

KORPORA GESPROCHENER SPRACHE UND DEUTSCH ALS FREMD- UND ZWEITSPRACHE: EINE CHANCENREICHE BEZIEHUNG

Christian Fandrych, Universität Leipzig
Cordula Meißner, Universität Innsbruck
Franziska Wallner, Universität Leipzig

Abstract

In diesem Beitrag wird das Potenzial von Korpora der gesprochenen Sprache für die DaF/DaZ-Forschung und -Didaktik näher beschrieben. Besondere Berücksichtigung finden dabei die im Projekt *ZuMult* entwickelten Zugangswege zu mündlichen Korpora. Es werden drei Perspektiven eingenommen. Zunächst werden die aktuellen Einsatzmöglichkeiten von Korpora gesprochener Sprache für den Unterricht Deutsch als Fremdsprache und Deutsch als Zweitsprache vorgestellt. Im nächsten Schritt wird gezeigt, inwieweit Korpora als Grundlage für die Ermittlung von vermittlungsrelevanten Phänomenen in der gesprochenen Sprache genutzt werden können. Schließlich wird auf Möglichkeiten und Grenzen der Analyse (fortgeschrittener) gesprochener Lerner/-innensprache auf der Grundlage von mündlichen Korpora eingegangen.

Keywords: Deutsch als Fremd- und Zweitsprache; Korpora gesprochener Sprache; Mündlichkeitsdidaktik; Gesprochene-Sprache-Forschung; Forschungsprojekt ZuMult

Abstract

This article discusses the potential benefits that spoken language corpora provide for research and teaching methodology in the context of German as a Foreign and Second Language (GFL/GSL), with special emphasis on the innovative user functions for such corpora developed by the *ZuMult* project. We focus on three different aspects. First, we outline new approaches and possibilities of implementing spoken language corpora into the GFL/GSL language classroom. Second, we discuss the ways in which spoken language phenomena that are particularly relevant to language teaching can be identified and described using corpora of spoken German. Finally, we discuss some of the benefits and methodological challenges that (advanced) spoken language learning corpora posit for research in the field of GFL/GSL.

Keywords: German as a Foreign and Second Language; spoken language corpora; teaching spoken language; spoken language research; research project ZuMult

Einleitung

Gesprochene Sprache ist aus korpusmethodischer Perspektive ein besonderer Gegenstand, insofern sie nur multimodal analysiert werden kann: Sie tritt auditiv (und meist auch visuell) in Erscheinung, muss jedoch für ihre Erforschung durch Transkription ins Medium der Schrift übertragen werden. Dementsprechend sind Korpora der gesprochenen Sprache an sich multimodal angelegt, indem sie einerseits Audio- und/oder Videoaufnahmen, andererseits deren Transkription umfassen. Durch den notwendigen Übertragungsschritt der Transkription ist der Aufbau mündlicher Korpora mit großem Aufwand verbunden. Das hat zur Folge, dass sich existierende Korpus-sammlungen in ihrem Umfang deutlich von schriftsprachlichen Korpora unterscheiden. Die Entwicklung mündlicher Korpora hat jedoch, sowohl was den Umfang, als auch was die Aufbereitung betrifft, in den letzten Jahren große Fortschritte gemacht (vgl. z.B. Fandrych / Meißner / Wallner 2017; Schmidt 2018). Dabei werden mit den aktuell vorliegenden und öffentlich zugänglichen mündlichen Korpus-sammlungen viele

Facetten der gesprochenen Sprache der Forschung und Lehre zugänglich gemacht. Unter anderem umfasst dies neben vielfältigen Sprechereignissen aus alltäglichen, institutionellen und öffentlichen Kommunikationssituationen (etwa im Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK), vgl. Kaiser 2018) auch unterschiedliche regionale Varietäten sowie verschiedene Professionalisierungsgrade bis hin zu Sprachproduktionen von Sprecher/-innen mit Deutsch als L2 (vgl. hierzu auch die über die Datenbank für Gesprochenes Deutsch zugänglichen Korpora¹). Dieses breite Spektrum bietet für Forschung und universitäre Lehre im Bereich Deutsch als Fremd- und Zweitsprache ein überaus vielfältiges und vielversprechendes Angebot. Auch aus der Perspektive des praktischen DaF/DaZ-Unterrichts stellt die gesprochene Sprache noch immer einen Gegenstand dar, der intensiv – und häufig kontrovers – diskutiert wird (vgl. etwa Rösler 2016). Merkmale der Mündlichkeit werden dabei häufig immer noch anhand einer an Schriftsprache orientierten Normvorstellung betrachtet. In Lehrmaterialien werden diese Merkmale bislang kaum systematisch thematisiert, Beispiele für authentisches sprachliches Handeln fehlen oft (vgl. Günthner / Wegner / Weidner 2013). Dies wurde in der Fachdiskussion zwar bereits aufgegriffen und es wurden erste Vorschläge für die Einbeziehung authentischer Sprachbeispiele auch anhand der Didaktisierung von Transkriptausschnitten erarbeitet (vgl. z.B. Moraldo / Missaglia 2013; Imo / Weidner 2018). Allerdings sind Fragen wie die Auswahl der zu behandelnden Phänomene, die damit verbundenen Lernziele für bestimmte Niveaustufen und Zielgruppen und die konkrete Einbindung in den Unterricht immer noch nicht hinreichend geklärt worden; daneben fehlt es an geeigneten Sprachbeispielen, die eine solche Einbindung erst ermöglichen würden.

Auch die direkte Nutzung der wachsenden mündlichen Korpora für die Entwicklung von Lehr- und Lernmaterialien, Curricula und für den Unterricht selbst wurde bisher kaum verfolgt. Ein wichtiger Grund dafür dürfte darin liegen, dass die Zugriffswege auf die Korpusdaten nicht auf die Nutzungsbedürfnisse der Fremdsprachendidaktik ausgerichtet sind. Mit dem Projekt *ZuMult* (Zugänge zu multimodalen Korpora gesprochener Sprache)² wurden nun für diese Zielgruppe spezielle Zugänge zu mündlichen Korpora geschaffen, durch die gezielt nach verschiedenen didaktisch relevanten Aspekten gesprochener Sprache in den Korpora gesucht werden kann. So kann das didaktische Potenzial mündlicher Korpora genutzt werden, nämlich eine situations- und adressatenangemessene Auseinandersetzung mit mündlicher Kommunikation zu ermöglichen sowie mündlichkeitsspezifische sprachliche Phänomene zu fokussieren. Dies wird, und das ist die Besonderheit, nicht nur für einzelne vorausgewählte Transkripte angeboten, sondern für gesamte Korpora und damit für ein weites Spektrum an vielfältigen und nach individuellen Bedürfnissen selektierbaren Sprechereignissen. Realisiert wird es dadurch, dass für die Entwicklung der *ZuMult*-Zugänge Verfahren der automatischen Analyse von Sprachdaten genutzt wurden, welche auf neu hinzukommende Korpusdaten unmittelbar angewendet werden können. Es ist somit eine generisch applizierbare Form der Vorstrukturierung von Korpusdaten geschaffen, die zugleich demonstriert, wie korpus- und computerlinguistische Verfahren eingesetzt werden können, um die direkte Arbeit mit mündlichen Korpora für die Fremdsprachendidaktik zu verbessern. Daneben wurde im Projekt *ZuMult* auch ein Zugangsweg geschaffen, der es gestattet, Korpora gesprochener Sprache mit Hilfe einer speziell für die Korpusrecherche entwickelten Suchanfragesprache zu analysieren. Diese ermöglicht es, komplexe und für die DaF/DaZ-Didaktik hochrelevante Suchanfragen zu stellen und macht damit auch unter anderem neue oder bisher nicht zugängliche Annotationen wie z.B. Code-Switching und Metakomentierungen des Korpus „Gesprochene Wissenschaftssprache kontrastiv“ (GWSS) zugänglich.

¹ Vgl. https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.sys_inv?v_session_id= (26.07.2021).

² *ZuMult* ist ein von der DFG gefördertes Kooperationsprojekt zwischen dem Archiv für gesprochenes Deutsch (AGD), dem Leibniz-Institut für Deutsche Sprache in Mannheim, dem Hamburger Zentrum für Sprachkorpora (HZSK) an der Universität Hamburg und dem Herder-Institut der Universität Leipzig. Vgl. <https://zumult.org> (26.07.2021).

Anliegen des vorliegenden Beitrags ist es, das Potenzial für die DaF/DaZ-Didaktik näher zu beschreiben, welches sich aus diesen neuen Nutzungsmöglichkeiten und Zugangswegen zu mündlichen Korpora ergibt. Hierzu werden drei Perspektiven eingenommen. Zunächst werden die aktuellen Einsatzmöglichkeiten von Korpora gesprochener Sprache für den Unterricht Deutsch als Fremdsprache und Deutsch als Zweitsprache vorgestellt. Dabei stehen insbesondere die im Projekt *ZuMult* entwickelten Werkzeuge *ZuMal* (Zugang zu Merkmalsauswahl von Gesprächen) und *ZuViel* (Zugang zu Visualisierungselementen für Transkripte) im Fokus (1). Im nächsten Schritt wird darauf eingegangen, inwieweit Korpora als Grundlage für die Ermittlung von vermittlungsrelevanten Phänomenen in der gesprochenen Sprache genutzt werden können (2). Exemplarisch wird dies demonstriert anhand ausgewählter Ergebnisse aus einer Studie zu Klitisierungen, die mit Hilfe des in *ZuMult* entwickelten Werkzeugs *ZuRecht* (Zugang zur Recherche in Transkripten) umgesetzt wurde (2.1). Zudem wird ein Ansatz vorgestellt, wie die typische Realisierung sprachlicher Handlungen korpusbasiert ermittelt werden kann (2.2). Schließlich wird auf Möglichkeiten (und Grenzen) der Analyse gesprochener Lerner/-innensprache auf der Grundlage von Korpora eingegangen (3).

1. Nutzung von mündlichen Korpora für den Unterricht

Es werden zunächst aktuelle Einsatzmöglichkeiten verfügbarer mündlicher Korpora³ für den Unterricht beschrieben. Daran anknüpfend werden die Erweiterungen der Zugriffsmöglichkeiten durch die in *ZuMult* entwickelten Zugänge *ZuMal* und *ZuViel* vorgestellt. Schließlich wird auf die mit der Nutzung mündlicher Korpora und der Anwendung der Zugänge verbundenen Anforderungen an eine Corpus Literacy seitens der Lehrpersonen eingegangen.

1.1 Einsatzmöglichkeiten mündlicher Korpora für die Fremdsprachenvermittlung

Über die Datenbank für Gesprochenes Deutsch (DGD)⁴ am Leibniz-Institut für Deutsche Sprache in Mannheim sind verschiedene Korpora mündlicher Interaktionen zugänglich, die für den Deutsch als Fremd- und Zweitsprache-Unterricht interessante Quellen bilden. Für den DaF/DaZ-Kontext sind insbesondere zwei Korpora von besonderem Interesse: Zum einen das Forschungs- und Lehrkorpus gesprochenes Deutsch (FOLK) (vgl. Deppermann / Hartung 2012; Schmidt 2014), welches ein weites Spektrum von Gesprächen aus unterschiedlichen privaten, institutionellen und öffentlichen Kontexten abdeckt und aktuell 374 Sprechereignisse mit über 1.200 Sprecher/-innen in 314 Aufnahmestunden umfasst⁵. Zum anderen bietet das GWSS-Korpus domänenspezifische Daten aus dem Bereich der Wissenschaftskommunikation. In der aktuell über die DGD zugänglichen Version umfasst dies Prüfungsgespräche, studentische Vorträge und Expertenvorträge im Deutschen als L1 und L2 mit insgesamt 257 Sprechereignissen, 480 Sprecher/-innen in 92 Aufnahmestunden (vgl. Fandrych / Meißner / Slavcheva 2014; Fandrych / Meißner / Wallner 2017). Die Korpora beinhalten die Audioaufnahmen der Gespräche und z.T. auch Videoaufnahmen, daneben umfangreiche Metadaten zu den Sprechereignissen und den beteiligten Sprecher/-innen. Die Audiodaten sind transkribiert und liegen in der DGD in den auf dem GAT2-Minimaltranskript (vgl. Selting et al. 2009)

³ Unter mündlichen Korpora bzw. Korpora gesprochener Sprache werden Sprachdatensammlungen gefasst, die in jedem Fall medial mündlich, in den meisten Fällen auch konzeptionell mündlich sind. In verschiedenen institutionellen und öffentlichen Gesprächsarten variiert jedoch der Grad der konzeptionellen Mündlichkeit, insbesondere bei mehr oder weniger stark vorformulierten Genres (etwa in wissenschaftlichen Vorträgen oder öffentlichen Reden).

⁴ Vgl. <https://dgd.ids-mannheim.de> (26.07.2021).

⁵ Version 2.16 (17.05.2021); vgl. <https://dgd.ids-mannheim.de> (26.07.2021).

basierenden cGAT-Konventionen vor. Dies umfasst eine aussprachenaher Verschriftung, die alle Abweichungen von der Standardlautung notiert, u.a. Klitisierungen (z.B. *gehste* [*gehst du*]), regionale Einflüsse (z.B. *net* [*nicht*]) oder umgangssprachliche Schnellsprechformen (z.B. *ham* [*haben*]). Die aussprachenahere Transkription macht somit Merkmale der Mündlichkeit im Medium der Schrift sichtbar und versucht auf diese Weise authentisches sprachliches Handeln rekonstruierbar zu machen. Darüber hinaus sind für die Transkripte der Korpora FOLK und GWSS in der DGD Aufbereitungsebenen verfügbar, die ihre orthographische Normalisierung, ihre Lemmatisierung und ihre Wortartenannotation betreffen (vgl. Westpfahl et al. 2017). Diese Aufbereitung erlaubt verschiedene Formen des Zugriffs und der Nutzung. So können die Sprechereignisse einerseits anhand der Metadaten ausgewählt und als Volltranskript aussprachenaher oder normalisiert betrachtet werden. Andererseits ist es möglich, gezielt sprachliche Phänomene abzufragen. Durch die Suche auf der Ebene der orthographisch normalisierten Form oder der Lemmaform können alle Varianten ihrer mündlichen Realisierungen abgerufen werden.

