Audio-CD*. Stuttgart: Ernst Klett.

- Fandrych, Christian / Thurmair, Maria (2021): *Grammatik im Fach Deutsch als Fremd- und Zweitsprache. Grundlagen und Vermittlung*. 2., neu bearbeitete und erweiterte Auflage. Berlin: Erich Schmidt Verlag.
- Fandrych, Christian / Wallner, Franziska (2023): Das GeWiss-Korpus: Neue Forschungs- und Vermittlungsperspektiven zur mündlichen Hochschulkommunikation. In: Deppermann, Arnulf / Fandrych, Christian / Kupietz, Marc / Schmidt, Thomas (Hrsg.): *Korpora in der germanistischen Sprachwissenschaft: Mündlich, schriftlich, multimedial*. Berlin: De Gruyter, 129-160.
- Frick, Elena / Helmer, Henrike / Schmidt, Thomas (2022): Querying Interaction Structure: Approaches to Overlap in Spoken Language Corpora. In: Calzolari, Nicoletta / Béchet, Frédéric / Blache, Philippe / Choukri, Khalid / Cieri, Christopher / Declerck, Thierry / Goggi, Sara / Isahara, Hitoshi / Maegaard, Bente / Mariani, Joseph / Mazo, Hélène / Odijk, Jan / Piperidis, Stelios (Hrsg.): *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*. Marseille, Frankreich. Paris: European Language Resources Association (ELRA), 715–722. <http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.75.pdf> (07.10.2022).
- Helmer, Henrike / Deppermann, Arnulf / Reineke, Silke (2017): Antwort, epistemischer Marker oder Widerspruch? Sequenzielle, semantische und pragmatische Eigenschaften von *ich weiß nicht*. In: Deppermann, Arnulf / Proske, Nadine / Zeschel, Arne (Hrsg.): *Verben im interaktiven Kontext. Bewegungsverben und mentale Verben im gesprochenen Deutsch*. Tübingen: Narr, 377-405.
- ISO 24624:2016. Language resource management — Transcription of spoken language.
- Kaiser, Julia (2023): ZuHand: Zugang zu Handlungssequenzen und handlungsbezogenen Themenausschnitten in einem qualitativ annotierten FOLK-Subkorpus. In: *Korpora Deutsch als Fremdsprache* 3: 1, 92-111.
- Kaiser, Julia / Schedl, Evi (2021): Das Forschungs- und Lehrkorpus Gesprochenes Deutsch als Ressource für den handlungsorientierten DaF-Unterricht – Potentiale und Herausforderungen. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 26: 1, 45-83. <https://d-nb.info/1231275278/34> (07.10.2022).
- König, Katharina (2017): *Question tags* als Diskursmarker? – Ansätze zu einer systematischen Beschreibung von *ne* im gesprochenen Deutsch. In: Hardarik Blühdorn, Arnulf Deppermann, Henrike Helmer, Thomas Spranz-Fogasy (Hrsg.): *Diskursmarker im Deutschen. Reflexionen und Analysen*. Mannheim: Verlag für Gesprächsforschung, 233-258.
- Kupietz, Marc / Schmidt, Thomas (2015): Schriftliche und mündliche Korpora am IDS als Grundlage für die empirische Forschung. In: Eichinger, Ludwig M. (Hrsg.): *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*. (Jahrbuch des Instituts für Deutsche Sprache 2014). Berlin / Boston: de Gruyter, 297-322.
- Maier, Elisabeth / Sadowski, Sabrina / Schmidt, Julia (2015): Annotation der Verweise und Zitate. https://gewiss.uni-leipzig.de/fileadmin/documents/VZ_Dokumentation.pdf (07.10.2022).
- Meißner, Cordula (2017): Gute Kandidaten. Ein Ansatz zur automatischen Ermittlung von Belegen für sprachliche Handlungen auf der Basis manueller pragmatischer Annotation. In: Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (Hrsg.): *Gesprochene Wissenschaftssprache – digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg, 165-213.
- Meißner, Cordula / Wallner, Franziska (2019): *Das gemeinsame sprachliche Inventar der Geisteswissenschaften. Lexikalische Grundlagen für die wissenschaftspropädeutische Sprachvermittlung*. Berlin: Erich Schmidt Verlag (Studien Deutsch als Fremd- und Zweitsprache 6).
- Moroni, Manuela Caterina (2021): Zum Einsatz von Daten aus FOLK und DeReKo für die Untersuchung und Didaktisierung der Formen *nach dem Motto* und *wegen*. In: *Korpora Deutsch als Fremdsprache* 1: 1, 51-78.
- Pomerantz, Anita (1984): Pursuing a response. In: Atkinson J. Maxwell / Heritage, John (Hrsg.): *Structures of social action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, 152-164.

