

**ERIC FUSS / MAREK KONOPKA / ANGELIKA WÖLLSTEIN**

## **EINLEITUNG**

Am Institut für Deutsche Sprache in Mannheim (IDS) wird an der Erstellung einer neuen, korpuslinguistisch fundierten Grammatik des Standarddeutschen gearbeitet (für entsprechende Präliminarien vgl. Bubenhofer/Konopka/Schneider 2014). Vorbereitend wurden eine Reihe von Pilotstudien durchgeführt, deren Ziel es war, anhand von detaillierten Untersuchungen beispielhafter Variationsphänomene aus verschiedenen Grammatikbereichen korpuslinguistische Fragestellungen und Herangehensweisen exemplarisch herauszuarbeiten. Dabei galt es auch, einen passenden Methodensatz zu entwickeln, der das methodologische Grundgerüst für die Unternehmung der eigentlichen „Korpusgrammatik“ bilden soll. Ein wesentlicher Teil der Erträge dieser Pilotphase ist in Donalies (2011), Konopka/Fuß (2016) sowie dem vorliegenden Band zusammengetragen. Daneben entstand eine Reihe von kleineren Pilotstudien. Alle Untersuchungen der Pilotphase führen zwei Aspekte linguistischer Forschung zusammen, die in der jüngsten Vergangenheit stärker in den Mittelpunkt des Erkenntnisinteresses gerückt sind. Dabei handelt es sich zum einen um das Bestreben, der modernen Linguistik ein besseres empirisches Fundament zu geben, wobei die Nutzung großer digitalisierter Textkorpora eine zunehmend wichtige Rolle spielt (vgl. z.B. Viana/Zyngier/Barnbrook (Hg.) 2011). Zum anderen ist hier das verstärkte Interesse der Grammatikforschung an den Formen, Bedingungen und Ursachen grammatischer Variation zu nennen, einem Phänomenbereich also, der traditionell primär Gegenstand der Soziolinguistik und Sprachwandelforschung ist. Im Rahmen dieses einführenden Kapitels möchten wir zunächst einige Perspektiven, Möglichkeiten und Herausforderungen eines korpusorientierten Zugangs zur Untersuchung grammatischer Variation aufzeigen. Im Weiteren sollen hier die Untersuchungen der Pilotphase vor dem Hintergrund des Projekts „Korpusgrammatik – grammatische Variation im standardsprachlichen und standardnahen Deutsch“, in dessen Rahmen sie entstanden sind, eingeordnet werden. Abschließend geben wir einen genaueren Überblick über die Inhalte des vorliegenden Bandes.

### **1. Die korpusorientierte Untersuchung grammatischer Variation**

Die Verfügbarkeit großer, elektronisch durchsuchbarer Korpora hat in den letzten Jahren zu einem neuen Interesse an empirischen Studien zur Grammatik natürlicher Sprachen geführt. Ein bedeutender Nebeneffekt der Beschäftigung mit größeren Mengen von Sprachdaten ist, dass die Variabilität der Sprache

stärker in den Blickwinkel der Grammatikforschung gerückt ist (vgl. Konopka 2018). Zwar ist die Beobachtung, dass die Sprache inhärent variabel ist, keineswegs neu; so erkannte bereits Pāṇini ca. 600 v. Chr., dass das grammatische System des Sanskrit in bestimmten Bereichen anscheinend gleichberechtigte Varianten zuließ (für solche Fälle benutzte er den Begriff *anyatarasyām*, vgl. Kiparsky 1979).<sup>1</sup> Allerdings blieb dieser Aspekt von Pāṇinis Lehren ohne größeren Einfluss auf die Ideengeschichte der Linguistik (vgl. Chambers 2002). Die Untersuchung sprachlicher Variation etablierte sich erst in der zweiten Hälfte des 20. Jahrhunderts als Teilbereich der modernen Linguistik. Ausgehend von der Beobachtung, dass abhängig vom Äußerungskontext oder gesellschaftlichen Status des Sprechers eine sog. linguistische Variable, d.h. eine bestimmte Bedeutung oder grammatische Funktion, durch verschiedene sprachliche Varianten realisiert werden kann, konzentrierte sich die wissenschaftliche Untersuchung sprachlicher Variation dabei zunächst vor allem auf die soziale Dimension von Sprache (vgl. Labov 1966, 2008).

Traditioneller Gegenstand der Soziolinguistik sind Fälle außergrammatisch determinierter Variation, die sowohl Sprechergruppen (sog. Intersprechervariation) als auch individuelle Sprecher (sog. Intrasprechervariation) betreffen kann. Variationsphänomene, die sich auf grammatische Faktoren zurückführen lassen, fallen hingegen in das Gebiet der allgemeinen Linguistik bzw. der Grammatikforschung. Auch hier kann zwischen Intra- und Intersprechervariation unterschieden werden. Kennzeichnend für die meisten Variationsphänomene ist es dabei generell, dass wir Faktoren identifizieren können, die zwar die Verteilung der Varianten beeinflussen, sie aber nicht vollständig bestimmen. Es sind in der Regel auch mehrere Faktoren im Spiel. Bei der Untersuchung solcher Einflussgrößen können tendenziell verschiedene Typen sprachlicher Variation unterschieden werden, die sich anhand der Verteilung der Varianten und der Natur der relevanten Faktoren (grammatisch vs. außergrammatisch) klassifizieren lassen, vgl. die Übersicht in Tabelle 1.

---

<sup>1</sup> Kiparsky (1979) weist darauf hin, dass Pāṇini zur Beschreibung grammatischer Variation drei unterschiedliche Begriffe verwendet, die mit verschiedenen Stufen der Präferenz für eine Variante assoziiert sind: *vā* bringt zum Ausdruck, dass es sich um eine eindeutig bevorzugte Variante handelt; *vibāṣi* wird für marginale, weniger präferierte Varianten verwendet, während *anyatarasyām* Fälle beschreibt, in denen die beiden Varianten mehr oder weniger gleichwertig sind.

	<b>grammatisch bestimmt</b>	<b>außergrammatisch bestimmt</b>
Intersprechervariation	abhängig von phonologischem/morpho-syntaktischem Kontext, Satzprosodie, Diskursfunktion etc.	abhängig von Alter, sprachgeografischem Hintergrund, Sozialstatus etc.
Intrasprechervariation		abhängig von Äußerungskontext, Stilebene, Textsorte, Medium o.Ä.

Tab. 1: Typen sprachlicher Variation (vgl. Fuß 2018)

Die typischen außergrammatischen Einflussgrößen gehen bei der Inter- und der Intrasprechervariation deutlich auseinander: Während sie im ersteren Fall primär als Eigenschaften des Sprechers erscheinen, betreffen sie im letzteren Fall Aspekte der Kommunikationssituation. Aber auch die grammatischen Einflussgrößen sind bei den beiden Variationsarten unterschiedlich zu interpretieren: Während sie bei der Intrasprechervariation dafür stehen, dass ein Sprecher je nach grammatischem Kontext zu der einen oder zu der anderen Variante neigt, bedeutet ihre Feststellung in einem Fall der Intersprechervariation erst einmal nur, dass die Sprecher, die die eine Variante verwenden, auch häufiger zu bestimmten grammatischen Konfigurationen im Kontext der Variable greifen als die Sprecher, die die andere Variante verwenden.<sup>2</sup> Dabei ist die Entscheidung, ob es sich um eine Inter- und/oder eine Intrasprechervariation handelt, in maschinell gestützten Untersuchungen sehr umfangreicher Korpora oft gar nicht trivial. Als Indikator für die eine oder die andere Variationsart kann allerdings die Beschaffenheit der außergrammatischen Faktoren dienen, worauf wir weiter unten noch genauer eingehen werden.

