

Sandra Hansen/Thilo Weber

Diskurs, Grammatik, Korpus: Auf dem Weg zu einer Korpusgrammatik des Deutschen

1 Einleitung

Die Diskurslinguistik lässt sich als „Ansatz zur Beschreibung des Sprachgebrauchs unter seinen je spezifischen sozialen, medialen und institutionellen Bedingungen“ charakterisieren (Ziem 2018, 112). Konstitutiv für die Korpuslinguistik ist die Arbeit mit der „Sammlung authentischer sprachlicher Äußerungen, die in der Regel nicht gesteuert vom Korpuslinguisten entstanden sind“ (Konopka 2018, 152). Daraus ergeben sich grundlegende Konvergenzen zwischen Diskurslinguistik und Korpuslinguistik und folglich zwischen Diskursgrammatik und Korpusgrammatik. Der vorliegende Beitrag setzt sich zum Ziel, diese Konvergenzen anhand ausgewählter Studien aus dem am Leibniz-Institut für Deutsche Sprache in Mannheim (IDS) beheimateten Projekt *Korpusgrammatik* und seinen Anschlussprojekten zu veranschaulichen. Darüber hinaus diskutieren wir die Frage, wie sich die Korpuslinguistik in den Metadiskurs **über** Grammatik einbringen kann und welchen Beitrag zum Forschungsdiskurs speziell das IDS-Projekt zu leisten anstrebt.

2 Die Korpuslinguistik als Grundlage für eine Grammatik

In der Korpuslinguistik werden aufgezeichnete Resultate von Kommunikationsprozessen als empirische Grundlage genutzt. Die Datenquelle ist demnach eine Sammlung authentischer Sprachereignisse, die in Form von Texten bzw. Textauschnitten vorliegen. Die Korpuslinguistik ist also nach ihrer Datenquelle benannt worden. Was der Begriff Korpuslinguistik genau umfasst, ist nach wie vor nicht eindeutig definiert. Konopka (2018, 152) zufolge kann unter der Bezeichnung letztendlich jeder Arbeitsschritt gefasst werden, „der auf der Strecke von der Idee eines Korpus bis zur linguistischen Erkenntnis liegt“. Hierzu zählen Aufbau und Bereitstellung von Korpora inklusive linguistischer Annotationen, die Erarbeitung von Methoden und Werkzeugen, um die Daten untersuchbar zu machen, und die Analyse konkreter linguistischer Fragestellungen auf Basis dieser Korpora.

Aus unserer Sicht ist das wissenschaftliche Programm der Korpuslinguistik eine sehr geeignete Grundlage für eine Grammatik, insbesondere dann, wenn

dabei Variationsfälle im Fokus stehen, die nicht nur deskriptiv erfasst werden, sondern deren grammatikinterne und -externe Einflussfaktoren ermittelt und erklärt werden sollen. Die Korpuslinguistik geht vom Gebrauch natürlicher Sprachen aus und ist somit eine empirische Methode zum Gewinn von Wissen über Sprache. Für Linguist*innen bieten Korpora die Möglichkeit, gezielt nach bestimmten linguistischen Phänomenen zu suchen. Damit ist die Möglichkeit geschaffen, nicht mehr auf Grund des eigenen Sprachgefühls bzw. der Introspektion, sondern auf Basis der im Korpus befindlichen Daten zu linguistischen Schlussfolgerungen zu kommen. Es gibt mittlerweile zahlreiche Arbeiten, die zeigen, wie man auf Basis von (annotierten oder nicht annotierten) Korpora gezielt nach grammatischen Phänomenen suchen kann (vgl. z. B. Meurers/Müller 2009). In Konopka (2018) wird darüber hinaus besprochen, was die Korpuslinguistik für die Grammatiktheorie leisten kann und was nicht.¹ Konopka skizziert drei Einsatzbereiche, für die die Korpuslinguistik eingesetzt werden kann: die Exemplifizierung theoretischer Feststellungen (Evidenz aus dem Korpus), die Überprüfung theoretischer Feststellungen (Validierung am Korpus) und die Verwendung eines Korpus als Quelle für zu beschreibende und erklärungsbedürftige Phänomene (Induktion mithilfe des Korpus). Die Korpuslinguistik kann demnach zur Theoriebildung beitragen, auch wenn sie kein bestimmtes theoretisches Paradigma verfolgt.

3 Konvergenzen von Diskursgrammatik und Korpusgrammatik

Versteht man den Begriff ‚Diskurs‘ gemäß der weiten Definition von ‚Sprachgebrauch‘ (Fasold 1990; Müller 2018, 77), dann ergeben sich grundlegende Konvergenzen von Diskurslinguistik und Korpuslinguistik und somit eben auch von Diskursgrammatik und Korpusgrammatik: Eine Grundannahme der Diskursgrammatik ist, dass grammatische Strukturen aus dem Sprachgebrauch heraus entstehen. So schließen sich z. B. Müller (2018, 76) und Ziem (2018, 114) der berühmten Feststellung Haspelmaths (2002, 284) an, wonach Grammatik „geronnener Diskurs“ sei. Grammatik, so Müller (2018, 76), „muss etwas sein, das in der Gebrauchsgeschichte sprachlicher Zeichen entsteht und somit als Epiphänomen semiotisch reifizierter

¹ Z.B. hat die Korpuslinguistik ihre Schwächen, wenn es um Fragen wie Akzeptabilität von Äußerungen im Allgemeinen und negative Evidenz im Besonderen oder auch um niedrigfrequente Phänomene geht (vgl. Konopka 2018, 153).

Akte des Meinens und Verstehens anzusehen ist.“ Ähnlich bezeichnet Ziem (2018, 114) Grammatik als „Epiphänomen kommunikativer Interaktion“. Aus dieser Grundannahme lässt sich ableiten, dass auch die Untersuchung grammatischer Regularitäten anhand authentischer Sprachgebrauchsdaten erfolgen sollte. In der Korpuslinguistik geschieht genau das (vgl. Abschnitt 2).

Die Diskursgrammatik geht ferner davon aus, dass sich Grammatik wandelt und dass sie über „thematische, situative, soziale und mediale Kontexte“ hinweg variiert (Müller 2018, 76). Damit einhergehend sieht Ziem (2018, 111) „eine genuin diskurslinguistische Aufgabe darin, den möglichen Variationsbereich grammatischer sowie lexikalisch-semantischer Muster diskursvergleichend zu eruieren und auszuweisen.“ Die Korpuslinguistik ermöglicht genau das, indem sie mit stratifizierten Korpora arbeiten kann, d. h. mit Sprachdaten, die sich hinsichtlich ihres Alters und ihrer thematischen, situativen, sozialen und medialen Entstehungs- und Gebrauchsbedingungen unterscheiden. Aus konstruktionsgrammatischer Perspektive stellt Ziem (2018, 111) ebenfalls fest, dass die Erfassung der o. g. Muster „bislang vorrangig an der Schnittstelle von Korpus- und Diskurslinguistik gelungen ist“.

Charakteristisch für die Diskursanalyse (und damit die Diskursgrammatik) ist aber auch ein spezifisches Interesse an gesellschaftlichen Debatten: „[S]ie reagiert auf mediale In-Wert-Setzungen von Themen und bringt sich in aktuelle Diskussionen (mehr oder weniger) deutlich ein“ (Warnke 2018, XXVIII). Für die Korpuslinguistik oder auch speziell die Korpusgrammatik ist dies kein konstitutives Merkmal. Viele bekannte korpusgrammatische Studien (wie z. B. Bresnan et al. 2007 zur Dativalternation im Englischen, vgl. *Susan gave toys to the children* vs. *Susan gave the children toys*) lassen keinen Ausgangspunkt in öffentlichen Debatten erkennen. Umgekehrt scheint aber zu gelten: **Wenn** eine Arbeit einen solchen Debattenbezug sucht, dann arbeitet sie oft korpuslinguistisch. Ein Beispiel dafür ist die Untersuchung von Ziem et al. (2013) zur Konzeptualisierung von *Krise*: Dort wird die These vertreten, „dass sich in den Krisen-Diskursen der Bundesrepublik eine diskursive Dynamik offenbart, die an variierenden semantischen Prägungen sprachlicher Muster erkennbar ist“, und dass diese Muster „sich korpusgesteuert identifizieren lassen“ (2013, 331). Die Korpuslinguistik ist also zwar nicht grundsätzlich an Fragestellungen mit aktuellem Debattenbezug interessiert; sie stellt aber Daten und Methoden zur Verfügung, die für entsprechende Fragestellungen genutzt werden können.

Zuletzt kann die Korpuslinguistik auch wertvolle Beiträge zum Metadiskurs **über** Grammatik leisten. Und zwar in mindestens zweierlei Hinsicht: Zum einen kann sie sich in den Diskurs zur grammatischen Norm einbringen (vgl. Abschnitt 7.1). Zum anderen versteht sich beispielsweise das IDS-Projekt *Korpusgrammatik* auch als Beitrag zur Erweiterung des wissenschaftlichen Diskurses

über Grammatik (vgl. Abschnitt 7.2). So werden nicht nur wissenschaftliche Texte veröffentlicht, sondern auch die ihnen zugrundeliegenden Sprachdaten in Form von annotierten Datensammlungen. Damit leistet das Projekt einen Beitrag zur offenen Wissenschaft im Sinne der Open-Science-Bewegung.