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 (Hrsg.): *Korpora in der germanistischen Sprachwissenschaft*. Berlin / Boston: de Gruyter, 71-102.

Sadowski, Sabrina (2017): Die Annotation von Zitaten und Verweisen im GeWiss-Korpus. In: Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (Hrsg.): *Gesprochene Wissenschaftssprache – digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg, 147-166.

Schegloff, Emanuel A. (2000): Overlapping Talk and the Organization of Turn-Taking for Conversation. In: *Language in Society* 29: 1, 1-63.

Schmidt, Thomas (2017): DGD – die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. In: *Zeitschrift für germanistische Linguistik* 45: 3. Berlin / Boston: de Gruyter, 451-463.

Schmidt, Thomas (2023): FOLK. Das Forschungs- und Lehrkorpus für Gesprochenes Deutsch. In: *Korpora Deutsch als Fremdsprache* 3: 1, 166-169.

Schmidt, Thomas / Fandrych, Christian / Frick, Elena / Schwendemann, Matthias / Wallner, Franziska / Wörner, Kai (2023): Zugänge zu mündlichen Korpora für DaF und DaZ. Projekt, Datengrundlagen, technische Basis. In: *Korpora Deutsch als Fremdsprache* 3: 1, 1-12.

Schmidt, Thomas / Schwendemann, Matthias / Wallner, Franziska (2023): ZuViel: Transkriptvisualisierung und Arbeiten mit Transkripten. In: *Korpora Deutsch als Fremdsprache* 3: 1, 72-91.

Selting, Margret / Couper-Kuhlen (2000): Argumente für die Entwicklung einer ‚interaktionalen Linguistik‘. In: *Gesprächsforschung Online* 1, 76-95.

Stift, Ulf-Michael / Schmidt, Thomas (2014): Mündliche Korpora am IDS: Vom Deutschen Spracharchiv zur Datenbank für Gesprochenes Deutsch. In: Institut für Deutsche Sprache (Hrsg.): *Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*. Mannheim: Institut für Deutsche Sprache, 360-375.

Torres Cajo, Sarah (2017): „das is SO lächerlich; ohne SCHEISS jetz ma“ – Zur affektiven Äußerungsmodalisierung durch *ohne Scheiß*-Konstruktionen im gesprochenen Deutsch. In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 18. Jg. (= Gesprächsforschung 2017). Mannheim: Verlag für Gesprächsforschung, 223-240.

Wallner, Franziska (2023): GeWiss – ein Korpus der gesprochenen Wissenschaftssprache. In: *Korpora Deutsch als Fremdsprache* 3: 1, 159-165.

Biographische Notiz: Elena Frick ist wissenschaftliche Mitarbeiterin der Abteilung „Pragmatik“ am Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim. Sie ist im Programmbereich „Mündliche Korpora“ tätig und beschäftigt sich mit der Entwicklung digitaler Korpusanwendungen für sprachwissenschaftliche Forschung. Sie war an der Entwicklung der Korpusanalyseplattform KorAP beteiligt und ist aktuell für das Design und die Implementierung von *ZuRecht* zuständig.

Kontaktanschrift:

Elena Frick
Leibniz-Institut für Deutsche Sprache
R5, 6-13
D-68161 Mannheim
Deutschland
frick@ids-mannheim.de

Biographische Notiz: Henrike Helmer ist wissenschaftliche Mitarbeiterin der Abteilung „Pragmatik“ am Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim. Sie leitet den Programmbereich „Mündliche Korpora“ und interessiert sich im Rahmen ihrer Forschung in verschiedenen Projekten für die Grammatik des gesprochenen Deutsch, Interaktionale Semantik sowie quantitative und qualitative Datenanalysemethoden.

Kontaktanschrift:

Henrike Helmer
Leibniz-Institut für Deutsche Sprache
R5, 6-13
D-68161 Mannheim
Deutschland
helmer@ids-mannheim.de

Biographische Notiz: Franziska Wallner ist wissenschaftliche Mitarbeiterin am Herder-Institut der Universität Leipzig. Ihre Forschungsschwerpunkte sind u.a. das Deutsche als fremde Bildungs- und Wissenschaftssprache, die korpusbasierte Erforschung der gesprochenen Sprache, Mündlichkeitsdidaktik sowie die Nutzung von Korpora im Kontext von Deutsch als Fremd- und Zweitsprache. Sie ist Mitglied der Redaktion der Zeitschrift *Deutsch als Fremdsprache*.

Kontaktanschrift:

Franziska Wallner
Herder-Institut
Universität Leipzig
Beethovenstr. 15
04107 Leipzig
Deutschland
f.wallner@uni-leipzig.de