Obige Klassifizierung verschiedener Typen sprachlicher Variation berücksichtigt zunächst nur Phänomene, bei denen die Wahl zwischen zwei oder mehreren Ausprägungen einer Variable von bestimmten, klar identifizierbaren Faktoren beeinflusst wird (dazu Fuß/Konopka/Wöllstein 2017). Daher kann man hier von „abhängiger Variation“ sprechen. Darüber hinaus wird in der Literatur die Möglichkeit freier Variation diskutiert. Freie Variation soll dann vorliegen, wenn die Verteilung von Varianten nicht auf bestimmte grammatische oder außergrammatische Einflussgrößen zurückgeführt werden kann. Eine solche Form sprachlicher Variation wird in der Regel mit Sprachwandelprozessen assoziiert (vgl. z.B. Kroch 1989). Dabei liegt die Idee zugrunde, dass Sprachwandel nur dann möglich ist, wenn die Distribution der Varianten so

<sup>2</sup> Ob dann die Annahme disparater grammatischer Systeme notwendig ist, sei an dieser Stelle dahingestellt.

unspezifisch bzw. zufällig geworden ist, dass Sprachlerner nicht mehr in der Lage sind, die Faktoren zu erkennen, die in der Zielgrammatik die Verteilung der entsprechenden Varianten regeln.

Im Gegensatz zu früheren, primär kompetenzgestützten Ansätzen in der Grammatikschreibung begreifen wir Variabilität, wie bereits angedeutet, als wesentliche Eigenschaft der Sprache, die sich nicht nur in der Vielfältigkeit des allgemeinen Sprachgebrauchs manifestiert, sondern die auch in der sprachlichen Kompetenz individueller Sprecher verankert ist (Stichwort „Intrasprechervariation“). Daraus folgt zum einen, dass eine adäquate Beschreibung grammatischer Variation nicht nur außergrammatische Aspekte der Distribution einer bestimmten Variante (z.B. ihre Abhängigkeit vom sprachgeografischen Hintergrund oder von der Textsorte) erfassen sollte, sondern auch – soweit vorhanden – die grammatischen Bedingungen herausarbeiten muss, die die Realisierung unterschiedlicher Ausprägungen der relevanten Variable beeinflussen. Eine Gesamtbeschreibung der Grammatik einer Einzelsprache soll ferner, wann immer möglich, Inter- und Intrasprechervariation auseinanderhalten und Letztere zu einem Schwerpunkt der Betrachtung machen. Dementsprechend befassen sich die Beiträge zu diesem Sammelband mit Phänomenen, in denen nicht nur der allgemeine Sprachgebrauch variabel ist, sondern grammatische Variation auch auf der Ebene des einzelnen Sprechers zu erwarten ist.

Die Identifikation von genuin grammatischen Einflussfaktoren, die als Eigenschaften des lautlichen oder morphosyntaktischen Kontexts erscheinen, kann im Rahmen eines korpuslinguistischen Zugangs sehr arbeitsintensiv sein, schon weil sie mit einem erheblichen Annotationsaufwand verbunden ist. Ein wesentlicher Beitrag dieses Bandes besteht darin, Wege aufzuzeigen, wie derartigen Herausforderungen mit neuen korpuslinguistischen und statistischen Methoden begegnet wird (vgl. auch Brandt/Fuß 2014; Bubenhofer/Hansen-Morath/Konopka 2014; Konopka/Fuß 2016; Fuß/Konopka/Wöllstein 2017). Im Mittelpunkt stehen Aspekte der morphosyntaktischen Variation, die am IDS in entsprechend annotierten Korpora (auf der Basis des Deutschen Referenzkorpus, *DeReKo*<sup>3</sup>) untersucht werden. Die traditionelle monoperspektivische Auffassung, nach der linguistisch interessante Variation dann vorliegt, wenn verschiedene Zeichen oder Zeichenketten zur Realisierung einer bestimmten semantischen oder grammatischen Funktion herangezogen werden können, wird dabei um eine inverse Perspektive erweitert, die von der Form ausgeht und betrachtet, welche unterschiedlichen Funktionen mit einem Zeichen bzw. einer Zeichenkette assoziiert sein können. Die „Variation der Funktion“ eines grammatischen Gebildes wurde in der bisherigen, weitgehend an der Soziolinguistik orientierten Variationsforschung kaum als linguistisch interessant wahr-

---

<sup>3</sup> Vgl. <http://ids-mannheim.de/kl/projekte/korpora.html> (Stand: 5.3.2018).

genommen bzw. bewusst fokussiert, wozu entscheidend Labovs (1966) frühe Einschränkung des Begriffs „linguistische Variable“ auf funktional definierte Größen mit unterschiedlichen formalen Ausprägungen beigetragen haben mag. Dennoch ist diese Perspektive bereits in den üblichen Forschungspraktiken angelegt. Jetzt soll der wiederkehrende, bewusste Perspektivenwechsel operationalisiert werden, und zwar als systematisch anzuwendende Methode, die die Erkundung der Grammatik weiter voranbringen kann (dazu Fuß/Konopka/Wöllstein 2017, S. 229–235). Die von uns hier vertretene Doppelperspektivik – von der Form zur Funktion und von der Funktion zur Form – setzt im Übrigen eine Tradition der Grammatikschreibung am IDS fort, die mit der IDS-Grammatik von Zifonun/Hoffmann/Strecker (1997, S. 7–8) begründet wurde.<sup>4</sup>

Die Unterscheidung zwischen Intersprechervariation und Intrasprechervariation, die neben der Ermittlung grammatischer Einflussfaktoren als wichtiger Aspekt der adäquaten Beschreibung grammatischer Variation hervorgehoben wurde, ist methodologisch schwierig. Tabelle 1 weiter oben könnte so verstanden werden, dass es bei außergrammatischen Einflussgrößen nur wenig Berührungspunkte zwischen den beiden Variationstypen gibt: Varianten, die mit Faktoren wie Alter, sozialer Hintergrund und regionale Herkunft assoziiert sind, sind typisch für verschiedene Sprechergruppen, während die Wahl unterschiedlicher Ausdrucksweisen im Zusammenhang mit dem außersprachlichen Kontext einer Äußerung ein Charakteristikum der Sprachproduktion einzelner Sprecher ist.<sup>5</sup> Im Gegensatz dazu scheinen bei grammatischen Faktoren die jeweils beteiligten Einflussgrößen an sich nichts über die Unterscheidung zwischen Intersprechervariation und Intrasprechervariation auszusagen: Ist bei einer Korpusuntersuchung die Verteilung einer Variante von einem bestimmten grammatischen Faktor wie der phonologischen Umgebung abhängig, so kann die daraus resultierende Variation mit verschiedenen Sprechergruppen assoziiert sein wie auch in der Sprachproduktion eines einzelnen Individuums auftreten. Eigentlich ist die Situation aber noch wesentlich komplizierter, weil die für verschiedene Sprechergruppen typischen Varianten ebenfalls in der Sprachproduktion eines einzelnen Individuums (nicht zuletzt eines sprachlich versierten) auftreten können, das über das Wissen um die Sprechergruppenabhängigkeit einer Variante verfügt und dieses bewusst einsetzt etwa als Autor eines literarischen Werkes. Die Begriffe Intersprechervariation und Intrasprechervariation stehen also nur für eine prototypische Zuordnung. Dennoch ist es für die grammatische Beschreibung und theoretische

<sup>4</sup> Weitergehende Problemstellungen betreffen mögliche (systematische) Zusammenhänge zwischen Form und Funktion (vgl. z.B. Wurzel 1984; Wiese 1999; Brandt 2016) wie die Frage, wann (und warum) eine bestimmte Form überhaupt zur Realisierung einer bestimmten Funktion genutzt werden kann.