Für den DaF-/DaZ-Unterricht bieten die Korpusdaten Anwendungsmöglichkeiten für die Reflexion und Vermittlung von gesprochener Sprache auf allen Ebenen: gesprächsstrukturell, interaktional/handlungsbezogen wie auch grammatisch-lexikalisch. So kann etwa die Struktur von kommunikativen Ereignissen anhand von Volltranskripten oder ausgewählten Ausschnitten erarbeitet werden, wie dies Bies (2020) am Beispiel des WG-Castings zeigt. Daneben lassen sich Handlungssequenzen fokussieren. Kaiser / Schedl (2021) stellen Ergebnisse eines Projekts vor, in dem anhand ausgewählter Sprechereignisse aus dem FOLK-Korpus für die Handlungsbereiche (Nach-)Fragen, Vorschläge/Angebote und Ratschläge auf Modalverben basierende sprachliche Formate (z.B. *Kann ich X?*) in ihrer sequentiellen Verwendung im Gespräch ermittelt und annotiert wurden, um so für die Verwendung im DaF-Unterricht eine Beispielsammlung authentischen Sprachhandelns bereitzustellen. Für die Thematisierung der sprachlichen Handlungen des Metakommentierens und Verweisens bzw. Zitierens in der mündlichen Wissenschaftskommunikation bieten die für das GWSS-Korpus vorliegenden Annotationen eine gute Materialbasis (vgl. Fandrych / Meißner / Wallner 2018). Für die Vermittlung von Spezifika der gesprochenen Sprache auf lexikogrammatischer Ebene eröffnen die Ergebnisse des Projekts „Lexik des gesprochenen Deutsch“ (LeGeDe) wertvolle Ansatzpunkte. Hier wurde ein Prototyp für die lexikographische Aufarbeitung der Lexik des gesprochenen Deutsch vorgelegt, in dem sprachliche Einheiten basierend auf Daten des FOLK-Korpus auch in ihrer interaktionalen Einbettung und Funktion erläutert werden (z.B. *guck mal* in der Funktion des Diskursmarkers, vgl. u.a. Meliss 2021)⁶.

Korpora gesprochener Sprache bieten zudem speziell für die Hörverstehensdidaktik Optionen zur Gestaltung von Übungen. Lernende sind beim Hörverstehen mit besonderen Anforderungen konfrontiert, da sie aus dem flüchtigen Lautstrom, in dem Einheiten zusammenhängend und in variabler Realisierungsform sowie im Hinblick auf die potenziell erkennbaren sprachlichen Einheiten mehrdeutig erscheinen, eine sprachliche Botschaft dekodieren müssen (vgl. Dietz 2017, 2021a). Diese Herausforderung wird in der Hörverstehensdidaktik durch die Arbeit mit Mikrohöraufgaben aufgegriffen, die auf eine Ausbildung der Dekodierfähigkeit abzielen. Korpora gesprochener Sprache halten mit ihren Daten aus verschiedenen Domänen und einer Vielfalt an soziodemographischen und regionalen Sprechereigenschaften eine reichhaltige Basis bereit, um solche Mikrohörübungen zu gestalten. So können etwa über die Suche nach einer normalisierten Form (z.B. *haben*) Belege und Hörproben für diverse Realisierungen (z.B. in FOLK *ham, ha, hän, haww, han, hewwe, habn* u.a.) abgerufen und hieraus Übungen erstellt werden (vgl. Dietz 2021b).

Daten aus mündlichen Korpora eignen sich zudem zur Förderung einer auf Spezifika des gesprochenen Mediums gerichteten Sprachbewusstheit. Anhand von Beispielen authentischer

⁶ Der LeGeDe-Prototyp ist verfügbar unter <https://www.owid.de/legede/> (26.07.2021).

Sprachverwendung können Merkmale eines mündlichen Gebrauchsstandards in Aussprache oder grammatischer Struktur (vgl. Schneider 2020; Lanwer / Schopf 2021) bewusst gemacht werden.

Die genannten Anwendungen setzten jedoch bislang eine Vertrautheit mit korpuslinguistischen Analysemethoden voraus, vor allem auch deshalb, weil die vorhandenen Zugriffsmöglichkeiten auf die mündlichen Korpusdaten an den Kompetenzen und Interessen von sprachwissenschaftlich forschenden Nutzer/-innen ausgerichtet waren. Die spezieller aus fremdsprachendidaktischer Perspektive sinnvollen und notwendigen Abfrage- und Zugriffsmöglichkeiten waren bisher nicht berücksichtigt worden; hierzu gehören etwa ganz grundlegend die Einstufung von Beispielen nach sprachlicher Schwierigkeit (z.B. bezogen auf den Wortschatz einer bestimmten Niveaustufe) oder nach Verständlichkeitsfaktoren (wie Sprechgeschwindigkeit oder die Nähe bzw. Ferne zu einem an der Schriftsprache orientierten Standard). Dieser Bedarf wurde im Projekt *ZuMult* aufgegriffen.

1.2 Zugänge zu mündlichen Korpora mit *ZuMult*

Mit dem Projekt *ZuMult* sollten neue und v.a. auf die spezifische Nutzergruppe der Fremdsprachendidaktiker/-innen zugeschnittene Möglichkeiten des Korpuszugriffs entwickelt werden. Eine Nutzungsstudie (vgl. Fandrych et al. 2016) ergab, dass viele interessierte Nutzer/-innen einerseits über geringe spezialisierte korpuslinguistische (Recherche-)Kenntnisse verfügen und so vielfältige vorhandene Werkzeuge nicht nutzen. Andererseits wurden Nutzungswünsche geäußert, die bisher so als Recherchemöglichkeit nicht umsetzbar waren. Durch das Projekt *ZuMult* sollten daher Zugänge, Recherche- und Anwendungsmöglichkeiten geschaffen werden, die in der Fremdsprachendidaktik, aber auch in verschiedenen weiteren Anwendungsfeldern von besonderem Interesse sind. Diese Zugänge sollten prinzipiell in den Korpusdaten schon vorhandene oder ermittelbare Informationen aufgreifen, um so auf Korpuserweiterungen übertragbar zu sein. Solche Informationen betreffen etwa die Metadaten zu Gesprächsart, Themen, Dauer, Region der Aufnahme und Herkunft der Sprecher/-innen. Daneben sollte der für die Fremdsprachendidaktik zentrale Faktor der sprachlich-kommunikativen Schwierigkeit dargestellt werden. Für dessen Operationalisierung wurden in den Korpusdaten enthaltene Merkmale identifiziert, die durch automatisierte Verfahren zugänglich gemacht werden können⁷. Zu den Verfahren zählt der Abgleich der Transkripte mit Wortschatzlisten verschiedener Niveaustufen bzw. Frequenzlisten. Außerdem gehört dazu die Ermittlung

- (1) des Anteils orthografisch normalisierter Einheiten im Transkript (Normalisierungsrate),
- (2) der Anzahl artikulierter (d.h. transkribierter) Silben pro Sekunde (Sprechgeschwindigkeit)⁸,
- (3) des Anteils längerer überlappend gesprochener Passagen sowie

⁷ Es ist uns bewusst, dass die ausgewählten Verfahren nur eine erste Annäherung an das Konzept ‚sprachlich-kommunikative Schwierigkeit‘ darstellen können. Zum einen ist ‚Schwierigkeit‘ nicht unabhängig von den Lernenden, deren Vorwissen und Handlungszielen ermittelbar, zum anderen kann ‚Schwierigkeit‘ auch nicht einfach an der sprachlichen Oberfläche abgelesen werden, sondern müsste lerner/-innenspezifisch empirisch erhoben werden (vgl. etwa Ender / Kaiser 2020). Im Rahmen des hier dargestellten Projekts war dies nicht möglich; entsprechend den Förderrichtlinien konnten auch keine neuen Annotationen erfolgen, so dass mit den hier dargestellten Verfahren die bereits in den Daten angelegten Informationen möglichst umfassend genutzt wurden.

⁸ Für die Bestimmung der Sprechgeschwindigkeit gibt es zwei Maße, die Artikulationsrate, die sich nur auf das pro Zeiteinheit geäußerte sprachliche Material (einschließlich gefüllter Pausen) bezieht, sowie die Sprechrate, welche stille Pausen in die Berechnung miteinschließt (vgl. Laver 1994: 539). Der Berechnung der Sprechgeschwindigkeit in *ZuMult* liegt die Artikulationsrate zugrunde.

- (4) des Anteils bestimmter für das Mündliche typischer Phänomene (wie Modalpartikeln, Diskursmarker, Klitisierungen u.a.).

Ausgehend von diesen Parametern wurden Zugänge für die Auswahl von Sprechereignissen (ZuMal) und die Unterstützung der didaktischen Arbeit mit einem ausgewählten Gespräch (ZuViel) entwickelt⁹.

ZuMal (Zugang zu Merkmalsauswahl von Gesprächen) kann für die Korpora FOLK und GWSS angewendet werden und ermöglicht eine Vorauswahl an Sprechereignissen, bei der die Metadaten mit den Einstellungen für schwierigkeitsbezogene Parameter kombiniert werden können. Zu den Metadaten gehören insbesondere die folgenden Kategorien: Gesprächstyp (privat, institutionell, öffentlich), Art des Gesprächs (bspw. Telefongespräch, Gespräch beim Arbeitsamt, Tischgespräch), Gesprächsthemen, Sprachregion und Dauer des Gesprächs. Die schwierigkeitsbezogenen Parameter umfassen Niveaustufe / Wortschatz, Standardnähe, Sprechgeschwindigkeit, Überlappungen sowie Häufigkeit mündlicher Phänomene. Es kann also bspw. nach Telefongesprächen von maximal halbstündiger Dauer, einer hohen Deckung durch B1-Wortschatz und geringer Normalisierungsrate gesucht werden (für konkretere Anwendungsszenarios vgl. Meißner / Wallner im E. sowie Fandrych / Schwendemann / Wallner 2021).

Mit Hilfe des Zugangs ZuViel (Zugang zu Visualisierungselementen für Transkripte) kann die Arbeit mit einem Transkript für Unterrichtszwecke oder auch qualitative Analysen unterstützt werden. Hat man in ZuMal ein bestimmtes Sprechereignis ausgewählt, kann man es in ZuViel öffnen, um damit weiter zu arbeiten¹⁰. Im Interface wird das Transkript in Listennotation angezeigt. Es ist verknüpft mit einem Audio- oder Videoplayer (mit einblendbaren Untertiteln), anhand dessen in der Aufnahme navigiert werden kann. Eine Lemmaliste führt alle im Transkript vorkommenden Einheiten mit ihren Frequenzen auf und kann mit Referenzwortlisten abgeglichen werden. Für die Transkriptansicht sind verschiedene Markierungen einstellbar (etwa die Deckung durch eine Wortschatzliste oder die vom schriftlichen Standard abweichend realisierten Einheiten). So kann etwa unbekannter Wortschatz visuell wahrnehmbar gemacht und direkt thematisiert werden (für eine Visualisierung vgl. Fandrych / Schwendemann / Wallner 2021). Unterstützt wird dies weiterhin durch die Option, entsprechende Wortlisten (z.B. nicht-gedeckter Wortschatz, von der Standardlautung abweichende Einheiten mit normalisierter und Lemmaform) herunterladen und so für die Vorentlastung nutzen zu können. Auch der Download ausgewählter Gesprächssequenzen ist möglich. Es kann zudem zwischen aussprachenaher und orthographisch normalisierter Transkriptansicht gewechselt sowie die Abspielgeschwindigkeit des Audios variiert werden. Dadurch wird eine binnendifferenzierende Arbeit mit dem Transkript möglich. So kann für den Einstieg die inhaltliche Erschließung anhand der orthographisch normalisierten Fassung erfolgen und anschließend anhand der aussprachenahen Transkription mit Hilfe der Hervorhebung der abweichend realisierten Einheiten gezielt auf Merkmale der Mündlichkeit eingegangen werden. Durch eine kompakte Übersicht über die zeitliche Anordnung der Gesprächsbeiträge der verschiedenen Sprecher/-innen (Transkript-Navigator) können Passagen mit verstärkter Interaktivität auf einen Blick identifiziert werden (vgl. Fandrych / Wallner im E.).

⁹ Die Zugänge sind in der Datenbank für gesprochenes Deutsch über den Reiter „Recherche“ erreichbar. Zur Illustration der in *ZuMult* entwickelten Zugänge gibt es auch ein Video (<https://www.youtube.com/watch?v=zZ-mtnVDUs8>, 26.07.2021).

¹⁰ Alternativ kann ZuViel auch ausgehend von der DGD und von dem auch im Projekt *ZuMult* entwickelten Prototypen ZuRecht aufgerufen werden.

1.3 Nutzung von mündlichen Korpora und Anforderungen an eine Corpus Literacy

Korpora gesprochener Sprache bieten vielfältige Einsatzmöglichkeiten für den Unterricht und eröffnen über die in *ZuMult* entwickelten Zugriffsmöglichkeiten für die Fremdsprachendidaktik spezifische Anwendungsoptionen. Nutzer/-innen für diese Ressourcen wären damit zunächst Lehrpersonen, welche dann wiederum Lernende mit den Anwendungsmöglichkeiten vertraut machen und so deren selbständige, individuellen Interessen folgende Nutzung der Angebote ermöglichen können (vgl. zu den Potenzialen der *ZuMult*-Zugänge für das autonome Lernen Meißner / Wallner im E.). Damit Lehrpersonen diese Vermittlungsaufgabe übernehmen können, sollten sie in ihrer Ausbildung Gelegenheit haben, Kompetenzen der Corpus Literacy zu erwerben, d.h. Kenntnisse über vorhandene Korpora, Konzepte der Korpuslinguistik (wie Metadaten, Lemmatisierung, Wortartenannotation) und Methoden der Korpusnutzung, aber auch die Fähigkeit, die Art der Daten und die Aussagekraft der Abfrageresultate bewerten zu können (etwa die Formbezogenheit der Verarbeitung, die unterschiedliche Lesarten unberücksichtigt lässt; vgl. z.B. Beißwenger / Storrer 2011; Callies 2019). In Bezug auf Korpora gesprochener Sprache sind diese Kenntnisse zu erweitern um Aspekte der Spezifik mündlicher Kommunikation und ihrer Überführung in Schrift mittels Transkription (Konventionen, Umfang der erfassten Phänomene der Sprechsituation), die Konzepte der orthographischen Normalisierung und der Annotation spezifisch mündlicher Wortarten (vgl. Westpfahl et al. 2017) sowie den Aspekt der Qualität der automatisch durchgeführten Aufbereitungsschritte für unterschiedliche mündliche Daten¹¹. Diese Kenntnisse bilden eine Voraussetzung auch für die Interpretation einiger der in *ZuMal* angebotenen schwierigkeitsbezogenen Auswahlparameter (z.B. zum formbezogenen Abgleich von Transkripten und Wortschatzlisten oder der Ermittlung der Normalisierungsrate). Schließlich stellt auch die Fähigkeit, die Korpusressourcen für den Unterricht nutzen zu können, einen Aspekt der Corpus Literacy für Lehrpersonen dar. Hierzu zählt zum einen die eigene Verwendung der Korpusressourcen zur Aufgaben- und Materialerstellung, zum anderen die Vermittlung von Korpusnutzungs Kompetenzen als Hilfsmittel für Lernende. So sollten Lernende in die Lage versetzt werden, sich bestimmte Phänomene mündlicher Kommunikation selbstständig erschließen zu können (etwa durch lexikographische Ressourcen wie LeGeDe sowie Korpora als Quelle für weitere Beispiele, aus denen die Bedeutung und Funktion einer Struktur erschlossen werden kann), Korpusressourcen für die Gestaltung der eigenen mündlichen Produktion hinzuziehen zu können (etwa um Einleitungssequenzen für Vorträge zu formulieren) oder die Unterschiede des mündlichen Sprachgebrauchs in verschiedenen privaten, institutionellen oder öffentlichen Kontexten zu reflektieren. Die Nutzung von Korpora eignet sich hierbei insbesondere für den sprachdidaktischen Ansatz des datengeleiteten Lernens, bei dem sich Lernende selbstständig mit authentischen Daten der Sprachverwendung auseinandersetzen, auf dieser Grundlage Gebrauchscharakteristika ableiten und somit eine aktive Rolle bei der Wissenskonstruktion einnehmen (vgl. u.a. Römer 2011). Dabei kann sich die Auswahl der Sprachdaten an individuellen Interessen und Bedürfnissen orientieren. Korpora wie FOLK oder GWSS bieten hierfür ein weites Spektrum des Sprachgebrauchs aus unterschiedlichen Domänen.

