

Der LeGeDe-Prototyp: Zur Erstellung eines korpusbasierten Online-Wörterbuchs zu lexikalischen Besonderheiten im gesprochenen Deutsch. Aktueller Stand und Perspektiven

Meike Meliss, *Universidad de Santiago de Compostela,
Spanien (meike.meliss@usc.es)*

und

Christine Möhrs, *Leibniz-Institut für Deutsche Sprache, Mannheim,
Deutschland (moehrs@ids-mannheim.de)*

Zusammenfassung: Im Beitrag steht das LeGeDe-Drittmittelprojekt und der im Laufe der Projektzeit entwickelte korpusbasierte lexikografische Prototyp zu Besonderheiten des gesprochenen Deutsch in der Interaktion im Zentrum der Betrachtung. Die Entwicklung einer lexikografischen Ressource dieser Art knüpft an die vielfältigen Erfahrungen in der Erstellung von korpusbasierten Onlinewörterbüchern (insbesondere am Leibniz-Institut für Deutsche Sprache, Mannheim) und an aktuelle Methoden der korpusbasierten Lexikologie sowie der Interaktionsanalyse an und nimmt als multimedialer Prototyp für die korpusbasierte lexikografische Behandlung von gesprochensprachlichen Phänomenen eine innovative Position in der modernen Onlinelexikografie ein. Der Beitrag befasst sich im Abschnitt zur LeGeDe-Projektpräsentation ausführlich mit projektrelevanten Forschungsfragen, Projektzielen, der empirischen Datengrundlage und empirisch erhobenen Erwartungshaltungen an eine Ressource zum gesprochenen Deutsch. Die Darstellung der komplexen Struktur des LeGeDe-Prototyps wird mit zahlreichen Beispielen illustriert. In Verbindung mit der zentralen Information zur Makro- und Mikrostruktur und den lexikografischen Umtexten werden die vielfältigen Vernetzungs- und Zugriffsstrukturen aufgezeigt. Ergänzend zum abschließenden Fazit liefert der Beitrag in einem Ausblick umfangreiche Vorschläge für die zukünftige lexikografische Arbeit mit gesprochensprachlichen Korpusdaten.

Stichwörter: INTERAKTIONSLINGUISTIK, KORPUSLINGUISTIK, LEXIK DES GESPROCHENEN, MAKROSTRUKTUR, MIKROSTRUKTUR, MULTIMEDIALITÄT, ONLINELEXIKOGRAFIE, UMTEXTE, VERNETZUNGSSTRUKTUREN

Abstract: The LeGeDe-prototype: The Creation of an Online Dictionary on Lexical Particularities in Spoken German. Current Status and Perspectives.

The article focuses on the LeGeDe third-party funded project and the corpus-based lexicographic prototype on the particularities of spoken German in interaction developed during the project period. The development of a lexicographic resource of this kind builds on the various experiences in the creation of corpus-based online dictionaries (especially at the Leibniz Institute for the German Language, Mannheim) and on methods of corpus-based lexicology as well as interaction analysis. As a multimedia prototype for the corpus-based lexicographic treatment of spoken language phenomena, it occupies an innovative position in the landscape of online lexicography. In the section on the presentation of the LeGeDe-project, the article deals in detail with research questions relevant to the project, the project goals, the corpus data basis and empirically determined expectations of a resource on spoken German. The presentation of the complex structure of the LeGeDe-prototype is illustrated with numerous examples. Apart from the central information on the macro- and microstructure and the lexicographical outer texts, the various linking and access structures are shown. In addition to the final conclusion, the article concludes with an outlook that provides extensive suggestions for future lexicographic work with spoken language corpus data.

Keywords: INTERACTIONAL LINGUISTICS, CORPUS LINGUISTICS, LEXICOLOGY OF SPOKEN LANGUAGE, MACROSTRUCTURE, MICROSTRUCTURE, MULTIMEDIALITY, ONLINE LEXICOGRAPHY, OUTER TEXTS, LINKING

1. Einleitung

Einsprachige Wörterbücher zum Deutschen gibt es in zahlreichen Varianten: Verlagspublikationen wie der *Duden* (in verschiedenen Reihen und Publikationsformen) oder lexikografische Nachschlagewerke des Verlags von *Langenscheidt* (wie das *Großwörterbuch Deutsch als Fremdsprache*) stehen neben Wörterbüchern, die in wissenschaftlichen Akademien oder Forschungsinstituten erstellt werden, wie z.B. das *DWDS* der Berlin-Brandenburgischen Akademie der Wissenschaften in Berlin oder *ellexiko*, das am Leibniz-Institut für Deutsche Sprache (=IDS) in Mannheim konzipiert wurde. Die Bandbreite an Wörterbuchtypen reicht dabei von Standardwörterbüchern (z.B. *Duden-Universalwörterbuch*) bis hin zu ausgewiesenen Spezialwörterbüchern (z.B. *Wörterbuch der deutsch-lothringischen Mundarten=LothWB*). Ebenso vielfältig ist das Angebot für unterschiedliche Zielgruppen (Muttersprachler/-innen, Fachspezialist/-innen, Lerner/-innen des Deutschen als Fremd- oder Zweitsprache etc.). Neben dem Printformat werden die unterschiedlichen Wörterbuchtypen des Deutschen zunehmend online angeboten. Einige der neu konzipierten Wörterbücher, die nicht auf einer vorausgehenden Printversion basieren, sind sogar ausschließlich online zugänglich (vgl. *ellexiko*)*.

In dieser Fülle von verschiedenen Nachschlagewerken zur deutschen Sprache gibt es trotz unterschiedlichem inhaltlichen Fokus eine Gemeinsam-

keit: Geht es um ein ausgewiesenes korpusgestütztes Wörterbuch, so handelt es sich bei der empirischen Datengrundlage in der Regel um Sprachkorpora geschriebensprachlicher Daten. Die herangezogene Datenbasis kann insbesondere bei Verlagswörterbüchern und frei zugänglichen lexikografischen Ressourcen (wie z.B. *Linguee*) jedoch nicht immer vollständig transparent nachvollzogen werden (vgl. Tarp 2019: 227). Meist findet sich in den Umtexten eine ungenaue Formulierung wie: „Grundlage für die möglichst exakte Darstellung der Wortbedeutungen sind das umfangreiche Dudenkorpus sowie die nahezu unbegrenzten Recherchemöglichkeiten des Internets“ (aus: Vorwort *Duden-DaF*: 5). Beim Wörterbuch *lexiko* lässt sich hingegen sehr transparent nachvollziehen, mit welchen Quellen gearbeitet wurde, da in den Umtexten detailliert auf die Korpusbasis, die für die Wortartikelarbeit herangezogen wurde, eingegangen wird.¹

Im Gegensatz zur deutschen Schriftsprache als korpusbasierte Datengrundlage wurde Material zum gesprochenen Deutsch insbesondere Daten aus authentischen mündlichen Interaktionssituationen in den meisten Wörterbüchern bislang eher vernachlässigt. Zwar wird in den bislang existierenden Wörterbüchern vereinzelt mit Angaben wie „gesprochen“ oder „umgangssprachlich“ (vgl. *Duden-online* und *DWDS* zu *gucken*) auf Besonderheiten im Gesprochenen hingewiesen, doch bleibt die Datenbasis oft eine unbekannte bzw. unpräzise Variable.

Demzufolge gibt es eine lexikografische Beschreibungslücke hinsichtlich lexikalischer Besonderheiten des gesprochenen Deutsch. Diese Tatsache hängt in sehr starkem Maße mit der Verfügbarkeit geeigneter Korpora zusammen, denn nur eine ausreichend große Datenbasis kann es ermöglichen, korpusgestützte Methoden zur Erstellung einer lexikografischen Ressource zu entwickeln.

Mit dem „Forschungs- und Lehrkorpus Gesprochenes Deutsch“ (=FOLK) wird am IDS seit einigen Jahren ein Korpus aufgebaut, das authentisches Gesprächsmaterial aus natürlicher und spontaner Interaktion überwiegend zum standardnahen Deutsch über eine online verfügbare Datenbank (DGD= Datenbank für Gesprochenes Deutsch) recherchierbar macht. Das Vorhandensein von FOLK als einer adäquaten Datengrundlage zur Dokumentation und Erforschung des gesprochenen Deutsch in der Interaktion in Verbindung mit der am IDS existierenden Expertise im Aufbau von Online-Wörterbüchern erschließt neue Wege. Im von der Leibniz-Gemeinschaft² geförderten Drittmittelprojekt „Lexik des gesprochenen Deutsch“ (=LeGeDe³) wurde in drei Jahren (von September 2016 bis September 2019) ein erster lexikografischer Prototyp zu lexikalischen Besonderheiten im gesprochenen Deutsch entwickelt. Der LeGeDe-Prototyp ist in die Plattform OWID^{plus} eingebettet und über die URL <https://www.owid.de/legede/kostenfrei> seit September 2019 abrufbar. Die erstellte lexikografische Ressource verfolgt zunächst primär das Ziel der Sprachdokumentation und spricht daher als mögliche Zielgruppe Sprachwissenschaftler/-innen verschiedener Disziplinen aber mit Einschränkung auch

Sprachlehrende als Übermittler/-innen an. In diesem Sinne werden auch Sprachlernende als Zielgruppe der Ressource indirekt mitangedacht (vgl. Meliss, Möhrs und Ribeiro Silveira 2018, 2019) (vgl. Abschnitt 2.2 (iii) und (iv)).

Der vorliegende Artikel bietet einen Überblick zum LeGeDe-Projekt (vgl. Abschnitt 2) und eine ausführliche Darstellung der Projektergebnisse (vgl. Abschnitt 3). In Abschnitt 2 werden Informationen zum Projektrahmen, zu Annahmen und Forschungsfragen sowie zu Projektzielen und der Datengrundlage geliefert. Außerdem wird ein Einblick in die Ergebnisse von empirischen Studien zu Erwartungshaltungen an eine zukünftige lexikografische Ressource zu Besonderheiten des Gesprochenen geboten. Im Zentrum von Abschnitt 3 steht insbesondere die Präsentation des LeGeDe-Prototyps. Dabei werden vor allem Aspekte in Verbindung mit der komplexen lexikografischen Struktur (Makrostruktur, Mikrostruktur, Vernetzungs- und Verweisstrukturen, Umtexte) thematisiert. Neben einem Fazit wird in Abschnitt 4 ein Ausblick angeboten, der umfangreiche Vorschläge für die zukünftige lexikografische Arbeit mit gesprochenen sprachlichen Korpusdaten liefert.

2. Das LeGeDe-Projekt

2.1 Projektrahmen

Das LeGeDe-Projekt hatte zum Hauptziel, einen korpusbasierten, lexikografischen Prototyp zu Besonderheiten des gesprochenen Deutsch in der Interaktion zu erstellen (vgl. Abschnitt 2.3). Die Kooperation zwischen den Abteilungen Pragmatik und Lexik des IDS ermöglichte während der Projektlaufzeit eine Verbindung der entsprechenden interdisziplinären Fachkompetenzen.

Für die Entwicklung des LeGeDe-Prototyps wurden im Projekt sowohl quantitative als auch qualitative Methoden entwickelt, mit denen die Spezifika gesprochenen sprachlicher Lexik des Deutschen auf der Basis der am IDS erstellten mündlichen Korpora (im Programmbereich „Mündliche Korpora“ der Abteilung Pragmatik) im Vergleich zu Teilkorpora der geschriebenen Sprache identifiziert, analysiert und für die lexikografische Anwendung aufbereitet werden konnten. Neuartige lexikografische Beschreibungsformate für lexikalische Daten aus Gesprächskorpora in audioelektronischer Form, ebenso wie innovative lexikografische Angabetypen, die auf die Funktion lexikalischer Einheiten in interaktionalen Kontexten Bezug nehmen, mussten daher im Laufe der Projektzeit für die Erfüllung der Projektziele entwickelt werden (vgl. dazu insbesondere These 10 aus Engelberg, Klosa-Kückelhaus und Müller-Spitzer 2019: 33). Bei diesem Prozess konnten auch einige der Erwartungen von zukünftigen Nutzer/-innen, die durch Umfragen empirisch erhoben wurden, berücksichtigt werden (vgl. Abschnitt 2.5).

2.2 Annahmen und Forschungsfragen

Das LeGeDe-Projekt basierte auf folgenden Hauptannahmen und Beobachtungen, aus denen sich entsprechende Forschungsfragen ableiten ließen:

(i) Es existieren Unterschiede auf verschiedenen sprachlichen Ebenen zwischen dem gesprochenen und dem geschriebenen Deutsch. Bezüglich der Lexik können sich die Divergenzen sowohl im Bestand als auch in Verbindung mit deren Form, Bedeutung und Verwendung auswirken (vgl. Deppermann, Proske und Zeschel 2017; Fiehler 2016; Imo 2007; Schwitalla 2012). Aus dieser Beobachtung heraus ergab sich die Frage, welche korpusbasierten Methoden geeignet sind, um diese Divergenzen zu identifizieren und für eine lexikografische Ressource nutzbar zu machen.

(ii) Die lexikografische Kodifizierung der interaktionstypischen Besonderheiten der gesprochensprachlichen Lexik des Deutschen ist bis jetzt unzureichend, obwohl eine beachtliche Menge an lexikalischen Studien mit interaktionslinguistischem Ansatz vorliegen (vgl. Meliss 2016; Meliss und Möhrs 2017, 2018, 2019; Moon 1998; Siepmann 2015; Trap-Jensen 2004). Es existieren zurzeit kaum korpusbasierte lexikografische Projekte zur Lexik der gesprochenen Sprache. Lediglich zum Dänischen wurde ein kleines Projekt zu Interjektionen (vgl. Hansen und Hansen 2012) durchgeführt. Die Frage, die sich aus dieser Beobachtung ergab, ist, wie die identifizierten lexikalischen Besonderheiten in einer lexikografischen Ressource auf adäquate Weise beschrieben werden können.

Sowohl die Ergebnisse zweier im Projekt durchgeführter Umfragen zu den Erwartungen und Anforderungen an eine lexikografische Ressource für Spezifika des gesprochenen Deutsch (vgl. Meliss, Möhrs und Ribeiro Silveira 2018, 2019), als auch entsprechende lexikografische Untersuchungen bestätigen, dass die lexikografische Kodifizierung der Merkmale der gesprochenen Sprache in der Interaktion in den aktuellen Wörterbüchern bisher nicht zufriedenstellend berücksichtigt wurde (Meliss 2016: 195; Eichinger 2017: 283). Trotz einiger jüngster Fortschritte in der Erstellung von gesprochensprachlichen Korpora zu verschiedenen Sprachen und Sprachvarietäten (vgl. Barcala Rodríguez et al. 2018; Fandrych, Meißner und Wallner 2017; Schmidt 2014a, 2014b, 2017, 2018; Verdonik und Sepesy Maučec 2017) sind Erfahrungen mit ihrer Nutzung in der Lexikografie bis jetzt kaum bekannt. Daher konnte sich der LeGeDe-Prototyp kaum auf bestehende Modelle stützen, die u.a. bei der Erstellung einer geeigneten Stichwortliste, der Entwicklung von neuen Angabeklassen und deren Darstellung als Orientierungshilfe hätten dienen können.