<sup>5</sup> Allerdings bedeutet das natürlich nicht, dass es keine Unterschiede zwischen Sprechergruppen hinsichtlich stilebenenabhängiger Varianten geben kann.

Analyse der empirischen Befunde von Bedeutung, dass zwischen den beiden Variationsarten unterschieden wird. Sonst besteht die Gefahr, dass man ein Beschreibungsmodell entwirft, das zwar die Verhältnisse im Korpus korrekt erfasst, aber dennoch an der sprachlichen Realität vorbeigeht, da es suggeriert, dass einzelne Sprecher über die ganze Bandbreite an Varianten (und Einflussgrößen) verfügen, die im Korpus auftreten.<sup>6</sup> Insbesondere bei der Untersuchung großer Korpora, bei denen sich die Datengrundlage aus den sprachlichen Produkten einer Vielzahl verschiedener Sprecher zusammensetzt, wird die Unterscheidung wichtig – sie wirft aber gerade dann besondere Probleme auf. Diesen kann begegnet werden, indem man zunächst solche Einflussgrößen isoliert, die außergrammatischer Natur sind und relativ eindeutig verschiedenen Gruppen von Sprechern (= Intersprechervariation) zugeordnet werden können (z.B. über Einflussfaktoren wie Land oder Region).<sup>7</sup> Für die Hypothese, dass ein bestimmtes Variationsphänomen einen Fall von Intrasprechervariation darstellt, können dann zum einen einschlägige Befunde aus der Forschungsliteratur herangezogen werden; zum anderen kann auch Introspektion als heuristisches Mittel zur Aufdeckung von Intrasprechervariation dienen. In jedem Fall sollten entsprechende Vermutungen durch unabhängige Korpusstudien<sup>8</sup> und andere empirische Untersuchungen validiert werden. Allerdings liegt es in der Natur der Dinge, dass eine Beschreibung grammatischer Variation, die sich an großen Korpora orientiert, insgesamt zwischen einer Beschreibung des im Korpus erfassten allgemeinen Sprachgebrauchs und einer Beschreibung des grammatischen Wissens der Sprachteilhaber oszilliert. Sie soll es aber wann immer möglich anstreben, die beiden Variationsarten auseinanderzuhalten, wie es auch im vorliegenden Band versucht wird.

Zuletzt muss noch eine Einschränkung genannt werden, die sich unmittelbar aus der Datengrundlage aller unserer Untersuchungen ergibt. Sie betrifft die Tatsache, dass die in DEREKO enthaltenen, geschriebenen Texte nur einen Teil der Variationsbandbreite zeigen, die die standardnahen Varietäten des Deutschen insgesamt aufweisen. Um ansatzweise Eigenschaften der standardnahen Mündlichkeit in den Blick zu bekommen, kann der Umweg über die Unter-

<sup>6</sup> Umgekehrt könnte es auch sein, dass das Sprachverhalten und die Grammatik einzelner Sprecher invariant sind und jede Variante einer bestimmten Gruppe von Sprechern zuzuordnen ist. In diesem Fall würde eine Instanz von Intersprechervariation vorliegen.

<sup>7</sup> Bei der Zuordnung von Varianten zu Sprechern bzw. Sprechergruppen sind verschiedene Fälle zu unterscheiden. Zum einen kann es sein, dass eine bestimmte Variante nur in einer bestimmten Varietät/Sprechergruppe auftritt, wie es im Fall von *parkieren* für das Schweizerdeutsche gilt (Faktor: „Land“). Zum anderen kann die Sprechergruppe nur einer von mehreren Faktoren sein, der die Wahl einer Variante günstig beeinflusst (vgl. Brandt/Fuß in diesem Band zur Verteilung des Relativums *das* nach dem Quantor *alles*).

<sup>8</sup> Es könnte beispielsweise überprüft werden, ob die betreffenden Varianten in ein und demselben Text bzw. Satz auftreten.

suchung von Textsorten genommen werden, die der konzeptuellen Mündlichkeit nahestehen und in denen stärker mit Abweichungen von der Sprachnorm zu rechnen ist (v.a. Internettexpte wie Foren, Chats, Blogs, Wikipedia-Diskussionen, vgl. z.B. Rabanus 1996 für eine frühe Bestandsaufnahme). Eine Einschätzung der Einflussgröße 'Nähe zur Mündlichkeit' kann dann geleistet werden, wenn entsprechende Befunde mit Ergebnissen aus Texten verglichen werden, die eindeutig als konzeptionell schriftlich einzuschätzen sind (vgl. dazu Brandt/Fuß, dieser Band; Fuß 2018 für eine Untersuchung von Kongruenzalternationen mit Subjektreihungen, die primär in Internettexpten beobachtet werden können).

## 2. Die bisherigen Untersuchungen

Zunächst sollen hier die grammatischen Ergebnisse bisheriger Pilotuntersuchungen skizziert werden. Danach werden die Beiträge zu theoretischen Fragestellungen umrissen. Schließlich werden die bisherigen Ergebnisse im Bereich der korpuslinguistischen Methodik zusammengefasst. Die erste größere Pilotstudie entstand im Bereich **Wortbildung** und wurde von Donalies (2011) verfasst. Sie untersuchte variierende Wortformen und Fugenelemente von Nomina wie *Tagtraum*, *Tageslicht*, *Tagedieb*. Donalies prüfte das Zusammenwirken von fünf konkurrierenden Regeln für die Wahl der Formen (der Paradigmenregel, der Bedeutungsregel, der Euphonieregel, der Strukturierungsregel und der Gewohnheitsregel). Sie zeigte die Sprachgebrauchstendenzen auf und gab einen fundierten Überblick über die große Materialfülle. Die Problematik der Fugenelemente in nominalen Komposita wurde von Bubenhofer/Hein/Brinckmann (2014) wiederaufgenommen, die mithilfe eines Verfahrens des maschinellen Lernens einen Entscheidungsbaum generierten, der die Fugenelemente in 400.000 Komposita mit einer Trefferquote von 95% vorhersagt. Bei 75% der Komposita konnte die Fugewahl auch mithilfe grammatischer Regeln aus der Forschungsliteratur erklärt werden. Die linguistische Interpretation des Entscheidungsbaums zeigte u.a., dass die Wahl des Fugenelements grundsätzlich vom Charakter des Erstgliedes der Komposition abhängt (wobei zwischen konsonantisch und vokalisch auslautenden Erstgliedern unterschieden werden muss) und dass das Flexionsparadigma des Erstglieds ein sehr 'mächtiges' Entscheidungskriterium darstellt. Der Bereich Wortbildung wurde auch in einigen anderen Untersuchungen relevant, etwa bei Konopka und Waßner (2013), die die Frequenz und Varianz der negativ-konditionalen Konnektoren *sonst*, *ansonsten*, *ansonst*, *sonsten* untersuchten. Die Wahl eines Konnektors aus dieser Gruppe erwies sich einerseits als von den außergrammatischen Faktoren Zeit, Raum und Register (Textsorte) abhängig. Andererseits zeigten einzelne Konnektoren besondere Affinitäten zu unterschiedlichen Verwendungen, so z.B. *sonst* zu 'Adverbkonnektor im Mittelfeld', *ansonsten* zu 'Adverbkonnektor

im Vorfeld' und *ansonst* zu 'phraseninternem Adverbkonnektor'. Variation bei Konnektoren wird schließlich auch von Waßner (2018) thematisiert, der Lücken in Wortbildungsparadigmen konzessiver Konnektoren (*obwohl* – ?*wennwohl* – *wiewohl*, *obgleich* – *wengleich* – ?*wiegleich* etc.) und Bildungen auf *-falls* (z.B. *ander(e)nfalls*) untersucht.