4 Das IDS-Projekt *Korpusgrammatik*

Das Projekt *Korpusgrammatik*² untersucht systematisch und korpusorientiert morphologische und syntaktische Variationsphänomene des Deutschen.³ Gegenstand ist die Grammatik des geschriebenen Deutsch, wie sie sich in Texttypen manifestiert, die typischerweise in der deutschen Standardsprache verfasst werden, wobei selbstverständlich davon ausgegangen wird, dass die Standardsprache Variation zulässt (vgl. Bubenhofer et al. 2014, 23–32). Die Korpusgrammatik als grammatikographisches Werk strebt eine empirisch fundierte Deskription an und stützt sich dabei größtenteils auf ein speziell aufbereitetes Teilkorpus des Deutschen Referenzkorpus (DeReKo, vgl. Kupietz et al. 2018).⁴ Dieses Teilkorpus umfasst zurzeit 23 Mio. Dokumente, 5,8 Mrd. Wörter sowie 7 Mrd. Tokens. Es enthält vor allem klassische Druckerzeugnisse und wurde mit Metadaten wie *Land*, *Region*, *Register*, *Medium*, (*inhaltliche*) *Domäne* und *Entstehungszeit* angereichert und stratifiziert (Bubenhofer et al. 2014, 54–76) – Daten also, die insbesondere auch für diskurslinguistische Fragestellungen von großem Interesse sind.⁵

Da die Abgrenzung der Standardsprache problematisch ist (vgl. Bubenhofer et al. 2014, 21–32), werden ergänzend Übergänge zum Substandard erfasst, indem Teilkorpora des Web-Korpus DECOW16 (Schäfer/Bildhauer 2012; Schäfer 2015) herangezogen werden, die substandardsprachliche Phänomenrealisierungen an der Grenze zum Standard erwarten lassen. Ergebnisse der Korpusgrammatik werden u. a. in einer neuen, „variationssensitiven“ Grammatik des Standarddeutschen dokumentiert. Mit modernen korpuslinguistischen Methoden wird dabei die Vielfalt und Variabilität der deutschen Grammatik in großer Detailschärfe erfasst (vgl. Konopka 2020, 7). Die Korpusgrammatik soll – im Vergleich zu vorherigen Grammatiken des Deutschen – ein höheres Maß an Deskriptionstiefe erreichen, einen Beitrag zum aktuellen Forschungsdiskurs leisten und somit das

² <https://www.ids-mannheim.de/gra/projekte/korpusgrammatik/> (Letzter Zugriff: 27.03.2023).

³ Zur Auswahl der Phänomene vgl. Konopka (2020, 13).

⁴ <https://www.ids-mannheim.de/digspra/kl/projekte/korpora> (Letzter Zugriff: 27.03.2023).

⁵ Genauere Informationen zu diesem Untersuchungskorpus stehen unter <https://grammis.ids-mannheim.de/korpusgrammatik/6616> (Letzter Zugriff: 28.03.2023) zur Verfügung.

Desiderat einer Grammatik des Deutschen auf korpuslinguistischer Basis erfüllen (vgl. Konopka 2020, 10–11). Der Fokus liegt auf der Variabilität der deutschen Standardsprache und ihrem tatsächlichen Gebrauch. Die Beschreibung von Variationsphänomenen ist dabei zentrales Thema. Sie werden mithilfe korpuslinguistisch-statistischer Verfahren im Hinblick auf ihre Abhängigkeit von verschiedenen sprachinternen und/oder -externen Einflussfaktoren erklärt (vgl. Konopka 2018, 166). Die Einflüsse möglicher sprachinterner Faktoren werden durch die Annotation grammatischer Merkmale untersucht. Die Erfassung externer Faktoren geschieht durch die Einbeziehung der oben beschriebenen Metadaten (vgl. Konopka 2020, 13). Neben den statistischen Analysen werden umfangreiche und vielschichtige qualitative Untersuchungen der Daten vorgenommen, die zur Exploration für quantitative Analysen dienen (vgl. Konopka 2020, 16).

Die Korpusgrammatik strebt zwar theoretische Fundiertheit an, will aber im Großen und Ganzen theorieneutral sein (vgl. Konopka 2020, 13–14). Ziele sind u. a., theorierelevante Daten zu generieren und Brücken zwischen konkurrierenden theoretischen Perspektiven zu schlagen. Die Datenerhebung, Datenaufbereitung und die statistischen Analysen werden dokumentiert und transparent als Open-Access-Publikationen der Leserschaft zur Verfügung gestellt, sodass die Validierbarkeit der zugrundeliegenden Untersuchungen gesichert ist und direkte Anschlussforschung ermöglicht werden kann (vgl. Konopka 2020, 11–12). Die primär anvisierte Zielgruppe der Korpusgrammatik sind Sprachwissenschaftler*innen.

In der Korpusgrammatik soll es verschiedene Kapiteltypen geben, die die Themenbereiche in der Abfolge vom Allgemeinen zum Besonderen behandeln. Zum einen werden Überblickskapitel unter Verwendung deskriptiv-statistischer Methoden größere Phänomenfelder abdecken und Variationsbereiche identifizieren (siehe Abschnitt 5). Zum anderen wird es Kapitel geben, die Fallstudien beschreiben (siehe Abschnitt 6). In letzteren Untersuchungen werden Einflussgrößen von ausgewählten Variationsphänomenen ermittelt und mit inferenzstatistischen Methoden untersucht (vgl. Konopka 2020, 17).

Die Korpusgrammatik soll idealerweise sowohl grammatisch bestimmbare Ausdrucksklassen behandeln, die unterhalb der Satzebene liegen (z. B. Substantiv oder Nominalphrase), als auch Phänomene, die typischerweise auf der Satzebene zu beobachten sind (z. B. syntaktische Funktionen und Relationen). Diese Aufteilung verdeutlicht die angestrebte Doppelperspektivik: Die eine Perspektive äußert sich darin, „dass formbezogene Einheiten in Bezug auf ihre verschiedenen Funktionen untersucht werden, d. h. eine Art semasiologische Vorgehensweise verfolgt wird, die andere Perspektive [darin], dass für bestimmte Funktionen verschiedene Realisierungsformen ermittelt werden, d. h. quasi onomasiologisch vorgegangen wird“ (Konopka 2020, 20).

In Konopka (2020, 18 ff.) wird ausführlich die tentative Gliederung beschrieben. Wir stellen hier aus Platzgründen lediglich stichpunktartig die beiden Großbereiche vor. Der erste Bereich ist im Großen und Ganzen als eine „Wort- und Phrasengrammatik“ konzipiert und behandelt Aspekte der Wortbildung, Wörter und Gruppen, Substantiv, Nomen und Nominalphrase, Adjektiv und Adjektivphrase, Adverb und Adverbphrase, andere nicht flektierbare Wortarten sowie Verb, Verbalkomplex und Verbgruppe (Verbalphrase). Der zweite Bereich bezieht sich auf den Satz als eine prototypische Realisierung der kommunikativen Minimaleinheit – er ist demnach als eine Art „Satzgrammatik“ konzipiert und umfasst folgende Themen: Satz und Satzkomponenten, Komplementation, Supplementation, Kongruenz und Rektion, Wortstellung, Informationsstruktur, komplexere Strukturen (Asyndese, Syndese, Koordination, Subordination, komplexe Sätze). Im zweiten Bereich sollen unterschiedliche Realisierungsformen syntaktischer Funktionen und Relationen, die verschiedenen grammatischen Ebenen angehören können, miteinander verglichen werden (vgl. Konopka 2020, 18–19). Die Gliederung ist modifizierbar, ergänzbar und flexibel, damit auf aktuelle Entwicklungen, die sich durch Forschungsdynamiken ergeben, reagiert werden kann. Auch die Formulierungen der Kapitelüberschriften sind vorläufig. Im Projekt *Korpusgrammatik* wird zurzeit die Realisierung des ersten Bereichs vorgenommen.

Im Folgenden werden die unterschiedlichen Kapiteltypen anhand von Beispieluntersuchungen aus der Projektarbeit genauer veranschaulicht.

5 Quantitative Überblicke: Modifikation in der Nominalphrase

Als Beispiel für ein Überblickskapitel, in dem ein ganzer Phänomenbereich unter Verwendung deskriptiv-statistischer Methoden behandelt wird, stellen wir im Folgenden die Befunde aus Münzberg (2020) zur Modifikation in der Nominalphrase vor.