(iii) Der Informationsbedarf zu typisch gesprochensprachlicher Lexik ist in den letzten Jahren besonders in unterschiedlichen Anwendungsbereichen, wie z.B. in Unterricht und Lehre (speziell im Sekundarbereich und in den Bereichen Deutsch als Fremd- und/oder Zweitsprache) sowie im Verlagswesen in Ver-

bindung mit der Erstellung von geeigneten Unterrichtsmaterialien gestiegen (vgl. Albert und Diao-Klaeger 2018; Handwerker, Bäuerle und Sieberg 2016; Imo und Moraldo 2015; Moraldo und Missaglia 2013; Reeg, Gallo und Moraldo 2012; Sieberg 2013). So wird z.B. im „Gemeinsamen europäischen Referenzrahmen für Sprachen“ (=GeR) beim Beurteilungsraster zur mündlichen Kommunikation und beim Parameter „Interaktion“ für Niveau C1 explizit darauf hingewiesen, dass der Lernende „[...] aus einem ohne weiteres verfügbaren Repertoire von Diskursmitteln eine geeignete Wendung auswählen [kann], um seine/ihre Äußerung angemessen einzuleiten, wenn er/sie das Wort ergreifen oder behalten will, oder um die eigenen Beiträge geschickt mit denen anderer Personen zu verbinden“ (Trim et al. 2001: 37). Sowohl Meliss und Möhrs (2018) als auch Fandrych, Meißner und Wallner (2018) weisen darauf hin, dass Korpora der gesprochenen Sprache von großem Interesse für viele Forschungs- und Anwendungsszenarien sind und gewinnbringend für die Sprachdidaktik eingesetzt werden können, wenn den Lehrenden und Lernenden ein benutzergerechter Zugang zu dem Material angeboten wird. Die sich ergebenden Fragen standen diesbezüglich v.a. in Zusammenhang mit dem Komplexitätsgrad der Informationsdarstellung und dem Nutzen des virtuellen, multimedialen Raums für die Mündlichkeitsdidaktik.

(iv) Außerdem zeigen die Ergebnisse der im LeGeDe-Projekt durchgeführten empirischen Studien zu den Erwartungen künftiger Nutzer/-innen an eine Ressource zum gesprochenen Deutsch in der Interaktion (vgl. Abschnitt 2.5), dass sowohl bei den L1- als auch bei den L2-Sprechenden des Deutschen zu jeweils über 70% Bedarf an einem Wörterbuch zu Spezifika des gesprochenen Deutsch vorhanden ist (vgl. Meliss, Möhrs und Ribeiro Silveira 2018, 2019). Diese Beobachtung bestätigt die grundsätzliche Annahme zum ansteigenden Bedarf sowohl in der Forschung als auch in der Lehre. Die sich daraus ergebenden Fragen betreffen Bereiche der Vermittlung von gesprochener Sprache im Unterricht und die Bereitstellung bzw. Ausarbeitung von geeignetem Material, auf das für Lehrwerke und Unterrichtsstoff zurückgegriffen werden kann (vgl. Meliss und Möhrs 2018).

2.3 Projektziele

Im LeGeDe-Projekt wurden auf Basis der in Abschnitt 2.2 beschriebenen Forschungsfragen und theoretischen Annahmen die folgenden zentralen theoretischen, methodologischen und anwendungsorientierten Ziele verfolgt, die sich bei der Beschäftigung mit dem Thema in der Projektarbeit ergaben:

- Entwicklung von quantitativen und qualitativen Verfahren zur Identifizierung von interaktionstypischen, gesprochensprachlichen lexikalischen Elementen und deren spezifischen Merkmalen im Vergleich zur Lexik der geschriebenen Sprache (vgl. Meliss und Möhrs 2017, 2018, 2019),

-
- Erstellung einer Stichwortkandidatenliste und Auswahl geeigneter Stichwörter für die Ressource (vgl. Meliss et al. 2018; Möhrs i. Dr.),
 - Entwicklung von korpusbasierten quantitativen und qualitativen Verfahren für die lexikalische und interaktional ausgerichtete Datenanalyse und Beschreibung auf verschiedenen sprachlichen Ebenen (vgl. Meliss et al. 2018, 2019; Möhrs und Torres Cajo i. Dr.),
 - Bestimmung der Besonderheiten von mündlichem Sprachgebrauch auf verschiedenen Ebenen (Form, Inhalt/Funktion, Situation etc.) im lexikalischen Bereich (vgl. Meliss 2020b; Meliss und Möhrs 2017, 2018, 2019),
 - Entwicklung neuartiger lexikografischer Angabeklassen, die u.a. auf die Funktion lexikalischer Einheiten in Interaktionskontexten Bezug nehmen (vgl. Meliss et al. 2019; Meliss i. Dr. [b]),
 - Entwicklung innovativer lexikografischer Beschreibungsformate in multi-medialer Form für hochgradig kontextualisierte lexikalische Daten (vgl. Meliss et al. 2019),
 - Entwicklung weiterer korpuslinguistischer Methoden und Tools zur Abfrage, Analyse und Strukturierung von automatisch generierten korpusbasierten Daten (vgl. Möhrs, Meliss und Batinić 2017, Lemmenmeier-Batinić 2020). Insbesondere verweisen wir hier auf das Korpusanalysetool „Lexical Explorer“, welches quantitative Analysen ermöglicht und in Abschnitt 2.4 genauer vorgestellt wird.

2.4 Datengrundlage

Die qualitativen Untersuchungen zum Forschungsgegenstand des LeGeDe-Projektes, der gesprochenen Sprache des Deutschen in der Interaktion, erfolgen ausschließlich auf Basis von FOLK (vgl. Schmidt 2014a, Kupietz und Schmidt 2015), das als erstes großes Gesprächskorpus des Deutschen seit 2008 am IDS aufgebaut und ständig erweitert wird. Parallel dazu wird die DGD (Schmidt 2014c, 2018), über die FOLK recherchierbar ist, permanent weiterentwickelt und mit innovativen korpus technologischen Funktionalitäten ausgestattet. Die Gesprächsaufnahmen und Transkripte (teils auch Videoaufnahmen) stammen aus unterschiedlichen privaten, institutionellen und öffentlichen Kontexten aus dem deutschsprachigen Raum. Die Daten zeichnen sich durch die Merkmale medial mündlich, authentisch, spontan (nicht eliziert), größtenteils standardnah und aktuell aus. FOLK ist in seiner Art das größte Korpus für das gesprochene Deutsch in der Interaktion. Zurzeit liegt FOLK in der DGD-Version 2.14 mit rund 285h/2,7 Mio. Tokens und 332 unterschiedlichen Sprechereignissen vor (Stand: 27.04.2020). Die DGD bietet als Korpusanalysetool vielfältige Möglichkeiten der Erschließung mündlicher Daten nach linguistischen und interaktionalen Merkmalen. Über die Anwendungsoberfläche der DGD können strukturierte Tokensuchen realisiert und über vier Annotationsebenen (cGAT-

Transkript, Normalisierung, Lemmatisierung, PoS) recherchiert werden (vgl. Westpfahl 2014; Westpfahl und Schmidt 2016). Außerdem können zu den Gesprächen sowohl Metadaten zu den Sprechenden als auch zu dem Gesprächsereignis abgerufen werden. Ausgehend von der Ausrichtung von FOLK, authentisches, standardnahes Deutsch in der Interaktion zu erfassen, hat sich die Analyse im LeGeDe-Projekt auf diese Sprachvariante fokussiert: „In besonderem Maße sollen jene Phänomene, die als ‚standardnah‘ charakterisiert werden können, behandelt werden. Regionale, soziale, funktionale oder idiolektale Sprachvarietäten werden daher ausgeklammert.“ (Meliss und Möhrs 2017: 44). Über die gesamte Projektlaufzeit wurde die Herausforderung, eine solche Abgrenzung für die konkrete Projektarbeit und nicht zuletzt auch für die Entwicklung der Ressource zu operationalisieren, im Austausch mit Fachkollegen/-innen thematisiert.

Für die im Rahmen der Projektarbeit durchgeführten Analysen wurde einerseits auf die Recherchemöglichkeiten in der DGD zurückgegriffen. Andererseits wurde im LeGeDe-Projekt auch an weitergehenden quantitativen Suchmöglichkeiten gearbeitet, um beispielsweise auch Kollokations- und Kookkurrenzanalysen über gesprochensprachliche Daten durchzuführen. Diese Verfahren sind in der lexikografischen Praxis, in der mit geschriebensprachlichen Daten und Recherchertools gearbeitet wird, inzwischen etabliert (vgl. Geyken und Lemnitzer 2016). Für die FOLK-Daten konnte über die DGD eine Analyse dieser Art bislang nicht durchgeführt werden. Um dem Projektziel „Entwicklung von quantitativen und qualitativen Verfahren zur Identifizierung von interaktionstypischen, gesprochensprachlichen lexikalischen Elementen und deren spezifischen Merkmalen im Vergleich zur Lexik der geschriebenen Sprache“ (vgl. Abschnitt 2.3) näher zu kommen, wurden in der ersten Projektphase Möglichkeiten quantitativer Vergleiche mit DEREKO (dem Deutschen Referenzkorpus; vgl. Kupietz und Keibel 2009; Kupietz et al. 2018) eruiert und ausgearbeitet. Diese Überlegungen und Datenanalysen sind neben weiteren quantitativen Messungen in das Tool „Lexical Explorer“ geflossen, das über die Plattform OWID^{plus} abrufbar ist (vgl. Lemmenmeier-Batinić 2020). Die Anwendung ermöglicht, die quantitativen Daten von zwei Korpora aus der DGD zu durchsuchen und abzufragen: FOLK und GeWiss (Gesprochene Wissenschaftssprache, vgl. Fandrych, Meißner und Wallner 2017). Die hinter dem Tool entwickelten Recherchemethoden wurden für anfängliche Lemmastudien im LeGeDe-Projekt entwickelt und für den Stichwortkandidatenansatz ausgearbeitet. Mit dem „Lexical Explorer“ können quantitative Korpusdaten mit Hilfe von Häufigkeitstabellen bezüglich der Wortverteilung über Wortformen, Kookkurrenzen und Metadaten erforscht werden, was über die Einbindung des Tools in OWID^{plus} auch der interessierten Öffentlichkeit möglich ist. Im LeGeDe-Projekt wurden die Analysen über FOLK maßgeblich für die Definition potenzieller Stichwortkandidaten (vgl. dazu auch Abschnitt 3.2) und die Detektion typischer Wortverbindungen zu einem spezifischen Analysestichwort genutzt. Auch die Daten zu quantitativen Recherchen zu einem Stichwort

und seinem Auftreten in Interaktionstypen (definiert über die Kategorien „private“, „public“, „non-private/non-public“ und „other“) wurden, neben anderen Metadaten, als Analysewerte über den „Lexical Explorer“ herangezogen und in die lexikografische Beschreibung eingebunden. Über ein Verlinkungsangebot können quantitative Angaben aus den lexikografischen Artikeln im LeGeDe-Prototyp direkt nachvollzogen werden (vgl. hierzu auch Abschnitt 3.4).

2.5 Empirische Forschung: Erwartungshaltungen

2.5.1 Hintergrund zu den Studien

In Ergänzung zu den konzeptionellen Arbeiten im LeGeDe-Projekt wurden auch verschiedene Arten empirischer Studien in Verbindung mit den Erwartungshaltungen potenzieller Nutzer/-innen an eine neuartige lexikografische Ressource zu Besonderheiten des Gesprochenen durchgeführt. Dieser Arbeitsbereich wurde in Kooperation mit dem Projekt „Empirische Methoden“⁴ in der Abteilung Lexik des IDS entwickelt und durchgeführt. Die Erwartungen potentieller Nutzer/-innen zu einem recht frühen Zeitpunkt der Projektarbeit abzufragen, sollte ermöglichen, die Konzeption der Ressource speziell daraufhin auszurichten. Zwischen Februar und Mai 2017 wurden dementsprechend zwei Studien durchgeführt, die bezüglich der befragten Zielgruppe und des Formats unterschiedlich gestaltet waren.⁵ Zum einen äußerten sich Expert/-innen in einem Interview gezielt, in geschlossenen und offenen Fragenformaten zu ihren Erwartungen. Und zum anderen wurde eine Online-Befragung konzipiert, die sich an eine breitere Probandengruppe gerichtet hat. Eine der größten Herausforderungen beider empirischer Studien war die Tatsache, dass zu dem Zeitpunkt der Befragungen eine zukünftige Ressource thematisiert wurde, die zum Befragungszeitpunkt noch gar nicht existierte und für die es auch keine Vorbilder gab. Daraus ergab sich die einmalige Chance, geäußerte Erwartungen bei der Erstellung der Ressource zu berücksichtigen.

2.5.2 Ergebnisse der Erwartungsstudien

Zu beiden Studien sind in Meliss, Möhrs und Ribeiro Silveira (2018) die Ergebnisse in umfassender Art dargestellt. Der Abgleich der Antworten bzw. Einschätzungen mit den soziodemografischen Daten ließ bei der Interpretation der Daten auch Rückschlüsse zu bestimmten Probandengruppen zu. So wurde in Meliss, Möhrs und Ribeiro Silveira (2019) besonders die Befragtengruppe der L2-Lerner/-innen in den Blick genommen. Die spezifischen anwendungsorientierten Bedürfnisse in den Bereichen DaF und DaZ in Verbindung mit dem Erwerb und der Anwendung gesprochensprachlicher Lexik in der Interak-

tion, auf die schon in Meliss und Möhrs (2018) hingewiesen wurde, konnten ebenfalls mit den Ergebnissen der Befragungen abgeglichen werden.

Es zeigte sich, dass sowohl bei L1- als auch L2-Sprecher/-innen Bedarf an einem Nachschlagewerk zu gesprochener Sprache in der Interaktion existiert, das Spezifika des gesprochenen Deutsch in Bedeutung, Verwendung, Kombinatorik, Angemessenheit und weiteren Bereichen im Vergleich zum geschriebenen Deutsch erfasst. Dieser Bedarf kann nicht oder nur teilweise durch gängige einsprachige Wörterbücher wie dem *Duden-online* oder auch dem *DWDS* befriedigt werden. Auch zweisprachige Werke wie *Leo*, *Linguee* oder *PONS-online* decken in ihrem lexikografischen Angebot Fragen zu Besonderheiten der Lexik im gesprochenen Deutsch nicht oder nur ungenügend ab. Eine Ressource, in der über den Einbezug authentischer Beispiele zusammen mit Audiomaterial und Transkripten Informationen bereitgestellt werden, stellt laut Meinung der Befragten eine nützliche Quelle für unterschiedliche Rechercheziele dar. Im Laufe der Projektarbeit wurden daher die Erwartungen an „besondere“/„gesprochensprachlich spezifische“ Stichwörter umfassend in den Blick genommen und an deren Bedeutungs- und Funktionsbeschreibung für die lexikografische Umsetzung gearbeitet. Geäußerte Erwartungen nach Verknüpfungen zu unterschiedlichen Ressourcen, wie u.a. zu dem zugrundeliegenden Korpus, wurden ebenfalls umgesetzt (vgl. Abschnitt 3.4).

Nicht alle Erwartungen und Wünsche der Befragten und auch nicht alle im Projekt entwickelten Ideen konnten im Rahmen der Projektlaufzeit im LeGeDe-Prototyp realisiert werden. Das Spektrum von potenziellen Möglichkeiten, das in den unterschiedlichen Frageformaten zum Ausdruck kam, zeigt die Vielfältigkeit der Anknüpfungspunkte für eine Weiterarbeit im Bereich der korpusbasierten Lexikografie gesprochen sprachlicher Besonderheiten auf (vgl. Abschnitt 4).