**Flexionsmorphologische Variation** und ihre möglichen Ursachen wurden detailliert am Beispiel der Variation der starken Genitivmarkierung des Nomens von Konopka und Fuß (2016) untersucht. In der großen Doppelstudie wurde einerseits die Alternation zwischen verschiedenen Endungen wie in *Hauses*, *Vaters* oder *Entwurfes/Entwurfs* und andererseits die Variation zwischen dem Vorhandensein und dem Fehlen einer Genitivmarkierung wie in (*des*) *Iraks/Irak* analysiert. Dabei wurden die zahlreichen lexikalischen, phonologischen, morphologischen, syntaktischen und außersprachlichen Einflussgrößen, die von der Forschung postuliert worden waren, geprüft und hierarchisiert. Die Einzelergebnisse wurden abschließend in einem Gesamtbild auf einen Hintergrund projiziert, in dem als wichtigste Triebfedern der Variation – aus diachroner Perspektive – die Dynamik der wortphonologischen Reduktion im Grundwortschatz (*-es* → *-s*) und der flexionsmorphologischen Anpassung im Sonderwortschatz<sup>9</sup> (Nullmarkierung → *-s*) und – aus synchroner Perspektive – der variierende Integrationsgrad des Sonderwortschatzes erscheinen. Die Entstehung der Studie wurde von einigen Publikationen begleitet, die Sonderfragen vertieften oder anfallende methodische Probleme erörterten. So wurde in einer frühen Untersuchungsphase auch auf die Variation der Genitivmarkierung ein Verfahren des maschinellen Lernens angewendet, das eine Grundlage für die Berechnung eines Entscheidungsbaums bildete, der die Markierungsform vorhersagte. Dabei zeigten die Analysen von Bubenhofer/Hansen-Morath/Konopka (2014), dass die Markierungsvariation durch hohe Systematizität gekennzeichnet ist. Die Visualisierung des Entscheidungsbaums machte besonders auf die Faktoren Lexemhäufigkeit, Silbenanzahl, Sonderwortschatzzugehörigkeit und Wortausgang sowie auf die außergrammatischen Einflussgrößen Zeit und Raum aufmerksam. Die Datenbasis, die allen Untersuchungen zur Genitivmarkierung zugrunde lag, wurde der Allgemeinheit in Form einer Genitivdatenbank (*GenitivDB*)<sup>10</sup> zur Verfügung gestellt (Bubenhofer/Hansen-Morath/Konopka/Schneider 2014, 2015). Die Datenbank umfasst ca. sieben Millionen Satzbelege mit Genitivnomina, die vielfach annotiert und nach Metadaten abfragbar sind. Die Entstehung, Beschreibung und Evaluation dieser Datenbank sind Thema in Hansen/Schneider (2013) und Schneider (2014).

<sup>9</sup> Fremdwörter, Fachwörter, Eigennamen etc.

<sup>10</sup> <https://grammis.ids-mannheim.de/genitivdb> (Stand: 26.2.2018).

**Morphosyntaktische Fragestellungen** sind Gegenstand der Studien in diesem Band (die im Abschnitt 3 umrissen werden) sowie einer ganzen Reihe von Beiträgen, die Zwischenetappen dokumentiert haben. So wurden der Gebrauch und das syntaktische Verhalten der konditionalen Konnektoren *angenommen* und *vorausgesetzt*, die oft als Varianten betrachtet werden, bereits von Volodina (2018) thematisiert. Sie stellt durch einen Wechsel von der Funktion-Form-Perspektive zu der Form-Funktion-Perspektive entscheidende Unterschiede zwischen den beiden Konnektoren fest, die genau betrachtet gegen deren Auffassung als Varianten sprechen. Die Variation zwischen den Relativpronomina *das* und *was* wurde zunächst von Brandt und Fuß (2014) korpuslinguistisch gesichtet, die die zentralen grammatischen Bedingungen der Variation aufarbeiteten und bereits die These aufstellten, dass *was* den Normalfall darstellt und für alle unterspezifizierten Verwendungen infrage kommt, die keine Identifikation mit einem nominalen Antezedens benötigen, während *das* auf diese prinzipiell angewiesen ist. Fuß (2017) und Brandt/Fuß (2017) zeigten dann, dass bei nominalisierten Adjektiven Positive sowohl *das* als auch *was* erlauben (*das Gute, das/was*), was zum Ausdruck subtiler semantischer Unterschiede genutzt werden kann, wohingegen Superlative *was* präferieren (*das Beste, was*), was, wie die Autoren argumentierten, damit zusammenhängt, dass der Quantifizierer *alles* quasi einen Teil der Superlativstruktur bildet. Die erste Phase der Untersuchungen zur Alternation zwischen *dass*-Komplementsätzen und gleichwertigen Infinitivkonstruktionen mit *zu*<sup>11</sup> wurde von Wöllstein (2015) beschrieben. Sie prüfte mögliche Einflussfaktoren der Variation und brachte u.a. empirische Evidenz dafür, dass die Eindeutigkeit hinsichtlich der Interpretation des (impliziten) Subjekts im Komplement das Auftreten der Infinitivkonstruktionen fördert. In einer komparativen Studie verglichen dann Brandt, Trawiński und Wöllstein (2016) Infinitivkonstruktionen in Antikontrollstrukturen<sup>12</sup> im Deutschen mit entsprechenden Konstruktionen im Polnischen, die mit einem overt Komplementierer *żeby* markiert sind. Sie diskutierten dabei die Frage, ob die Konstruktionen als Satzkomplemente zu analysieren sind.

Neben den auch in diesem Band behandelten Variationsphänomenen bildeten sich Probleme der Subjekt-Verb-Kongruenz bei Koordinationsstrukturen zu einem Schwerpunkt der Untersuchungen heraus. Münzberg und Hansen-Morath (2018) behandelten die Variation des Numerus des finiten Verbs bei zwei mit *und* koordinierten singularischen Nomina als Subjekt. Sie zeigten, dass in Sätzen, in denen beide Nomina mit einem Artikel erscheinen, der Plural signifikant überrepräsentiert ist und in Sätzen, in denen das zweite Nomen

<sup>11</sup> Z.B. *Alex leugnet/erzählt/gibt (es) zu, dass er am Marmeladenglas genascht hat/am Marmeladenglas genascht zu haben.*