¹¹ Die Fehlerquote der Wortartenannotation liegt derzeit bei 5 % (vgl. Westpfahl / Schmidt 2016).

2. Erforschung der mündlichen Kommunikation zur Bestimmung relevanter Vermittlungsgegenstände

Korpora der gesprochenen Sprachen sind nicht nur eine Ressource für authentische Sprachdaten, die im Unterricht genutzt werden, sondern bilden auch eine gute Grundlage für die Erforschung der mündlichen Kommunikation. Mit der zunehmenden Verfügbarkeit mündlicher Korpora ist auch eine wachsende Anzahl an Forschungsarbeiten, die sich korpusbasiert mit den Charakteristika der gesprochenen Sprache befassen, zu beobachten. Im Fokus stehen dabei zum einen Phänomene auf der Einzelwortebene – so u.a. der Gebrauch von Konnektoren (vgl. Slavcheva 2018), von Diskursmarkern (vgl. Slavcheva / Meißner 2014; Wallner 2017), von Verschmelzungen (etwa von Präposition und Artikel, vgl. Lenort / Pohle / Sakhno im E.), von Modalpartikeln (vgl. Gutzmann / Turgay 2016) oder auch von sprachlichen Realisierungen eines einzelnen Lemmas (wie bspw. *haben* bei Lanwer / Schopf 2021). Zum anderen werden aber auch über die Einzelwortebene hinausgehende Phänomene und Fragestellungen auf der Grundlage mündlicher Korpora untersucht – so etwa syntaktische Aspekte wie die Verbspitzenstellung und Apokoinu-Strukturen (vgl. Schneider / Butterworth / Hahn 2018) oder auch die Realisierung sprachlicher Handlungen (vgl. Meißner 2016, 2017). Aus sprachdidaktischer Perspektive sind diese Arbeiten von besonderem Interesse, da auf ihrer Grundlage diejenigen gesprochensprachlichen Phänomene identifiziert werden können, die Gegenstand der Vermittlung im DaF- und DaZ-Unterricht sein sollten. Neben der Bestimmung relevanter Vermittlungsgegenstände können korpusbasierte Forschungsarbeiten auch verlässliche Erkenntnisse im Hinblick auf Vorkommenshäufigkeit und Verbreitung sowie bezüglich der sprachlichen Eigenschaften und Gebrauchsbesonderheiten von gesprochensprachlichen Phänomenen liefern. Die DaF/DaZ-Didaktik kann davon enorm profitieren, da sich anhand dieser Untersuchungen die Auswahl und Gewichtung von Vermittlungsgegenständen sowie Informationen zur sprachlichen Ausgestaltung und Kontextbedingungen ableiten lassen. Dies soll im Folgenden anhand ausgewählter Ergebnisse von zwei Studien genauer erläutert werden. Dabei geht es zunächst um Klitisierungen in der mündlichen Kommunikation, die auf der Grundlage des FOLK-Korpus mit Hilfe des in *ZuMult* entwickelten Werkzeugs *ZuRecht* analysiert wurden. Dieses als Ergänzung zur Datenbank für Gesprochenes Deutsch (DGD) entwickelte Werkzeug ermöglicht es, umfassende quantitative Informationen zum Gebrauch und zur konkreten sprachlichen Realisierung von Klitisierungen für ein gesamtes Korpus zu erheben und unter Berücksichtigung sprachdidaktischer Informationsbedürfnisse systematisch auszuwerten. Daran anschließend stehen Möglichkeiten zur korpusbasierten Beschreibung der typischen Realisierung sprachlicher Handlungen im Fokus. Am Beispiel von Metakomentierungen wird gezeigt, wie sich Spezifik und Musterhaftigkeit im Gebrauch anhand mündlicher Korpora empirisch bestimmen lassen.

2.1 Klitisierungen in der mündlichen Kommunikation

Unter Klitisierungen werden gemeinhin Verschmelzungen von lexikalischen Einheiten verstanden. Dazu zählen sowohl grammatikalisierte Einheiten wie etwa *zum* [*zu + dem*] oder *ins* [*in + das*], die in komplementärer Distribution zu den entsprechenden Vollformen stehen, als auch Verschmelzungen wie *kannste* [*kannst + du*] oder *inner* [*in + der*], die als (noch) nicht vollständig lexikalisiert gelten und als Besonderheit der mündlichen Kommunikation beschrieben werden (vgl. Duden Grammatik 2009: 1213-1214). Obwohl die als (noch) nicht lexikalisiert eingestuften Klitisierungen vielerorts als typisches Charakteristikum der gesprochenen Sprache erwähnt werden, existieren nur wenige Arbeiten, die sich auf der Grundlage mündlicher Korpora mit diesem Phänomen befassen. Ausnahmen bilden hier etwa die Arbeiten von Lanwer (2015), Augustin (2018)

und Lenort / Pohle / Sakhno (im E.). Im Fokus stand dabei bislang stets nur ein Teilbereich der Klitisierungen, wie etwa die Verschmelzung von Präposition + Artikel. Aus DaF/DaZ-didaktischer Perspektive ist jedoch auch ein Gesamtüberblick über das Verschmelzungsverhalten sprachlicher Einheiten in der gesprochenen Sprache von Interesse¹². Das FOLK-Korpus bietet hierfür eine geeignete Grundlage, da es eine große Vielfalt unterschiedlicher Sprechereignisse aus dem privaten, institutionellen und dem öffentlichen Bereich umfasst¹³. Zudem wird durch die Bereitstellung sowohl aussprachenaher Transkriptionen als auch der orthografisch normalisierten Fassungen eine empirische Ermittlung aller Vorkommen von Klitisierungen im Korpus ermöglicht.

Erste Analysen haben dabei gezeigt, dass Klitisierungen in fast allen der aktuell verfügbaren 374 Sprechereignisse vorkommen. Ausnahmen bilden lediglich vier überwiegend kurze institutionelle Sprechereignisse¹⁴. Im Durchschnitt sind 1,74 % der Token eines Sprechereignisses von Klitisierung betroffen. Die Standardabweichung ist allerdings mit $SD=0,99$ % recht hoch. Beim Vergleich der Interaktionsdomänen fällt auf, dass in den öffentlichen Sprechereignissen ($n=16$) der Anteil der Klitisierungen mit durchschnittlich 0,56 % ($SD=0,33$ %, Median (Mdn)=0,49 %) deutlich geringer ausfällt als in den privaten Sprechereignissen ($n=140$) mit 1,76 % ($SD=0,85$ %, $Mdn=1,61$ %), in den institutionellen Sprechereignissen ($n=167$) mit 1,79 % ($SD=1,04$ %, $Mdn=1,60$ %) und in den unter ‚sonstige‘ gefassten Sprechereignissen ($n=51$) mit 1,91 % ($SD=1,10$ %, $Mdn=1,67$ %). Ein Kruskal-Wallis-Test¹⁵ ergibt signifikante Unterschiede zwischen den Interaktionsdomänen ($H(3)=31,908$; $p<0,001$). Die paarweisen Vergleiche der Interaktionsdomänen (Dunn-Bonferroni-Test) mit angepassten p -Werten (Bonferroni-Korrektur) haben gezeigt, dass sich die öffentlichen Sprechereignisse hinsichtlich des Anteils an Klitisierungen von allen anderen Interaktionsdomänen signifikant mit annähernd mittleren Effektstärken unterscheiden (öffentlich vs. institutionell: $z=5,399$; $p<0,001$; $r=0,28$; öffentlich vs. privat: $z=5,429$; $p<0,001$; $r=0,28$; öffentlich vs. sonstige: $z=5,275$; $p<0,001$; $r=0,27$)¹⁶. Zwischen den Interaktionsdomänen institutionell, privat und sonstige ergeben sich keine signifikanten Unterschiede. Abbildung 1 visualisiert die jeweilige Anzahl an Klitisierungen auf 100 Token in den institutionellen, privaten, öffentlichen und sonstigen Sprechereignissen im FOLK. Anhand des Boxplots wird die Streuung der Daten deutlich. Demnach gibt es unter den institutionellen, privaten und sonstigen Sprechereignissen sowohl solche mit nur wenigen Klitisierungen pro 100 Token als auch solche, die über fünf Klitisierungen pro 100 Token aufweisen¹⁷. Die Boxen, welche jeweils 50 % der Sprechereignisse abbilden, verdeutlichen jedoch,

¹² Hierzu wird derzeit von Franziska Wallner und Matthias Schwendemann eine empirische Studie durchgeführt, die im Jahr 2022 abgeschlossen sein wird.

¹³ Zu den privaten Sprechereignissen zählen bspw. Tischgespräche, Telefonate und Spielinteraktionen, zu den institutionellen gehören u.a. Verkaufsgespräche, Teambesprechungen, Beratungen; öffentliche Sprechereignisse umfassen Podiumsdiskussionen, TV-Debatten, Schlichtungsgespräche. Hinzu kommen noch verschiedene Interviewsituationen und Kommunikationsspiele (Maptask), die unter sonstige Sprechereignisse gefasst werden. Zwar ist das FOLK-Korpus bezüglich verschiedener diatopischer, diastratischer und situativer Merkmale noch keinesfalls ausgewogen, das dynamische Wachstum lässt aber hoffen, dass auch die Ausgewogenheit weiter zunimmt und die zur Verfügung stehenden Kommunikationsdomänen auch aus DaF/DaZ-Sicht zunehmend alle relevanten Sprachverwendungsbereiche abdecken.

¹⁴ Bei diesen Sprechereignissen handelt es sich um zwei Gespräche im Polizeirevier von lediglich 19 bzw. 36 Sekunden, um ein Verkaufsgespräch von einer Minute und 18 Sekunden sowie um eine Sprachfördersituation im Kindergarten von 14 Minuten und acht Sekunden.

¹⁵ Der Kruskal-Wallis-Test ist ein nicht-parametrischer Test, der angewendet wurde, da die Voraussetzungen für eine einfaktorielles ANOVA nicht erfüllt waren (keine Varianzhomogenität und Normalverteilung der Daten).

¹⁶ Der Wert r gibt hier die Effektstärke und damit die Relevanz der Ergebnisse an. Er quantifiziert den Unterschied in den Mittelwerten der Gruppen. Nach Cohen (1988) bedeutet $r\leq 0,1$ einen geringen Effekt, $r\leq 0,3$ einen mittleren Effekt und $r\geq 0,5$ einen großen Effekt.

¹⁷ Die meisten Klitisierungen zeigt ein unter ‚sonstige‘ gefasstes Sprechereignis mit 6,29 Klitisierungen pro 100 Token. Es handelt sich dabei um ein Kommunikationsspiel (Maptask).

dass die Sprechereignisse in diesen drei Domänen hinsichtlich ihrer Anzahl an Klitisierungen pro 100 Token vergleichbar sind. Auch der Median (visualisiert durch die Linie in der Box) ist annähernd gleich. Bei den öffentlichen Sprechereignissen fällt die Streuung hingegen deutlich geringer aus. Zudem liegen die Sprechereignisse mit ihrer jeweiligen Anzahl an Klitisierungen pro 100 Token sowie der Median in einem niedrigeren Bereich. Darüber hinaus wird anhand der Violin-Plots sichtbar, wie sich die Sprechereignisse in den einzelnen Domänen je nach Anzahl an Klitisierungen verteilen. So ist beispielsweise erkennbar, dass in der privaten Domäne etwas mehr Sprechereignisse einen Klitisierungswert unterhalb des Medians aufweisen, während in der institutionellen Domäne Sprechereignisse mit einem Klitisierungswert unterhalb und oberhalb des Medians ähnlich verteilt sind.