3. Der LeGeDe-Prototyp

Im folgenden Abschnitt werden hauptsächlich Informationen zu der lexikografischen Struktur des LeGeDe-Prototyps angeboten. Bezüglich der Makrostruktur wird besonders der Stichwortansatz thematisiert (Abschnitt 3.1). In Verbindung mit der Mikrostruktur (Abschnitt 3.2) wird ausführlich die Methode der Datenanalyse und -strukturierung vorgestellt, um daran anschließend die komplexe lexikografische Angabenstruktur auf den drei Informationsebenen (i) Überblicksartikel, (ii) Modul 1 und (iii) Modul 2 darzustellen. Zur Illustration dienen ausgewählte Einträge aus dem LeGeDe-Prototyp. Außerdem werden die Zugriffsmöglichkeiten und Vernetzungsstrukturen (Abschnitt 3.3) und das Angebot an unterschiedlichen lexikografischen Umtexten (Abschnitt 3.4) vorgestellt.

3.1 Makrostruktur: Stichwortansatz

Eine der zentralen Forschungs- und Methodikfragen, mit denen sich das LeGeDe-Projekt beschäftigt hat, ist die korpusbasierte Identifizierung typischer gesprochensprachlicher Lexik. In einem direkten Bezug zu den distinktiven Merkmalen der Lexik der geschriebenen vs. der gesprochenen Sprache in der Interaktion steht die Erstellung einer Stichwortliste (vgl. Klosa 2013a, 2013b; Klosa und Tiberius 2016; Klosa, Schnörch und Schoolaert 2010; Schnörch 2005; Stadler 2014; Wiegand 1983) mit Kandidaten für den LeGeDe-Prototyp, die möglichst typische Phänomene des Gesprochenen in der natürlichen, spontanen Interaktion aufweisen und sich quantitativ deutlich von der geschriebenen Sprache abgrenzen. Für die Entwicklung des LeGeDe-Prototyps wurden so neben Einwortlemmata, die im Vergleich zur Schriftsprache in der gesprochenen Sprache in der Interaktion eine spezifische Bedeutung oder Verwendung bzw. Funktion einnehmen (z.B. Interjektionen, Diskurspartikeln), auch Mehrwortausdrücke und Konstruktionen mit spezifischen Funktionen in der Interaktion (z.B. *was weiß ich*, *ich weiß nicht*, *keine Ahnung*, *guck mal*) als mehrteilige Stichwortkandidaten integriert (vgl. Bergmann 2017, Günthner 2018, Helmer und Deppermann 2017, Helmer, Deppermann und Reineke 2017, Imo 2007, Möhrs 2020, Torres Cajo 2019, Zeschel 2017).

Für die Erstellung einer Stichwortkandidatenliste wurde daher im LeGeDe-Projekt eine korpusbasierte und interpretative Methode entwickelt (vgl. Meliss et al. 2018), mit der die wichtigsten Kandidaten der typischen gesprochenen Lexik in der Interaktion aufgedeckt werden konnten. Der Vergleich zur Schriftsprache des Deutschen erfolgte über eine Teilmenge aus DEREKO. Das frequenzgesteuerte korpusbasierte Verfahren, bei dem die Häufigkeitsklassen der Lemmata aus FOLK und aus DEREKO ermittelt und diese dann in einem direkten Vergleich betrachtet werden können, ist über den „Lexical Explorer“ nachvollziehbar. Die Stichwortkandidatenliste (Stand: 23.05.2018), die den Ausgangspunkt für die Auswahl der ausführlich bearbeiteten Lemmata im LeGeDe-Prototyp darstellt, kann mit den 322 Kandidaten vollumfänglich über die Umtexte der Online-Ressource abgerufen werden (Anhang im Download-Dokument unter „Über LeGeDe“).⁶

Die Rangliste der 25 in der Abfrage zuerst angezeigten möglichen Stichwortkandidaten (vgl. Abb. 1) vermittelt einen ersten Eindruck von denjenigen Kandidaten, die für den LeGeDe-Prototyp relevant waren. Die hohe Häufigkeitsklassendifferenz bei den Interjektionen (*ah*, *ach*, *oh*), Modalpartikeln (*ja*, *halt*), Gesprächspartikeln (*okay*, *ja*, *na*) und Verben (*gucken*, *kriegen*) weist mit der beschriebenen quantitativen Perspektive auf Besonderheiten im Gesprochenen vs. Geschriebenen hin, die in weiteren Schritten durch qualitative Studien im LeGeDe-Projekt genauer analysiert wurden.

■ **FOLK vs. DeReKo**

Column visibility CSV Show 25 entries

Lemma	FOLK HK	DeReKo HK	HK Diff	Filter	PoS
okay	4	14	10	1	NG
ah	4	14	10	1	NG
ach	4	13	9	1	NG
ja	0	8	8	1	PTK/NG
oh	5	13	8	1	NG
gucken	5	13	8	1	V
halt	4	12	8	1	PTK/NG
du	2	9	7	1	P
nachher	7	14	7	1	ADV
danke	7	14	7	1	NG
irgendetwas	7	14	7	1	P
na	5	12	7	1	NG
irgendwie	5	12	7	1	ADV
kriegen	5	12	7	1	V
nein	3	10	7	1	NG
mal	2	8	6	1	ADV/PTK
eh	7	13	6	1	ADV
Mama	7	13	6	1	N
cool	7	13	6	1	NG/ADJ
drin	6	12	6	1	ADV/PTK
drauf	6	12	6	1	ADV/PTK
dran	6	12	6	1	ADV/PTK
raus	6	12	6	1	PTK/ADV
sozusagen	6	12	6	1	ADV/NG
dein	5	11	6	1	P

Showing 1 to 25 of 322 entries (filtered from 52,966 total entries)

Previous 1 2 3 4 5 ... 13 Next

Abb. 1: Rangliste der 25 in der Abfrage zuerst angezeigten Stichwortkandidaten mit der höchsten Häufigkeitsklassendifferenz zwischen FOLK und DEREKO (Stand 2018)

Die Kombination aus automatisierten Verfahren und manueller Analyse der Korpusdaten hat sich als effektiver Weg erwiesen, um die Stichwortkandidatenauswahl für eine Ressource zu Spezifika der Lexik des gesprochenen Deutsch anzugehen. Mit Hilfe des „Lexical Explorers“ kann die automatische Sortierung und Filterung der Stichwortkandidaten nach Wortarten, Frequenz und anderen Merkmalen sowie die Auswahl weiterer Stichwörter in weiteren Schritten des Projekts nachvollzogen werden.

3.2 Mikrostruktur: Methodik und Informationsangebot

3.2.1 Datenanalyse und Datenstrukturierung

Verschiedene quantitative und qualitative Verfahren zur lexikalischen Bedeu-

tungsdisambiguierung und zur Entwicklung von Wortprofilen, die als methodologische Ansätze in der korpusbasierten Lexikologie und Lexikografie gelten (vgl. u.a. Engelberg 2015, 2018, 2019; Engelberg et al. 2011), wurden mit interaktionslinguistischen Analysen sprachlicher Einheiten und der Beschreibung ihrer Formen und Funktionen in der Interaktion (vgl. u.a. Deppermann 2007; Couper-Kuhlen und Selting 2018) vereint. Unterschiedliche quantitative Informationen, wie u.a. automatisch generierte Daten (Frequenzdaten zum Formenbestand, zu Kollokationen, Kookkurrenzen etc.), wurden nicht nur bei der ersten korpusbasierten Annäherung an die Daten durch entsprechende Hypothesenbildung, sondern auch bei der Interpretation der Analyseergebnisse unterstützend genutzt. Auf der Grundlage von qualitativen lexikalisch-semantischen und interaktionslinguistischen Einzelbeleganalysen wurden für die unterschiedlichen grammatischen Kategorien Kodierschemata entwickelt.

Die Datengrundlage für die Analyse eines Lemmas ist jeweils grundsätzlich eine Zufallsstichprobe aus FOLK, aus der die ersten 100 gültigen Belege manuell kodiert wurden. Der Prozess des Kodierens wurde durch qualitative Analysen gestützt. Neben Metadaten zum Treffer und zum Transkript, die automatisch extrahiert vorliegen, wurden v.a. formale (Person, Numerus, Modus etc.), inhaltliche (Bedeutung, Bedeutungsumschreibung etc.), kombinatorische (Strukturmuster, Kookkurrenzen etc.), morphosyntaktische (Realisierungsmöglichkeiten der Aktanten etc.), funktionale, sequenzbezogene sowie prosodische Aspekte betrachtet und festgehalten. Da sich schon in der Stichprobe aus 100 analysierten Treffern die Notwendigkeit ergab, einerseits die Ebene der Bedeutung und andererseits die funktionale Ebene der ausgewählten Stichwörter mit teilweise unterschiedlichen Analyseparametern und Methoden zu untersuchen, zu strukturieren und lexikografisch darzustellen (Modul 1 vs. Modul 2), wurden für die weiteren Untersuchungen, Strukturierungen und Darstellungen der Ergebnisse zwei unterschiedliche Analyse- und Beschreibungsverfahren angesetzt (vgl. Meliss i. Dr. [b]).

Im Zentrum von Modul 1 steht die semantisch und syntaktisch motivierte Disambiguierung der einzelnen Lesarten, während im Zentrum von Modul 2 die Beschreibung der interaktionalen Funktionen steht. Entsprechend wurden die weiteren Analyse-, Strukturierungs- und Beschreibungsverfahren den Anforderungen des jeweiligen Moduls angepasst.

(a) Modul 1

Im Hinblick auf das Informationsangebot für Modul 1 wurden die Kodierergebnisse der Lemma-Stichprobe in ihren jeweiligen Gesprächssequenzen nach Form, Bedeutung und Kombinatorik (Strukturmuster, Kollokationen, feste Wendungen, interaktionale Einheiten etc.) sowie unter Berücksichtigung von Kontext und Metadaten analysiert (vgl. Abb. 2) und auf der Grundlage einer semantisch-syntaktischen Bedeutungsdisambiguierung einzelne Lesarten identifiziert. Als Orientierungshilfe für die jeweilige Lesartendisambiguierung

dienten dafür entsprechende lexikografische Informationen ausgewählter Wörterbücher (z.B. *E-Valbu*, *LGWB DaF*, *Duden-online*, *DWDS*). Durch den Vergleich der Daten aus der LeGeDe-Stichprobe mit denen aus gängigen lexikografischen Werken, die hauptsächlich auf der geschriebenen Sprache basieren, wurden u.a. folgende Ziele verfolgt:

(i) Identifizierung von typisch gesprochen sprachlichen Lesarten: Mögliche, besonders interessante Beobachtungen beziehen sich auf solche Lesarten, die nur in den FOLK-Daten belegt werden konnten, aber nicht in den konsultierten Wörterbüchern kodifiziert sind (so konnten z.B. für *gucken* in FOLK wesentlich mehr Lesarten identifiziert werden als in den konsultierten Wörterbüchern). Diese Fälle erlauben den Schluss, dass es sich zumindest um Lesarten handelt, die vor allem in der gesprochenen Interaktion auf der Grundlage der LeGeDe-Stichprobe aufgedeckt werden konnten.

(ii) Identifizierung von eher nicht typisch gesprochen sprachlichen Lesarten: Die Identifizierung von solchen Lesarten, die nicht in den FOLK-Daten belegt werden konnten, aber in den konsultierten Wörterbüchern kodifiziert sind, erlaubt den Schluss, dass es sich zumindest um Lesarten handelt, die in der gesprochenen Interaktion auf der Grundlage der LeGeDe-Stichprobe eher selten sind. Da sie nicht in der LeGeDe-Stichprobe belegt werden konnten, wurden sie für den LeGeDe-Prototyp ausgeschlossen. Allerdings bieten wir in den Überblicksartikeln (vgl. Abschnitt 3.2.2) Links zu Wörterbuchquellen an, sodass ein/-e Nutzer/-in zu weiteren Lesarten, die in anderen Wörterbüchern verzeichnet sind, hier nachschlagen kann (vgl. auch Abschnitt 3.3).

(iii) Information zu formalen, morphosyntaktischen Eigenschaften bzw. Distributionsbeschränkungen (z.B. Präferenz für Modus Imperativ bei *gucken* in der Bedeutung: ‚richtungsorientiert visuell wahrnehmen‘).

(iv) Identifizierung von typisch gesprochen sprachlichen, kombinatorischen Eigenschaften: Mögliche Fälle sind solche mit unterschiedlichen kombinatorischen Eigenschaften, die explizit in den FOLK-Daten belegt werden konnten, aber in den konsultierten Wörterbüchern nicht oder nur unzureichend explizit kodifiziert sind. Diese Fälle umfassen verschiedene Bereiche:

- Informationen zu den Strukturmustern und nicht realisierten Aktanten (z.B. für *kriegen* in der Bedeutung ‚bekommen‘ konnte die Rolle „Sender“ nicht belegt werden),
- Informationen zu den Strukturmustern und der Verwendung von zusätzlichen deiktischen Elementen (vgl. *gucken* und *schauen* in der Bedeutung ‚richtungsorientiert visuell wahrnehmen‘),
- Informationen zu typischen Verbindungen/Kollokationen (So zeigt z.B. *wissen* in der Bedeutung ‚informiert sein‘ eine sehr häufige Verbindung mit den Partikeln *ja* und *aber* auf. Das Adjektiv *gut* in der Bedeutung von ‚positiv bewertet‘ tritt häufig mit *echt* und das Adverb *eben* mit *gerade* auf.),

- Informationen zu festen Wendungen (z.B. *keine Ahnung haben*, *gucken wo jemand bleibt*, *frei kriegen*, *Besuch kriegen*, *die Kurve kriegen*) oder den Routiniformeln (z.B. *guten Appetit*, *danke schön*).

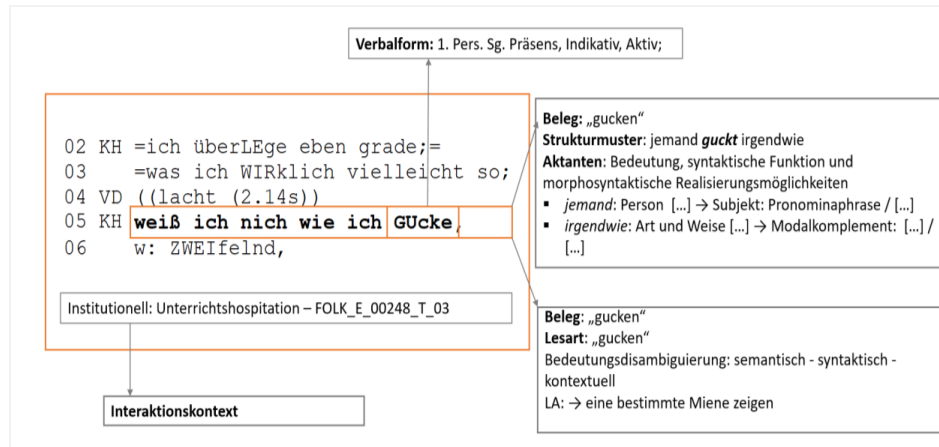


Abb. 2: Illustration ausgewählter Analyseparameter für die Kodierung (Modul 1)

(b) Modul 2

Die Analysen für Modul 2 basieren auf der Lemma-Stichprobe von 100 gültigen Belegen, die Hinweise auf relevante Phänomene liefert (vgl. Abb. 3). Die Belege wurden daraufhin geprüft, ob und welche Kandidaten sich für eine interaktional ausgerichtete Beschreibung, so wie sie in Modul 2 erfolgt, eignen, d.h. Einheiten mit einer spezifischen kommunikativen Funktion im Gespräch wie z.B. die verschiedenen Muster zum Verb *wissen*: *ich weiß nicht*, *weißt du*, *ich weiß*, *wer weiß*, *was weiß ich*. In den Fällen, in denen die Stichprobe zwar Hinweise auf verfestigte Muster gibt, diese jedoch in nicht ausreichender Anzahl vorliegen, um belastbare Aussagen treffen zu können, wurden separate, gezielte Abfragen in FOLK generiert (z.B. *was weiß ich*: 2 Belege in der Stichprobe; eine Suche in der DGD ergibt insgesamt weit mehr Belege). Für die Analysen für Modul 2 wurde das grundlegende Kodierschema für die qualitativen Einzelfallanalysen angepasst und um weitere, für das jeweilige Phänomen relevante formale und funktionale Parameter ergänzt. Anschließend wurden die ermittelten typischen, rekurrenten und verfestigten Form-Funktions-Zusammenhänge unter Berücksichtigung der Sequenz und Interaktionskontexte beschrieben. Dabei wurde versucht, das Informationsangebot sprachlich, inhaltlich und gestalterisch an die Erfordernisse einer lexikografischen Umgebung anzupassen.