<sup>12</sup> *dass*-Sätze bzw. Infinitivkonstruktionen bei sog. obviativen Verben wie *anordnen, billigen, gutheißen.*

ohne Artikel erscheint, der Singular überrepräsentiert ist (z.B.: *Die Wucht und Strömung war immens*). Dieser sog. Ellipseneffekt ist stärker als der Einfluss eines hohen Abstraktionsgrades der Nomina, scheint sich allerdings in schweizerischen Texten abzuschwächen, wodurch ein außergrammatischer Parameter ins Spiel kommt, der tendenziell für Intersprechervariation spricht. Fuß (2018) beschäftigt sich wiederum mit Kongruenz bei koordinierten Pronomina. Er zeigt, dass bei Subjekttrihungen aus 2. und 3. Person Sg. (z.B. *du und er*) am Verb die Endung der 2. Person Sg. (-st) in der Regel nur dann auftritt, wenn das Verb dem komplexen Subjekt vorangeht, während die Benutzung der Endungen der 2. oder der 3. Person Plural von mehreren Faktoren beeinflusst wird. Insgesamt scheint es sich um einen genuinen Fall von Intrasprechervariation zu handeln. Im Weiteren wurden auch AcI-Konstruktionen untersucht und mit V2-, *dass*- und *wie*-Sätzen verglichen, die üblicherweise als Varianten der AcI-Konstruktionen erwogen werden. Fuß, Konopka und Wöllstein (2017) fanden dabei Hinweise darauf, dass die Häufigkeit der AcI-Konstruktionen textsorten- bzw. registerabhängig ist und zeitlich gesehen insgesamt abnimmt. In theoretischer Hinsicht postulierten sie, dass als Varianten nur AcI-Konstruktionen<sup>13</sup> und *wie*-Sätze betrachtet werden können. Schließlich lieferten Konopka und Hansen-Morath (2017) einen statistischen Vergleich der beiden Konstruktionen, der nahelegt, dass sie ähnliche außergrammatische Gebrauchsbedingungen haben.

**Theoretische Fragestellungen** der korpuslinguistischen Herangehensweise wurden ausführlich zunächst in Bubenhofer/Konopka/Schneider (2014) behandelt. Die Autoren diskutierten darin u.a. verschiedene Verwendungen des Begriffs 'Standardsprache', Konzepte wie 'Variation' und 'Varianz', verschiedene Betrachtungsweisen der Frequenz von Phänomenen, Fragen der Zusammensetzung eines Korpus für die Untersuchungen zu einer wissenschaftlichen Grammatik des Deutschen und die Relevanz von Metadaten, um außergrammatische Bedingungen für den Gebrauch der Phänomene zu untersuchen. Später thematisierten Fuß, Konopka und Wöllstein (2017) detailliert den Ansatz der Doppelperspektivik in der deskriptiven Grammatikforschung. Sie postulierten gleichzeitig die Notwendigkeit der heuristischen Suspendierung des Konzepts der „freien Variation“ und führten die Unterscheidung zwischen Intersprechervariation und Intrasprechervariation ein, die dann von Fuß (2018) ausführlich thematisiert wurde. Zuletzt analysierte Konopka (2018) das Zusammenwirken von Korpuslinguistik, Grammatiktheorie und Grammatikografie. Er argumentierte im Ergebnis, dass die empirische Wende in der Grammatikforschung eine wissenschaftliche Grammatik des Deutschen überfällig

<sup>13</sup> Sofern Letztere als propositionsbezeichnende Komplemente aufgefasst werden, die auf direkte Perception bzw. zeitliche Überlappung der bezeichneten Proposition mit der durch das Wahrnehmungsverb im Matrixsatz bezeichneten Handlung beschränkt sind.

macht, die zu einer Theoriebildung beiträgt, die kompetenz- und performanzbezogene Überlegungen zu einer integrativen Gesamtheorie weiterentwickelt.

Die genannten theoretischen Gesichtspunkte fanden ihren Niederschlag in Einzeluntersuchungen, die von den grammatischen Inhalten her teilweise schon weiter oben umrissen wurden. So etwa setzten sich Konopka und Waßner (2013) mit den Möglichkeiten auseinander, der Argumentation für die Zugehörigkeit einer Variante zum Standard, einem Teilstandard<sup>14</sup> oder zu einer Nonstandardvarietät korpuslinguistisch-statistische Anhaltspunkte zu geben. In beinahe allen Untersuchungen seit 2013 wurde die Abhängigkeit der Variationsphänomene von in Bubenhofer/Konopka/Schneider (2014) theoretisch diskutierten Metadaten geprüft, was etwa bei Konopka/Waßner (2013), Bubenhofer/Hansen-Morath/Konopka (2014) oder Fuß/Konopka/Wöllstein (2017) zu einigen überraschenden vor allem sprachhistorischen und sprachgeografischen Zuordnungen führte und des Öfteren einen Hinweis auf *Intersprechervariation* gab. So war auch die übliche Perspektive der Variationsbetrachtung, von der Funktion zur Form, regelmäßig um den Blick ergänzt worden, der von den Formen ausgeht und ihre funktionale Ausdifferenzierung betrachtet. Dies resultierte häufig in der Feststellung, dass die ursprünglich als Varianten betrachteten Phänomene nach einer eingehenden Analyse nicht mehr als solche zu betrachten sind. Auf eine integrative Betrachtung der kompetenz- und performanzbezogenen Aspekte der Variation schließlich wurde laufend hingearbeitet, indem systematisch sowohl grammatische als auch außergrammatische Faktoren der grammatischen Variation erkundet und in Gesamtbildern (exemplarisch in Konopka/Fuß 2016, S. 244–264) aufeinander bezogen wurden.

**Korpuslinguistisch** gesehen sind im Projekt „Korpusgrammatik“ zahlreiche Methoden ausprobiert worden, wovon einige jetzt auch zum methodologischen Grundgerüst der Unternehmung zählen dürfen. Zu Beginn des Projekts ging es zunächst um die Zusammenstellung eines Untersuchungskorpus, die Konzeption der Metadaten-Anreicherung der Korpus-texte aus den TreeTagger-, Connexor- und früher auch Xerox-annotierten Ressourcen des DEREKo sowie um die Entwicklung eines ersten adäquaten Abfrage- und Analyseportals. Alle diese Schritte wurden in Bubenhofer/Konopka/Schneider (2014) beschrieben.<sup>15</sup> Bubenhofer diskutierte darin auch schon die grundlegenden statistischen Verfahren, die dazu dienen, die Verteilung eines Phänomens in den definierten Teilkorpora zu untersuchen und ein Phänomen mit einem oder mehreren anderen Phänomenen zu vergleichen, wie Chi-Quadrat-Tests, Effektstärke-Messungen und Berechnungen der Dispersion von Phänomenen. Danach wurde in Bubenhofer/Konopka/Schneider (2014), Bubenhofer/Hansen-Morath/Konopka (2014)

<sup>14</sup> Z.B. im Sinne eines areal definierten Gebrauchsstandards (vgl. Elspaß/Dürscheid 2017).

<sup>15</sup> Zu aktualisierten Informationen zur Zusammensetzung des Untersuchungskorpus vgl. <https://grammis.ids-mannheim.de/korpusgrammatik/4755> (Stand: 4.3.2018).

und Konopka/Fuß (2016) eine automatische Abfrageroutine für die phonologischen und morphologischen Informationen der Celex-Datenbanken entwickelt. Diese werden bei Korpusrecherchen herangezogen, die grammatische Merkmale betreffen, die in den Korpus-texten nicht annotiert sind. Im Weiteren wurden auch avancierte korpuslinguistische bzw. statistische Methoden benutzt und in Einzelstudien beschrieben, so die Modellierung von Entscheidungsbäumen mithilfe maschinellen Lernens in Bubenhofer/Konopka/Schneider (2014) und Bubenhofer/Hansen-Morath/Konopka (2014), Clusteranalysen in Brandt/Fuß (2017) und logistische Regressionen in Konopka/Fuß (2016) sowie Münzberg/Hansen-Morath (2018). Ein wichtiges Ergebnis – nicht zuletzt im Hinblick auf die Durchführbarkeit des Projekts „Korpusgrammatik“ – war letztendlich der Entschluss, ein Paket an grundlegenden statistischen Methoden zu standardisieren und deren routinemäßige und (soweit nur möglich) automatische Durchführung zu sichern (vgl. weiter unten). Aufwendigere statistische Verfahren, die nicht nur mehr Vorbereitung und Zeitaufwand benötigen, sondern bei denen auch die Interpretation der Ergebnisse mit diffizilen und umfangreichen Erklärungen verbunden ist, sollten dann nur in theoretisch besonders interessanten Fragen hinzukommen.