Nominalphrasen, deren Kopf ein Substantiv (und nicht etwa ein Pronomen) ist, können sehr weit mit modifizierenden Attributen ausgebaut werden (Admoni 1973, 36–76).⁶ Münzberg (2020) unterscheidet insgesamt dreizehn verschiedene Attributstypen, darunter u. a. Adjektivattribute (*grünes Licht*), Genitivattribute (*die Sehenswürdigkeiten Hamburgs*) und Relativsätze (*eine Dimension, der man sich so kaum bewusst ist*). Anhand von Stichproben von mehreren Tausend zufällig

⁶ Münzberg (2020) spricht von ‚Supplementen‘ statt von ‚Modifikatoren‘.

ausgewählten Nominalphrasen untersucht Münzberg die Frequenz und Distribution dieser Typen. Dabei werden Nominalphrasen in verschiedenen Satzgliedfunktionen (Subjekt, Prädikativ, Akkusativobjekt, Dativobjekt bzw. freier Dativ) berücksichtigt, und neben Daten aus DeReKo werden auch Texte aus dem Subkorpus „Forum“ des DECOW-Webkorpus untersucht. Münzbergs Auswertung ergibt u. a., dass sowohl in den DeReKo- als auch in den Forendaten das attributive Adjektiv den häufigsten Attributstyp darstellt. Und sowohl in DeReKo als auch in den Foren tritt dieser am häufigsten in Nominalphrasen in prädikativer Funktion auf. Dieser Befund steht im Einklang mit der Beobachtung Pons (2011, 168, 266), wonach Prädikative besonders umfangreich sind. Ein Beispiel ist der Beleg in (1). Die prädikative Nominalphrase ist grau hinterlegt, die Adjektivattribute sind durch Fettdruck hervorgehoben:

- (1) Obwohl Montenegro **die einzige ehemalige jugoslawische** Teilrepublik ist, **die abgesehen von Serbien je als unabhängiger Staat (1878 und 1918) existierte**, (...). (Kleine Zeitung, 25.04.2000, ohne Seite)

Münzberg erklärt die hohe Komplexität von Prädikativen funktional: Nämlich damit, dass Prädikative häufig den Charakter einer Definition haben: Zu der vom Subjekt bezeichneten Entität (in (1): *Montenegro*) wird ein Oberbegriff („Genus proximum“) genannt (*Teilrepublik*), und mit Hilfe von Attributen (neben den Adjektivattributen *einzige*, *ehemalige* und *jugoslawische* auch der Relativsatz *die ... existierte*) spezifiziert, inwiefern sich die vom Subjekt bezeichnete Entität von anderen Entitäten unterscheidet („Differentiae specificaе“), die vom selben Oberbegriff bezeichnet werden können.

Neben grundlegenden Gemeinsamkeiten zwischen den DeReKo-Daten und den Forendaten lassen sich auch einige Unterschiede nachweisen. So kommen z. B. die Adjektivattribute in den Forendaten insgesamt seltener vor als in den DeReKo-Daten; das Gleiche gilt u. a. auch für Genitivattribute. Was Letztere betrifft, zeigen Kopf/Bildhauer (2024) in einer enger fokussierten Variationsstudie, dass die Webforen unter ansonsten vergleichbaren Bedingungen in stärkerem Maße zu Präpositionalattributen mit *von* tendieren (*die Sehenswürdigkeiten Hamburgs* vs. *die Sehenswürdigkeiten von Hamburg*).

Überblicke wie Münzberg (2020) können eine erste Auskunft darüber geben, welche Formen innerhalb einer funktionalen Domäne wie oft zum Einsatz kommen, und einige Grundtendenzen ihrer Verteilung offenlegen. Sie ermöglichen eine Fülle von deskriptiven Beobachtungen, denen in Form von stärker fokussierten Folgestudien genauer nachgegangen werden kann.

6 Fallstudien

Neben Überblickskapiteln setzt sich die anvisierte Korpusgrammatik aus Fallstudien zusammen, in denen grammatische Variationsphänomene mit ihren internen und externen Einflussfaktoren in den Blick genommen werden. Als Beispiel dafür stellen wir in Abschnitt 6.1 ein syntaktisches und in Abschnitt 6.2 ein morphologisches Phänomen vor. In Abschnitt 6.1 geht es um die Variation zwischen finiten und infiniten Attributsätzen (basierend auf Weber/Bildhauer/Münzberg 2024). In Abschnitt 6.2 geht es um Variation bei der Flexion der sog. schwachen Maskulina (basierend auf Weber/Hansen 2024). Anhand der schwachen Maskulina geben wir zugleich einen Ausblick auf Anschlussprojekte, die den korpusgestützten Ansatz von der Variation innerhalb der Schriftlichkeit auf die Variation zwischen schriftlichem und mündlichem Sprachgebrauch sowie die Variation innerhalb der Mündlichkeit ausweiten (basierend auf Hansen/Weber im Erscheinen).

6.1 Finite vs. infinite Attributsätze

Unter bestimmten Bedingungen alternieren im Deutschen finite, von *dass* eingeleitete Nebensätze mit *zu*-Infinitiven. Während vorherige Studien wie z. B. Brandt (2019) diese Variation dort in den Blick nehmen, wo die Nebensätze als Komplement eines Verbs fungieren (*Doch ich habe **geschworen**, dass ich beim Malen sterbe* vs. *Ich hatte **geschworen**, die Wahrheit zu suchen*, vgl. Brandt 2019, 211), untersuchen Weber/Bildhauer/Münzberg (2024) die *zu*-/*dass*-Variation bei Substantiven, vgl. Beispiele (2–3). Das Substantiv ist in den Beispielen durch Fettdruck hervorgehoben, der Attributsatz ist grau hinterlegt.

- (2) Mauersegler, die sich von Fluginsekten ernähren, haben die **Eigenart**, dass sie Hochdruckgebieten praktisch hinterherfliegen.
(Schweriner Volkszeitung, 14.07.2009, S. 16)
- (3) Europäische Würger haben die lebenswerte **Eigenart** entwickelt, lebende Beute auf Dornen aufzuspießen.
(Süddeutsche Zeitung, 08.12.2008, S. 25)

Weber/Bildhauer/Münzberg (2024) setzen sich zum Ziel, die Bedingungen herauszuarbeiten, unter denen die beiden Realisierungsformen überhaupt gegeneinander austauschbar sind, und Faktoren zu isolieren, die innerhalb der variablen Kontexte eher zur Wahl der einen oder der anderen Variante führen. Im Folgenden greifen wir zwei Faktoren heraus, die die Rolle des tatsächlichen Sprachge-

brauchs bezeugen, d. h. die Rolle der Verwendung einer konkreten Attributsatz-Instanz durch eine konkrete Sprachbenutzerin bzw. einen konkreten Sprachbenutzer in einer konkreten Schreibsituation.

Zum einen ermitteln Weber/Bildhauer/Münzberg (2024), dass komplexe Attributsätze (verstanden als solche, die selbst wiederum mindestens einen weiteren Nebensatz enthalten oder mit einem anderen Attributsatz koordiniert sind bzw. zumindest zwei koordinierte Prädikate enthalten) etwas stärker zur *dass*-finiten Variante tendieren als Attributsätze, die nicht in diesem Sinne komplex sind. Dieser Befund steht im Einklang mit Rohdenburgs (1996, 151) *complexity principle*, wonach Sprecher*innen explizitere Strukturen verwenden, wenn sie kognitiv anspruchsvollen Situationen ausgesetzt sind. Die *dass*-finite Variante lässt sich nämlich insbesondere insofern als die komplexere ansehen, als dort das Subjekt des Attributsatzes explizit gemacht wird, vgl. z. B. das Pronomen *sie* als Subjekt von *hinterherfliegen* in (2). Beim *zu*-Infinitiv hingegen bleibt das Subjekt implizit: Dass es z. B. in (3) die Europäischen Würger sind, die lebende Beute auf Dornen aufspießen, muss vom Leser bzw. von der Leserin erst rekonstruiert werden.

Zum anderen gibt es Hinweise darauf, dass auch die Unterscheidung zwischen Mündlichkeit und Schriftlichkeit sowie Registerwechsel einen Einfluss auf die Wahl zwischen finiten und infiniten Strukturen haben können. Für Adverbialsätze im Englischen etwa zeigen Biber et al. (1999, 826), dass finite Sätze in gesprochener Sprache (*conversation*) häufiger vorkommen als in geschriebener. Und in manchen Dialekten des Deutschen, d. h. primär gesprochenen Varietäten, gibt es keine dem Standarddeutschen entsprechenden *zu*-Infinitive (vgl. Bayer/Brandner 2004 zum Mittelbairischen und Reichenauerischen). Daraus lässt sich ableiten, dass auch medial schriftliche Texte, die ein gewisses Maß an konzeptueller Mündlichkeit aufweisen, in bestimmten Fällen eine Präferenz für finite Strukturen zeigen. Für die *zu-dass*-Variation bei Substantiven ermitteln Weber/Bildhauer/Münzberg (2024), dass Texte aus Webforen, die häufig in nächstsprachlichen Situationen, spontan und unter geringerem Normdruck entstehen, tatsächlich etwas stärker zur *dass*-finiten Variante tendieren als die überwiegend pressesprachlichen DeReKo-Daten. Illustrativ ist hier ein Beispiel wie (4).