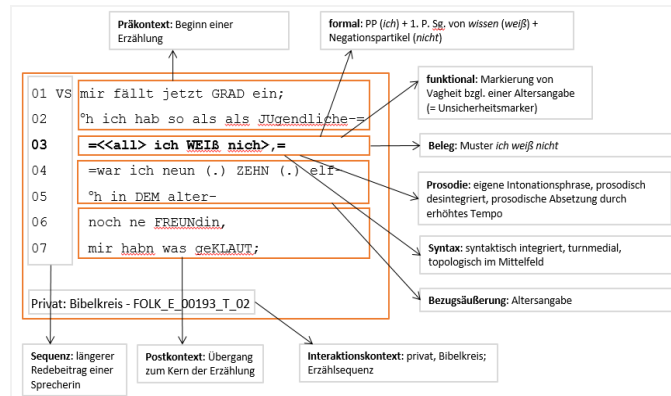


Abb. 3: Illustration ausgewählter Parameter für die Kodierung (Modul 2)

Da sich die Analysen beider Module zunächst auf die jeweilige Lemma-Stichprobe (100 gültige Belege) beziehen, wurden auch nur solche Phänomene (Bedeutungen, kommunikative Muster, Funktionen) beschrieben, die hier belegt werden konnten. Solche Phänomene, die ggf. Gegenstand von Beschreibungen in anderen wissenschaftlichen Einzelstudien und/oder lexikografischen Werken sind, aber nicht in der Stichprobe dokumentiert werden konnten, sind daher nicht Teil des LeGeDe-Informationsangebots.

3.2.2 Überblicksartikel

schauen

Das Vollverb **schauen** konnte in der LeGeDe-Stichprobe in acht verschiedenen Lesarten und in zwei verschiedenen interaktionalen Mustern belegt werden. Die Grundbedeutung von **schauen** steht in Zusammenhang mit der visuellen richtungsorientierten Wahrnehmung. Diese konnte auch in anderen Bedeutungsvarianten von **schauen** in Zusammenhang mit den Lesarten 'überprüfen', 'ansehen' und 'suchen' nachgewiesen werden. Die visuelle Wahrnehmungsbedeutung verliert deutlich in den Bedeutungsvarianten, in denen **schauen** in den Lesarten 'abwarten', 'abwägen', 'sich bemühen' und 'achten auf' auftritt. In diesen Fällen dominiert teilweise eine kognitiv-mentale Bedeutungskomponente. Neben dem Muster **schau mal** als Diskursmarker tritt das Muster **mal schauen** zur Verschiebung einer Handlung in die Zukunft teilweise als Beendigungssignal auf.

Frequenz
FOLK 2.11 (absolut): 669

FOLK-Zugriff
schauen (in FOLK)
schauen (im Lexical Explorer)

Bedeutungen	Funktionen im Gespräch
schauen [VOLLVERB] <ul style="list-style-type: none"> ◦ richtungsorientiert visuell wahrnehmen ◦ überprüfen ◦ ansehen ◦ abwarten ◦ abwägen ◦ sich bemühen ◦ suchen ◦ achten auf 	schau mal [VERBALSYNTAGMA] <ul style="list-style-type: none"> ◦ Diskursmarker mal schauen [VERBALSYNTAGMA] <ul style="list-style-type: none"> ◦ Verschiebung einer Handlung in die Zukunft

Information in anderen lexikografischen Ressourcen

- Online-Wortschatz-Informationssystem Deutsch (OWID)
- Elektronisches Wörterbuch zur Verbalenz (eValbu)
- Digitales Wörterbuch der Deutschen Sprache (DWDS)

Forschungsliteratur

- Günthner (2017)
- Imo (2007)
- Wegner (2015)

Abb. 4: Überblicksartikel zum Lemma *schauen*

Für jedes Stichwort stehen allgemeine Übersichtsinformationen zur Verfügung und bieten in beschreibender Form bedeutungs- und funktionsorientierte Informationen (z.B. *schauen*, vgl. Abb. 4). Eine klare modulare Aufteilung der Informationen ermöglicht einerseits die Darstellung lexikalisch-semantischer Informationen, die sich an der jeweiligen Bedeutung der entsprechenden Lesarten eines Lemmas (=Modul 1) orientieren, und andererseits das Angebot funktionspezifisch interaktiv orientierter Informationen (=Modul 2). Für beide Module wurden spezifische lexikografische Angabeklassen verwendet oder neu entwickelt, die neben herkömmlichen Wörterbuchinformationen völlig neue Erkenntnisse und Formate bieten. Unterschiedliche Querverbindungen zwischen den beiden Modulen stehen durch eine interne Verknüpfung explizit zur Verfügung (vgl. Abschnitt 3.4). Ein erweitertes externes Informationsangebot wird durch Links bereitgestellt, die einerseits zu weiteren lexikografischen Ressourcen und andererseits zu FOLK und dem „Lexical Explorer“ führen (vgl. Abschnitte 2.4, 3.4). Zusätzlich wird die berechnete korpusbasierte Häufigkeitsklassendifferenz zwischen den Stichwörtern in den jeweiligen Korpora (geschrieben: DEREKO, gesprochen: FOLK) visualisiert (vgl. Abb. 5).

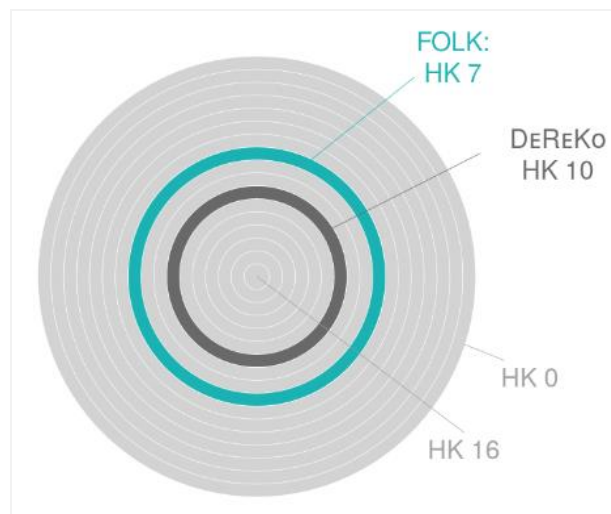


Abb. 5: Frequenzvisualisierung (FOLK/DEREKO) zum Lemma *schauen*

3.2.3 Modul 1

Bewegen sich Nachschlagende aus der alphabetisch angeordneten Stichwortliste (siehe linker Menübalken) oder aus dem Überblicksartikel (vgl. Abschnitt 3.2.2) zum Hauptlemma in Modul 1 (Bsp. *schauen*), so werden im Wortartikelkopf zunächst lesartübergreifende Informationen (zu Form, Varianz und sehr zen-

trale, lemmabezogene Forschungsliteratur) angezeigt. Darunter sind die unterschiedlichen Bedeutungen mit einem kurzen Label (z.B. ‚abwarten‘) und einer kurzen Bedeutungsbeschreibung aufgelistet (vgl. Abb. 6). Über das Pfeilsymbol neben dem Label klappt sich auf Klick die detaillierte, lesartspezifische Beschreibung auf.

The screenshot shows a dictionary entry for the verb 'schauen'. At the top, it is labeled 'schauen [VOLLVERB]'. Below this, there are three columns: 'Form' (with '---' below it), 'Varianz' (with '---' below it), and 'Forschungsliteratur' (with a list of references: Günthner (2017), Imo (2007), and Wegner (2015)). The main content consists of four entries, each with a red square icon and a right-pointing arrow:

- richtungsorientiert visuell wahrnehmen** ▶
Mit der Grundbedeutung von **schauen** wird ausgedrückt, dass jemand etwas durch einen visuellen Eindruck wahrnimmt, indem er oder sie den Blick dabei bewusst auf etwas oder jemanden richtet.
- überprüfen** ▶
Mit **schauen** wird ausgedrückt, dass jemand etwas durch eigene visuelle Wahrnehmung in Verbindung mit einer bewussten Rechercheaktivität überprüft bzw. feststellt oder beobachtet.
- ansehen** ▶
Bei **schauen** in der Bedeutung 'sich etwas ansehen/anschauen' richtet sich die visuelle Wahrnehmung darauf, sich etwas (z.B. einen Film) bewusst in einem entsprechenden Medium anzusehen. Wenn konkret das Medium 'Fernseher' thematisiert wird, dann erhält das Verb die spezifische Bedeutung 'fernsehen'.
- abwarten** ▶
Bei **schauen** zum Ausdruck von einem zukunftsorientierten Abwarten in Bezug auf das ungewisse Eintreten eines Sachverhaltes oder der Durchführung einer zukünftigen Handlung verblasst die visuelle Wahrnehmungsbedeutung deutlich.

Abb. 6: Ausschnitt aus Modul 1 zu *schauen* mit übergreifenden Informationen („Kopf“) und einer Kurzübersicht zu den einzelnen Lesarten

Eine weitere Möglichkeit des Zugriffs direkt auf eine bestimmte Lesart kann über den Überblicksartikel erfolgen. Hier können Nachschlagende auf das Label einer Lesart klicken und gelangen zum ausgeklappten kompletten Wortartikelinhalt.

Der Wortartikel beinhaltet neben Label und einer Bedeutungsbeschreibung die folgenden Bereiche:

- **Belegblock:** Ein illustrierender Belegausschnitt aus FOLK mit Belegtitel, Transkriptausschnitt als Minimaltranskript aus der DGD, Belegkontext und optional einer weiteren Erläuterung zum Belegausschnitt ergänzen die Bedeutungsbeschreibung. Unterhalb des Transkripts steht die FOLK-ID, über die angemeldete DGD-Nutzer/-innen direkt in die DGD zum entsprechenden Transkriptausschnitt gelangen, darin vertieft recherchieren und die Audiodatei anhören können (vgl. Abb. 7).

[1] **schauen** in der Bedeutung 'abwarten'

NI 0103 **AZ** weiß
 NI 0104 (0.84)
 NI 0105 **DZ** ja weißwein auch
 NI 0106 (5.23)
 NI 0107 **DZ** packst du die für mich ein
 NI 0108 (1.8)
 NI 0109 **AZ** ja
 NI 0110 (1.04)
 NI 0111 **AZ** kannst hier einfach reinstecken
 NI 0112 (1.46)
 NI 0113 **AZ** **schau**n wie lang das hier dauert
 NI 0114 (1.54)
 NI 0115 **DZ** die ham
 NI 0116 (10.15)
 NI 0117 **AZ** nja ja
 NI 0118 (1.89)
 NI 0119 **AZ** na des is h° genau is kein gemütliches thema is gut formuliert
 NI 0120 (3.96)
 NI 0121 ((Ein Mann bezahlt seine Getränke, 7,16 Sek.))
 NI 0122 (1.49)
 NI 0123 **AZ** hätt gern_n weißwein

Belegkontext
 Während einer Theaterpause warten Mutter (AZ) und Tochter (DZ) darauf, ihre Getränkebestellung aufgeben zu können, und unterhalten sich dabei.

Privat: Pausenkommunikation im Theater - FOLK_E_00246_T_01

Abb. 7: Belegblock im Artikel *schauen*, Bedeutung ‚abwarten‘

- **Formbesonderheiten:** Ein optionaler lexikografischer Hinweistext erläutert Besonderheiten auf formaler und kombinatorischer Ebene (vgl. Abb. 8).

Der auffordernde Gebrauch in der 1. Person Plural (**schau(e)n wir mal**) verweist auf eine Art zuständigkeitsorientierte Einladung an alle Interagierenden, abzuwarten, was in Zukunft passieren wird. Gleichzeitig konnte in allen Belegen eine Kombination mit der Modalpartikel **mal** dokumentiert werden, die der Äußerung eine gewisse Unverbindlichkeit verleiht.

Abb. 8: Formbesonderheiten im Artikel *schauen*, Bedeutung ‚abwarten‘

- **Kombinatorik:** Diese Information wird durch die Beschreibung der Strukturmuster, der festen Verbindungen (im weiten Sinne) sowie der Kollokationen und der interaktionalen Einheiten differenziert. Kurze Belegblöcke illustrieren in allen Fällen die aufgeführten Verwendungsweisen.

(i) Die **Strukturmuster** (vgl. Abb. 9a) werden durch eine abstrahierte, formelhafte Information angeboten. Die einzelnen Aktanten, aus denen die Strukturmuster zusammengesetzt sind, werden bezüglich ihrer semantischen Rolle, ihrer syntaktischen Funktion und den morphosyntaktischen Realisierungsmöglichkeiten genauer erfasst, wobei Besonderheiten angegeben werden.

jemand schaut, (was sein wird) [STRUKTURMUSTER]

- **jemand:** Person, die durch ihre kognitiv-mentale Tätigkeit abwartet, ob/wie etwas bzw. was sein wird
SYNTAKTISCHE POSITION: Subjekt
- **was sein wird:** ungewisser zukünftiger Sachverhalt, auf den sich eine Person in abwartender Haltung bezieht
SYNTAKTISCHE POSITION: satzförmiges Komplement: fakultativ
REALISIERUNGSFORMEN: **ob** oder **was/wie**-Satz
ANMERKUNG: Wenn dieses satzförmige Komplement nicht realisiert wird, wird das Ende einer thematischen Sequenz markiert.

mal **schauen**, was dann passiert

MI 1169 SK mal **schauen** was dann pa[ssiert]

Privat: Spielinteraktion zwischen Erwachsenen - FOLK_E_00021_T_11

mal **schauen**, wie er es macht

MI 1248 BP könnt man ja so linien machen und dann fotos damit verbinden zum beispiel

MI 1249 (0.29)

MI 1250 BP hab ich ihm geschrieben

MI 1251 (0.29)

MI 1252 BP *hh ja s hat er

MI 1253 (0.46)

MI 1254 BP dankend angenommen un (.) mal **schauen** wie er_s macht

Privat: Tischgespräch - FOLK_E_00271_T_01

Abb. 9a: Strukturmuster zu *schauen*, Bedeutung ‚abwarten‘

(ii) Unter dem Oberbegriff **Feste Wendungen/Kollokationen** (vgl. Abb. 9b) werden im LeGeDe-Prototyp verschiedene Arten von mehr oder weniger festen (idiomatischen) lexikalischen Einheiten, ohne eine weitere Spezifizierung bzw. terminologische Präzisierung vorzunehmen, zusammengefasst. Die entsprechenden Kommentare umfassen Aspekte der Bedeutung und des Gebrauchs.

schön wär's [FESTE WENDUNGEN / KOLLOKATIONEN]

Die Verbindung **schön wär's** kommt in der LeGeDe-Stichprobe als alleinstehende Äußerung vor. Sie wird von einem Sprechenden A als Reaktion auf eine davorstehende Äußerung eines Sprechenden B eingesetzt. Der/Die Sprechende A kommentiert die zuvor gebrachte Äußerung des Sprechenden B in einer Weise, dass es aus seiner/ihrer Sicht so ist, dass der Sachverhalt, auf den referiert wird, so nicht zutrifft oder umgesetzt werden kann. Die Konjunktivform von **sein** in der Verbindung **schön wär's** deutet an, dass es der Sprechende A theoretisch als erfreulich einstufen würde, wenn die Äußerung doch zutreffen würde oder umgesetzt werden könnte. Mit der Verbindung **schön wär's** schließt er/sie dies aber im Sinne von 'es wäre positiv, wenn der Fall einträfe, dies scheint aber nicht realistisch' und damit eines hypothetischen Wunsches aus.

schön wär's

MI 0044 KA zum zweiten mal schon das gleiche falsch gedacht (.) vor allm

MI 0045 (0.25)

MI 0046 PA wenn das zusammen nimmt hast dann richtig gedacht oder

MI 0047 KA ((Lachansatz)) schön wär_s ((lacht))

MI 0048 AM minu[s] mal minus macht plus

MI 0049 KA [*h ich glaub das w]ird nix

Privat: Spielinteraktion zwischen Erwachsenen - FOLK_E_00132_T_16

Abb. 9b: Feste Wendungen/Kollokationen zu *schön* (Adj.), Bedeutung ‚erfreulich‘

(iii) Die Auflistung von festen Verwendungsmustern, die als **interaktionale Einheiten** identifiziert werden konnten (vgl. Abb. 9c), ermöglicht die Einrichtung einer internen Verlinkung (siehe mal schauen unterstrichen in Abb. 9c) zu interaktionalen Funktionalitäten, die in Modul 2 ausführlich beschrieben werden.

mal schauen : Verschiebung einer Handlung in die Zukunft [INTERAKTIONALE EINHEIT]

Mit **mal schauen** schieben Sprechende eine Handlung, Entscheidung, Einschätzung, Bewertung, Planung o.Ä. zeitlich auf. Damit dient **mal schauen** häufig als (vorläufiges) Beendigungssignal.