### **3. Die Pilotstudien in diesem Band**

Bei den in diesem Band versammelten Arbeiten handelt es sich um Pilotstudien, die anhand ausgewählter, variationsträchtiger morphosyntaktischer Phänomene exemplarisch aufzeigen, wie ein korpuslinguistischer Zugang genutzt werden kann, um die Vielfalt und Variabilität des Sprachgebrauchs in standardnahen Texten aufzuzeigen und in einer größeren Detailschärfe zu beschreiben, als dies bisher möglich war. Die Erkenntnisse der Studien sollen in die zu verfassende Grammatik einfließen, deren wesentliche Ziele und Gegenstände auch im Mittelpunkt der hier vorgelegten Arbeiten stehen. Ausgangspunkt ist die Überlegung, dass sprachliche Variation im Gegensatz zu früheren normativen Auffassungen als integraler Bestandteil der Sprache (insbesondere auch der Standardsprache) anzusehen ist und somit auch grammatikografisch erfasst werden muss. Dabei geht es zunächst um eine möglichst genaue Beschreibung der Verteilung und Häufigkeit der verschiedenen Ausprägungen ausgewählter Variablen. Eine umfassende Beschreibung eines Variationsphänomens beinhaltet zudem die Ermittlung und Gewichtung der Faktoren, die die Distribution der Varianten steuern. In diesem Zusammenhang werden auch existierende Hypothesen aus der einschlägigen Forschungsliteratur unter Verwendung moderner statistischer Verfahren überprüft und evaluiert. Neben der Beschreibung und Evaluation beinhalten die vorliegenden Studien eine explorative Komponente, die sich mit der Aufdeckung neuer Muster, Regularitäten und linguistischer Zusammenhänge befasst. Aus methodologischer Sicht liefert

der vorliegende Band ebenfalls wichtige Vorarbeiten zur entstehenden Korpusgrammatik, indem verschiedene korpuslinguistische und statistische Ansätze und Verfahren erprobt und evaluiert werden; besonders hervorzuheben ist dabei, dass innerhalb der Projektarbeit eine Auswahl standardisierter statistischer Instrumente entwickelt wurde, die als KoGra-R auch online<sup>16</sup> zur Verfügung stehen und zur Datenexploration und Validierung linguistischer Hypothesen herangezogen werden können (vgl. das Kapitel von Hansen-Morath/Schmitz/Schneider/Wolfer).

Das Kapitel von Anna Volodina („Variation im Sprachgebrauch – *angenommen* und *vorausgesetzt* als einbettende Prädikatsausdrücke“) ist befasst mit dem Gebrauch und syntaktischen Verhalten der konditionalen Konnektoren *angenommen* und *vorausgesetzt*, die aus den formidentischen Perfektpartizipien hervorgegangen sind. Während bisherige Beschreibungen in der Regel davon ausgegangen sind, dass diese Konnektoren mehr oder weniger identische semantische und syntaktische Eigenschaften aufweisen (vgl. Pasch et al. 2003 und Breindl/Volodina/Waßner 2014), wird in der vorliegenden Arbeit gezeigt, dass sich auf der Basis einer umfassenden Korpusstudie deutliche Unterschiede im Gebrauch der beiden Elemente nachweisen lassen. Dabei wurden aus dem Korpus ca. 40.000 einschlägige Belege extrahiert, die anschließend semi-automatisch für Eigenschaften wie Wortstellung, Modus und unmittelbare syntaktische Umgebung annotiert wurden. Mithilfe der statistischen Werkzeuge und Visualisierungsmittel, die von der KoGra-R-Schnittstelle zur Verfügung gestellt werden, konnte eine Reihe von Kontrasten zwischen den beiden Konnektoren herausgearbeitet werden. So unterscheiden sich *angenommen* und *vorausgesetzt* hinsichtlich a) ihrer Präferenz für die Einbettung von V2- vs. *dass*-Komplementsätzen, b) Modus (Indikativ vs. Konjunktiv) und Position des eingebetteten Satzes relativ zum Matrixsatz sowie c) der Auswahl der (adverbialen) Elemente, die bevorzugt zur Modifikation des Konnektors herangezogen werden. Die beobachteten Kontraste wurden anschließend im Rahmen eines pragmatisch-funktionalen Ansatzes gedeutet, der auf der Idee aufbaut, dass die beiden Konnektoren die Aufmerksamkeit des Adressaten auf verschiedene Weise lenken: Mit dem *angenommen*-Konnekt wird in der Regel eine neue, hypothetische Situation eingeführt, in der der Matrixsatz evaluiert wird. Dabei wird die Aufmerksamkeit des Adressaten eher auf die Möglichkeit gerichtet, dass die im *angenommen*-Konnekt ausgedrückte Proposition wahr ist. Im Gegensatz dazu schränkt *vorausgesetzt* den Geltungsbereich des Matrixsatzes ein, wobei die Möglichkeit, dass der Sachverhalt des *vorausgesetzt*-Konnekts falsch sein könnte, in den Mittelpunkt rückt. Eine wichtige Konsequenz dieser Analyse ist, dass *angenommen* und *vorausgesetzt* entgegen bisherigen Annahmen bei allen Gemeinsamkeiten keine vollständig identi-

<sup>16</sup> <http://ids-mannheim.de/kogra-r> (Stand: 6.3.2018).

sche Semantik aufweisen, was die Grundlage für die subtilen Unterschiede beim Sprachgebrauch bildet.

Gegenstand des Kapitels von Patrick Brandt und Eric Fuß („Relativpronomen-selektion und grammatische Variation: *was* vs. *das* in attributiven Relativsätzen“) ist die Verteilung von d- und w-Pronomen in attributiven Relativsätzen. Im Mittelpunkt steht die Beobachtung, dass im Kontext Neutrum Singular anstatt des kanonischen Relativpronomens unter bestimmten Umständen (insbes. nach Indefinita, Demonstrativa sowie substantivierten Adjektiven) auch die w-Form *was* auftritt. Der Kern der Studie umfasst ca. 18.000 einschlägige Belege, die aus der Korpusgrammatik-Datenbank (KoGra-DB)<sup>17</sup> extrahiert wurden und für Eigenschaften wie Kategorie und Identität des Bezugslements semi-automatisch annotiert wurden. Zusätzlich wurden zu spezifischen Fragestellungen weitere Korpusrecherchen mittels der webbasierten COSMAS-II-Schnittstelle<sup>18</sup> zum DEREKo durchgeführt. Bei den quantitativen Analysen kam zunächst eine Auswahl an deskriptiv- und inferenzstatistischen Verfahren und Visualisierungsmethoden der KoGra-R-Schnittstelle (insbes. Assoziations- und Mosaikplots) zur Anwendung. Ferner wurden Clusteranalysen durchgeführt, die zur Kategorisierung von Kontexten genutzt wurden, die *das* oder *was* in ähnlicher Weise präferieren (bzw. dispräferieren). Auf der Basis der auf diese Weise gewonnenen Daten entwickeln die Autoren ein neuartiges Beschreibungsmodell dieses Variationsphänomens, das auf der Idee aufbaut, dass die Form des relativischen Anschlusses (mittels d- vs. w-Formen) im Kontext Neutrum Singular wesentlich von der syntaktischen Kategorie des Antezedens bestimmt wird. Handelt es sich bei dem Bezugslement um ein lexikalisches Nomen, wird der Relativsatz durch ein d-Pronomen eingeleitet. In allen anderen Kontexten erscheint *was* als Default-Relativpronomen. Diese Analyse wird durch Korpusbefunde gestützt, die zeigen, dass die Präsenz eines lexikalischen Bezugsnomens i.d.R. Relativierung durch *das* auslöst – auch bei Elementen, die ansonsten *was* als Relativum verlangen (z.B. Quantoren und Indefinita). Diese Perspektive ermöglicht überdies eine einheitliche Erklärung für das Auftreten von w-Pronomina in anderen relativsatzähnlichen Konstruktionen, die ebenfalls kein substantivisches Antezedens aufweisen (freie Relativsätze, weiterführende Relativsätze sowie attributive Relativsätze, die sich auf Zitate beziehen). Im Anschluss diskutieren die Autoren eine Reihe von Fällen, in denen die Distribution von *das* und *was* abweichende Eigenschaften aufweist. Darunter fallen (a) Kontexte, die *das* verlangen, obwohl anscheinend kein substantivisches Bezugslement vorliegt (vor allem individuierende Quantoren und Indefinita wie *jedes*, *eines* oder *keines*), (b) Kontexte, die *was* als Relativum auslösen, obwohl anscheinend ein substantivisches Bezugslement vorliegt