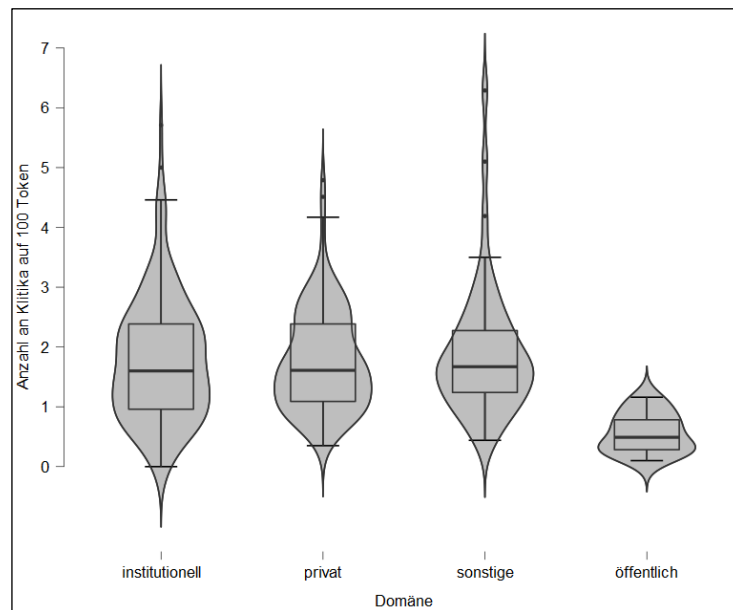


Abbildung 1

Anzahl an Klitisierungen auf 100 Token in den verschiedenen Interaktionsdomänen im FOLK-Korpus

Aus Perspektive der DaF / DaZ-Didaktik macht dies deutlich, dass öffentliche Sprechereignisse, auf die in Vermittlungskontexten aufgrund ihrer leichteren Zugänglichkeit gern zurückgegriffen wird (man denke da etwa an den Einsatz von Podcasts, Radio- oder Fernsehsendungen), im Bereich der Klitisierungen offensichtlich von anderen Sprechereignissen aus dem privaten oder institutionellen Bereich erheblich abweichen können¹⁸. Bei der Auswahl von Sprechereignissen für die Mündlichkeitsdidaktik sollte dieser Befund berücksichtigt werden. Die Frage nach Sprechereignissen mit besonders vielen Klitisierungen ergibt hingegen ein heterogenes Bild. Unter den Sprechereignissen mit einem Anteil an Klitisierungen auf alle Token von mindestens 4 % finden sich private Sprechereignisse (Paargespräch und ein Familiengespräch beim Umräumen eines Kinderzimmers), institutionelle Sprechereignisse (Schichtübergabe in einem Krankenhaus, Verkaufsgespräche im Gartencenter und in der Apotheke, Gespräch im Polizeirevier) sowie ein Maptask¹⁹, das dem Bereich sonstige Sprechereignisse zugeordnet wird. Viele von der Gesprächsart

¹⁸ Hierbei ist jedoch zu beachten, dass der Anteil an öffentlichen Gesprächen im FOLK-Korpus mit 16 Sprechereignissen relativ gering ausfällt. Sie sind jedoch vergleichsweise lang und stellen mit knapp 44h 14 % der insgesamt 314 Aufnahmestunden.

¹⁹ Unter Maptask versteht man ein Kommunikationsspiel „bei de[m] jedem der zwei Gesprächsbeteiligten eine Karte vorliegt, auf der Gegenstände oder Personen abgebildet sind. Auf der Karte eines Gesprächsteilnehmers befindet sich

vergleichbare Sprechereignisse zeigen jedoch einen eher durchschnittlichen Anteil an Klitisierungen. Eine Empfehlung für die DaF/DaZ-Didaktik im Hinblick auf eine etwaige Eignung bestimmter Gesprächsarten für die Thematisierung von Klitisierungen lässt sich daher aufgrund dieser Heterogenität nicht ableiten.

Neben der Interaktionsdomäne und der Gesprächsart sind aus didaktischer Perspektive die konkreten sprachlichen Realisierungen von Klitisierungen von Interesse. Zu den häufigsten Vorkommen von Klitisierungen im FOLK gehören aktuell *son* [*so ein*], *gibts* [*gibt es*] und *gehts* [*geht es*]. Interessant ist dabei, dass diese im Korpus meist deutlich frequenter sind als die in eckigen Klammern aufgeführten analytischen Formen. Unter den häufigsten zehn gibt es diesbezüglich aber auch eine Ausnahme: So kommt die analytische Form *mit dem* deutlich häufiger als ihre Klitisierung *mim* vor. Dies konnte auch bei anderen Klitisierungen des Typs Präposition und Artikel beobachtet werden – etwa bei der Verschmelzung *fürn* [*für den*] mit 434 Treffern und 621 Treffern bei der analytischen Form²⁰.

Tabelle 1 zeigt die häufigsten zehn Vorkommen von Klitisierungen im FOLK-Korpus. Darin wird die Anzahl der Treffer für die Klitisierungen²¹ und für die dazugehörigen analytischen Formen²² angegeben.

Klitisierung und analytische Form	Frequenz der Klitisierung	Frequenz der analytischen Form
<i>son</i> [<i>so ein</i>]	2481	313
<i>gibts</i> [<i>gibt es</i>]	1878	516
<i>gehts</i> [<i>geht es</i>]	1082	206
<i>ichs</i> [<i>ich es</i>]	946	89
<i>wars</i> [<i>war es</i>]	817	103
<i>wenns</i> [<i>wenn es</i>]	810	220
<i>isn</i> [<i>ist ein</i>]	633	195
<i>habs</i> [<i>hab es</i>]	564	21
<i>mim</i> [<i>mit dem</i>]	529	2171
<i>hats</i> [<i>hat es</i>]	502	91

Tabelle 1
Die zehn frequentesten Klitisierungen im FOLK

Die quantitativen Verhältnisse sprechen dafür, die klitisierten Formen – zumindest im Rahmen der Mündlichkeitsdidaktik – parallel zu den analytischen Formen zu vermitteln, wie das etwa auch Fandrych / Thurmair (2021: 100) für *son* vorschlagen.

zudem ein Weg, der dem anderen nicht bekannt ist. Dieser kann den Weg wegen eines Sichtschutzes nicht sehen. Der erste Gesprächsteilnehmer beschreibt diesen Weg, der andere zeichnet ihn in seine Karte ein. Anschließend vergleichen sie die Wege. Dann werden die Rollen getauscht.“ (vgl. <https://dgd.ids-mannheim.de>, 26.07.2021)

²⁰ Zu berücksichtigen ist hierbei, dass eine solche quantitative Gegenüberstellung nur einen ersten Eindruck im Hinblick auf die jeweiligen Realisierungstendenzen liefert, da die analytischen und die synthetischen Formen nicht überall austauschbar sind. Bei weiteren Analysen müsste diesbezüglich auch der jeweilige Kontext einbezogen werden.

²¹ Bei der Transkription wurden die Klitisierungen im FOLK entweder zusammengeschrieben (*gibts*) oder ein Teil als assimiliert gekennzeichnet (*gibt s* <assimiliert>). Für die vorliegende Auswertung wurden die entsprechenden Vorkommen aufsummiert und für eine bessere Darstellung die zusammengeschriebene Form verwendet.

²² Als analytische Form wurde diejenige gewählt, in die die jeweilige Klitisierung im FOLK am häufigsten im Rahmen der orthografischen Normalisierung zerlegt wurde. Bei der Klitisierung *ichs* ist dies die analytische Form *ich es* mit 869 Treffern. Daneben gibt es aber auch die analytischen Formen *ich das* mit 61 Treffern, *ich sie* mit 4 Treffern sowie einige Fehltreffer und Abbrüche. Zu berücksichtigen ist dabei, dass die orthografische Normalisierung ein automatischer Aufbereitungsschritt ist, der fehlerbehaftet sein kann.

Daneben stellt sich auch die Frage, welche Wortarten am häufigsten klitisiert werden. Unter den im FOLK-Korpus angetroffenen Klitisierungen finden sich am häufigsten Verschmelzungen von Verb + Personalpronomen (bspw. *haste* [*hast du*]) mit insgesamt 20197 Treffern. Berücksichtigt wurden dabei alle Klitisierungen mit mindestens 20 Vorkommen im Korpus. Weitere Klitisierungsmuster, die in FOLK häufig vorkommen, sind Verschmelzungen von Präpositionen + Artikel (bspw. *inner* [*in der*]) mit 4341 Treffern, Verschmelzungen von Fokus-/Intensitäts- und Gradpartikel + Artikel (bspw. *some* [*so eine*]) mit 3637 Treffern oder auch Verschmelzungen von zwei Pronomen (bspw. *dus* [*du es*]) mit 2983 Treffern.

Die vorgestellten Ergebnisse geben bereits einen interessanten Einblick in den Gebrauch von Klitisierungen in der gesprochenen Sprache. Lehrende sowie Autor/-innen von Lehrmaterialien können auf der Grundlage dieser und weiterer Befunde entscheiden, welche Klitisierungen im Rahmen der Mündlichkeitsdidaktik aufgegriffen werden sollten. So könnte etwa im Anschluss an eine erste Begegnung mit dem Phänomen Klitisierung im authentischen Hörtext bzw. im Transkript eine Sensibilisierung für typische Klitisierungen und Klitisierungsmuster erfolgen, die Lernende dazu befähigt, entsprechende Vorkommen zu erkennen, aufzulösen, adäquat zu gebrauchen sowie ggf. auch eigenständig zu bilden. Hierfür wären auch weitere Studien wünschenswert, die bspw. zeigen, welche Klitisierungen überregional relevant sind, zur Realisierung welcher sprachlicher Handlungen und unter welchen grammatisch-äußerungsstrukturellen Bedingungen Klitisierungen üblicherweise zum Einsatz kommen und in welchen konkreten Situationen und kontextuellen Einbettungen klitisierte bzw. analytische Formen bevorzugt werden.

2.2 Realisierung sprachlicher Handlungen in der mündlichen Kommunikation

Sprachliche Handlungen bilden seit dem kommunikativen Ansatz einen wesentlichen Vermittlungsgegenstand der Fremdsprachendidaktik. Nicht nur die interaktionalen Abläufe selbst sind häufig musterhaft, sondern sie werden – je nach Kontext und kommunikativem Ereignis in unterschiedlichem Maße – häufig mithilfe wiederkehrender, musterhafter Formulierungen (vgl. Wray 2002) bzw. gesprächsarttypischer Routinen realisiert. Zwar finden sich in Lehrwerken und curricularen Dokumenten verschiedene Redemittelsammlungen zu bestimmten sprachlichen Handlungen (z.B. Glaboniat et al. 2005), diese wurden jedoch bislang weder auf empirischer Basis gewonnen noch empirisch überprüft. Korpora gesprochener Sprache eröffnen hier die Möglichkeit, Lehrenden und Lernenden eine gebrauchsbasierte Auswahl und Beschreibung typischer Handlungsrealisierungen bereitzustellen. Als pragmatisches Phänomen sind sprachliche Handlungen jedoch nicht ohne weiteres über formale Merkmale in Korpora abfragbar. Methodisch wird daher in der Korpuslinguistik versucht, typische sprachlich-pragmatische Muster über interpretativ-manuelle Annotation (vgl. z.B. Alsop / Nesi 2014; Baur et al. 2014), angenommene typische Formmerkmale (vgl. z.B. Rühlemann 2010) oder die korpusweite Ermittlung von N-Grammen, die anschließend funktional interpretiert werden (z.B. Scharloth / Bubenhofer 2012), zu ermitteln. Ziel der korpuspragmatischen Analyse ist es, Zusammenhänge zwischen sprachlichen Mitteln und Ausdrucksweisen einerseits und (Sprach-)Handlungsfunktionen andererseits herauszuarbeiten (vgl. Felder / Müller / Vogel 2012). Vor diesem Hintergrund wurde in den Untersuchungen von Meißner (2016, 2017) ein Beschreibungsansatz für die Realisierung von Sprachhandlungen entwickelt, der auf Basis der manuellen Annotation eines Referenzkorpus typische sprachliche Oberflächenmerkmale der Handlung bestimmt und diese anschließend in Form eines Suchausdrucks auf nicht-annotierte Untersuchungskorpora anwendet, um potenzielle Vertreter der Handlung zu ermitteln. Anhand von Beschreibungsparametern, die aus dem Suchausdruck, seinen Bestandteilen und den jeweils erzielten

Treffern und Fehltreffern abgeleitet werden, lässt sich die Realisierung der Sprachhandlung im Vergleich zum Referenzmuster charakterisieren.

Ausgearbeitet wurde dieser Ansatz am Beispiel der Metakomentierungen, einem der Gliederung und Rezipientenorientierung dienenden Handlungstyp, der für wissenschaftliche Vorträge beschrieben (vgl. Fandrych 2014) und in den L1-Expertenvorträgen des GWSS-Korpus annotiert wurde (vgl. Baur et al. 2014)²³. Durch die korpuslinguistische Analyse dieser annotierten Belege konnten Formen ermittelt werden, deren Vorkommen im Korpus mit der Handlung des Metakomentierens assoziiert war (z.B. Modal- und Hilfsverbformen wie *möchte / ich möchte, ich will / wollt, werd(e)*; Vollverbformen wie *kommen wir, vorstellen, festzuhalten, eingehen*; temporale adverbiale Ausdrücke wie *als nächstes, vorhin, gleich noch, kurz*; diskontinuierliche Ausdruckskombinationen wie z.B. *ich* in Verbindung mit *jetz(t) / hier / nun / noch / ihnen etc.* im Abstand von bis zu fünf Wörtern und in unterschiedlicher Reihenfolge, oder die elliptische Form *nun zu* (wie in: *nun zu den Ergebnissen*). Die folgenden Beispiele zeigen zwei mit diesen typischen Merkmalen realisierte Metakomentierungen:

- (1) das **möcht ich ihnen hier** auch **kurz** °h **vorstellen** (GWSS_E_00020)
- (2) da **werd ich** wenn **ich** (.) zeit habe später auch **noch kurz** drauf **eingehen** (GWSS_E_00023)

Im ersten Beleg erfolgt die einleitende Ankündigung einer sich anschließenden Sprechhandlung, im zweiten Beleg wird ein Vorausverweis auf eine später folgende Handlung gegeben. Der auf der Basis der ermittelten Merkmale gebildete Suchausdruck erfasst 60 % der manuell annotierten Belege bei einer Korrektheit von 76 % (vgl. Meißner 2017: 194). Es konnte damit gezeigt werden, dass die Realisierung der Handlung in den untersuchten Daten variiert, aber zu großen Teilen auf der Verwendung wiederkehrender Formen basiert. Für die Fremdsprachenvermittlung liefern die ermittelten Indikatormerkmale insofern Erkenntnisse, als sich mit ihnen produktiv die Handlung musterhaft erkennbar realisieren lässt und sie rezeptiv Verstehenshilfen bilden, anhand derer die Handlung identifiziert werden kann²⁴. Durch die Anwendung des Suchausdrucks auf Vergleichskorpora konnten darüber hinaus Hypothesen über die Realisierung von Metakomentierungen bei anderen Sprechereigenschaften gebildet und Perspektiven für die Variationsforschung und Lernaltersprachanalyse eröffnet werden: So ergab die Auswertung in einem Vergleichskorpus aus Konferenzvorträgen polnischer Deutsch-L2-Sprecher/-innen (GWSS-Teilkorpus DEU_L2_PL_EV), dass hier reduzierte Formen (*wollt, komm*) sowie die elliptische Kurzform (*nun zu*) gemieden werden und sich gleichzeitig im Vergleich zu den L1-Vorträgen ein Mehrgebrauch unreduziert realisierter Formen (*möchte ich, kommen wir, haben sie*) und der genitivischen Form (*vortrags*) findet. Bei der Auswertung eines Vergleichskorpus britischer Deutsch-L2-Sprecher/-innen (GWSS-Teilkorpus DEU_L2_UK_EV) zeigten sich solche Tendenzen hingegen nicht. Dies deutet auf Unterschiede zwischen Diskursgemeinschaften im Gebrauch von Mündlichkeitsmerkmalen im wissenschaftlichen Vortrag hin (vgl. Meißner 2016). Durch die Auswertung in einem GWSS-Vergleichskorpus mit studentischen Vorträgen von L1- und L2-Sprecher/-innen des Deutschen konnten lernaltersprachliche Unterschiede in der Realisierung der Handlung beschrieben werden. So zeigte sich in den studentischen L1-Vorträgen eine größere Varianz in der Abfragekorrektheit des Suchausdrucks, während L2-Vorträge recht ähnliche Werte aufwiesen. Dies deutet auf Unterschiede in der Verknüpfung von Form und Funktion bei diesen

²³ In den Annotationen im GWSS-Korpus tragen Metakomentierungen die Bezeichnung „Diskurskommentierungen“.