Abb. 9c: Ausschnitt zu interaktionalen Einheiten zu *schauen*, Bedeutung ‚abwarten‘

— **weitere Auffälligkeiten:** In einem zweiten optionalen Textfeld werden auffällige Analyseergebnisse zu Metadaten und Frequenzen kommentiert (vgl. Abb. 10).



Über 90% der Belege der LeGeDe-Stichprobe konnten dieser Lesart zugeordnet werden. In dieser Bedeutungsvariante wird **wissen** nur selten in öffentlichen Interaktionssituationen belegt. Außerdem konnte **wissen** als Teil eines interaktionalen Musters in einem Viertel der Belege in Verbindung mit dieser Lesart dokumentiert werden. In diesen Fällen werden die Argumente des Strukturmusters nicht immer alle realisiert und es treten zusätzliche interaktionale Funktionen auf.

Abb. 10: Weitere Auffälligkeiten zu *wissen*, Bedeutung ‚informiert sein‘

3.2.4 Modul 2

In Modul 2 werden die Funktionen sowohl von ein- als auch von mehrteiligen Lemmata beschrieben. In einem Artikelkopf sind funktionsübergreifende Informationen zu Form, Varianz, Kombinatorik und Forschungsliteratur für Nachschlagende aufbereitet. Direkt angeschlossen werden die einzelnen Funktionalitäten mittels eines kurzen Labels und einer kurzen Funktionsbeschreibung angeboten (vgl. Abb. 11).

schön [DISKURSPARTIKEL]			
Form	Varianz	Kombinatorik	Forschungsliteratur
---	---	sehr schön, voll schön, ja schön, ah schön, oh schön, ach schön, na schön	• Willkop (1988)
<i>Die Partikel schön kann um vorangestellte Intensitätspartikeln (sehr, voll), Interjektionen (ah, oh, ach) oder Diskurspartikeln (ja) ergänzt werden. Diese können jeweils eigenständige Funktionen übernehmen (ja schön, ah schön, oh schön, ach schön) oder die Bedeutung/Funktion von schön modifizieren (ja sehr schön, voll schön, na schön).</i>			
Bewertung ▶			
Mit schön nehmen Sprechende Bezug auf eine vorausgegangene Äußerung und bewerten den Inhalt dieser Äußerung oder eine Handlung als positiv (vgl. dazu auch schön in den Bedeutungen passend und erfreulich).			
Abschlussmarker ▶			
Mit schön zeigen Sprechende an, dass eine vorausgegangene kommunikative Aufgabe ausreichend bearbeitet wurde, um diese abzuschließen und in der Interaktion fortzufahren.			

Abb. 11: Artikelkopf und Funktionsbeschreibungen im Überblick zu schön als Diskurspartikel

Die funktionsspezifischen Wortartikel in Modul 2 beinhalten das im Folgenden erläuterte Informationsangebot:

- **Belegblock:** Die Funktionsbeschreibung wird durch einen oder mehrere Belege illustriert (vgl. Abb. 12). Im Unterschied zu Modul 1 wird bei den Belegen zur Funktionsbeschreibung eine obligatorische sequenzanalytische Kommentierung angeboten. Für interaktionslinguistische, formfunktionsbasierte Analysen sind prosodische Parameter hochrelevant. Aus diesem Grund wurden die Transkripte, die in Modul 2 angeboten werden, um prosodische Informationen ergänzt und somit zu Basistranskripten ausgebaut. Wenn sich Nutzer/-innen auch mit dieser Transkriptnotation noch näher vertraut machen wollen, finden sie dazu in den Umtexten (vgl. Abschnitt 3.4) vertiefendes Material.

[1] schön zur Bewertung des Inhalts eines Berichts

MI 001 MD ^h aber [heut mo]rgen WAR er dann mal im gang,
MI 002 MP [^h]
MI 003 MD hat dann sein BETT auch frisch be[zogen gl]eisch heut morgen,
MI 004 MP [h^öi,]
MI 005 MD ^h ähm WA:R,
MI 006 (θ.ɜ)
MI 007 MD nicht (.) g der GANZ erste in der visite,
MI 008 MD aber der ZWEITE heut in der visite,]
MI 009 ME [((räuspert sich))]
MI 010 MD un nach der visite gleich nach HAUSE ab-
MI 011 MD ^h
MI 012 MP sch[ön];]
MI 013 MD [un kom]mt morgen abend um ACHT wieder;
MI 014 MD der is auch über NACHT weg.]
MI 015 MP [sch^ön];
MI 016 (6.θɜ)

Belegkontext
In einer Schichtübergabe im Krankenhaus berichtet die Krankenschwester MD den anderen Mitarbeitenden von einem Patienten, der aufgrund seiner Medikation immer Mühe hat, morgens früh aufzustehen.

Analyse
MD schildert detailliert das veränderte, positive Verhalten des Patienten. Während ihres Berichts gibt die Kollegin MP Rückmeldungen und bewertet in Überlappung mit den noch laufenden Darlegungen von MD die außergewöhnliche Tatsache, dass der Patient heute Morgen bereits früh aktiv war und erst am nächsten Tag wiederkommt, jeweils mit "schön;" (Z. 012, 015) als positiv.

Institutionell: Schichtübergabe in einem Krankenhaus - FOLK_E_00114_T_01

Abb. 12: Belegblock im Artikel *schön* (Diskurspartikel), Funktion ‚Bewertung‘

- **Funktionsabstrahierung:** Dieser lexikografische Kommentar fasst die jeweils beschriebene Funktion auf einer allgemeineren Ebene und in Rückgriff nicht nur auf die illustrierenden Belege, sondern auf zahlreiche analysierte relevante Belege, zusammen und bildet eine zentrale Informationseinheit in Modul 2 (vgl. Abb. 13).


Mit **schön** nehmen Sprechende Bezug auf eine vorausgegangene Äußerung und bewerten den Inhalt dieser Äußerung oder eine Handlung als positiv bzw. in Lehr-Lern-Interaktionen als korrekt. Mit dieser Bewertung auf Sachverhaltsebene drücken sie ihre Haltung dem jeweiligen Bewertungsgegenstand gegenüber aus, die vor allem durch eine entsprechende prosodische Realisierung unterschiedlich stark emphatisch aufgeladen werden kann. Dabei wird **schön** als positive, emphatische Bewertung häufig dazu genutzt, Anteilnahme sowie subjektive Stellungnahmen und Empfindungen expressiv auszudrücken, und trägt als solche zur Diskursorganisation bei. Dies geschieht am häufigsten als Reaktion auf längere Äußerungen wie Erzählungen, Berichte oder Beschreibungen sowie in dritter Position meist nach einer Frage-Antwort-Sequenz. Eine spezifische Verwendung tritt dabei in schulischen Interaktionen auf, wenn mit **schön** eine Schülerantwort innerhalb einer dreischrittigen Sequenz aus Frage-Antwort-Evaluierung als korrekt bewertet wird. Auffällig ist, dass **schön** sehr häufig als Bewertung verwendet wird. Dabei wird **schön** auch mit weiteren Partikeln wie **oh**, **ah** oder **och** und Intensitätspartikeln wie **sehr** oder **voll** kombiniert sowie häufig durch eine emphatische prosodische Gestaltung hervorgehoben.

Abb. 13: Funktionsabstrahierung im Artikel *schön* (Diskurspartikel), Funktion ‚Bewertung‘

- **Syntax-Sequenz-Realisierung** und **Prosodie:** Kurze Informationen zur Syntax- und Sequenz-Realisierung (vgl. Abb. 14) und zur Prosodie (vgl. Abb. 15), die wie bei *ja* und *eben* als Modalpartikel vs. Diskurspartikel auch distinktiv sein können, stehen zusammen mit illustrativen kurzen Belegausschnitten den Nachschlagenden zur Verfügung.

Syntax-Sequenz-Realisierung

mal schauen wird meistens ohne weitere Komplemente realisiert. Es kann turnfinal stehen und den Sprecher beenden.

 **mal schauen** turnfinal, ohne Komplement


MI 001 **JA** ich nehm erst mal ne KARTE;
MI 002 **JA** un dann mal SCHAUen;

Privat: Spielinteraktion zwischen Erwachsenen - FOLK_E_00132_T_12

Abb. 14: Syntax-Sequenz-Realisierung im Artikel *mal schauen*, Funktion ‚Verschiebung einer Handlung in die Zukunft‘

Prosodie

ja wird immer prosodisch integriert und unbetont realisiert.

 **ja** prosodisch integriert, unbetont

MI 001 **AC** ich kann **ja** mal FRAGn;
MI 002 **AC** aber eigentlich geht sie ned da HIN.

Privat: Kochinteraktion - FOLK_E_00332_T_01

Abb. 15: Prosodie im Artikel *ja* (Modalpartikel), Funktion ‚Abschwächung‘

- **weitere Auffälligkeiten:** Optionale Informationen zu weiteren Auffälligkeiten, die hauptsächlich in Verbindung mit den Metadaten stehen, reflektieren in Form eines lexikografischen Kommentares Beobachtungen, die aus der Analyse der Korpusbelege hervorgegangen sind (vgl. Abb. 16).


 **schön** als Abschlussmarker tritt überwiegend in handlungsschematischen Interaktionen wie Unterrichtsstunden, Planungen oder Spielinteraktionen auf, in denen einzelne Themen und Handlungen aufeinander aufbauen und systematisch nacheinander bearbeitet werden.

Abb. 16: Weitere Auffälligkeiten im Artikel *schön* (Diskurspartikel), Funktion ‚Abschlussmarker‘

3.3 Zugriffsmöglichkeiten und Vernetzungsstrukturen

Die Zugriffsmöglichkeiten auf die vielfältige multimediale Information im LeGeDe-Prototyp sind unterschiedlich und orientieren sich jeweils an der Art der Information. Die Umsetzung berücksichtigt dabei die aktuellen Anforderungen an elektronische Wörterbücher und wird auch den Erwartungen der Befragten aus den LeGeDe-Studien gerecht (vgl. Dziemianko 2018; vgl. insbesondere These 7 in Engelberg, Klosa-Kückelhaus und Müller-Spitzer 2019: 33; Engelberg, Müller-Spitzer und Schmidt 2016; Meliss, Möhrs und Ribeiro Silveira 2018). Zunächst lassen sich wörterbuchinterne von externen Zugriffsmöglichkeiten und Vernetzungsstrukturen unterscheiden.

(a) Die Nachschlagenden können sich über eine Vielzahl von **wörterbuchinterne** Zugriffs- und Vernetzungsstrukturen innerhalb des LeGeDe-Prototyps bewegen. Dadurch ist die Möglichkeit gegeben, Wörterbuchinhalte über verschiedene Zugriffspunkte und Verknüpfungen zwischen dem Bedeutungs- (Modul 1) und dem Funktionsmodul (Modul 2) anzusteuern sowie Glossar- und Literatureinträge aus den Umtexten (vgl. Abschnitt 3.4) direkt abzurufen (vgl. Tab. 1). So können Nachschlagende beispielsweise direkt von der Seite „Übersicht: Index“ auf einen Überblicksartikel oder Modul 1 bzw. 2 eines Lemmas gelangen. Die Lemmaliste ist beim Navigieren zwischen verschiedenen Wortartikeln immer sichtbar, sodass hierdurch bequem zwischen verschiedenen Wortartikeln gewechselt werden kann.

Beschreibung	Verlinkung/Verweis
<p>Übersicht Index Überblick zu allen LeGeDe-Lemmata, den jeweiligen Kurzlabeln der Module 1 und 2 zu Bedeutungen/Funktionen sowie kurze Auszüge aus der Mikrostruktur (z.B. Phonetik, Kombinatorik)</p>	<p><u>Ort:</u> LeGeDe-Startseite und immer sichtbar im linken Menüband <u>Ziel (nach Klick):</u> der jeweils ausgewählte Linkort</p>
<p>Modul 1 → Modul 2 Wörterbuchinterner Verweis zu einer Funktion in Modul 2; der Verweis deutet darauf hin, dass eine inhaltliche Verknüpfung zwischen den Modulen vorliegt.</p>	<p><u>Ort:</u> Kombinatorik → Interaktive Einheit als Form ist verlinkt <u>Ziel (nach Klick):</u> Wörterbuchartikelteil zur relevanten Funktion in Modul 2</p>
<p>Modul 2 → Modul 1 Wörterbuchinterner Verweis zu einer Bedeutung in Modul 1; der Verweis deutet darauf hin, dass eine inhaltliche Verknüpfung zwischen den Modulen vorliegt.</p>	<p><u>Ort:</u> Funktionsbeschreibung; am Ende des Satzes ist das relevante → Bedeutungslabel aus Modul 1 verlinkt <u>Ziel (nach Klick):</u> Wörterbuchartikelteil zur relevanten Bedeutung in Modul 1</p>

<p>Glossar-Stichwörter Erläuterungen eines Fachbegriffes</p>	<p><u>Ort</u>: innerhalb des Wörterbuchartikels (Modul 1, Modul 2); im Fließtext weist eine gestrichelte Linie auf den Link hin <u>Ziel (nach Klick)</u>: Pop-up Fenster mit der Erläuterung, die auch im Umtext „Glossar“ zu finden ist</p>
<p>Kurztitel: Forschungsliteratur Zentrale Forschungsliteratur, die für das jeweilige Lemma relevant ist</p>	<p><u>Ort</u>: 1. Überblicksartikel, 2. Artikelkopf in Modul 1 und 2 <u>Ziel (nach Klick)</u>: Pop-up Fenster mit dem vollständigen bibliografischen Hinweis, der auch im Umtext „Literatur“ zu finden ist</p>

Tab. 1: Wörterbuchinterne Zugriffs- und Vernetzungsstrukturen

(b) Die **externen** Zugriffs- und Vernetzungsstrukturen im LeGeDe-Prototyp führen zu weiteren Ressourcen und lassen sich in drei große Bereiche aufteilen (vgl. Tab. 2⁷): (i) Einerseits geht es um Ressourcen, die eine direkte Verbindung zu der Datengrundlage des LeGeDe-Prototyps herstellen (FOLK, die DGD und der „Lexical Explorer“). (ii) Andererseits handelt es sich um externe Onlinewörterbücher, über die die lexikologische Information zu den einzelnen Stichwörtern erweitert werden kann. (iii) Neben diesen Verknüpfungen wird außerdem ein externer Link zu einer fachterminologischen Datenbank in dem lexikografischen Umtext „Glossar“ angeboten. Die Tabelle 2 resümiert die Information in Verbindung mit dem Ort, an dem die Verlinkung angesteuert werden kann und den möglichen Erreichbarkeitsbedingungen.