<sup>17</sup> Vgl. Bubenhofer/Konopka/Schneider 2014, S. 79ff.

<sup>18</sup> <http://ids-mannheim.de/cosmas2/> (Stand: 5.3.2018).

(insbes. substantivierte Superlative), sowie (c) Kontexte, die Variation zwischen *das* und *was* zulassen (substantivierte Positive). Zudem stellen sie fest, dass die Alternation zwischen *das* und *was* in Kontexten, in denen der starke Faktor „substantivisches Bezugselement“ nicht eindeutig vorliegt, zusätzlich von schwächeren semantisch-pragmatischen und außergrammatischen Faktoren (Region, Nähe zur Mündlichkeit etc.) beeinflusst wird. Generell gilt dabei, dass *was*-Relative typischerweise mit generalisierenden Lesarten verbunden sind, während *das* tendenziell eine konkrete/individuiierende Interpretation des Bezugselements auslöst.

Gegenstand des Kapitels „Alternation von *zu*- und *dass*-Komplementen: Kontrolle, Korpus und Grammatik“ von Patrick Brandt (unter Mitwirkung von Felix Bildhauer) sind die Bedingungen der Wahl der einen oder anderen Realisierung des Komplements. Die Datenextraktionen enthielten knapp über eine Million Satztoken-Belege. Bei den quantitativen Analysen wurde – wie in anderen Pilotstudien ebenso – eine Auswahl an deskriptiv- und inferenzstatistischen Verfahren (logistische Regressionsanalysen) und Visualisierungsmethoden angewendet (insbes. Assoziations-, Effekt- und Mosaikplots der KoGra-R-Schnittstelle). Es wurde bereits beobachtet, dass erstens die Variation zwischen *dass*- und *zu*-Komplementen offenbar an semantische Bedingungen geknüpft ist, die zwischen einem Argument des Matrixprädikats und dem Subjekt des eingebetteten Prädikats Koreferenz ermöglichen, und dass zweitens Ähnlichkeiten in Selektion und Verwendung sich dort zeigen, wo vergleichbare Muster bei der Argumentstrukturalternation vorliegen – in diesem Sinne erlaubt die Untersuchung der betrachteten Variation Rückschlüsse auf gemeinsame semantische Eigenschaften von Verben bzw. der sie involvierenden Strukturen und trägt so zu deren empirisch fundierter und streng grammatisch motivierter Taxonomie bei. In diesem Rahmen wird eingehend geprüft, unter welchen Bedingungen *dass*- und *zu*-Komplemente konkurrieren bzw. variieren und welche Einflussgrößen für die Ausprägung der jeweilig realisierten Form verantwortlich sind. Die leitende Hypothese der Untersuchung zur Variation in der Realisierung des Komplementtyps ist folgende: Wenn Kontrollbedingungen vorliegen, wird *ceteris paribus* *zu*-infinitivisch komplementiert (z.B. Wöllstein 2015; Rapp et al. 2017). Fragen, die sich genau in diesem Zusammenhang stellen und hier auf korpuslinguistischer Basis untersucht werden, betreffen etwaige Unterschiede zwischen einzelnen Kontrolltypen oder aber die Frage, ob prototypische, ideale Fälle für Kontrolle typische Komplementrealisierungen zeigen und umgekehrt – insbesondere wird hier auch Fällen der Anti(subjekt)kontrolle (z.B. mit *anordnen*, *veranlassen*) nachgegangen – und ob die *zu*- vs. *dass*- Alternation für systemgrammatische Fragen fruchtbar gemacht werden kann (z.B. ob mit aus dem Sprachgebrauch erwachsenden pragmatischen Prinzipien wie „Ökonomie des Ausdrucks“ gerechnet werden kann). Die Studie ist hypothesengeleitet ausgerichtet und geht daher vorran-

gig mit Daten um, die als einschlägig für die entsprechenden Hypothesen gelten können. Dabei bestätigt sich die Hypothese, dass das Vorliegen von Kontrollbedingungen die Wahl eines *zu*-infinitivischen Komplements in hoch-signifikanter Weise begünstigt. Es zeigt sich unter anderem, dass Subjekt- oder Akkusativobjektkontrolle offenbar idealtypische Kontrollmuster darstellen, die auch mit lexikalisch anderweitig spezifizierten Verben vorkommen. In Richtung einer grammatisch begründeten Verbtaxonomie ist ferner bemerkenswert, dass Akkusativobjektkontrollverben (*bitten*, *zwingen*) und Antikontrollverben (*anordnen*, *billigen*) auf struktureller Ebene starke Gemeinsamkeiten aufweisen. Ein weiteres Ergebnis ist, dass grammatische Operationen wie Passivierung oder Modalisierung signifikant häufiger auftreten, wenn sie zur Herstellung von idealtypischen Kontrollbedingungen führen.