²⁴ Durch die mit dem Suchwerkzeug ZuRecht geschaffenen Abfragemöglichkeiten kann nun im GWSS-Korpus innerhalb der annotierten Metakomentierungen nach den genannten Indikatormerkmalen gesucht werden. Eine Anleitung dazu findet sich auf der Hilfeseite in ZuRecht. Es können so typische Handlungsrealisierungen abgefragt und im Unterricht thematisiert werden.

Sprechergruppen und eine festere, musterhaftere Verknüpfung bei L2-Sprechenden hin. Bezogen auf Präzision und Ertrag einzelner Merkmale zeigten sich in den L2-Vorträgen insgesamt weniger Merkmale des Suchausdrucks, diese wiesen jedoch größere Treffermengen auf, was darauf schließen lässt, dass in den L2-Daten Metakomentierungen mit (wenigeren) wiederkehrenden Merkmalen realisiert werden (vgl. Meißner 2017: 197-204).

Korpusanalysen wie diese zeigen das Potenzial mündlicher Korpora für die gebrauchsbasierte Erforschung sprachlicher Handlungen für die Fremdsprachenvermittlung. Anhand von Korpora können Spezifik und Musterhaftigkeit der Realisierung sprachlicher Handlungen empirisch bestimmt und so Grundlagen für die Vermittlung erarbeitet werden. Da die Verlässlichkeit der Zuordnung von Form und Bedeutung bzw. Funktion einen entscheidenden Faktor für Spracherwerb und sprachliches Lernen bildet (vgl. Ellis 2017), ist es eine wesentliche Aufgabe für die fremdsprachendidaktische Sprachbeschreibung, die verlässlich indizierenden Merkmale zu bestimmen, die sehr stark bis eindeutig mit der Realisierung einer bestimmten Handlung verknüpft sind. Korpuspragmatische Analysen auf der Basis mündlicher Korpora können hier einen wichtigen Beitrag zum Verständnis des Form-Funktion-Zusammenhangs im mündlichen Sprachhandeln leisten. So zeigt eine aktuelle Untersuchung anhand des FOLK-Korpus, dass sich der Geltungsbereich der auf der Basis wissenschaftlicher Vorträge ermittelten Indikatormerkmale für Metakomentierungen auch auf die gesprächssteuernden Sprecherrollen in den thematisch strukturierten Interaktionstypen von Lehrer/-innen in Unterrichts-, Prüfer/-innen in Prüfungs- und Schlichter/-innen in Schlichtungsgesprächen erstreckt (vgl. Meißner angenommen). Die hieraus ableitbaren formbasierten Beziehungen zwischen sprachlichen Handlungen bieten für die Fremdsprachenvermittlung Einsicht in die Breite und Spezifik der funktionalen Anwendbarkeit sprachlicher Mittel.

3. Möglichkeiten zur Analyse gesprochener Lerner/-innensprache auf der Grundlage von Korpora

Mündliche L2-Korpora können auch als Grundlage für die Analyse von Lerner/-innensprache herangezogen werden. Es handelt sich hierbei jedoch um einen recht jungen Forschungszweig; dies ist nicht zuletzt auf die Datenlage zurückzuführen, denn der Bestand an mündlichen Korpora mit L2-Daten ist begrenzt – insbesondere, was öffentlich zugängliche Korpora angeht. Im Folgenden wird daher zunächst ein kurzer Einblick in die korpusbasierte Erforschung gesprochener Lerner/-innensprache gegeben. Daran anschließend soll anhand ausgewählter Beispiele auf das Potenzial der korpusbasierten Erforschung von mündlichen L2-Daten sowie auf damit verbundene methodologische Fragen eingegangen werden.

3.1 Zur korpusbasierten Erforschung gesprochener Lerner/-innensprache

Wenngleich die Anzahl an öffentlich zugänglichen Korpora, die gesprochensprachliche Lerner/-innendaten enthalten, bei weitem nicht an das Angebot mündlicher Sprachdaten von Sprecher/-innen mit Deutsch als L1 heranreicht, ist insgesamt eine Zunahme an mündlichen L2-Korpora sowie ein wachsendes Interesse an der korpusbasierten Erforschung mündlicher L2-Daten zu verzeichnen. Dies wird auch daran deutlich, dass sich ein umfangreiches Themenheft der *Zeitschrift für germanistische Linguistik* diesem Thema widmet (Wisniewski Hrsg. im E. a). Neben einer Bestandsaufnahme zu aktuellen Forschungs- und Korpusprojekten zu diesem Thema wird hier auch ein online verfügbarer Überblick über (demnächst) öffentlich zugängliche mündliche L2-Korpora für Deutsch als Fremd-

und Zweitsprache publiziert (Wisniewski im E. b). Zu den größten aktuell öffentlich zugänglichen mündlichen L2 Korpora zählen u.a.²⁵

1. das GWSS-Korpus (Gesprochene Wissenschaftssprache kontrastiv, auch GeWiss), welches studentische Vorträge und Expertenvorträge sowie Prüfungsgespräche enthält, die u.a. von Sprecherinnen mit Deutsch als L2 stammen²⁶,
2. das BeMaTaC (Berliner Map Task Corpus), welches Dialoge zu einer Map Task beinhaltet, die von erwachsenen Sprecherinnen mit Englisch als L1 produziert wurden²⁷,
3. das HMAT (Hamburg Map Task Corpus, auch HaMaTaC), enthält ebenfalls Dialoge zu einer Map Task, die von erwachsenen DaF-Lernenden mit verschiedenen L1 stammen²⁸.

Insgesamt ist die Landschaft mündlicher L2 Korpora recht heterogen und ebenso sind es die Fragestellungen, die bislang auf der Grundlage dieser Daten bearbeitet wurden. Hierzu zählen etwa Studien zu phonetisch-phonologischen Fragen (z.B. Trouvain im E.) sowie zu flüssigkeitsbezogenen Aspekten wie bspw. Pausen, Selbstreparaturen und Verzögerungsphänomenen in mündlichen L2-Produktionen (vgl. z.B. Belz 2013; Belz et al. 2017; Belz / Odebrecht im E.). Daneben sei an dieser Stelle auf die unter 2. bereits angesprochenen einzelwortbezogenen Studien verwiesen, in denen der Gebrauch von Diskursmarkern bzw. Konnektoren in mündlichen L1-Produktionen und in mündlichen L2-Produktionen verglichen wird (vgl. Slavcheva / Meißner 2014; Wallner 2016; Slavcheva 2018). Über die Einzelwortebene hinaus gehen Arbeiten zu Formelhaftigkeit in L2-Produktionen (vgl. Petkova-Kessanlis 2014; Fandrych / Wallner im E.) oder auch zur Realisierung wissenschaftssprachlicher Handlungen wie Diskurskommentierungen und Verweis und Zitat (vgl. Meißner 2016, 2017; Sadowski 2017; Fandrych / Wallner im E.). Weitere Studien zu mündlichen L2-Produktionen widmen sich Sprachwechselfhänomenen (vgl. Reershemius / Lange 2014) oder auch dem Einfluss von Aufgabenstellungen und -modalitäten (mündlich versus schriftlich) auf L2-Produktionen (vgl. Karges / Studer / Hicks im E.). Für die DaF/DaZ-Didaktik ergeben sich daraus wertvolle Einblicke in den Erwerb und die Entwicklung mündlicher Handlungsfähigkeit. Zudem geben die Studien Hinweise auf erwerbsbedingte Herausforderungen und Fehlerschwerpunkte. Anzumerken ist dabei allerdings, dass sich vorliegende korpusbasierte Arbeiten zu mündlichen L2-Produktionen aufgrund der Datenlage hauptsächlich mit sehr weit fortgeschrittenen Lerner/-innen (sogenannten „top language performers“, vgl. zu diesem Begriff Pallotti 2019) befassen²⁹.

Auch die Methoden, die in den bisherigen korpusbasierten Studien zu mündlichen L2-Produktionen genutzt werden, unterscheiden sich teilweise erheblich. Sie reichen von eher exemplarischen Analysen über quantitative Studien mit deskriptiv-statistischen Angaben bis hin zu Arbeiten, die auf fortgeschrittenere multivariate statistische Analysemethoden zurückgreifen (vgl. hierzu auch Wisniewski im E. b). Zudem steht die Methodenreflexion im Bereich der korpusbasierten Erforschung (fortgeschrittener) mündlicher L2-Produktionen noch vor größeren theoretischen und praktischen Herausforderungen (vgl. hierzu auch Fandrych / Wallner im E.). Ein wichtiger Aspekt ist dabei die Frage der Vergleichsnorm bzw. des Maßstabs, anhand dessen Lerner/-innendaten beurteilt werden können. Zwar stellt sich diese Problematik auch bei schriftsprachlichen L2-Daten, bei

²⁵ An dieser Stelle seien für einen ersten Eindruck nur drei Beispiele genannt, ausführlicher siehe Wisniewski (im E. b).

²⁶ Vgl. ausführlicher dazu die Abschnitte 1.1 und 1.2.

²⁷ Für weitere Informationen vgl. Sauer / Lüdeling (2016); der Zugang zum Korpus erfolgt über ANNIS (https://korpling.german.hu-berlin.de/annis3/#_q=9ybQ&_c=QmVNYVRhQ19MMI8zLjA&cl=5&cr=5&s=0&l=10,26.07.2021).

²⁸ Weitere Informationen finden sich bei Hedeland / Schmidt (2012), der Zugang erfolgt über die DGD. Zudem kann HaMaTaC mit ZuRecht analysiert werden.

²⁹ Eine Ausnahme bildet hier jedoch die Studie von Karges / Studer / Hicks (im E.), die mündliche L2-Daten aus dem schulischen Kontext der Klassenstufen 8 - 10, die sich auf mittleren bis höherem Leistungsniveau befinden, analysieren.

mündlichen Lerner/-innenproduktionen ist die Identifikation einer verlässlichen Orientierungsgröße jedoch noch deutlich schwieriger, da sich Konzepte wie „gesprochener Standard“ in der Gemeinsprache (vgl. hierzu bspw. Butterworth et al. 2018) oder in stärker domänenspezifischen Verwendungskontexten noch nicht etabliert haben³⁰. Abgesehen von den erwerbsbezogenen Erkenntnissen, die korpusbasierte Untersuchungen von mündlichen L2-Produktionen liefern, kann die DaF/DaZ-Didaktik hier auch in Hinblick auf die Diskussion um Normfragen und die Einschätzung bzw. auch Bewertung mündlicher Produktionen profitieren.

Im Folgenden soll anhand ausgewählter Belege aus dem GWSS-Korpus aufgezeigt werden, welche sprachlichen Phänomene in lernersprachlichen Daten untersucht werden können, welche methodischen Fragen dabei auftreten und wie solche Forschungsprojekte die DaF-Didaktik und -Forschung inspirieren können. Wir konzentrieren uns hierbei auf die im Korpus enthaltenen studentischen und Expertenvorträge von L2-Sprecher/-innen, die im Fach Germanistik / Deutsch als Fremdsprache an den Standorten Wrocław, Birmingham, Leipzig und Sofia erhoben wurden³¹.

3.2 Die Erforschung fortgeschrittener L2-Daten: Methodische Überlegungen und Potenziale

Wie bereits angesprochen, stellt sich bei lernersprachlichen Daten generell die Frage nach dem Maßstab, anhand dessen die Angemessenheit von sprachlichen Äußerungen eingeschätzt werden kann. Bei gesprochensprachlichen Genres in einem spezialisierten Domänenbereich, wie dies ein Hochschulstudium darstellt, kommt zu der bereits erwähnten Problematik der Einschätzung von gesprochensprachlichen Daten hinzu, dass wir es mit sehr spezifischen, domänen-, fach- und kontextspezifischen Angemessenheitserwartungen zu tun haben. Ob etwa ein studentisches Referat in einem Seminar im Bachelor in sprachlich-kommunikativer Hinsicht gelungen ist, hängt vom Fach und den dort geltenden Erwartungen und Konventionen ebenso ab wie von den (fachlichen oder allgemein akademischen) Diskurstraditionen, die an einem Standort gelten. Außerhalb des universitären Kontexts mag ein solches Referat auch ganz anders beurteilt werden als in einem Seminarkontext. Es kommt hinzu, dass auch die L1-Studierenden hinsichtlich wichtiger sprachlich-fachlicher Aspekte Lernende sind, denn auch sie durchlaufen einen Sozialisierungsprozess, in dessen Verlauf sie sich Kompetenzen bezüglich des wissenschaftlichen und fachbezogenen Arbeitens sowie der damit verbundenen rezeptiven und produktiven kommunikativen Anforderungen sukzessive erst aneignen³². Allerdings ist dieser Aneignungsprozess – anders als etwa bei bestimmten schriftlichen Genres (vgl. Steinhoff 2007) – für mündliche Genres noch nicht weiter untersucht worden, weshalb hierüber noch kaum Aussagen getätigt werden können. Es liegt aber auf der Hand, dass die Sprachdaten von L1-Studierenden schon aus diesem Grund nicht einfach als Vergleichs- oder Bewertungsmaßstab für L2-Sprachdaten herangezogen werden können – sie sind selbst hinsichtlich ihrer Angemessenheit erst noch einzustufen. Es kommt hinzu, dass die L2-Daten im GWSS-Korpus an unterschiedlichen Standorten und damit auch in verschiedenen akademischen Diskurstraditionen

³⁰ Es gibt aber durchaus Hinweise darauf, dass sich domänenspezifische – z.B. wissenschaftssprachliche – mündliche Formulierungsroutinen etabliert haben, die sich von den domänenbezogenen schriftsprachlichen Routinen unterscheiden; hierzu gehören etwa explizite argumentative Positionierungsroutinen des Typs *ich glaub, ich denk, ich mein*, vgl. Fandrych (2021).