Ressource	Verlinkung	Erreichbarkeit
<p>DGD (Datenbank)</p>	<p><u>Ort</u>: LeGeDe-Menüleiste <u>Ziel (nach Klick)</u>: Anmeldeseite der DGD</p>	<p>mit Anmeldung über persönlichen DGD-Account</p>
<p>FOLK in DGD (Korpus abrufbar über Datenbank)</p>	<p><u>Ort</u>: Überblicksartikel <u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis in FOLK über die DGD zum jeweiligen Lemma</p>	

	<p><u>Ort</u>: Belegblock innerhalb der einzelnen Bedeutungen (Modul 1) oder Funktionen (Modul 2)</p> <p><u>Ziel (nach Klick)</u>: belegbezogen: Suchergebnis in FOLK über die DGD zum jeweiligen Transkriptausschnitt</p>	
Lexical Explorer (Tool)	<p><u>Ort</u>: Überblicksartikel</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis im "Lexical Explorer" zum jeweiligen Lemma</p>	ohne Anmeldung
OWID (Wörterbuchportal)	<p><u>Ort</u>: Überblicksartikel</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis in OWID zum jeweiligen Lemma</p>	
eValbu (elekt. Valenzwörterbuch)	<p><u>Ort</u>: Überblicksartikel bei Verben, wenn das Stichwort im eValbu verzeichnet ist</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis im eValbu zum jeweiligen Lemma</p>	
DWDS (lexikografisches Informationssystem)	<p><u>Ort</u>: Überblicksartikel</p> <p><u>Ziel (nach Klick)</u>: lemmabezogen: Suchergebnis im DWDS zum jeweiligen Lemma</p>	
grammis (grammatisches Informationssystem)	<p><u>Ort</u>: Glossar</p> <p><u>Ziel (nach Klick)</u>: stichwortbezogen: Suchergebnis in grammis zum jeweiligen Glossarstichwort</p>	

Tab. 2: Verlinkungen aus dem LeGeDe-Prototyp zu externen Ressourcen

3.4 Umtexte

Die verschiedenen Umtexte im LeGeDe-Prototyp bieten ein umfassendes Informationsangebot an, welches sich an unterschiedliche Nutzergruppen wendet und die verschiedensten Möglichkeiten einer internetbasierten Ressource ausnutzt (vgl. Klosa und Gouws 2015).

Der Informationstext „Über LeGeDe“ stellt in ausführlicher Form die theoretischen und methodologischen Grundlagen des LeGeDe-Projekts dar und bietet detailliert Information zu der lexikografischen Mikrostruktur und den einzelnen Angabeklassen. In seiner Ausführlichkeit und Komplexität ist dieser

Text besonders an ein Fachpublikum gerichtet, das neben dem Interesse an der Ressource an sich, Interesse an metalexikografischen Fragestellungen hat.

Im Unterschied dazu sind die „Benutzungshinweise“ so konzipiert, dass sie dem Nachschlagenden anschaulich und zusammenfassend den Zugang zu der Ressource und zu den einzelnen Strukturelementen durch eine „Guided Tour“ ermöglichen. Ein separates Dokument führt zudem in die wichtigsten Transkriptkonventionen ein.

Das „Glossar“ beinhaltet in der Ressource relevante linguistische Fachbegriffe, die zum einen mit einer Kurzdefinition erläutert sind und zum anderen mit einem Verlinkungsangebot zu dem Modul „Wissenschaftliche Terminologie“, das in *grammis* integriert ist, versehen sind. Im Vordergrund steht dabei hauptsächlich das Ziel, den für den LeGeDe-Prototyp verfolgten terminologischen Ansatz zu skizzieren. Der Zugriff auf den entsprechenden Glossareintrag ist dabei entweder direkt über einen entsprechenden Begriff in der Ressource (vgl. Abschnitt 3.3) oder über das alphabetisch angeordnete vollständige Glossar möglich.

Der Umtext „Literatur“ bietet einen Einblick in für die Lemmata im LeGeDe-Prototyp relevante Forschungsliteratur. Zugriffsoptionen existieren entweder über Kurzverweise in den einzelnen Wörterbuchartikeln (vgl. Abschnitt 3.3) oder alphabetisch über eine komplette Liste.

4. Fazit und Ausblick

Abschließend sollen einerseits in Form eines Fazits die Leistungen des LeGeDe-Prototyps aufgezeigt werden. Andererseits wird ein umfangreicher Ausblick geliefert, der auf verschiedene und relevante Aspekte verweist, die bei der Ausarbeitung des LeGeDe-Prototyps nicht berücksichtigt werden konnten, die aber bei zukünftigen lexikografischen Projekten unter Einbindung von korpusbasierten gesprochensprachlichen Daten von Interesse sein können.

4.1 Fazit

Es konnte verdeutlicht werden, dass die neuartigen Aspekte des LeGeDe-Prototyps vielfältig sind und sowohl methodologische als auch inhaltliche, strukturelle und darstellungsorientierte Perspektiven umfassen. Die konkreten innovativen Merkmale der Ressource lassen sich in folgenden Bereichen ausmachen:

- (i) **Datengrundlage:** Die dargebotene Information im LeGeDe-Prototyp basiert ausschließlich auf korpusbasierten gesprochensprachlichen Daten und stellt in diesem Sinne ein Novum dar (vgl. Abschnitt 2.4).
- (ii) **Stichwortkandidatenliste:** Die Erstellung einer Stichwortkandidatenliste auf der Basis von Häufigkeitsklassenvergleichen zwischen zwei Korpora

(DEREKO als Referenzkorpus für das geschriebene Deutsch und FOLK als Korpus des gesprochenen Deutsch in der Interaktion) bietet eine korpusbasierte Grundlage zur Identifizierung von relevanten standardnahen, gesprochen sprachlichen lexikalischen Elementen (vgl. Abschnitt 3.1).

(iii) **Erwartungshaltungen:** Die Neukonzipierung einer lexikografischen Ressource zur Darstellung von gesprochensprachlichen Spezifika, für die kaum auf Vorbilder zurückgegriffen werden konnte, ermöglichte die konkrete Berücksichtigung von bestimmten empirisch erhobenen Erwartungshaltungen der zukünftigen Benutzer/-innen (vgl. Abschnitt 2.5).

(iv) **Lexikografische Angabeklassen:** Das Informationsangebot zu den angebotenen lexikografischen Angabeklassen kombiniert klassische und neuartige Formen. Die entwickelten Vorschläge für die Darstellung lexikalischer Phänomene des Gesprochenen in der Interaktion ermöglichen, diese für lexikografische Zwecke adäquat zu strukturieren und zu beschreiben. Die korpusbasierten Daten wurden durch eine spezifische Methode quantitativ ermittelt und qualitativ analysiert und strukturiert (vgl. Abschnitt 3.2.1).

(v) **Verknüpfung mit Korpusdaten:** Das Angebot an authentischen Korpusbelegen erfolgt mittels ausgewählter Transkriptausschnitte, über die eine Schnittstelle zu den Audiodateien und den ausführlichen Informationen zu den Metadaten der DGD hergestellt wird. Damit wird der LeGeDe-Prototyp den technischen Herausforderungen gerecht (vgl. Abschnitte 3.2.3, 3.2.4 und 3.3).

(vi) **Multimedialität:** Der multimediale Charakter der Ressource zeichnet sich v.a. dadurch aus, dass für die Korpusdaten neben den Transkripten Audiodateien und teilweise auch entsprechende Videodateien über den Zugriff auf die DGD zur Verfügung stehen. Die Verknüpfung zu dem Analysetool „Lexical Explorer“ ermöglicht den Benutzer/-innen zusätzlich einen direkten Zugriff auf die der Ressource zugrundeliegenden Korpusdaten und eigene erweiterte Analysemöglichkeiten (vgl. u.a. Abschnitte 2.4 und 3.3).

(vii) **Komplexe Struktur der Informationsvernetzung:** Durch diverse interne und externe Verlinkungsmöglichkeiten wird eine komplexe Informationsvernetzung und -strukturierung angeboten. Der LeGeDe-Prototyp wird auf diese Weise dem Internetmedium mit all seinen Vorteilen gerecht (vgl. Abschnitt 3.3).

4.2 Ausblick

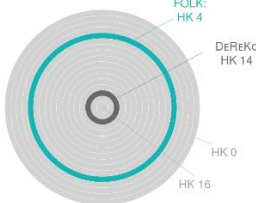
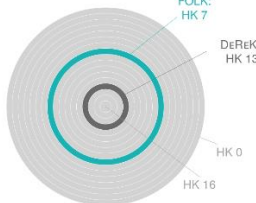
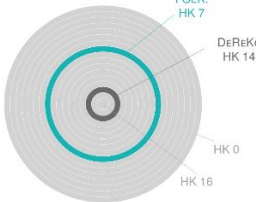
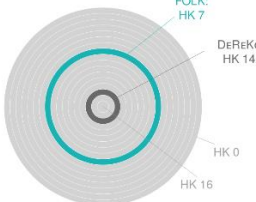
Die vorgestellte lexikografische Ressource liegt zurzeit als Prototyp vor. Die theoretischen und methodologischen Annahmen für die Konzipierung der Ressource sowie die entsprechende technische Umsetzung erlauben jederzeit eine Erweiterung des lexikografischen Angebots. Als Ausblick sollen daher einige Ausbaumöglichkeiten aufgezeigt werden, die nicht nur für eine Erwei-

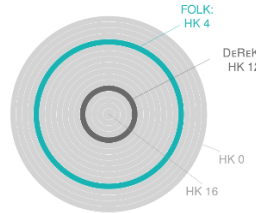
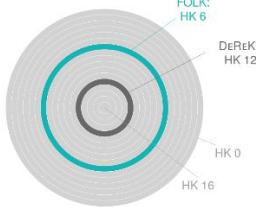
terung des LeGeDe-Prototyps von Interesse sind, sondern auch bei der lexikografischen Aufbereitung von anderen mündlichen Korpusdaten für zukünftige Projekte relevant sein könnten.

Diese potenziellen Ausbaumöglichkeiten betreffen verschiedene Ebenen, aus denen folgende zwei für eine genauere Betrachtung ausgewählt wurden: (i) die Makrostruktur und konkrete Vorschläge für die Erweiterung der Anzahl der (nicht) redaktionell bearbeiteten Stichwörter und (ii) unterschiedliche Recherche- und Zugriffsmöglichkeiten auf die zahlreichen Informationsmodule. Im Folgenden sollen diese Möglichkeiten eingehender beschrieben werden.

4.2.1 Art und Anzahl der Stichwörter

Über den korpusbasierten Stichwortansatz, der für den LeGeDe-Prototyp entwickelt wurde (vgl. Abschnitt 3.1), konnten 322 für die gesprochene Sprache relevante Stichwortkandidaten ermittelt werden, die verschiedenen grammatischen Kategorien angehören (z.B. Partikeln, Interjektionen, Adverbien) und/oder verschiedene Bereiche abdecken (z.B. Routineformeln, Vagheitsausdrücke, deiktische Ausdrücke, Lemmata mit Besonderheiten hinsichtlich Kombinatorik, Stil und Register). Da die LeGeDe-Ressource zunächst als lexikografischer Prototyp konzipiert wurde, erfolgte die Auswahl der behandelten Stichwörter aus der Stichwortkandidatenliste nach Kriterien in Zusammenhang mit dem Anspruch, eine Bandbreite von unterschiedlichen Phänomenen bzw. Erscheinungen abzudecken und dabei gleichzeitig auch die Ergebnisse wissenschaftlicher Teilstudien zu berücksichtigen. Dennoch konnten in der Projektlaufzeit nicht alle Besonderheiten erfasst werden. Bereits die Liste der 25 Stichwortkandidaten aus der Rangliste (vgl. Abb. 1) verweist auf unterschiedliche Ausschnitte, wie z.B. Interjektionen, die im LeGeDe-Prototyp bislang nicht beschrieben worden sind. Die Bearbeitung der Bereiche, die in den LeGeDe-Prototyp der ersten Arbeitsphase nicht aufgenommen werden konnten, bildet zusammen mit der quantitativen Erweiterung der Lemmaanahl zu den schon behandelten Phänomenen ein breites zukünftiges lexikografisches Arbeitsfeld. Daher wäre es durchaus denkbar, neben den redaktionell bearbeiteten Stichwörtern auch eine Auswahl relevanter Stichwörter anzubieten, die redaktionell nicht ausführlich bearbeitet wurden, für die jedoch entsprechende Verlinkungsangebote in der Ressource bereitgestellt werden könnten. Ein solches Angebot würde es interessierten Nutzer/-innen ermöglichen, eine eigenständige Recherche durchzuführen. In Tab. 3 wird exemplarisch zu Stichwörtern der 25-er Rangliste, die unterschiedlichen Phänomenbereichen zugeordnet wurden, ein Vorschlag für ein Informationsangebot mit verschiedenen Verlinkungsangeboten und der Visualisierung des Häufigkeitsklassenvergleichs präsentiert, das den Recherchierenden ohne größeren redaktionellen Aufwand zur Verfügung gestellt werden könnte, um schnell und gezielt quantitative Informationen zu verschiedenen grammatikalischen und metalinguistischen Daten zu erhalten.

Weitere Lemmakandidaten: nicht redaktionell ausgearbeitetes Informationsangebot	
<p>ah</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Interjektionen (<i>ah, ach, oh</i> etc.)</p>	 <p>FOLK-Zugriff ah (in FOLK) ah (im Lexical Explorer)</p>
<p>cool</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Diskurspartikel/ Adjektive// Entlehnungen (<i>okay, cool</i> etc.)</p>	 <p>FOLK-Zugriff cool (in FOLK) cool (im Lexical Explorer)</p>
<p>nachher</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Adverbien//temporale und lokale Deixis (<i>nachher, jetzt; hier, da</i>, etc.)</p>	 <p>FOLK-Zugriff nachher (in FOLK) nachher (im Lexical Explorer)</p>
<p>danke</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Partikeln//Routineformeln (<i>danke, bitte</i> etc.)</p>	 <p>FOLK-Zugriff danke (in FOLK) danke (im Lexical Explorer)</p>

<p>irgendwie</p> <p>Information in anderen lexikografischen Ressourcen OWID DWDS</p> <p>Partikeln//Vagheitsmarker (<i>irgendwie, irgendwas</i> etc.)</p>	 <p>FOLK-Zugriff irgendwie (in FOLK) irgendwie (im Lexical Explorer)</p>
<p>raus</p> <p>Information in anderen lexikografischen Ressourcen OWID</p> <p>Adverbien/Trennbare Verbpartikeln// Formreduktion (<i>raus, rein, rauf</i> etc.)</p>	 <p>FOLK-Zugriff raus (in FOLK) raus (im Lexical Explorer)</p>

Tab. 3: Mögliches Informationsangebot für nicht redaktionell bearbeitete Artikel⁸

4.2.2 Zugriffs- und Suchfunktionen

Bislang erfolgt der Zugriff auf die Wortartikel im LeGeDe-Prototyp alphabetisch-semasiologisch. Die Datenanalysen und -modellierungen wurden jedoch so angelegt, dass auch Zugriffsmöglichkeiten über formale und funktionale Kategorien technisch realisierbar wären. Ein formal-kategoriales, syntaxsequenz- und prosodieorientiertes sowie funktional-strukturiertes Informationsangebot und eine entsprechende Suchanfrage für diese erweiterten Zugriffsarten wurden von den Probanden unserer Umfragen erwünscht und stellen v.a. für anwendungsorientierte Zwecke im DaF- und DaZ-Bereich in Zusammenhang mit dem Nachschlageverhalten für Produktionszwecke einen interessanten Mehrwert dar (vgl. Meliss, Möhrs und Ribeiro Silveira 2019: 111-112; Sieberg 2013), wobei außerdem auch den Anforderungen des GeR zur Kompetenzentwicklung in der mündlichen Interaktion Rechnung getragen würde.