- (4) Ich baue auf seine **Beteuerung**, dass er so etwas nicht wieder tut.
(<http://forum.express.de/archive/index.php/t-7664.html>)

Der Beleg ist insofern auffällig, als das Subjekt des Attributsatzes (*er*) hier korreferent mit einem Ausdruck innerhalb derselben Nominalphrase ist, in der auch der Attributsatz selbst steht (*seine*). Wie Weber/Bildhauer/Münzberg zeigen, ist dies eine Konstellation, die insgesamt stärker als andere die Wahl des *zu*-Infinitivs begünstigt; die Forenbenutzerin bzw. der Forenbenutzer in (4) wählt hier dennoch

die *dass*-finite Variante. Es ließe sich spekulieren, dass eine etwaige *dass*-Präferenz gesprochener Sprache im Zusammenhang mit dem *complexity principle* steht. Denn wie Auer (2000) herausstellt, stellen Aspekte der Zeitlichkeit mündlicher Sprache, wie z. B. ihre Flüchtigkeit, die Sprachbenutzer*innen vor kognitive Herausforderungen, die im schriftlichen Sprachgebrauch nicht bestehen. In jedem Fall erscheinen Beobachtungen zum Zusammenhang von Medialität/Register und Attributsatzrealisierung aus der Perspektive einer „explorativen Diskurslinguistik“ interessant. Müller (2018, 94) bezeichnet damit ein Forschungsprogramm, das danach fragt, „inwiefern erstens spezifische Diskurse als Erklärung für die Präferenz für bestimmte grammatische Typen verstanden werden können und zweitens Grammatik im einzelsprachlichen Sinne als Emergenzphänomene aus diskursiven Makrokonstellationen erklärt werden kann.“

6.2 Variation bei schwachen Maskulina

Das zweite Variationsphänomen stammt aus dem Bereich der substantivischen Kasusflexion. Es betrifft die sogenannten schwachen Maskulina. Traditionell werden Substantive dieser Flexionsklasse im Akkusativ (5), Dativ (6) und Genitiv (7) Singular (sowie im Plural) mit *-(e)n* markiert:

- (5) Man muss ihn nicht mögen, **den Berliner Bären**.
(die tageszeitung, 25.02.2005, S. 25)
- (6) Stimmen die Merkmale überein, handelt es sich **beim Bären** aus dem Müns-
tertäl um ein eineinhalb Jahre altes Jungtier.
(Saale-Zeitung, 20.11.2013, S. 6)
- (7) Der polnische Präsident Walesa mahnt einen raschen Beitritt an, weil der Erfolg
Schirínówskijs die Gefährlichkeit **des russischen Bären** zeige.
(Süddeutsche Zeitung, 05.01.1994, S. 4)

Wie allgemein bekannt, besteht jedoch eine Tendenz dazu, die Substantive im Akkusativ (8) und Dativ (9) endungslos zu verwenden. Im Genitiv ist – wenngleich seltener – eine Ersetzung von *-(e)n* durch *-(e)s* zu beobachten (10) (vgl. z. B. Dudengrammatik 2016, 210–216; Dudenwörterbuch 2021, 906–908; Schäfer 2019). Nicht nur Genitivformen auf *-(e)s*, sondern auch die endungslosen Akkusativ- und Dativformen lassen sich als Ausdruck einer analogischen Angleichung an das

starke Flexionsmuster sehen, das „Standardmuster“ für Maskulina (vgl. z. B. *dem Tag_*⁷, *den Tag_ des Tages*) (Dudengrammatik 2016, 213).

- (8) Sie sprach trotzdem auf der Gedenkveranstaltung für **den erschossenen Bär** Bruno auf dem Münchner Marienplatz [...].
(Süddeutsche Zeitung, 21.09.2007, S. 35)
- (9) Da bei Kronstadt die höchste Braunbärdichte weltweit mit **einem Bär** pro Quadratkilometer herrscht, reagieren die Menschen gelassen.
(Die Südostschweiz, 15.08.2005, ohne Seite)
- (10) Die Heimat **des größten Bärs** der Welt sind die Harburger Berge
(Der Prignitzer, 25.02.2013, S. 25)

Bei der Variation zwischen schwachen und starken (einschließlich der endungslosen) Formen handelt es sich um einen klassischen Zweifelsfall: Bei manchen Substantiven gelten die endungslosen Akkusativ- und Dativformen als standardsprachlich anerkannt (z. B. *mit einem **Magneten/Magnet***), bei anderen hingegen nicht (*die Prinzessin küsste den **Prinzen***) (vgl. Dudengrammatik 2016, 973). Das Zweifelsfälle-Wörterbuch (2021, 907) geht davon aus, dass das Weglassen der Endung „auch im geschriebenen Standarddeutsch so weit verbreitet“ sei, „dass es nicht einfach als inkorrekt bezeichnet werden kann.“

Zum fraglichen Variationsphänomen wurden zwei Fallstudien durchgeführt. Die erste Fallstudie (Weber/Hansen 2024) fand im Rahmen des Projekts *Korpusgrammatik* statt. Sie basiert auf Daten des DeReKo und dem Webforen-Teil des DECOV und verfolgt das Ziel, neben grammatischen Einflussfaktoren auch Registerunterschiede in den Blick zu nehmen. Die zweite Fallstudie (Hansen/Weber im Erscheinen), auf die weiter unten detaillierter eingegangen wird, fand im Rahmen der IDS-Nachfolgeprojekte *Grammatische Strukturen im Gesprochenen Deutsch* und *Methoden der Grammatikforschung* statt und legt den Schwerpunkt auf medial mündliche Daten. Wir skizzieren im Folgenden kurz einige zentrale Erkenntnisse der bisherigen Forschung zu den grammatischen Einflussfaktoren und gehen anschließend auf Register- und Medialitätsunterschiede ein.

Köpcke (1995) zeigt, dass viele schwache Maskulina prototypische phonotaktische und semantische Merkmale aufweisen. So enden z. B. viele von ihnen auf Schwa und bezeichnen Menschen oder zumindest höhere Lebewesen (z. B. *Mat-*

⁷ Das mehr oder weniger relikthaft erhaltene Dativ *-e* wird hier nicht weiter berücksichtigt, siehe hierzu z. B. Konopka (2012).

rose, Kollege, Löwe). Darauf aufbauend zeigt Schäfer (2019) im Rahmen einer breit angelegten Korpusstudie, dass im DECOW-Webkorpus vor allem diejenigen Substantive dazu tendieren, stark flektiert zu werden, die wenig prototypisch sind, wie z. B. Einsilber wie das Substantiv *Bär* (5–10). Wie eingangs erwähnt, ist die Flexionsschwankung darüber hinaus kasusabhängig. Die starke Flexion tritt primär in Form von endungslosen Akkusativ- und Dativformen auf, seltener in Form von starken Genitivendungen (vgl. Thieroff 2003, 109, 114; Schäfer 2019, 405; Zweifelsfälle-Wörterbuch 2021, 907). Thieroff erklärt dies damit, dass die abstrakte Struktur des starken Paradigmas (markierter Genitiv vs. unmarkierter Nicht-Genitiv, vgl. z. B. *des Tag(e)s* vs. *der/den/dem Tag_*) bereits dann erreicht ist, wenn lediglich das *-(e)n*-Suffix im Akkusativ und Dativ weggelassen wird, der Genitiv aber weiterhin mit *-(e)n* (statt *-(e)s*) markiert wird (*des Menschen* vs. *der/den/dem Mensch_*). Neben diesem Struktur-bezogenen Faktor dürfte aber auch ein soziolinguistischer Grund eine Rolle spielen: Der Genitiv an sich ist im Deutschen ein stilistischer Marker für Hoch- oder Schriftsprachlichkeit geworden (vgl. Szczepaniak 2014; Pickl 2020; Seiler 2022, 58). Daraus lässt sich ableiten, dass Sprecher*innen, wenn sie **überhaupt** den Genitiv verwenden, diesen tendenziell auch normgerecht (also schwach) realisieren. Für die DeReKo- und die DECOW-Daten können Weber/Hansen (2024) in ihrer ersten Fallstudie die Ergebnisse zu den grammatischen Einflussfaktoren aus Schäfer (2019) bestätigen und zudem u. a. Registerunterschiede nachweisen: So neigen Texte aus dem Webforen-Teil des DECOW-Webkorpus in stärkerem Maße zu starken Formen als Texte aus dem pressesprachlichen DeReKo. Dennoch überwiegen auch in den Webforen die schwachen Formen noch sehr deutlich (s. u.).

In der zweiten Fallstudie wird schwerpunktmäßig der mündliche Sprachgebrauch in den Blick genommen und mit dem schriftlichen Sprachgebrauch verglichen. Als Datengrundlage für den mündlichen Sprachgebrauch dient das Forschungs- und Lehrkorpus (FOLK) (Deppermann/Hartung 2012; Kaiser 2018). FOLK ist ein kontinuierlich wachsendes Korpus, das Gesprächsdaten aus unterschiedlichsten Bereichen des gesellschaftlichen Lebens, wie z. B. Arbeit, Freizeit, Bildung, öffentliches Leben, Dienstleistungen usw., im deutschen Sprachraum beinhaltet. Die Gespräche in FOLK sind nach verschiedenen Metadaten kategorisiert. Ein Parameter ist die „Interaktionsdomäne“, der die einzelnen Gespräche zugeordnet werden, wobei zwischen „privat“, „institutionell“, „öffentlich“ und „Sonstiges“ unterschieden wird. Wir fassen im Folgenden einige Ergebnisse der Fallstudie zusammen, die speziell aus diskursgrammatischer Perspektive interessant erscheinen.

Insgesamt treten die schwachen Maskulina in ca. 15% der ausgewerteten FOLK-Belege in starken Formen auf. Das ist mehr als in Vergleichsstichproben aus DeReKo (ca. 2%) und dem Webforen-Teil des DECOW-Webkorpus (ca. 8%)

(vgl. Abb. 1). In den Webforen tritt darüber hinaus in sehr seltenen Fällen das Genitivsuffix *-(e)ns* auf („misch“ in Abb. 1, z. B. *des Pilotens*, *des Soldatens*), in dem man eine Kombination aus schwachem und starkem Suffix sehen kann (Krischke 2012).