³¹ Ausführlichere Informationen zu den Daten finden sich in Fandrych et al. (2014) sowie in Fandrych / Wallner (im E.). Die L2-Daten wurden an allen genannten Standorten erhoben.

³² Anders als bei den frühen Spracherwerbsstufen in L1 und L2 erscheint es uns bei domänenspezifischen Sprachkompetenzen, die auch eng mit fachlichen Kompetenzen verbunden sind, angemessener, von ‚Aneignung‘ statt von ‚Erwerb‘ zu sprechen, da die mit ‚Erwerb‘ und ‚Erwerbssequenz‘ meist verbundenen Annahmen (etwa einer kognitiv prädisponierten Abfolge von zueinander in einem Inklusionsverhältnis stehenden Erwerbsstufen) hier wohl nicht zutrifft (oder zumindest noch nicht nachgewiesen ist).

und Lehr-Lernsituationen erhoben wurden; daher sind die mit ihnen verbundenen Erwartungen an die Angemessenheit möglicherweise unterschiedlich. Dies betrifft sowohl die Makrostruktur des Genres ‚Vortrag‘, als auch die Erwartungen an einzelne funktionale Abschnitte sowie das stilistische und lexikalische ‚Design‘ von sprachlichen Handlungen und Formulierungsroutinen³³. Auch ein Abgleich mit den Vorträgen von L1- und L2-Expert/-innen (die sich ebenfalls im GWSS-Korpus finden) ist methodisch nicht unbedingt angemessen, da hierbei ja Maßstäbe an die mündliche Sprachperformanz von Studierenden angelegt würden, die diese zum einen noch nicht erfüllen können; zum anderen aber unterscheiden sich wissenschaftliche Vorträge von Expert/-innen auch kontextuell von studentischen Referaten, denn letztere erfolgen in einem geschützten hochschuldidaktischen Raum, während wissenschaftliche Vorträge bereits eine Art Veröffentlichungscharakter haben und damit viel grundlegender zum Gegenstand einer kritischen diskursiven Auseinandersetzung werden können³⁴.

Somit ist die Problematik eines Bewertungsmaßstabs für fortgeschrittene Lerner/-innendaten bei mündlichen wissenschaftlichen Genres nicht einfach zu lösen, insbesondere, wenn man auch stilistisch-rhetorische, lexikalisch-idiomatische und funktional-genrespezifische Aspekte mit in die Analyse einbezieht.

Im Folgenden wollen wir exemplarisch den Blick auf einige von L2-Sprecher/-innen des Deutschen produzierte Äußerungen aus studentischen und Expertenvorträgen im GWSS-Korpus werfen, deren Status als angemessen oder abweichend man diskutieren könnte; dabei ist es das Ziel, die mit solchen domänenbezogenen, in authentischen Kontexten erhobenen gesprochensprachlichen Korpusdaten verbundenen methodischen Probleme bei der Bewertung von Lernendenäußerungen zu reflektieren und mögliche Konsequenzen für die zukünftige Forschung zu diskutieren.

Im engeren morpho-syntaktischen und morphophonetischen Bereich finden sich schwer einzuordnende Fälle, die ohne einen Abgleich mit kontextuellen Angemessenheitserwartungen kaum endgültig zu beurteilen sind³⁵. So sind etwa Realisierungen wie *hauppproblem* [‚Hauptproblem‘], *wirklich* [‚wirklich‘], *jetz wer ich* [‚jetzt werde ich‘] oder *geprächt* [‚geprägt‘] auch in alltäglicher (teils regional geprägter) L1-Umgangssprache üblich³⁶. Da Seminarreferate gerade im deutschsprachigen Kontext häufiger auch durch deutliche Merkmale alltäglicher Mündlichkeit geprägt sind, kann zumindest nicht von vornherein auf eine Abweichung von einer genrespezifischen Angemessenheit geschlossen werden. Der Übergang zu vermutlich genuin L2-induzierten

³³ Unterschiede zwischen dem funktionalen und stilistischen Aufbau von Einleitungen in chinesischen und deutschen wissenschaftlichen Vorträgen haben etwa Günthner / Zhu (2014) herausgearbeitet.

³⁴ Des Weiteren muss natürlich auch innerhalb der recht breiten Genres „wissenschaftlicher Vortrag“ und „studentisches Referat“ nach dem je spezifischen Kontext differenziert werden (Größe der Veranstaltung, Bekanntheitsgrad unter den Teilnehmenden, institutionelle Vorgeschichte – z.B. vorherige Seminarreferate, Sektionsbeiträge etc., thematisch-fachliche Ausrichtung – z.B. Projektbericht, Grad der theoretischen / methodischen Ausrichtung etc.). Gerade in der Germanistik im nicht-deutschsprachigen Raum finden sich häufig auch Lehrveranstaltungen, bei denen fachliche und sprachliche Lernziele miteinander verbunden werden – in einem solchen Kontext haben dann Referate häufig einen weniger starken fachwissenschaftlichen Bezug; vgl. dazu auch Fandrych / Wallner (im E.).

³⁵ Zu berücksichtigen ist hier grundsätzlich, dass die durchgeführte Minimaltranskription nach GAT2 bezüglich der genauen phonetischen Realisierung nicht immer verlässlich ist. Schon die Qualität der Audioaufnahmen ist nicht immer ausreichend für eine genaue phonetische Transkription; wichtiger ist aber, dass die Transkriptionskonventionen von GAT2 zwar eine aussprachenahere Verschriftung ermöglichen sollen, aber nicht für eine phonetische Analyse konzipiert wurden. Es ist daher mit Variationen bei der Transkription von Ausspracheabweichungen zu rechnen – die Transkribend/-innen wurden entsprechend der Zielsetzungen des Projekts nicht für eine genaue phonetische Transkription geschult. Eine Nutzung des Korpus für die Analyse von Ausspracheabweichungen müsste also die zu untersuchenden Phänomene nach Möglichkeit nochmals genauer nachtranskribieren. – Die hier diskutierten Fälle wurden auf eine angemessene Transkription hin nochmals überprüft, verstehen sich aber zunächst als erste Hinweise auf einen interessanten Untersuchungsbereich.

³⁶ Bei allen Belegen wurden die Originaltranskriptionsversionen und die normalisierten Versionen beibehalten.

Aussprachevarianten ist teils fließend; hierzu gehören wohl schon ungewöhnliche Koronalisierungen des *ich*-Lauts (auch wenn sie teils regional üblich sind, vgl. *möschte* ‚möchte‘; *isch* ‚ich‘), die Realisierung von Frikativen statt Plosiven (*augenblich* ‚Augenblick‘; *einblich* ‚Einblick‘) oder Variation im Bereich von Stimmlosigkeit / Stimmhaftigkeit von Plosiven oder Frikativen (*prismatischen bilter* ‚prismatischen Bilder‘; *äusern* ‚äußern‘). Im Bereich der Wortbildung finden sich einerseits viele Unsicherheiten bezüglich der Form der Erstglieder sowie der Fugenelemente, die allerdings als recht minimale Abweichungen eingestuft werden können (*wortebücher* ‚Wörterbücher‘; *belegevarianten* ‚Belegvarianten‘; *bestandaufnahme* ‚Bestandsaufnahme‘), daneben ungewöhnliche, aber nicht unverständliche Wortbildungen (*sinn- und nutzvoll* ‚sinnvoll und nützlich‘; *problemhaft* ‚problematisch‘; *devulgarisation* ‚Devulgarisierung‘; *faszinierung* ‚Faszination‘) sowie teils Unsicherheiten bei der Univerbierung (*kurze lebigkeit* ‚Kurzlebigkeit‘; *regierungen unabhängige* ‚regierungsunabhängige‘), die sicher eher als Abweichungen empfunden werden. Angemerkt werden muss hier wiederum, dass die Transkription natürlich Ergebnis der Interpretation der Transkribend/-innen ist, so dass etwa die Entscheidung, ob *wortebücher* oder *kurze lebigkeit* als Wortbildung oder Wortgruppe transkribiert wurden, nicht unbedingt der Sprecher/-innenintention entsprechen muss und für die Transkribend/-innen nicht immer leicht zu treffen ist.

Selbstverständlich finden sich in den studentischen L2-Vorträgen auch morphosyntaktische Abweichungen, die in mündlichen L1-Produktionen nicht auftreten würden, insbesondere Genus-/Kasusabweichungen, Probleme bei der Wahl des Artikels oder der passenden Präposition, Wort-/Satzgliedstellungsprobleme sowie Abweichungen bezüglich der gewählten syntaktischen Konstruktion (fehlende oder nicht passende Realisierung von Satzgliedern etc.). Diese sind vergleichsweise einfach zu identifizieren, auch wenn solche Abweichungen in der gesprochenen Sprache keinesfalls ausschließlich auf L2-Sprecher/-innen beschränkt sind – als Phänomen der *online*-Syntax (vgl. Auer 2000) können sie auch in L1-Äußerungen auftreten, wenn auch wohl nicht so systematisch. Insbesondere im Bereich der Satzgliedstellung allerdings findet sich auch bei deutschen L1-Sprecher/-innen eine recht große, auch funktional und medial bedingt Variabilität, gerade, was Expansionen angeht (vgl. dazu ausführlicher Salzmann 2017). Dennoch sind hier Akzeptabilitätseinschätzungen insgesamt einfacher als im lexikalischen und phonetischen Bereich, der oben bereits angesprochen wurde, etwa in Fällen wie dem folgenden:

- (3) (1.4) °h äh chomsky (0.2) äh vertritt die these für eine universalgrammatik und das bedeutet **dass** jeder mensch **ist** (.) von geburt an mit einem grammatischen (.) programm °h ausgestattet (GWSS_E_00246)

Der für die fortgeschrittene domänenspezifische Sprachkompetenz besonders relevante Bereich der Formulierungs- und Gesprächsroutinen sowie der rhetorisch-textuellen Strukturierung von domänenspezifischen Genres (also im mündlichen Bereich etwa Seminarreferate, Prüfungsgespräche, Semindiskussionen) stellt die Lerner Sprachenanalyse und die Didaktik hingegen vor besondere Probleme, da hier stärker stilistisch-rhetorische, idiomatische, genre- und kontextspezifische Faktoren eine Rolle spielen. Erste Analysen (vgl. Fandrych / Wallner im E.) zeigen etwa Abweichungen im Bereich der Diskursmarker (4) und Deiktika (5): Sie treten zwar an Stellen auf, wo sie erwartbar und typisch für die jeweilige sprachliche Handlung sind, sind aber semantisch-funktional nicht immer (ganz) passend, vgl.:

- (4) ja ich weiß dass es zwei teile hier gibt °h aber °h (.) es gibt eigentlich (1.1) ja man könnte sagen dass es zwei teile gibt °h ähm (0.5) **so** die erste äh das erste tol [Teil] ist (0.6) ähm (1.1) zu englisch in deutschen (0.4) und dann de z zweite teil ist ähm (.) °h (0.5) produktkommunikots kommunikation in deutschland (GWSS_E_00211)
- (5) äh (.) zu dem (0.57) äh empirisches arbeiten in der linguistik °h äh **danach** äh komme ich zu den methoden der sprachwissenschaft (0.61) also äh (0.65) es gibt äh (.) drei methoden (.) ähm (1.0)

((schnalzt)) erstens die befragung von sprechern °hh experimente und äh die korpuslinguistik selbst (GWSS_E_00243)

Während man bei (4) den Diskursmarker *also* erwarten würde (und hier englischer Einfluss von *so* vermutet werden kann), steht bei (5) statt einer Nähedexis (z.B. *jetzt, nun*) mit *danach* ein Ausdruck, der üblicherweise auf eine weiter entfernt liegende Passage im Vortrag verweisen würde.

Auch im Bereich verbaler Formulierungsroutinen zeigen sich häufiger Abweichungen, die letztlich als idiomatisch-genrespezifisch eingestuft werden können, vgl.:

- (6) so jetzt **gehe** ich zu dem (0.4) übersicht ich hoffe das is (1.1) naja (.) °h (0.3) gut ich werd es einfach ((schnalzt)) vorlesen (1.2) (GWSS_E_00166)
- (7) sein argument wird anhand der [...]°h **beleuchtet** (GWSS_E_00215)

In (6) handelt es sich um eine typische Metakomentierung (siehe Abschnitt 2.2), bei der auch ein Verb (*gehen*) gewählt wurde, das dem für solche Überleitungen typischen Bildbereich „Bewegung im Raum“ entstammt – nur ist hier im Deutschen das telische Verb *kommen* üblich, *gehen* ist selbstverständlich gut verständlich, aber nicht idiomatisch. In (7) wird ebenfalls ein Bildbereich bemüht, der für das Feld des Erkenntnisprozesses einschlägig ist: das Feld der guten Sichtbarkeit / Wahrnehmbarkeit. Im Deutschen ist allerdings *beleuchten* in der allgemeinen Wissenschaftssprache auf Verwendungsweisen beschränkt, in denen ein spezifischer Zugang oder eine bestimmte Perspektive thematisiert wird, aus der ein Gegenstand oder eine Herangehensweise betrachtet wird (*etwas aus einer anderen Perspektive beleuchten*); im vorliegenden Fall aber handelt es sich um die Stützung eines Arguments anhand eines Beispiels, daher wäre hier *illustrieren* oder *verdeutlichen* angemessener. Hier mag auch Einfluss der L1 (Englisch) der Sprecherin vorliegen, da *illustrate* an vergleichbarer Stelle im Englischen durchaus angemessen wäre.