Hierfür schlagen wir u.a. folgende Möglichkeiten vor, die (fast) alle bereits in den aufbereiteten LeGeDe-Daten entsprechend ausgezeichnet sind:

- (i) **rekurrente Muster:** z.B. <mal + visuelles Verb> (*mal schauen, mal gucken, mal sehen*), <Imperativform eines visuellen Verbs + mal> (*guck mal, schau mal*) oder <weiß X> (*weiß der Geier / der Kuckuck*)
- (ii) **grammatische Kategorien:** z.B. formal-kategoriale Auszeichnungen der Wortarten (Nomen, Verb, Diskurspartikel etc.)
- (iii) **Syntax-Sequenz-Merkmale:** z.B. Merkmale zur syntaktischen Integration (desintegriert, integriert etc.), Sequenzposition (initial, final etc.)
- (iv) **Prosodie-Merkmale:** z.B. prosodische Merkmale wie Betonung und prosodische Integration
- (v) **Funktionen in der Interaktion:** Im LeGeDe-Prototyp sind über die einzelnen Funktionslabel die Ansätze für eine funktionale Suchanfrage gelegt. In Meliss (2020a) wird ein erweiterter Vorschlag diesbezüglich vorgelegt, der die Funktionskategorien im LeGeDe-Prototyp mit Informationen aus einschlägiger Fachliteratur verbindet. In Tab. 4 wird exemplarisch eine Auflistung der möglichen sprachlichen Mittel zum „Ausdruck von Vagheit“ zusammen mit weiteren Subfunktionalitäten aufgeführt.

Funktion	Subfunktionalitäten	Beispiele (fett = in LeGeDe-Prototyp beschrieben)
Ausdruck von Vagheit	Unschärfemarker: epistemic hedge	<ul style="list-style-type: none"> • <i>(ich) weiß nicht</i> • <i>was weiß ich</i> • <i>keine Ahnung</i> • <i>weiß der Teufel was</i> • <i>was weiß ich was</i> • <i>irgendwie</i> • <i>oder so</i> • <i>ungefähr</i> • <i>oder so (was)</i> • <i>glaub ich</i>
	Anzeige von Unschärfe eines Beispiels	<ul style="list-style-type: none"> • <i>was weiß ich</i> • <i>keine Ahnung</i> • <i>zum Beispiel</i> • <i>beispielsweise</i> • <i>sagen wir mal</i>
	Anzeige von Unschärfe einer Listung: ‚etcetera-Formeln‘	<ul style="list-style-type: none"> • <i>(und) was weiß ich</i> • <i>keine Ahnung</i> • <i>weiß der Geier (was)</i> • <i>weiß der Kuckuck (was)</i> • <i>und so (weiter) (und so fort)</i> • <i>und und und</i> • <i>bla bla bla</i> • <i>etcetera pp</i>

Tab. 4: Sprachliche Mittel zum „Ausdruck von Vagheit“ (Auswahl)

Die Liste (vgl. Tab 4) setzt sich aus den Lemmata zusammen, die im LeGeDe-Prototyp beschrieben sind, und aus weiteren, die u.a. in Fachstudien angeführt werden. Über ein Informationsangebot dieser Art könnten die Benutzer/-innen aus einer Palette von mehreren Ausdrucksmöglichkeiten auswählen.

Abschließend kann festgehalten werden, dass der LeGeDe-Prototyp dank der unterschiedlichen Zugriffe auf die Korpusdaten und dem strukturierten und selektierten multimedialen Informationsangebot neben seinem primären Ziel der sprachwissenschaftlichen Dokumentation vielfältige Möglichkeiten für anwendungsorientierte Szenarien bietet. Das Glossar und die Benutzungshinweise, die als Umtexte zur Verfügung stehen, leisten dazu u.a. eine adäquate Hilfe. Bedingung dafür ist allerdings, dass die Sprachdidaktiker/-innen diese Nutzungsmöglichkeiten kennen, um eine entsprechende Vermittlerrolle einnehmen zu können (vgl. Meliss und Möhrs 2018, Meliss i. Dr. [a]). Daher soll an dieser Stelle auch für eine intensivere Beschäftigung mit der anwendungsorientierten Korpuslinguistik und der Lexikografie im Studium und für Weiterbildungsangebote plädiert werden. Darüber hinaus zeichnen sich auch für die Lexikografie des Gesprochenen Ansatzpunkte für neue technologische Herausforderungen in Zusammenhang mit der künstlichen Intelligenzforschung, der Spracherkennung und Sprachproduktion ab, die für die künftige Lexikografie neue Wege bahnen und u.a. eine mobile, multimediale und interaktive Anwendung zum Ziel haben können. In diesem Sinne schließen wir uns der folgenden Einschätzung von Tarp und Gouws (2019: 266) an:

Modern-day lexicographers are in a position to make some of the unfulfilled dreams of the past a reality. The challenge of the future is to make the impossible possible. We have work to do.

Endnoten

- * Wir danken den anonymen Gutachter/-innen für die vielen wertvollen Hinweise und Anmerkungen.
- 1. Vgl. *lexiko*-Glossareintrag zum *lexiko*-Korpus:
<https://www.owid.de/wb/lexiko/glossar/lexiko-Korpus.html>
[zuletzt abgerufen: 14.08.2020].
- 2. Leibniz-Wettbewerb 2016, Förderlinie „Innovative Vorhaben“:
<https://www.leibniz-gemeinschaft.de/forschung/leibniz-wettbewerb/gefördertevorhaben.html>
[zuletzt abgerufen: 14.08.2020].
- 3. LeGeDe-Projektwebseite:
<https://www1.ids-mannheim.de/lexik/lexik-des-gesprochenen-deutsch.html>
[zuletzt abgerufen: 14.08.2020].
- 4. Empirische Methoden-Projektwebseite:
URL: <https://www1.ids-mannheim.de/lexik/empirische-methoden.html>
[zuletzt abgerufen: 14.08.2020].

5. Die Eckdaten und Rahmenbedingungen der Studien können in Meliss, Möhrs und Ribeiro Silveira (2018: 107-110) genauer nachgelesen werden. Vgl. auch die Informationen zu den Studien auf der LeGeDe-Projektwebseite:
<https://www1.ids-mannheim.de/lexik/lexik-des-gesprochenen-deutsch/projektbeschreibung/empirische-forschung.html>
[zuletzt abgerufen: 14.08.2020].
6. Umtext „Über LeGeDe“:
URL: <https://www.owid.de/legede/about.jsp>
[zuletzt abgerufen: 14.08.2020].
7. Die Links zu allen in den Tabellen aufgeführten Ressourcen können in der Bibliographie eingesehen werden.
8. Bei den Links zu den in der Tabelle angegebenen Lemmata zu FOLK ist (genauso wie im LeGeDe-Prototyp) zu beachten, dass das Abfrageergebnis in der DGD nur nach vorheriger kostenloser Anmeldung eingesehen werden kann (vgl. dazu auch die Information in Tab. 2 in Abschnitt 3.3).

Bibliografie

- Albert, G. und S. Diao-Klaeger (Hrsg.).** 2018. *Mündlicher Sprachgebrauch zwischen Normorientierung und pragmatischen Spielräumen*. Stauffenburg Linguistik 101. Tübingen: Stauffenburg.
- Barcala Rodríguez, F.M. et al.** 2018. El corpus ESLORA de español oral: diseño, desarrollo y explotación. *CHIMERA: Romance Corpora and Linguistic Studies* 5(2): 217-237.
- Bergmann, P.** 2017. Gebrauchsprofile von *weiß nich* und *keine Ahnung* im Gespräch. Ein Blick auf nicht-responsive Vorkommen. Blühdorn, H. et al. (Hrsg.). 2017. *Diskursmarker im Deutschen. Reflexionen und Analysen*: 157-182. Göttingen: Verlag für Gesprächsforschung.
- Couper-Kuhlen, E. und M. Selting.** 2018. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.
- Deppermann, A.** 2007. *Grammatik und Semantik aus gesprächsanalytischer Sicht*. Linguistik — Impulse und Tendenzen 14. Berlin/New York: De Gruyter.
- Deppermann, A., N. Proske und A. Zeschel (Hrsg.).** 2017. *Verben im interaktiven Kontext. Bewegungsverben und mentale Verben im gesprochenen Deutsch*. Studien zur deutschen Sprache 74. Tübingen: Narr.
- Dziemianko, A.** 2018. Electronic Dictionaries. Fuertes-Olivera, P.A. (Hrsg.). 2018. *The Routledge Handbook of Lexicography*: 663-683. London/New York: Routledge.
- Eichinger, L.M.** 2017. Gesprochene Alltagssprache. Deutsche Akademie für Sprache und Dichtung/ Union der deutschen Akademien der Wissenschaften (Hrsg.). 2017. *Vielfalt und Einheit der deutschen Sprache. Zweiter Bericht zur Lage der deutschen Sprache*: 283-331. Tübingen: Stauffenburg.
- Engelberg, S. et al.** 2011. Argumentstrukturmuster als Konstruktionen? Identität — Verwandtschaft — Idiosynkrasien. Engelberg, S., A. Holler und K. Proost (Hrsg.). 2011. *Sprachliches Wissen zwischen Lexikon und Grammatik*: 71-112. Jahrbuch des Instituts für Deutsche Sprache 2010. Berlin: De Gruyter.

- Engelberg, S.** 2015. Gespaltene Stimuli bei Psych-Verben. Kombinatorische Mustersuchen in Korpora zur Ermittlung von Argumentstrukturverteilungen. Engelberg, S. et al. (Hrsg.). 2015. *Argumentstruktur zwischen Valenz und Konstruktion*: 469-491. Studien zur deutschen Sprache 68. Tübingen: Narr.
- Engelberg, S.** 2018. The Argument Structure of Psych-verbs: A Quantitative Corpus Study on Cognitive Entrenchment. Boas, H. und A. Ziem (Hrsg.). 2018. *Constructional Approaches to Syntactic Structures in German*: 47-84. Trends in Linguistics Studies and Monographs 322. Berlin: De Gruyter.
- Engelberg, S.** 2019. Argumentstrukturmuster. Ein elektronisches Handbuch zu verbalen Argumentstrukturen im Deutschen. Czicza, D., V. Dekalo und G. Diewald (Hrsg.). 2019. *Konstruktionsgrammatik VI. Varianz in der konstruktionalen Schematizität*: 13-38. Tübingen: Stauffenburg.
- Engelberg, S., A. Klosa-Kückelhaus und C. Müller-Spitzer.** 2019. Lexikographie zwischen Grimm und Google? *Sprachreport* 35(2): 30-34.
- Engelberg, S., C. Müller-Spitzer und T. Schmidt.** 2016. Vernetzungs- und Zugriffsstrukturen. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016. *Internetlexikografie. Ein Kompendium*: 153-196. Berlin: De Gruyter.
- Fandrych, C., C. Meißner und F. Wallner (Hrsg.).** 2017. *Gesprochene Wissenschaftssprache — digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg.
- Fandrych, C., C. Meißner und F. Wallner.** 2018. Das Potenzial mündlicher Korpora für die Sprachdidaktik. Das Beispiel GeWiss. *Deutsch als Fremdsprache* 55(1): 3-13.
- Fiehler, R.** 2016. Gesprochene Sprache. Wöllstein, A. (Hrsg.). 2016. *Duden — Die Grammatik. Unentbehrlich für richtiges Deutsch*: 1181-1260. Berlin: Dudenverlag.
- Geyken, A. und L. Lemnitzer.** 2016. Automatische Gewinnung von lexikografischen Angaben. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016. *Internetlexikografie. Ein Kompendium*: 197-248. Berlin: De Gruyter.
- Günthner, S.** 2018. Routinisierte Muster in der Interaktion. Deppermann, A. und S. Reineke (Hrsg.). 2018. *Sprache im kommunikativen, interaktiven und kulturellen Kontext*: 29-50. Berlin/New York: De Gruyter.
- Handwerker, B., R. Bäuerle und B. Sieberg (Hrsg.).** 2016. *Gesprochene Fremdsprache Deutsch. Perspektiven Deutsch als Fremdsprache* 32. Baltmannsweiler: Schneider.
- Hansen, C. und M.H. Hansen.** 2012. A Dictionary of Spoken Danish. Fjeld, R.V. und J.M. Torjusen (Hrsg.). 2012. *Proceedings of the 15th EURALEX International Congress, 7–11 August 2012, Oslo, Norway*: 929-935. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.
- Helmer, H. und A. Deppermann.** 2017. ICH WEIß NICHT zwischen Assertion und Diskursmarker: Verwendungsspektren eines Ausdrucks und Überlegungen zu Kriterien für Diskursmarker. Blühdorn, H. et al. (Hrsg.). 2017. *Diskursmarker im Deutschen. Reflexionen und Analysen*: 131-156. Göttingen: Verlag für Gesprächsforschung.
- Helmer, H., A. Deppermann und S. Reineke.** 2017. Antwort, epistemischer Marker oder Widerspruch? Sequenzielle, semantische und pragmatische Eigenschaften von *ich weiss nicht*. Deppermann, A., N. Proske und A. Zeschel (Hrsg.). 2017: 377-406.
- Imo, W.** 2007. *Construction Grammar und Gesprochene-Sprache-Forschung. Konstruktionen mit zehn matrixsatzfähigen Verben im gesprochenen Deutsch*. Germanistische Linguistik 275. Berlin: De Gruyter.