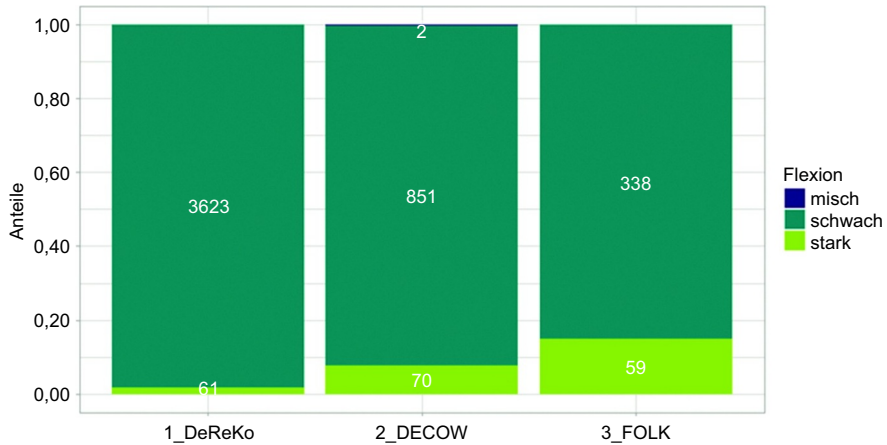


Abb. 1: Anteile starker und schwacher Formen in FOLK im Vergleich mit DeReKo und DECOW-Webforen (vgl. Hansen/Weber im Erscheinen).

Der in Abb. 1 dargestellte Befund steht im Einklang mit der Annahme, dass im mündlichen Sprachgebrauch die Norm-divergenten, starken Formen (12) häufiger vorkommen als in schriftlichen, insbesondere pressesprachlichen Texten. Dennoch überwiegen auch in FOLK die schwachen Formen (11) noch deutlich.⁸

(11) wenn man **en en toten menschen** sieht den man gekannt hat
(FOLK_E_00339_SE_01_T_02)

(12) und geruch ist für **n mensch** instinktiv extrem wichtig
(FOLK_E_00426_SE_01_T_01)

Hansen/Weber (im Erscheinen) diskutieren die Frage, wie die unterschiedlich hohen Anteile starker Formen in FOLK, DECOW und DeReKo genau zu erklären

⁸ Die Ergebnisse werden detaillierter in Hansen/Weber (im Erscheinen) dargestellt, wo auch Ergebnisse zur Verteilung starker und schwacher Formen in FOLK in Abhängigkeit der Faktoren Kasus, Phonotaktik und Semantik erörtert werden.

sind: In welchem Maße kommen sie dadurch zustande, dass möglicherweise bestimmte grammatische Bedingungen, die die starke Flexion grundsätzlich (also unabhängig von der Medialität) begünstigen, in FOLK und den Webforen stärker vertreten sind als in DeReKo? Und in welchem Maße neigen Sprecher*innen bzw. Schreiber*innen in den drei Korpora auch unter **vergleichbaren grammatischen Bedingungen** unterschiedlich stark zur starken Flexion? So lässt sich z. B. zeigen, dass die schwachen Maskulina in DeReKo häufiger im Genitiv stehen als in DECOW, und in DECOW wiederum häufiger als in FOLK. Und wie oben beschrieben, neigt der Genitiv insgesamt stärker als Akkusativ und Dativ zur schwachen Form. Außerdem lässt sich u. a. zeigen, dass der Anteil von Mehrsilbern mit Endbetonung (*Präsident*, *Konkurrent* etc.) in DeReKo höher ausfällt als in DECOW und FOLK. Und wie Schäfer (2019, 404) zeigt, neigen Substantive dieses Typs insgesamt stärker zu schwachen Formen als Einsilber (z. B. *Bär*, *Mensch*) oder Mehrsilber mit Nicht-Endbetonung (z. B. *Bauer*, *Nachbar*).

Um bezüglich der grammatischen Faktoren kontrollieren zu können, werden die Daten mit einer multivariaten statistischen Methode analysiert. Dabei wird das Korpus, aus dem der Beleg stammt, gemeinsam mit den grammatischen Faktoren berücksichtigt, die (potentiell) einen Einfluss auf die Wahl zwischen starker und schwacher Flexion ausüben. Die Analyse ergibt, dass neben grammatischen Faktoren tatsächlich erwartungsgemäß auch das Korpus einen Einfluss hat: Wie sich zeigt, tendieren insbesondere die FOLK-Daten, in geringerem Maße aber auch die Webforen, stärker zur starken Flexion als DeReKo. M.a.W.: Die Verteilungsunterschiede aus Abb. 1 lassen sich zum Teil, aber eben nicht allein, dadurch erklären, dass z. B. Substantive in DeReKo häufiger im Genitiv stehen oder es sich bei ihnen häufiger um Mehrsilber mit Endbetonung handelt.

In einem weiteren Schritt betrachten wir mögliche diskursive Einflüsse auf die Variation der Flexion innerhalb der mündlichen Daten genauer. Dazu greifen wir auf die „Interaktionsdomäne“ zurück, der die jeweilige FOLK-Interaktion zugeordnet ist (vgl. Kaiser 2018, 521–522): „Öffentliche Interaktionen“ sind Gespräche, die im Rahmen öffentlich zugänglicher und/oder massenmedial vermittelter Anlässe stattfinden. Sie umfassen u. a. die Schlichtungsgespräche zu Stuttgart 21 sowie TV-Debatten. „Private Interaktionen“ sind definiert als informelle Gespräche mit Familie und/oder Freunden. „Institutionelle Interaktionen“ umfassen Gespräche, in denen Personen als institutionelle bzw. professionelle Vertreter*innen agieren (z. B. Gespräche am Arbeitsplatz oder in Ausbildungsstätten). Die Kategorie „Sonstiges“ bezeichnet Interviews und experimentelle Interaktionen. Abb. 2 zeigt die Anteile starker und schwacher Formen über die verschiedenen Interaktionsdomänen im FOLK. Wie sich zeigt, ist der Anteil der starken Formen in den privaten Interaktionen höher als in den institutionellen und öffentlichen.

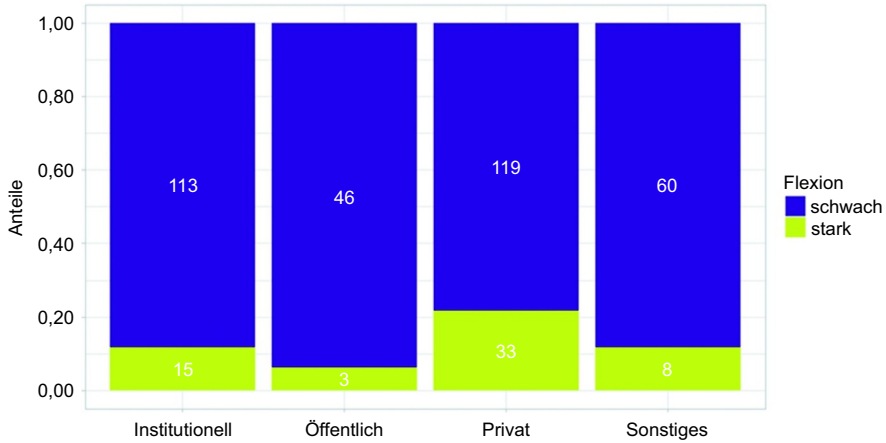


Abb. 2: Anteile starker und schwacher Formen in FOLK nach Interaktionsdomäne (vgl. Hansen/Weber im Erscheinen).

Der Mosaikplot in Abb. 3 visualisiert die Häufigkeitsverteilung über Rechtecke, deren Größe proportional zur Fallzahl in den Zellen ist. Die Farbe der Rechtecke bezieht sich auf die standardisierten Residuen. Das grün eingefärbte Rechteck zeigt an, dass die entsprechende Zelle signifikant überrepräsentiert ist.⁹

In den öffentlichen Interaktionen kann man einen Indikator für den sog. „mündlichen Gebrauchsstandard“ (im Sinne von Deppermann et al. 2013) sehen (siehe auch Abschnitt 7.1). Wenn das der Fall ist, dann scheinen endungslose Formen insbesondere in dieser Sprachform eher untypisch zu sein, häufiger sind sie in privater Kommunikation vorzufinden.

Die hier dargestellte Variation bei der Flexion der sog. schwachen Maskulina ist ein Beispiel dafür, wie mit einem korpusgestützten Ansatz die Beschreibung grammatischer und außergrammatischer Variationsparameter zunächst innerhalb der Schriftlichkeit auf die Beschreibung variationssteuernder Parameter zwischen konzeptionell schriftlichem und mündlichem Sprachgebrauch sowie auf die Variationsbeschreibung medial mündlicher Daten ausgeweitet werden kann. Nachdem wir in den Abschnitten 6.1 und 6.2 konkrete Beispiele für korpuslinguistische Untersuchungen grammatischer Phänomene dargestellt haben, gehen wir im Folgenden darauf ein, welche Beiträge die Korpusgrammatik zum Diskurs **über** Grammatik leisten kann.