Als letzter Phänomenbereich sei noch der sprachliche Bezug auf Forscher/-innenrollen und wissenschaftliche Publikationen genannt. Auch hier finden sich in den studentischen L2-Korpora in GWSS Abweichungen, die evtl. mit den Standort- und Kontextbedingungen zu tun haben könnten:

- (8) unser referat stützt (0.6) ähm (0.5) sich auf (0.2) **professor doktor angelika** (0.3) **storrers** (.) äh schriften (GWSS_E_00242)
- (9) dafür ist stötzel und eitz das (.) wörterbuch der vergangenheitsbewältigung (0.3) ganz gut geeignet (0.6) ((schnalzt)) ähm da andere **werke** wie ähm stanford (0.3) äh sechsunneunzig da eine (0.8) äh einem großen vergleich ähm (0.2) ((schnalzt)) (.) von ähm (0.6) ((schnalzt)) ja germanist ähm (.) germanisten (0.3) in englisch(er) sprache (GWSS_E_00216)
- (10) wenn wir den ausdruck verwalter zur kenn (.) zeichnung der rolle der juden im deutschen kulturleben **betrachten** °h so haben **wir** diesen ausdruck **nicht etwa selbst geprägt** sondern (.) **bedienen uns** einer bezeichnung °h die ... (GWSS_E_00171)

Die Erwähnung akademischer Titel (8) bei Bezügen auf andere Wissenschaftler/-innen wäre im deutschen Kontext unüblich, *Werke* (9) wecken eher die Vorstellung gesammelter (häufig literarischer) Publikationen größeren Umfangs und sind für den Bezug auf wissenschaftliche Einzelpublikationen nicht einschlägig, und die Verwendung des Pluralis Auctoris in (10), der sich auf nur eine Sprecherin (eine Studentin im polnischen Kontext) bezieht, wäre ebenfalls in einem deutschen Wissenschaftskontext stark markiert.

Bei der Beurteilung solcher und weiterer Fälle (vgl. Fandrych / Wallner im E.) gilt jedoch, dass die jeweiligen diskursiven, fachlichen und akademischen Traditionen und Konventionen des Standorts mitberücksichtigt werden müssten. Natürlich kann dann – aus einer didaktischen Perspektive – immer noch überlegt werden, welche der Genre- und Formulierungskonventionen in einem deutschsprachigen akademischen Kontext eher (un)angemessen wären und es könnten

alternative sprachlich-kommunikative Muster vermittelt werden. Allerdings benötigen wir auch für eine solche Vermittlungsnorm zunächst eine Art *Rating* der Textqualität auch von L1-Vorträgen, aus denen man dann überhaupt erst eine Vermittlungsnorm ermitteln kann. Auch die Angemessenheit der L2-Daten lässt sich, zumindest was die stilistisch-formulierungsbezogenen und genrespezifischen Aspekte angeht, nur über Akzeptabilitätsratings ermitteln. Es kommt hinzu, dass sich die fachwissenschaftliche und die sprachlich-kommunikative Ebene hier nicht immer strikt voneinander trennen lassen – gerade in den Geisteswissenschaften ist Sprache ja selbst ganz elementar Erkenntnisinstrument; Kategorien, Begriffe und Modelle sind wesentlich sprachlich gefasst, und auch der Gegenstand der Geisteswissenschaften ist elementar sprachlich.

4. Zusammenfassung

Der vorliegende Beitrag wollte anhand dreier Ebenen und Perspektiven die Relevanz gesprochen sprachlicher Korpora für das wissenschaftliche Fach DaF/DaZ sowie für die Sprachvermittlung selbst aufzeigen. Der erste Abschnitt machte deutlich, dass für den Einsatz gesprochen sprachlicher Korpora für die praktische Sprachvermittlung noch wichtige Aufbereitungsschritte von bestehenden Korpora notwendig sind, die es erst ermöglichen, relevante gesprochen sprachliche Phänomene und auch mündliche kommunikative Ereignisse als Ganze für die Entwicklung von Lehr- und Lernmaterialien, für die zielgruppengerechte Arbeit im Unterricht sowie für die eigenständige Nutzung durch die Lernenden angemessen zu berücksichtigen. *ZuMult* ist ein erster Versuch in diese Richtung; weitere Korpuserschließungen, die sicher auch mit zusätzlichem Annotationsaufwand einhergehen werden, sind notwendig. Wichtig ist, dass solche Entwicklungen nach aller Möglichkeit plattform- und korpusübergreifend angelegt sind und somit auch für neu entstehende Korpusbestände anwendbar sind, so, wie dies bei *ZuMult* exemplarisch erarbeitet wurde. In einem zweiten Abschnitt hat der vorliegende Beitrag anhand einiger Phänomenbereiche aufgezeigt, welcher Forschungsbedarf bezüglich der gesprochenen Sprache aus Perspektive des Fachs DaF/DaZ noch besteht. Dies wurde zum einen anhand morphophonetischer Phänomene gezeigt. Diese stellen schon rein rezeptiv eine spezifische Anforderung an Lernende dar, werden aber in der herkömmlichen Sprachvermittlung nicht immer in angemessener Weise behandelt, sondern – wenn überhaupt - eher als sprechsprachliche Verkürzungen thematisiert – obwohl sie in der gesprochenen Sprache häufig der Normalfall sind, während explizite analytische Formulierungen den Sonderfall darstellen. Aber auch sprachliche Handlungen, die jedenfalls teilweise genrespezifisch sind, und ihre sprachliche Musterhaftigkeit sind noch zu wenig erforscht. Im vorliegenden Beitrag wurde anhand der Metakommentierungen gezeigt, wie man versuchen kann, mithilfe korpuslinguistischer Methoden den Form-Funktionszusammenhang herauszuarbeiten und so typische Oberflächenmerkmale zur Exploration von nicht annotierten Korpora aus dem Wissenschaftsbereich, aber auch aus anderen Domänen zu nutzen. Im dritten Teil des vorliegenden Beitrags wurde anhand von fortgeschrittenen mündlichen L2-Daten aus dem philologisch-akademischen Kontext (studentische Referate) aufgezeigt, welches Potenzial eine solche domänen- und genrespezifische Lerner Sprachenanalyse hat, welchen methodischen Herausforderungen sie sich aber auch im Kontext eines internationalen, in verschiedene Diskurstraditionen eingebetteten Wissenschaftsbetriebs gegenübergestellt sieht. Hier gilt es, die verschiedenen Faktoren, die zu Akzeptabilitätsurteilen führen können, systematisch zu reflektieren und auch das methodische Vorgehen zu verfeinern und zu erweitern, etwa durch *Ratings* und die Einbeziehung von kontextspezifischen Bewertungskriterien und Angemessenheitseinschätzungen. Ein einfacher Abgleich mit der Intuition von Forscher/-innen, mit L1-Daten von vergleichbaren Studierendengruppen aus dem deutschsprachigen Kontext, mit Expert/-innendaten oder gar mit schriftsprachlichen Daten kann jedenfalls die situative und

kontextuelle Angemessenheit der Lerner/-innendaten nicht erbringen – allerdings können so sicherlich sprachliche Profile herausgearbeitet werden, die umso besser interpretierbar sind, je mehr Metadaten wir zu den Settings der mündlichen Daten haben (wie bspw. Grad der Öffentlichkeit, Themenbezug, Fachlichkeit, institutionelle Einbettung, Relevanz des kommunikativen Ereignisses im Studienverlauf etc.).

Literatur und Ressourcen

Alsop, Sian / Nesi, Hilary (2014): The pragmatic annotation of a corpus of academic lectures. In: Calzolari, Nicoletta et al. (Hrsg.): *Proceedings of LREC 2014. Ninth International Conference on Language Resources and Evaluation. Reykjavik: European Language Resources Association (ELRA)*, 1560-1563.

Auer, Peter (2000): On line-Syntax – Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen. In: *Sprache und Literatur in Wissenschaft und Unterricht* 85, 43-56.

Augustin, Hagen (2018): *Verschmelzung von Präposition und Artikel: Eine kontrastive Analyse zum Deutschen und Italienischen*. Berlin: de Gruyter.

Baur, Benedikt et al. (2014): Dokumentation zur Annotation der Diskurskommentierungen. https://gewiss.uni-leipzig.de/fileadmin/documents/Annotationsdokumentation_GeWiss.pdf (15.07.2021)

Beißwenger, Michael / Storrer, Angelika (2011): Digitale Sprachressourcen in Lehramtsstudiengängen: Kompetenzen – Erfahrungen – Desiderate. In: *Journal for Language Technology and Computational Linguistics* 26: 1, 119-139.

Belz, Malte (2013): *Disfluencies und Reparaturen bei Muttersprachlern und Lernern - eine kontrastive Analyse*. Berlin: Humboldt-Universität zu Berlin.

Belz, Malte et al. (2017): Fluently disfluent? Pauses and Repairs of Advanced Learners and Native Speakers of German. In: *International Journal of Learner Corpus Research* 3: 2, 118-148.

Belz, Malte / Odebrecht, Carolin (im E.): Abschnittsweise Analyse sprachlicher Flüssigkeit in der Lernersprache: Das Ganze ist weniger informativ als seine Teile. In: *Zeitschrift für germanistische Linguistik*. Themenheft: Gesprochene Lernerkorpora des Deutschen.

Bies, Andrea (2020): WG-Castings im DaF-Unterricht. In: *Deutsch als Fremdsprache* 57: 2, 88-101.

Butterworth, Judith / Hahn, Nadine / Schneider, Jan Georg (2018): Gesprochener Standard, da gibt es viel zu sagen. In: Albert, Georg / Diao-Klaeger, Sabine (Hrsg.): *Mündlicher Sprachgebrauch zwischen Normorientierung und pragmatischen Spielräumen*. Tübingen: Stauffenburg (= Stauffenburg Linguistik 101), 3-24.

Callies, Marcus (2019): Integrating corpus literacy into language teacher education. In: Götz, Sandra / Mukherjee, Joybrato (Hrsg.): *Learner corpora and language teaching*. Amsterdam: Benjamins, 245-263.

Cohen, Jacob (1988): *Statistical Power Analysis for the Behavioral Sciences*. New York.

Deppermann, Arnulf / Hartung, Martin (2012): Was gehört in ein nationales Gesprächskorpus? Kriterien, Probleme und Prioritäten der Stratifikation des 'Forschungs- und Lehrkorpus Gesprochenes Deutsch' (FOLK) am Institut für Deutsche Sprache (Mannheim). In: Felder, Ekkehard / Müller, Marcus / Vogel, Friedemann (Hrsg.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin: de Gruyter, 414-450.

Dietz, Gunther (2017): Mentale Prozesse beim mutter- und fremdsprachlichen Hören und Konsequenzen für die Hörverstehensdidaktik. In: Di Venanzio, Laura / Lammers, Ina / Roll, Heike (Hrsg.): *DaZu und DaFür -*

neue Perspektiven für das Fach Deutsch als Zweit- und Fremdsprache zwischen Flüchtlingsintegration und weltweitem Bedarf: 43. Jahrestagung des Fachverbandes Deutsch als Fremd- und Zweitsprache an der Universität Duisburg-Essen 2016. Göttingen: Universitätsverlag Göttingen, 97-116.

Dietz, Gunther (2021a): Fremdsprachliches Hörverstehen: Schwächen der traditionellen Hörverstehensdidaktik – Perspektiven der Vermittlung für Deutsch als Fremdsprache. In: *Deutsch als Fremdsprache* 58: 2, 67-75.

Dietz, Gunther (2021b): Authentische Hörmaterialien im DaZ-/DaF-Unterricht – Korpora gesprochener Sprache als Fundgrube für die Erstellung von Mikro-Höraufgaben. In: *Korpora Deutsch als Fremdsprache* 1: 1, 97-123.

Duden Grammatik (2009) = Eisenberg, Peter / Kunkel-Razum, Kathrin / Münzberg, Franziska (2009): *Duden - Die Grammatik: Unentbehrlich für richtiges Deutsch*; 8., überarb. Aufl. Mannheim: Dudenverl.

Ellis, Nick (2017): Cognition, Corpora, and Computing: Triangulating Research in Usage-Based Language Learning. In: *Language Learning* 67: 1, 40-65.

Ender, Andrea / Kaiser, Irmtraud (2020): Fressen oder gefressen werden? Rezeptive bildungssprachliche Kompetenzen bei ein- und mehrsprachigen Jugendlichen der Sekundarstufe. In: Langlotz, Miriam (Hrsg.): *Grammatikdidaktik – theoretische und empirische Zugänge zu sprachlicher Heterogenität*. Baltmannsweiler: Schneider Verlag Hohengehren, 117-144.

Fandrych, Christian (2014): Metakomentierungen in wissenschaftlichen Vorträgen. In: Fandrych, Christian / Meißner, Cordula / Slavcheva, Adriana (Hrsg.): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron, 95-111.

Fandrych, Christian (2021): Ich denke die Indizien sind eindeutig ...: Positionierungshandlungen als spezifisch mündliche Phänomene in wissenschaftlichen Vorträgen. In: Günthner, Susanne / Schopf, Juliane / Weidner, Beate (Hrsg.): *Gesprochene Sprache in der kommunikativen Praxis. Analysen authentischer Alltagssprache und ihr Einsatz im DaF-Unterricht*. Tübingen: Stauffenburg, 219-245.

Fandrych, Christian et al. (2016): User, who art thou? User Profiling for Oral Corpus Platforms. In: Nicoletta Calzolari et al. (Hrsg.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia*. Paris: European Language Resources Association (ELRA), 280-287.

Fandrych, Christian et al. (im E.): ZuMult: Neue Zugangswege zu Korpora gesprochener Sprache. In: Kämper, Heidrun et al. (Hrsg.): *Sprache in Politik und Gesellschaft: Perspektiven und Zugänge*. Jahrbuch des Instituts für Deutsche Sprache 2021. Berlin etc.: de Gruyter.

Fandrych, Christian / Meißner, Cordula / Slavcheva, Adriana (Hrsg.) (2014): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron-Verlag.

Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (Hrsg.) (2017): *Gesprochene Wissenschaftssprache – digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg.

Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (2018): Das Potenzial mündlicher Korpora für die Sprachdidaktik: Das Beispiel GeWiss. In: *Deutsch als Fremdsprache* 55: 1, 3-13.

Fandrych, Christian / Schwendemann, Matthias / Wallner, Franziska (2021): „Ich brauch da dringend ein passendes Beispiel ...“: Sprachdidaktisch orientierte Zugriffsmöglichkeiten auf Korpora der gesprochenen Sprache aus dem Projekt ZuMult. In: *Informationen Deutsch als Fremdsprache* 50: 6, 711-729.