- Imo, W. und S.M. Moraldo (Hrsg.).** 2015. *Interaktionale Sprache und ihre Didaktisierung im DaF-Unterricht*. Deutschdidaktik 4. Tübingen: Stauffenburg.
- Klosa, A.** 2013a. Aktuelle Tendenzen in der deutschen Lexikographie der Gegenwart. Stickel, G. und T. Váradi (Hrsg.). 2013. *Lexical Challenges in a Multilingual Europe. Contributions to the Annual Conference 2012 of EFNIL in Budapest*: 75-93. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 99. Frankfurt am Main: Peter Lang.
- Klosa, A.** 2013b. The Lexicographical Process (with Special Focus on Online Dictionaries). Gouws, R.H. et al. (Hrsg.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary volume: Recent Developments with Focus on Electronic and Computational Lexicography*: 517-524. Handbücher zur Sprach- und Kommunikationswissenschaft 5.4. Berlin: De Gruyter.
- Klosa, A. und R.H. Gouws.** 2015. Outer Features in e-Dictionaries. *Lexicographica* 31: 142-172.
- Klosa, A., U. Schnörch und S. Schoolaert.** 2010. Stichwort, Stichwortliste und Eigennamen in *lexiko*. Dykstra, A. und T. Schoonheim (Hrsg.). 2010. *Proceedings of the XIV EURALEX International Congress, 6–10 July 2010, Leeuwarden, the Netherlands*: 653-663. Leeuwarden: Afûk.
- Klosa, A. und C. Tiberius.** 2016. Der lexikografische Prozess. Klosa, A. und C. Müller-Spitzer (Hrsg.). 2016. *Internetlexikografie. Ein Kompendium*: 65-110. Berlin: De Gruyter.
- Kupietz, M. et al.** 2018. The German Reference Corpus DEREKO: New Developments — New Opportunities. Calzolari, N. et al. (Hrsg.). 2018. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, May 7–12, 2018, Miyazaki, Japan*: 4353-4360. Miyazaki: European Language Resources Association.
- Kupietz, M. und H. Keibel.** 2009. The Mannheim German Reference Corpus (DEREKO) as a Basis for Empirical Linguistic Research. Minegishi, M. und Y. Kawaguchi (Hrsg.). 2009. *Working Papers in Corpus-based Linguistics and Language Education*: 53-59. Tokyo: Tokyo University of Foreign Studies (TUFSS).
- Kupietz, M. und T. Schmidt.** 2015. Schriftliche und mündliche Korpora am IDS als Grundlage für die empirische Forschung. Eichinger, L.M. (Hrsg.). 2015. *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*: 297-322. Jahrbuch des Instituts für Deutsche Sprache 2014. Berlin: De Gruyter.
- Lemmenmeier-Batinić, D.** 2020. Lexical Explorer: Extending Access to the Database for Spoken German for User-specific Purposes. *Corpora* 15: 55-76.
- Meliss, M. i. Dr. [a].** Die LeGeDe-Ressource: korpusbasierte lexikografische Einblicke und anwendungsorientierte Ausblicke. *Zeitschrift DaF* 4/2020.
- Meliss, M. i. Dr. [b].** Das Spektrum der Angabeklassen in der lexikografischen Ressource zum gesprochenen Deutsch: LeGeDe. Piosik, M., J. Taborek und M. Woźnicka (Hrsg.). i. Dr. 10. *Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven, 19.–20. Oktober 2018, Adam-Mickiewicz-Universität in Poznań, Polen*. *Lexicographica: Series Maior*. Berlin/New York: De Gruyter.
- Meliss, M.** 2016. Gesprochene Sprache in DaF-Lernerwörterbüchern. Handwerker, B., R. Bäuerle und B. Sieberg (Hrsg.). 2016. *Gesprochene Fremdsprache Deutsch*: 179-199. *Perspektiven Deutsch als Fremdsprache* 32. Baltmannsweiler: Schneider.
- Meliss, M.** 2020a. Musterhaftigkeit in der Lexik des gesprochenen Deutsch: eine handlungs- und funktionsorientierte Annäherung mit lexikografischer Anwendungsperspektive. Mellado Blanco, C. et al. (Hrsg.). 2020. *Produktive Muster in der Phraseologie*: 181-202. *Studia Phraseologica et Paroemiologica* 4. Hamburg: Verlag Dr. Kovač.

- Meliss, M.** 2020b. Lexik des gesprochenen Deutsch in der Interaktion: Partikelverben und ihre Besonderheiten. García Adánez, I. et al. (Hrsg.). 2020. Dr. *Das Leben in einem Rosa Licht sehen. Festschrift für Rosa Piñel/Homenaje a Rosa Piñel*: 77-97. Perspektiven der Germanistik und Komparatistik in Spanien 16. Bern: Peter Lang.
- Meliss, M. et al.** 2018. Creating a List of Headwords for a Lexical Resource of Spoken German. Krek, S. et al. (Hrsg.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana, Slovenia*: 1009-1016. Ljubljana: Ljubljana University Press, Faculty of Arts.
- Meliss, M. et al.** 2019. A Corpus-based Lexical Resource of Spoken German in Interaction. Kosem, I. et al. (Hrsg.). 2019. *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal*: 783-804. Brno: Lexical Computing CZ, s.r.o.
- Meliss, M. und C. Möhrs.** 2017. Die Entwicklung einer lexikografischen Ressource im Rahmen des Projektes LeGeDe. *Sprachreport* 33: 42-52.
- Meliss, M. und C. Möhrs.** 2018. Lexik in der spontanen, gesprochensprachlichen Interaktion: Eine anwendungsorientierte Annäherung aus der DaF-Perspektive. *German as a Foreign Language* 3: 79-110.
- Meliss, M. und C. Möhrs.** 2019. Herausforderungen eines innovativen lexikografischen Projekts zu Besonderheiten des gesprochenen Deutsch in der Interaktion. Doval, I. und E. Liste Lamas (Hrsg.). 2019. *Germanistik im Umbruch — Linguistik, Übersetzung und DaF*: 13-26. Berlin: Frank & Timme.
- Meliss, M., C. Möhrs und M. Ribeiro Silveira.** 2018. Erwartungen an eine korpusbasierte lexikografische Ressource zur ‚Lexik des gesprochenen Deutsch in der Interaktion‘: Ergebnisse aus zwei empirischen Studien. *Zeitschrift für Angewandte Linguistik* 68: 103-138.
- Meliss, M., C. Möhrs und M. Ribeiro Silveira.** 2019. Anforderungen und Erwartungen an eine lexikografische Ressource des gesprochenen Deutsch aus der L2-Lernerperspektive. *Lexicographica* 34: 89-121.
- Möhrs, C.** i. Dr. Quantitative und qualitative Ansätze zu Stichwortkandidaten für die lexikografische Ressource zum gesprochenen Deutsch: LeGeDe. Piosik, M., J. Taborek und M. Woźnicka (Hrsg.). i. Dr. 10. *Kolloquium zur Lexikographie und Wörterbuchforschung. Korpora in der Lexikographie — Stand und Perspektiven, 19.–20. Oktober 2018, Adam-Mickiewicz-Universität in Poznań, Polen*. *Lexicographica: Series Maior*. Berlin/New York: De Gruyter.
- Möhrs, C.** 2020. „Hast du eine Ahnung, ...?“ Eine lexikografische und korpusbasierte Untersuchung am Beispiel des Lexems *Ahnung*. Behr, J. et al. (Hrsg.). 2020. *Schnittstellen der Germanistik. Festschrift für Hans Bickes*: 185-214. Berlin: Peter Lang.
- Möhrs, C. und S. Torres Cajo.** i. Dr. The Microstructure of a Lexicographical Resource of Spoken German: Meanings and Functions of the Lemma *eben*. *Rasprave: Journal of the Institute of Croatian Language and Linguistics*.
- Möhrs, C., M. Meliss und D. Batinić.** 2017. LeGeDe — Towards a Corpus-based Lexical Resource of Spoken German. Kosem, I. et al. (Hrsg.). 2017. *Electronic Lexicography in the 21st Century, Proceedings of eLex 2017 Conference, Leiden, the Netherlands, 19–21 September 2017*: 281-298. Brno: Lexical Computing CZ s.r.o.
- Moon, R.** 1998. On Using Spoken Data in Corpus Lexicography. Fontenelle, T. et al. (Hrsg.). 1998. *Proceedings of the 8th EURALEX International Congress, 4–8 August 1998, Liège, Belgium*: 347-355. Liège: English and Dutch Departments, University of Liège.

- Moraldo, S. M. und F. Missaglia (Hrsg.).** 2013. *Gesprochene Sprache im DaF-Unterricht. Grundlagen — Ansätze — Praxis.* Sprache — Literatur und Geschichte 43. Heidelberg: Winter.
- Reeg, U., P. Gallo und S. M. Moraldo (Hrsg.).** 2012. *Gesprochene Sprache im DaF-Unterricht. Zur Theorie und Praxis eines Lerngegenstandes.* Interkulturelle Perspektiven in der Sprachwissenschaft und ihrer Didaktik 3. Berlin: Waxmann.
- Schmidt, T.** 2014a. The Research and Teaching Corpus of Spoken German — FOLK. Calzolari, N. et al. (Hrsg.). 2014. *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, May 26–31, 2014, Reykjavik, Iceland:* 383-387. Reykjavik: European Language Resources Association.
- Schmidt, T.** 2014b. Gesprächskorpora und Gesprächsdatenbanken am Beispiel von FOLK und DGD. *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion* 15: 196-233.
- Schmidt, T.** 2014c. The Database for Spoken German — DGD2. Calzolari, N. et al. (Hrsg.). 2014. *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, May 26–31, 2014, Reykjavik, Iceland:* 1451-1457. Reykjavik: European Language Resources Association.
- Schmidt, T.** 2017. DGD — Die Datenbank für Gesprochenes Deutsch. Mündliche Korpora am Institut für Deutsche Sprache (IDS) in Mannheim. *Zeitschrift für germanistische Linguistik. Deutsche Sprache in Gegenwart und Geschichte* 45(3): 451-463.
- Schmidt, T.** 2018. Gesprächskorpora. Kupietz, M. und T. Schmidt (Hrsg.). 2018. *Korpuslinguistik: 209-230.* Germanistische Sprachwissenschaft um 2020, Bd. 5. Berlin/Boston: De Gruyter.
- Schnörch, U.** 2005. Die *lexiko*-Stichwortliste. Haß, U. (Hrsg.). 2005. *Grundfragen der elektronischen Lexikographie.* *lexiko — das Online-Informationssystem zum deutschen Wortschatz:* 71-90. Schriften des Instituts für Deutsche 12. Berlin/New York: De Gruyter.
- Schwittalla, J.** 2012. *Gesprochenes Deutsch.* Eine Einführung. 4. Aufl. Grundlagen der Germanistik 33. Berlin: Schmidt.
- Sieberg, B.** 2013. *Sprechen lehren, lernen und verstehen. Stufenübergreifendes Studien- und Übungsbuch für den DaF-Bereich.* Tübingen: Julius Groos.
- Siepmann, D.** 2015. Dictionaries and Spoken Language: A Corpus-based Review of French Dictionaries. *International Journal of Lexicography* 28(2): 139-168.
- Stadler, H.** 2014. *Die Erstellung der Basislemmaliste der neuhochdeutschen Standardsprache aus mehrfach linguistisch annotierten Korpora.* OPAL — Online publizierte Arbeiten zur Linguistik 5(2014). Mannheim: Institut für Deutsche Sprache.
- Tarp, S.** 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249. URL: <https://lexikos.journals.ac.za/pub/issue/view/87> [zuletzt abgerufen: 14.08.2020].
- Tarp, S. und R. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268. URL: <https://lexikos.journals.ac.za/pub/issue/view/87> [zuletzt abgerufen: 14.08.2020].
- Torres Cajo, S.** 2019. Zwischen Strukturierung, Wissensmanagement und Argumentation im Gespräch — Interaktionale Verwendungsweisen der Modalpartikeln *halt* und *eben* im gesprochenen Deutsch. *Deutsche Sprache* 47(4): 289-310.
- Trap-Jensen, L.** 2004. Spoken Language in Dictionaries: Does it really matter? Williams, G. und S. Vessier (Hrsg.). 2004. *Proceedings of the 11th EURALEX International Congress, 6–10 July 2004, Lorient, France:* 311-318. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud.

- Trim, J. et al.** 2001. *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Verdonik, D. und M. Sepesy Maučec.** 2017. A Speech Corpus as a Source of Lexical Information. *International Journal of Lexicography* 30: 143-166.
- Westpfahl, S.** 2014. STTS 2.0? Improving the Tagset for the Part-of-Speech-Tagging of German Spoken Data. Levin, L. und M. Stede (Hrsg.). 2014. *LAW VIII. The 8th Linguistic Annotation Workshop in conjunction with COLING 2014. Proceedings of the Workshop, August 23–24, 2014, Dublin, Ireland*: 1-10. Dublin: Association for Computational Linguistics.
- Westpfahl, S. und T. Schmidt.** 2016. FOLK-Gold — A GOLD Standard for Part-of-Speech Tagging of Spoken German. Calzolari, N. et al. (Hrsg.). 2016. *Proceedings of the Tenth International Conference on Language Resources and Evaluation. May 23-28, 2016. Portorož, Slovenia*: 1493-1499. Portorož: European Language Resources Association.
- Wiegand, H.E.** 1983. Was ist eigentlich ein Lemma? Ein Beitrag zur Theorie der lexikographischen Sprachbeschreibung. Wiegand, H.E. (Hrsg.). 1983. *Studien zur neuhochdeutschen Lexikographie III*: 401-474. Germanistische Linguistik 1–4. Hildesheim: Georg Olms.
- Zeschel, A.** 2017. *Denken und wissen im gesprochenen Deutsch*. Deppermann, A., N. Proske und A. Zeschel (Hrsg.). 2017. *Verben im interaktiven Kontext. Bewegungsverben und mentale Verben im gesprochenen Deutsch*: 249-336. Studien zur deutschen Sprache 74. Tübingen: Narr.

Wörterbücher und Online-Ressourcen

(Online-Ressourcen: alle zuletzt abgerufen am 14.08.2020)

DEREKO = Deutsches Referenzkorpus.

URL: <http://www.ids-mannheim.de/kl/projekte/korpora/>

DGD = Datenbank Gesprochenes Deutsch.

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

Duden-DaF = *Duden, Standardwörterbuch — Deutsch als Fremdsprache* / hrsg. von der Dudenredaktion. 2., neu bearbeitete und erweiterte Auflage. Berlin: Dudenverlag, 2010.

Duden-online. URL: <https://www.duden.de/>

Duden-Universalwörterbuch = *Duden, deutsches Universalwörterbuch* / hrsg. von der Dudenredaktion; red. Bearbeitung Melanie Kunkel u.a.; unter Mitarbeit von Dr. Laura Sturm und Lena Wagner. 9., vollständig überarbeitete und erweiterte Auflage. Berlin: Dudenverlag, 2019.

DWDS = Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart.

URL: <https://www.dwds.de/>

elexiko = Online Wörterbuch zur deutschen Gegenwartssprache im Verbund OWID (2003ff.).

Leibniz-Institut für Deutsche Sprache, Mannheim.

URL: <https://www.owid.de/docs/elex/start.jsp>

E-Valbu = Elektronisches Valenzwörterbuch deutscher Verben.

URL: <https://grammis.ids-mannheim.de/verbvalenz>

FOLK = Forschungs- und Lehrkorpus Gesprochenes Deutsch.

URL: <http://agd.ids-mannheim.de/folk.shtml>

GeWiss = Gesprochenes Wissenschaftsdeutsch.

URL: <https://gewiss.uni-leipzig.de/>

grammis = Grammatisches Informationssystem.

URL: <https://grammis.ids-mannheim.de/>

LEO. URL: <https://www.leo.org/>

LGWB DaF = Götz, Dieter (Hg.). 2015. *Langenscheidt Großwörterbuch Deutsch als Fremdsprache*. Neubearbeitung. München: Langenscheidt.

LeGeDe = Lexik des gesprochenen Deutsch.

URL (Projektwebseite): <http://www.ids-mannheim.de/lexik/lexik-des-gesprochenen-deutsch.html>

LeGeDe-Ressource. URL: <https://www.owid.de/legede/>

Lexical Explorer. URL: <https://www.owid.de/lexex/>

Linguee. URL: <https://www.linguee.de/>

LothWB = Wörterbuch der deutsch-lothringischen Mundarten. Bearb. von Michael Ferdinand Follmann. Leipzig 1909. (Quellen zur lothringischen Geschichte — Documents de l'Histoire de la Lorraine 12). Onlinefassung über das Wörterbuchnetz.

URL: http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=LothWB

OWID = Online-Wortschatz-Informationssystem Deutsch.

URL: <https://www.owid.de>

OWID^{plus} = Online-Wortschatz-Informationssystem Deutsch ^{plus}.

URL: <https://www.owid.de/plus/index.html>

PONS-online = PONS Online-Wörterbuch. Einfach nachschlagen und übersetzen.

URL: <https://de.pons.com/%C3%BCbersetzung>