⁹ Zu weitergehenden Informationen zur Funktionsweise und Lesbarkeit von Mosaikplots vgl. Friendly (1994) und Meyer et al. (2023).

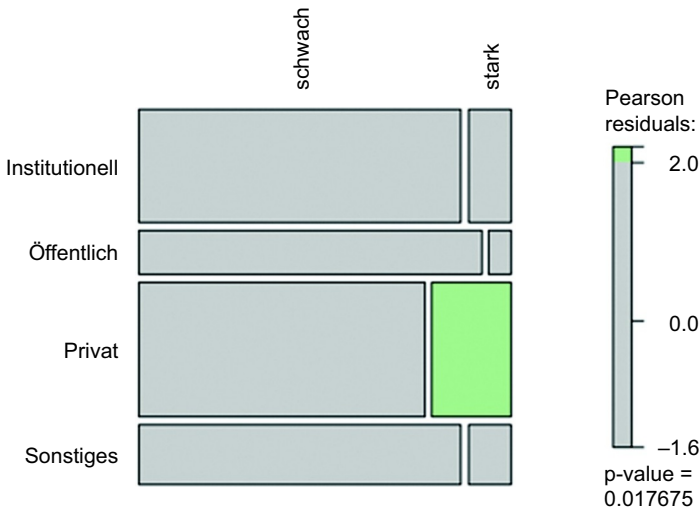


Abb. 3: Mosaikplot, Flexionsmuster in Abhängigkeit der Interaktionsdomäne in FOLK.

7 Die Korpuslinguistik und der Diskurs über Grammatik

7.1 Korpuslinguistik und sprachliche Norm

Die Korpuslinguistik hat keine normative Ausrichtung; sie verfolgt nicht das Ziel, grammatische Formen in Bezug auf Kategorien wie „richtig“ oder „falsch“ zu bewerten. Allerdings kann sie sich in Diskussionen um die sprachliche Norm einbringen. Wir gehen im Folgenden auf drei Punkte ein, die uns in dieser Hinsicht wichtig erscheinen, und kommen zu ihrer Illustration auf die in Abschnitt 6.2 vorgestellten schwachen Maskulina zurück.

Erstens: Dadurch, dass die Korpuslinguistik differenzierte Sprachdaten zum tatsächlichen Sprachgebrauch liefert, kann sie Auskunft darüber geben, welche Formen in welchen Kontexten eher üblich oder unüblich sind. Damit sind ihre Ergebnisse relevant für Ansätze, die dafür plädieren, „Standardsprache“ nicht als präskriptive Norm zu verstehen, sondern als deskriptiv beschreibbare Sprachform, die in bestimmten (semi-)formellen Gebrauchskontexten auftritt. Für die gesprochene Sprache wird ebendieser Ansatz von Deppermann et al. (2013, 90) unter der Bezeichnung „Gebrauchsstandard“ vertreten. Wie erwähnt (vgl. Abschnitt 6.2), dürften in FOLK insbesondere die öffentlichen Interaktionen auf-

schlussreich in Bezug auf die Frage sein, welche Formen als Teil des mündlichen Gebrauchsstandards anzusehen sind. Für die schwachen Maskulina hat sich gezeigt, dass starke Formen in den öffentlichen Interaktionen kaum zu beobachten sind. Das würde darauf hindeuten, dass sie nur eingeschränkt als Teil des mündlichen Gebrauchsstandards gelten können.

Zweitens können die von der Korpuslinguistik ermöglichten, differenzierten Sprachdaten als Korrektiv für – bisweilen möglicherweise zu stark pauschalisierende – Urteile aus der Sprachkritik herangezogen werden. So sind etwa Bastian Sick (2005, 65) zufolge „Sätze wie „Dem Patient geht’s gut“ und „Lukas, lass den Elefant in Ruhe“ [...] mittlerweile häufiger zu hören als die korrekt formulierten Aussagen „Dem Patienten geht’s gut“ und „Lukas, lass den Elefanten in Ruhe““. Wie sich gezeigt hat (vgl. Abb. 1), kommen starke Formen in FOLK zwar tatsächlich häufiger vor als in den schriftlichen Korpora (DECOW und insbesondere DeReKo). Allerdings überwiegen die Formen auf *-(e)n* auch in FOLK immer noch deutlich. Und zwar auch innerhalb der Untergruppe der Mehrsilber mit Endbetonung, zu denen Sicks Beispielsubstantive (*Patient*, *Elefant*) gehören.

Drittens kann die Korpuslinguistik – bzw. die Linguistik als wissenschaftliche Disziplin insgesamt – Erklärungen dafür liefern, **warum** Sprecher*innen in bestimmten Fällen auf bestimmte Art und Weise von der kodifizierten Norm abweichen. Häufig wird dadurch offengelegt, dass bestimmte „Fehler“ in der Systematik der Sprache selbst angelegt sind und z. B. als Abbau von Irregularitäten aufgefasst werden können. So zeigt Schäfers (2019) Untersuchung ja, dass die starken Formen am häufigsten bei solchen schwachen Maskulina auftreten, die lautlich und/oder semantisch untypisch für diese Klasse sind. Wenn man zeigen kann, dass es für bestimmte Normverstöße somit „gute Gründe“ gibt, kann dies zu einer Destigmatisierung dieser Verstöße beitragen.

7.2 Die IDS-Korpusgrammatik als Beitrag zum Forschungsdiskurs

7.2.1 Bausteine einer Korpusgrammatik des Deutschen

Eine deskriptive, korpuslinguistisch fundierte Grammatik des Deutschen zu erstellen, ist eine enorme und damit zeitaufwendige Aufgabe. Folgerichtig kann es eine Korpusgrammatik als abgeschlossenes Werk nicht so schnell geben (vgl. Konopka 2020, 7, 14–15). Aus diesem Grund werden einzelne Bausteine der Grammatik vorab in der Open-Access-Reihe „Bausteine einer Korpusgrammatik des Deut-

schen“ (Konopka/Wöllstein/Felder 2020, 2021, 2024) veröffentlicht.¹⁰ Ziel dabei ist es, korpusgestützte Untersuchungen zeitnah in Form von einzelnen Kapiteln der Korpusgrammatik vorzulegen und diese im Sinne der Open-Science-Bewegung der Öffentlichkeit zugänglich zu machen und dem Wissenschaftsdiskurs frühzeitig zur Diskussion zu stellen. Auf diese Weise werden Bestandteile der Korpusgrammatik in fortlaufend erscheinenden Einzelteilen veröffentlicht, die sich zu einem immer vollständigeren Bild der deutschen Grammatik zusammensetzen.¹¹ Die „Bausteine“ werden in Kooperation mit dem Germanistischen Seminar der Universität Heidelberg unter der Trägerschaft des Europäischen Zentrums für Sprachwissenschaften (EVS) herausgegeben.

7.2.2 Materialien des Projekts Korpusgrammatik

Publikationen, die im Rahmen der Projektarbeit entstehen, bestehen nicht nur aus Kapiteln für die Korpusgrammatik, Aufsätzen und Monographien, sondern auch aus den zugehörigen Materialien, die wir über das grammatische Informationssystem grammis (<https://grammis.ids-mannheim.de/korpusgrammatik>¹², zuletzt eingesehen am 29.03.2023) der Öffentlichkeit zugänglich machen. Auf dieser Plattform werden u. a. die Datensammlungen und Dokumentationen zur Korpusgrammatik der Öffentlichkeit zur Verfügung gestellt (<https://grammis.ids-mannheim.de/korpusgrammatik/6568>, zuletzt eingesehen am 29.03.2023). Die Datensätze bestehen aus einschlägigen Korpusbelegen aus DeReKo, teilweise ergänzt durch Internetbelege aus DECOW¹³. Sie sind durchgängig nach relevanten sprachlichen Eigenschaften auf verschiedenen linguistischen Beschreibungsebenen und nach außersprachlichen Merkmalen (Metadaten wie *Register*, *Region*, *thematische Domäne*, *Medium*) annotiert. Die Datensätze sind mit ausführlichen Dokumentationen herunterladbar und werden zum Teil zusätzlich als abfragbare Datenbanken online präsentiert. Die Ergebnisse der Studien im Rahmen des Projekts *Korpusgrammatik* sind auf diese Art und Weise von externen Wissenschaftler*innen überprüfbar, und eine eigene Anschlussforschung mit weitergehenden Forschungsfragen ist möglich.

¹⁰ <https://heiup.uni-heidelberg.de/journals/index.php/bkgd/index> (zuletzt eingesehen am 27.03.2023).

¹¹ Vgl. hierzu auch die Beschreibung der „Bausteine“ in Konopka (2020, 15).

¹² DOI: 10.14618/korpusgrammatik.

¹³ Zur Datenbasis des Projekts *Korpusgrammatik* vgl. Abschnitt 4 in diesem Beitrag.