Fandrych, Christian / Thurmair, Maria (2021): *Grammatik im Fach Deutsch als Fremd- und Zweitsprache. Grundlagen und Vermittlung*. Berlin: Erich Schmidt Verlag (2. Auflage).

- Fandrych, Christian / Wallner, Franziska (im E.): Funktionale und stilistische Merkmale gesprochener fortgeschrittener Lerner:innensprache: Methodische und konzeptionelle Überlegungen am Beispiel von GeWiss. In: *Zeitschrift für germanistische Linguistik*. Themenheft: Gesprochene Lernerkorpora des Deutschen.
- Felder, Ekkehard / Müller, Marcus / Vogel, Friedemann (2012): Korpuspragmatik. Paradigma zwischen Handlung, Gesellschaft und Kognition. In: Felder, Ekkehard / Müller, Marcus / Vogel, Friedemann (Hrsg.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analyse*. Berlin: de Gruyter, 3-32.
- Glaboniat, Manuela et al. (2005): *Profile Deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen, Kannbeschreibungen, kommunikative Mittel, Niveau A1-A2, B1-B2, C1-C2*. Berlin: Klett.
- Günthner, Susanne / Wegner, Lars / Weidner, Beate (2013): Gesprochene Sprache im DaF-Unterricht - Möglichkeit der Vernetzung der Gesprochene-Sprache-Forschung mit der Fremdsprachenvermittlung. In: Moraldo, Sandro / Missaglia, Federica (Hrsg.): *Gesprochene Sprache im DaF-Unterricht. Grundlagen – Ansätze – Praxis*. Heidelberg: Winter, 113-150.
- Günthner, Susanne / Zhu, Qiang (2014): Wissenschaftsgattungen im Kulturvergleich - Analysen von Eröffnungssequenzen chinesischer und deutscher Konferenzvorträge. In: Meier, Simon / Rellstab, Daniel / Schiewer, Gesine (Hrsg.): *Dialog und (Inter-)Kulturalität*. Tübingen: Narr, 175-196.
- Gutzmann, Daniel / Turgay, Katharina (2016): Zur Stellung von Modalpartikeln in der gesprochenen Sprache. In: *Deutsche Sprache* 44: 2, 97-122.
- Hedeland, Hanna / Schmidt, Thomas (2012): Technological and methodological challenges in creating, annotating and sharing a learner corpus of spoken German. In: Schmidt, Thomas / Wörner, Kai (Hrsg.): *Multilingual Corpora and Multilingual Corpus Analysis*. (= Hamburg Studies on Multilingualism 14). Amsterdam: Benjamins, 2012. 25-46.
- Imo, Wolfgang / Weidner, Beate (2018): Mündliche Korpora im DaF- und DaZ-Unterricht. In: Schmidt, Thomas / Kupietz, Marc (Hrsg.): *Korpora in der Linguistik*. Berlin: de Gruyter, 231-251.
- Kaiser, Julia (2018): Zur Stratifikation des FOLK-Korpus: Konzeption und Strategien. In: *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 19, 515-552. <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2018/px-kaiser.pdf> (15.07.2021).
- Kaiser, Julia / Schedl, Evi (2021): Das Forschungs- und Lehrkorpus Gesprochenes Deutsch als Ressource für den handlungsorientierten DaF-Unterricht – Potentiale und Herausforderungen. In: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 26: 1, 45-83. <http://tujournals.ulb.tu-darmstadt.de/index.php/zif> (15.07.2021).
- Karges, Katharina / Studer, Thomas / Hicks, Nina Selina (im E.): Lerner Sprache, Aufgabe und Modalität: Beobachtungen zu Texten aus dem Schweizer Lernerkorpus SWIKO. In: *Zeitschrift für germanistische Linguistik*. Themenheft: Gesprochene Lernerkorpora des Deutschen.
- Lanwer, Jens Philipp (2015): Aber da ham wi son bärenstarken Pernot gekriegt. Zur didaktischen Relevanz des Unterschiedes zwischen Aus- und Alltagssprache. In: Imo, Wolfgang / Moraldo, Sandro M. (Hrsg.): *Interaktionale Sprache und ihre Didaktisierung im DaF-Unterricht*. Tübingen: Stauffenburg, 83-112.
- Lanwer, Jens Philipp / Schopf, Juliane (2021): „Hölzernes Deutsch“ im DaF-Unterricht. Zum Zusammenhang von phonologischer und konversationeller Kompetenz. In: *Deutsch als Fremdsprache* 58: 1, 16-24.
- Laver, John (1994): *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Lenort, Lisa / Pohle, Anna / Sakhno, Anna (im E.): Zur Verschmelzung von Präposition und bestimmtem Artikel im gesprochenen Deutsch. In: *Deutsch als Fremdsprache* 59: 1.

- Meißner, Cordula (2016): Die Realisierung mündlicher wissenschaftssprachlicher Handlungen im Deutschen als L1 und L2: Eine gebrauchsbasierte Analyse. In: Kontutyte, Eglè / Žeimantienė, Vaiva (Hrsg.): *Sprache in der Wissenschaft. Germanistische Einblicke*. Frankfurt a.M.: Peter Lang (Duisburger Arbeiten zur Sprach- und Kulturwissenschaft, 111), 175-187.
- Meißner, Cordula (2017): Gute Kandidaten. Ein Ansatz zur automatischen Ermittlung von Belegen für sprachliche Handlungen auf der Basis manueller pragmatischer Annotation. In: Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (Hrsg.): *Gesprochene Wissenschaftssprache – digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg, 165-213.
- Meißner, Cordula (im E.): Indikatormerkmale in metakomentierenden Sprechhandlungen thematisch strukturierter Interaktionen: Eine korpuspragmatische Untersuchung zur Beziehung zwischen Funktion und Form. Erscheint in: Bülow, Lars et al. (Hrsg.): *Digitale Pragmatik*. Stuttgart: Metzler (= Digital Linguistics).
- Meißner, Cordula / Wallner, Franziska (im E.): Korpora gesprochener Sprache als virtuelle Lernräume der Mündlichkeitsdidaktik: Affordanzen eines außerunterrichtlichen Sprachlernsettings. Erscheint in: Feick, Diana / Rymarczyk, Jutta (Hrsg.): *Fremdsprachenunterricht im virtuellen Raum*. Peter Lang.
- Meliss, Meike (2021): Die LeGeDe-Ressource: korpusbasierte lexikografische Einblicke und anwendungsorientierte Ausblicke. In: *Deutsch als Fremdsprache* 58: 1, 3-15.
- Moraldo, Sandro / Missaglia, Federica (2013) (Hrsg.): *Gesprochene Sprache im DaF-Unterricht. Grundlagen – Ansätze – Praxis*. Heidelberg: Winter.
- Pallotti, Gabriele (2019): An Approach to Assessing the Linguistic Difficulty of Tasks. In: *Journal of the European Second Language Association*, 3:1, 58-70. <https://doi.org/10.22599/jesla.61> (15.07.2021).
- Petkova-Kessanlis, Mikaela (2014): Grade sprachlicher Formelhaftigkeit bei der Realisierung der Textsorte ‚Studentisches Referat‘ in der Fremdsprache Deutsch. In: Fandrych, Christian / Meißner, Cordula / Slavcheva, Adriana (Hrsg.): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron, 177-192.
- Reershemius, Gertrud / Lange, Daisy (2014): Sprachkontakt in der mündlichen Wissenschaftskommunikation. In: Fandrych, Christian / Meißner, Cordula / Slavcheva, Adriana (Hrsg.): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron, 57-74.
- Römer, Ute (2011): Corpus research applications in second language teaching. In: *Annual Review of Applied Linguistics* 31, 205-225.
- Rösler, Dietmar (2016): Nähe und Distanz zur Mündlichkeit in der fremdsprachendidaktischen Diskussion. Versuch einer Annäherung. In: *Deutsch als Fremdsprache* 53: 3, 135-149.
- Rühlemann, Christoph (2010): What can the corpus tell us about pragmatics? In: O'Keeffe, Anne / McCarthy, Michael (Hrsg.): *The Routledge handbook of corpus linguistics*. London: Routledge, 288-301.
- Salzmann, Katharina (2017): *Expansionen in der deutschen und italienischen Wissenschaftssprache. Kontrastive Korpusanalyse und sprachdidaktische Überlegungen*. Berlin: Erich Schmidt Verlag.
- Sadowski, Sabrina (2017): Die Annotation von Zitaten und Verweisen im GeWiss-Korpus. In: Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (Hrsg.): *Gesprochene Wissenschaftssprache – digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg, 143-163.
- Sauer, Simon / Lüdeling, Anke (2016): Flexible Multi-Layer Spoken Dialogue Corpora. In: *International Journal of Corpus Linguistics* 21: 3, Special Issue on Spoken Corpora, 419-438.
- Scharloth, Joachim / Bubenhofer, Noah (2012): Datengeleitete Korpuspragmatik. Korpusvergleiche als Mittel der Stilanalyse. In: Felder, Ekkehard / Müller, Marcus / Vogel, Friedemann (Hrsg.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analyse*. Berlin: de Gruyter, 195-230.

- Schmidt, Thomas (2014): Gesprächskorpora und Gesprächsdatenbanken am Beispiel von FOLK und DGD. In: *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 15, 196-233. <http://www.gespraechsforschung-ozs.de/fileadmin/dateien/heft2014/px-schmidt.pdf> (15.07.2021).
- Schmidt, Thomas (2018): Gesprächskorpora. In: Kupietz, Marc / Schmidt, Thomas (Hrsg.) (2018): *Korpuslinguistik*. (=Germanistische Sprachwissenschaft um 2020, Bd. 5). Berlin / Boston: de Gruyter, 209-230.
- Schneider, Jan Georg / Judith Butterworth / Nadine Hahn (2018): *Gesprochener Standard in syntaktischer Perspektive. Theoretische Grundlagen – Empirie – didaktische Konsequenzen*. Tübingen: Stauffenburg (= Stauffenburg Linguistik 99) [Projekt-Publikation zum DFG-Projekt „Gesprochener Standard“].
- Schneider, Jan Georg (2020): Sprechen die meisten Deutschen grammatisch nicht korrekt? Das DFG-Projekt „Gesprochener Standard“ und seine Bedeutung für den DaF-Unterricht. In: *Deutsch als Fremdsprache* 57: 4, 206-218.
- Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion* 10, 353-402. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (15.07.2021).
- Slavcheva, Adriana (2018): Zum Konnektorengebrauch in der gesprochenen Wissenschaftssprache Deutsch durch fortgeschrittene Lerner/innen. In: *Moderna språk* 112: 1, 84-105, <http://ojs.ub.gu.se/ojs/index.php/modernasprak/article/view/4347/3509> (15.07.2021).
- Slavcheva, Adriana / Cordula Meißner (2014): Also und so in wissenschaftlichen Vorträgen. In: Fandrych, Christian / Meißner, Cordula / Slavcheva, Adriana (Hrsg.): *Gesprochene Wissenschaftssprache: Korpusmethodische Fragen und empirische Analysen*. Heidelberg: Synchron, 113-132.
- Steinhoff, Torsten (2007): *Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Tübingen: Niemeyer
- Trouvain, Jürgen (im E.): Das IFCASL-Korpus als phonetisches Lernerkorpus. In: *Zeitschrift für germanistische Linguistik*. Themenheft: Gesprochene Lernerkorpora des Deutschen.
- Wallner, Franziska (2016): Diskursmarker in wissenschaftlichen Vorträgen. In: Kontutyté, Eglè / Žeimantienė, Vaiva (Hrsg.): *Sprache in der Wissenschaft. Germanistische Einblicke* [= Duisburger Arbeiten zur Sprach- und Kulturwissenschaft 111]. Frankfurt a. M.: Peter Lang, 189-204.
- Wallner, Franziska (2017): Diskursmarker funktional. Eine quantitativ-qualitative Beschreibung annotierter Diskursmarker im GeWiss-Korpus. In: Fandrych, Christian / Meißner, Cordula / Wallner, Franziska (Hrsg.): *Gesprochene Wissenschaftssprache – digital. Verfahren zur Annotation und Analyse mündlicher Korpora*. Tübingen: Stauffenburg, 107-122.
- Westpfahl, Swantje / Schmidt, Thomas (2016): FOLK-Gold – A GOLD standard for Part-of-Speech-Tagging of Spoken German. In: Nicoletta Calzolari et al. (Hrsg.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia*. Paris: European Language Resources Association (ELRA), 1493-1499.
- Westpfahl, Swantje et al. (2017): *STTS 2.0. Guidelines für die Annotation von POS-Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS)*. Arbeitspapier. Mannheim: Institut für Deutsche Sprache.
- Wisniewski, Katrin (Hrsg.) (im E. a): Gesprochene Lernerkorpora. In: *Zeitschrift für germanistische Linguistik*. Themenheft der Zeitschrift für germanistische Linguistik.
- Wisniewski, Katrin (im E. b): Gesprochene Lernerkorpora des Deutschen: Ein Überblick. In: *Zeitschrift für germanistische Linguistik*. Themenheft: Gesprochene Lernerkorpora des Deutschen.

Wray, Alison (2002): *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Kurzbios:

Christian Fandrych ist Professor für Deutsch als Fremdsprache mit Schwerpunkt Linguistik am Herder-Institut, Universität Leipzig. Er arbeitet u.a. zu den Themen Lexik- und Grammatikvermittlung, Korpuslinguistik, Text- und Gesprächslinguistik, Wissenschaftssprache und Sprachenpolitik. Er ist Chefredakteur der Zeitschrift „Deutsch als Fremdsprache“ und Mitherausgeber der Zeitschrift „Fremdsprache Deutsch“.

Anschrift:

Christian Fandrych
Herder-Institut
Universität Leipzig
Beethovenstr. 15
04107 Leipzig
fandrych@uni-leipzig.de

Cordula Meißner ist Assistenzprofessorin am Institut für Germanistik an der Universität Innsbruck. Ihre Forschungsschwerpunkte umfassen u.a. die korpusbasierte Sprachbeschreibung als Basis für das Lehren und Lernen von Deutsch als fremder Sprache, Korpuspragmatik und Lexikogrammatik geschriebener und gesprochener Sprache sowie die Sprachverwendung in Wissenschaft, Bildung und Beruf.

Anschrift:

Cordula Meißner
Institut für Germanistik
Universität Innsbruck
Innrain 52d
A-6020 Innsbruck
Österreich
cordula.meissner@uibk.ac.at

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

Anschrift:

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