Neben den Datensätzen werden Werkzeuge und Materialien online zur Verfügung gestellt, die methodische Aspekte der Korpusuntersuchungen betreffen. Zurzeit werden zwei Anwendungen angeboten:

- 1) KoGra-R (Falke et al. 2020) ist eine web-basierte Schnittstelle, mit der vordefinierte, in R (R Core Team 2022)¹⁴ programmierte statistische Auswertungen basierend auf Häufigkeitstabellen durchgeführt werden können. Nutzer*innen können u. a. Reintext-Tabellen zur Analyse hochladen. Von KoGra-R werden sowohl Analysen für Einzelabfragen als auch vergleichende Auswertungen bereitgestellt. Die in KoGra-R implementierten statistischen Tests sind geeignet, um mindestens einen ersten statistisch abgesicherten Eindruck einer Datenlage zu erlangen. Im Rahmen von Pilotstudien (z. B. Brandt/Fuß 2019; Konopka/Fuß 2016; Bubenhofer et al. 2014, 125 ff.; Konopka/Waßner 2013) wurden verschiedene statistische Analysen erprobt. Um einige von denen, die sich für die erste Datenexploration als sinnvoll erwiesen haben, auf einfache Weise zugänglich zu machen, wurde das Tool KoGra-R entwickelt. Es ist über die Internetadresse <http://kograno.ids-mannheim.de/> (zuletzt eingesehen am 06.04.2023) frei zugänglich. Die Funktionsweise des Tools und die durchführbaren Tests sind ausführlich in Hansen-Morath et al. (2019) beschrieben.¹⁵ Die Analysefunktionen sind prinzipiell offen und können erweitert werden. Bislang können mit dem Tool sowohl deskriptive Statistiken und Visualisierungen als auch inferenzstatistische Verfahren auf der Basis von Häufigkeitstabellen durchgeführt werden. Außerdem ist die Berechnung eines Dispersionsmaßes, der DPnorm (vgl. Gries 2008, 2009; Lijffijt/Gries 2012), implementiert.
- 2) Die zweite Anwendung ist ein Tutorial zu Visualisierungsmöglichkeiten linguistischer Daten mit R (Hansen-Morath/Wolfer 2017). Das Tutorial vermittelt einen stark anwendungsorientierten Einstieg in die Visualisierung mit R und legt mithilfe von vielen linguistischen Anwendungsbeispielen die Grundlagen für ein eigenständiges Weiterentwickeln der eigenen Fähigkeiten im Umgang mit der Software. Das Tutorial ist frei zugänglich: <http://kograno.ids-mannheim.de/VisR-OnlinePub/> (zuletzt eingesehen am 06.04.2023). Es geht gezielt auf einige wenige Pakete ein, die insbesondere für die Visualisierung linguistischer Daten und Analyseergebnisse hilfreich sind. Die Auswahl dieser Pa-

¹⁴ R ist eine freie und flexible Software zur Umsetzung von statistischen Analysen, die u. a. zahlreiche Optionen zur Datenvisualisierung bereithält und sehr gut für große Datensätze geeignet ist. R ist mittlerweile in vielen Wissenschaftsbereichen zum Standard für quantitative Analysen geworden. Eine große Stärke von R liegt in der weltweiten Gemeinschaft der Anwender*innen, die über Zusatzpakete immer neue Funktionen zur Verfügung stellen.

¹⁵ In Hansen-Morath/Wolfer (2017) wird außerdem gezeigt, inwiefern das Tool dazu genutzt werden kann, Variationsphänomene auf verschiedenen linguistischen Ebenen zu untersuchen.

kete ist dabei einerseits aus der täglichen Praxis bei der linguistischen Arbeit des Projektes *Korpusgrammatik*, andererseits aus einer möglichst breiten Anwendbarkeit motiviert. Es gibt viele linguistische Visualisierungen, die durch das Tutorial nicht abgedeckt werden. Die vorgestellten Pakete liefern allerdings einen Grundstock an Handwerkszeug, das dazu befähigt, ansprechende Visualisierungen sprachlicher Daten und linguistischer Analyseergebnisse zu erstellen.

8 Fazit

Im vorliegenden Beitrag haben wir einige Konvergenzen von Diskurslinguistik und Korpuslinguistik (und folglich von Diskursgrammatik und Korpusgrammatik) herausgearbeitet und anhand von Studien veranschaulicht, die im Rahmen des IDS-Projekts *Korpusgrammatik* und seiner Nachfolgeprojekte entstanden sind bzw. im Entstehen begriffen sind. Für die Diskursanalyse sind Korpora zu einer unverzichtbaren Ressource geworden. Bubenhofer (2018) plädiert dafür, die Korpuslinguistik dabei nicht als bloße Hilfswissenschaft anzusehen, sondern als „Schlüssel zu einem neuen Verständnis des Umgangs mit Daten in den Geisteswissenschaften“ (ebd., 208). Umgekehrt bedeutet Korpuslinguistik zwar nicht zwingend Diskurslinguistik und somit Korpusgrammatik nicht notwendigerweise Diskursgrammatik. Im IDS-Projekt werden allerdings neben inner-grammatischen durchweg auch zahlreiche extra-grammatische Variationsdimensionen berücksichtigt. Damit spielen diskursgrammatische Fragestellungen immer auch eine Rolle. Darüber hinaus haben wir gezeigt, wie sich die Korpuslinguistik in den Metadiskurs **über** Grammatik einbringen kann und welchen Beitrag zum Forschungsdiskurs speziell das IDS-Projekt zu leisten anstrebt.

Literatur

- Admoni, Vladimir G. (1973): Die Entwicklungstendenzen des deutschen Satzbaus von heute. München: Hueber.
- Auer, Peter (2000): *On line*-Syntax – Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen. In: *Sprache und Literatur*, 31, 1, S. 43–56.
- Bates, Douglas/Mächler, Martin/Bolker, Ben/Walker, Stev (2015): Fitting Linear Mixed Effects Models Using lme4. In: *Journal of Statistical Software*, 67, 1, S. 1–48.
- Bayer, Josef/Brandner, Ellen (2004): Klitisiertes *zu* im Bairischen und Alemannischen. In: Patocka, Franz/Wiesinger, Peter (Hgg.): *Morphologie und Syntax deutscher Dialekte und Historische Dialektologie des Deutschen*. Wien: Praesens, S. 160–188.

- Biber, Douglas/Johansson, Stig/Leech, Geoffrey/Conrad, Susan/Finegan, Edward (1999): Longman Grammar of Spoken and Written English. Harlow: Longman.
- Brandt, Patrick (2019): Alternation von *zu-* und *dass-*Komplementen: Kontrolle, Korpus und Grammatik. In: Fuß, Eric/Konopka, Marek/Wöllstein, Angelika (Hgg.): Grammatik im Korpus: Korpuslinguistisch-statistische Analysen morphosyntaktischer Variationsphänomene. Tübingen: Narr Francke Attempto, S. 211–297.
- Brandt, Patrick/Fuß, Eric (2019): Relativpronomenselektion und grammatische Variation: ‘was’ vs. ‘das’ in attributiven Relativsätzen. In: Fuß, Eric/Konopka, Marek/Wöllstein, Angelika (Hgg.): Grammatik im Korpus. (Studien zur Deutschen Sprache 80). Tübingen: Narr, S. 91–209.
- Bresnan, Joan/Cueni, Anna/Nikitina, Tatiana, Baayen, R. Harald (2007): Predicting the dative alternation. In: Bouma Gerlof/Krämer, Irene/Zwarts, Joost (Hgg.): Cognitive foundations of interpretation. Amsterdam: Royal Netherlands Academy of Science, S. 69–94.
- Bubenhof, Noah/Konopka, Marek/Schneider, Roman (2013): Präliminarien einer Korpusgrammatik. Tübingen: Narr.
- Bubenhof, Noah (2018): Diskurslinguistik und Korpora. In: Warnke, Ingo H. (Hgg.): Handbuch Diskurs. Berlin/Boston: De Gruyter, S. 208–241.
- 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 (Hgg.): Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen. Berlin/Boston: De Gruyter, S. 415–450.
- Deppermann, Arnulf/Kleiner, Stefan/Knöbl, Ralf (2013): “Standard usage”: towards a realistic conception of spoken standard German. In: Caro Reina, Javier/Auer, Peter/Kaufmann, Göz (Hgg.): Language variation. European perspectives 4.: selected papers from the 6. International Conference on Language Variation in Europe (ICLaVE 6), Freiburg, June 2011. Amsterdam: Benjamins, S. 83–116.
- Wöllstein, Angelika/Dudenredaktion (Hgg.) (2016): Duden – Die Grammatik. 9. Aufl. Berlin: Dudenverlag.
- Falke, Stefan/Hansen-Morath, Sandra/Wolfer, Sascha (2020): KoGra-R 2.0: Standardisierte statistische Verfahren für korpusbasierte Häufigkeiten. 2. überarbeitete Version. Mannheim: Institut für Deutsche Sprache. Online-Ressource: <http://kograno.ids-mannheim.de/> (Stand: 24.04.2023).
- Friendly, Michael (1994): Mosaic displays for multi-way contingency tables. In: Journal of the American Statistical Association, 89, S. 190–200.
- Gries, Stefan Thomas (2008): Dispersions and adjusted frequencies in corpora. In: International Journal of Corpus Linguistics, 13, S. 403–437.
- Gries, Stefan Thomas (2009): Dispersions and adjusted frequencies in corpora: Further explorations. In: Language and Computers 71, 1, S. 197–212.
- Hansen, Sandra/Weber, Thilo (im Erscheinen): *Wenn sie den richtigen Kandidat wählen* – Zur Flexion der schwachen Maskulina im heutigen gesprochenen (und geschriebenen) Deutsch. In: Adriano Murelli (Hgg.): Das heutige gesprochene Deutsch zwischen Sprachkontakt und Sprachwandel. Berlin/New York: De Gruyter.
- Hansen-Morath, Sandra/Wolfer, Sascha (2017): Standardisierte statistische Auswertung von Korpusdaten im Projekt „Korpusgrammatik“ (KoGra-R). In: Konopka, Marek/Wöllstein, Angelika (Hgg.): Grammatische Variation. Empirische Zugänge und theoretische Modellierung. (Jahrbuch des Instituts für Deutsche Sprache 2016). Berlin/Boston: de Gruyter, S. 345–356.
- Hansen-Morath, Sandra/Schmitz, Hans-Christian/Schneider, Roman/Wolfer, Sascha (2019): KoGra-R: Standardisierte statistische Auswertung von Korpusrecherchen. In: Fuß, Eric/Konopka, Marek/

- Wöllstein, Angelika (Hgg.): *Grammatik im Korpus*. (= Studien zur Deutschen Sprache 80). Tübingen: Narr, S. 299–357.
- Hennig, Mathilde (Hgg.) (2021): *Dudenband 9 – Sprachliche Zweifelsfälle. Das Wörterbuch für richtiges und gutes Deutsch*, 9. Aufl. Berlin: Dudenverlag.
- Kaiser, Julia (2018): Zur Stratifikation des FOLK-Korpus: Konzeption und Strategien. In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 19, S. 515–552.
- Köpcke, Klaus-Michael (1995): Die Klassifikation der schwachen Maskulina in der deutschen Gegenwartssprache – ein Beispiel für die Leistungsfähigkeit der Prototypentheorie. In: *Zeitschrift für Sprachwissenschaft*, 14, 2, S. 159–180.
- Konopka, Marek (2018): Korpuslinguistik, Grammatiktheorie, Grammatikographie. In: Wöllstein, Angelika/Gallmann, Peter/Habermann, Mechthild/Krifka, Manfred (Hgg.): *Grammatiktheorie und Empirie in der germanistischen Linguistik*. (Germanistische Sprachwissenschaft um 2020, 1). Berlin/Boston: de Gruyter, S. 151–184.
- Konopka, Marek (2012): Dem Manne kann geholfen werden — Wann kommt das Dativ-e zum Einsatz? In: Konopka, Marek/Schneider, Roman (Hgg.): *Grammatische Stolpersteine digital — Festschrift für Bruno Strecker zum 65. Geburtstag*. Mannheim: Institut für deutsche Sprache, S. 115–124.
- Konopka, Marek (2020): Einleitung. In: Konopka, Marek/Wöllstein, Angelika/Felder, Ekkehard (Hgg.): *Bausteine einer Korpusgrammatik des Deutschen. Band 1*. Heidelberg: Heidelberg University Publishing, S. 1–28.
- Konopka, Marek/Fuß, Eric (2016): *Genitiv im Korpus. Untersuchungen zur starken Flexion des Nomens im Deutschen*. (Studien zur Deutschen Sprache 70). Tübingen: Narr.
- Konopka, Marek/Waßner, Ulrich Hermann (2013): Standarddeutsch messen? Frequenz und Varianz negativ-konditionaler Konnektoren. In: *Korpus – Grammatika – Axiologie* 8. Hradec Králové: Univerzita, S. 12–35.
- Konopka, Marek/Wöllstein, Angelika/Felder, Ekkehard (Hgg.) (2020): *Bausteine einer Korpusgrammatik des Deutschen*. Heidelberg: Heidelberg University Publishing.
- Konopka, Marek/Wöllstein, Angelika/Felder, Ekkehard (Hgg.) (2021): *Determination, Syntaktische Funktionen der Nominalphrase und Attribution. Bausteine einer Korpusgrammatik Band 2*. Heidelberg: Heidelberg University Publishing.
- Konopka, Marek/Wöllstein, Angelika/Felder, Ekkehard (Hgg.) (2024): *Substantivflexion, Attributsätze, Präfix- und Partikelverben. Bausteine einer Korpusgrammatik des Deutschen Band 3*. Heidelberg: Heidelberg University Publishing.
- Kopf, Kristin/Bildhauer, Felix (2024): The genitive alternation in German. *Corpus Linguistics and Linguistic Theory*. Published online November 13, 2024.
- Krischke, Wolfgang (2012): Des Menschens Genitive. Normabweichende Genitiv-Varianten bei schwachen Maskulina. In: *Linguistik Online*, 53, 3, S. 55–84.
- Kupietz, Marc/Lüngen, Harald/Kamocki, Paweł/Witt, Andreas (2018): The German Reference Corpus DeReKo: New Developments – New Opportunities. In: Calzolari, Nicoletta et al. (Hgg.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: ELRA, S. 4353–4360.
- Lijffijt, Jeffrey/Gries, Stefan Thomas (2012): Correction to „Dispersions and adjusted frequencies in corpora“. In: *International Journal of Corpus Linguistics*, 17, 1, S. 147–149.
- Meyer, David/Zeileis, Achim/Hornik, Kurt (2023). *vcd: Visualizing Categorical Data*. R package version 1.4-11, <https://CRAN.R-project.org/package=vcd> (Stand: 24.04.2023).
- Müller, Marcus (2018): *Diskursgrammatik*. In: Ingo H. Warnke (Hgg.): *Handbuch Diskurs*. Berlin/Boston: De Gruyter, S. 75–103.

- Münzberg, Franziska (2020): *Supplemente in der Nominalphrase*. Manuskript, Leibniz-Institut für Deutsche Sprache.
- Pickl, Simon (2020): Polarization and the emergence of a written marker. A diachronic corpus study of the adnominal genitive in German. In: *Journal of Germanic Linguistics*, 32, S. 145–182.
- Pon, Leonard (2011): *Nominalphrase in der deutschen Pressesprache von heute*. Dissertation, Josip-Juraj-Strossmayer-Universität Osijek.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> (Stand 24.04.2023).
- Schäfer, Roland (2015): Processing and Querying Large Web Corpora with the COW14 Architecture. In *Proceedings of Challenges in the Management of Large Corpora (CMLC-3)* (IDS publication server), S. 28–34.
- Schäfer, Roland (2019): Prototype-driven alternations: The case of German weak nouns. In: *Corpus Linguistics and Linguistic Theory*, 15, 2, S. 383–417.
- Schäfer, Roland/Bildhauer, Felix (2012): Building large corpora from the web using a new efficient tool chain. In: Calzolari, Nicoletta et al. (Hgg.): *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: ELRA, S. 486–493.
- Seiler, Guido (2022): Wie viele Kasus hat das Deutsche? In: Brommer, Sarah/Roth, Kersten Sven/Spitzmüller, Jürgen (Hgg.): *Brückenschläge. Linguistik an den Schnittstellen*. Tübingen: Narr, S. 39–64.
- Sick, Bastian (2009): *Der Dativ ist dem Genitiv sein Tod. Folge 2. Neues aus dem Irrgarten der deutschen Sprache*. Köln: Kiepenheuer & Witsch.
- Szczepaniak, Renata (2014): Sprachwandel und sprachliche Unsicherheit: Der formale und funktionale Wandel des Genitivs seit dem Frühneuhochdeutschen. In: Plewnia, Albrecht (Hgg.): *Sprachverfall? Dynamik – Wandel – Variation*. Berlin/Boston: de Gruyter, S. 33–49.
- Thieroff, Rolf (2003): Die Bedienung des Automaten durch den Mensch. Deklination der schwachen Maskulina als Zweifelsfall. In: *Linguistik Online*, 16, 4, S. 105–117.
- Weber, Thilo /Hansen, Sandra (2024): Schwankungen zwischen schwacher und starker Substantivflexion. In: Konopka, Marek/Wöllstein, Angelika/Felder, Ekkehard (Hgg.): *Substantivflexion, Attributsätze, Präfix- und Partikelverben (Bausteine einer Korpusgrammatik des Deutschen 3)*. Heidelberg: Heidelberg University Publishing, S. 13–58.
- Weber, Thilo/Bildhauer, Felix/Münzberg, Franziska (2024): Finite vs. infinite Attributsätze: zu-dass-Alternation bei Substantiven. In: Konopka, Marek/Wöllstein, Angelika/Felder, Ekkehard (Hgg.): *Substantivflexion, Attributsätze, Präfix- und Partikelverben (Bausteine einer Korpusgrammatik des Deutschen 3)*. Heidelberg: Heidelberg University Publishing, S. 59–113.
- Ziem, Alexander (2018): Diskurslinguistik und (Berkeley) Construction Grammar. In: Warnke, Ingo H. (Hgg.): *Handbuch Diskurs*. Berlin/Boston: De Gruyter, S. 104–133.
- Ziem, Alexander/Scholz, Ronny /Romer, David (2012): Korpuslinguistische Zugänge zum öffentlichen Sprachgebrauch: spezifisches Vokabular, semantische Konstruktionen und syntaktische Muster in Diskursen über „Krisen“. In: Felder, Ekkehard (Hgg.): *Faktizitätsherstellung in Diskursen. Die Macht des Deklarativen*. Berlin/Boston: de Gruyter, S. 329–358.