Im letzten Kapitel „KoGra-R: Standardisierte statistische Auswertung von Korpusrecherchen“ von Sandra Hansen-Morath, Hans-Christian Schmitz, Roman Schneider und Sascha Wolfer wird das Ergebnis der Bemühungen präsentiert, Techniken und Werkzeuge zu entwickeln, die routinemäßig für die erste statistische Exploration von Variationsphänomenen bei der Erstellung der Grammatik benutzt werden können. Die Autoren bauen dabei auf Erfahrungen auf, die in Pilotuntersuchungen gesammelt wurden. Vorgestellt wird das im Projekt entwickelte webbasierte Werkzeug KoGra-R, mit dessen Hilfe ein vorgegebener Satz an statistischen Analysen und Visualisierungen von Häufigkeitsdaten automatisch ausgeführt werden kann. Er ist in der Programmiersprache R (R Core Team 2016) implementiert und als die Menge derjenigen statistischen Verfahren definiert, die idealerweise auf alle Fragestellungen der geplanten Grammatik angewendet werden können, um den ersten Überblick über die Häufigkeit und Verteilung eines Phänomens zu bekommen. Input für KoGra-R sind Kontingenztabelle, die Häufigkeiten von Einzelphänomenen oder von mehreren Phänomenvarianten in Abhängigkeit von Ausprägungen unabhängiger Variablen (etwa grammatischer oder außergrammatischer Faktoren) enthalten. Die Kontingenztabelle können als Rechercheergebnisse direkt an KoGra-R übergeben werden, aber auch als CSV-Dateien hochgeladen oder manuell in ein Formular eingetragen werden. Die statistischen Verfahren umfassen u.a. Visualisierungen normierter und relativer Häufigkeiten, Chi-Quadrat-Tests, Berechnungen der Assoziationsstärke Phi/Cramérs V, Visualisierungen von Häufigkeitsverteilungen mithilfe von Assoziationsplots und Mosaikplots sowie Berechnungen zur Verteilung von Phänomenen über das Gesamtkorpus ( $DP_{norm}$ , vgl. Gries 2008). KoGra-R wurde zwei Evaluationsprozessen unterzogen. Das Werkzeug wurde dabei als nützlich und in der Bedienung einfach empfunden – einige Kritikpunkte konnten bei seiner Weiterentwicklung berücksichtigt werden. KoGra-R ist im Internet allgemein zugänglich<sup>19</sup> und kann

<sup>19</sup> Siehe <http://ids-mannheim.de/kogra-r/> (Stand: 6.3.2018).

auch in anderen Projekten benutzt werden. Das Tool bzw. die darin eingegangenen statistischen Verfahren wurden bei vielen der im Abschnitt 2 beschriebenen Untersuchungen und bei allen Pilotstudien in diesem Band für die ersten Explorationsschritte benutzt. Im weiteren Verlauf der Untersuchungen wurden dann oft avanciertere statistische Methoden (etwa logistische Regressionen) herangezogen, die je nach Beschaffenheit des Phänomens auszuwählen sind und sich nicht in dem Maße standardisieren lassen.

Die Herausgeber möchten sich an dieser Stelle bei den Autoren der Beiträge für die fruchtbare Zusammenarbeit bedanken. Sehr verbunden sind wir Frau Dr. Anja Steinhauer für die redaktionelle Arbeit und Frau Carolin Häberle für den Satz und die Betreuung der Veröffentlichung seitens des IDS.

## Literatur

- Brandt, Patrick (2016): Fehlkonstruktion und Reparatur in der Bedeutungskomposition. In: *Linguistische Berichte* 248, S. 395–433.
- Brandt, Patrick/Fuß, Eric (2014): Most questionable pronouns: Variation between *das*-vs. *was*-relatives in German. In: *Linguistische Berichte* 239, S. 297–329.
- Brandt, Patrick/Fuß, Eric (2017): A corpus-based analysis of pronoun choice in German relative clauses. In: De Cuyper, Ludovic/Vanderschueren, Clara/De Sutter, Gert (Hg.): *Current trends in analyzing syntactic variation*. (= *Belgian Journal of Linguistics* 31). Amsterdam: Benjamins, S. 189–212.
- Brandt, Patrick/Trawiński, Beata/Wöllstein, Angelika (2016): (Anti-)control in German: Evidence from comparative, corpus- and psycholinguistic studies. In: Reich, Ingo/Speyer, Augustin (Hg.): *Co- and subordination in German and other languages*. (= *Linguistische Berichte – Sonderheft* 21). Hamburg: Buske, S. 77–98.
- Breindl, Eva/Volodina, Anna/Waßner, Ulrich Hermann (2014): *Handbuch der deutschen Konnektoren*. Bd. 2: *Semantik der deutschen Satzverknüpfers*. (= *Schriften des Instituts für Deutsche Sprache* 13). Berlin/Boston: De Gruyter.
- Bubenhof, Noah/Hansen-Morath, Sandra/Konopka, Marek (2014): Korpusbasierte Exploration der Variation der nominalen Genitivmarkierung. In: *Zeitschrift für Germanistische Linguistik* 42, 3, S. 379–419.
- Bubenhof, Noah/Hansen-Morath, Sandra/Konopka, Marek/Schneider, Roman (2014): *GenitivDB 1.0 – Datenbank zur Genitivmarkierung* (Release vom 1.6.2014). Mannheim: Institut für Deutsche Sprache.
- Bubenhof, Noah/Hansen-Morath, Sandra/Konopka, Marek/Schneider, Roman (2015): *GenitivDB 2.0 – Datenbank zur Genitivmarkierung* (Release vom 1.9.2015). Mannheim: Institut für Deutsche Sprache.
- Bubenhof, Noah/Konopka, Marek/Schneider, Roman (2014), unter Mitwirkung von Carin Brinkmann, Katrin Hein und Bruno Strecker: *Präliminarien einer Korpusgrammatik*. (= *Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache* 4). Tübingen: Narr.

- Chambers, Jack K. (2002): Studying language variation: An informal epistemology. In: Chambers, Jack K./Trudgill, Peter/Schilling-Estes, Natalie (Hg.): The handbook of language variation and change. (= Blackwell Handbooks in Linguistics 11). Malden, MA u.a.: Blackwell, S. 3–14.
- Donalies, Elke (2011): *Tagtraum, Tageslicht, Tagedieb*. Ein korpuslinguistisches Experiment zu variierenden Wortformen und Fugenelementen in zusammengesetzten Substantiven. Mit einem Exkurs und zahlreichen Statistiken von Noah Bubenhofer. (= amades 42). Mannheim: Institut für Deutsche Sprache.
- Elspaß, Stephan/Dürscheid, Christa (2017): Areale grammatische Variation in den Gebrauchsstandards des Deutschen. In: Konopka/Wöllstein (Hg.), S. 85–104.
- Fuß, Eric (2017): Relativierungsverhalten und syntaktische Kategorie substantivierter Adjektive. In: Döring, Sandra/Geilfuß-Wolfgang, Jochen (Hg.): Probleme der syntaktischen Kategorisierung: Einzelgänger, Außenseiter und mehr. (= Stauffenburg Linguistik 90). Tübingen: Stauffenburg, S. 43–100.
- Fuß, Eric (2018): Sprachliche Variation. In: Wöllstein, Angelika/Gallmann, Peter/Habermann, Mechthild/Krifka, Manfred (Hg.): Grammatiktheorie und Empirie in der germanistischen Linguistik. Berlin/Boston: De Gruyter, S. 185–224.
- Fuß, Eric/Konopka, Marek/Wöllstein, Angelika (2017): Perspektiven auf grammatische Variation. In: Konopka/Wöllstein (Hg.), S. 229–254.
- Gries, Stefan T. (2008): Dispersions and adjusted frequencies in corpora. In: International Journal of Corpus Linguistics 13, 4, S. 403–437.
- Hansen, Sandra/Schneider, Roman (2013): Decision tree-based evaluation of genitive classification – An empirical study on CMC and text corpora. In: Gurevych, Iryna/Biemann, Chris/Zesch, Torsten (Hg.): Proceedings of GSCL 2013: The bi-annual meeting of the German Society for Computational Linguistics and Language Technology. Springer LNAI, S. 83–88.
- Kiparsky, Paul (1979): *Pāṇini as a variationist*. (= Centre of Advanced Study in Sanskrit 6). Cambridge, MA: The MIT Press.
- Konopka, Marek (2018): *Korpuslinguistik, Grammatiktheorie und Grammatikografie*. Ms. Mannheim: Institut für Deutsche Sprache.
- 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/Hansen-Morath, Sandra (2017): *AcI-Konstruktionen und Wie-Komplementsätze. Dokumentation der Parallelexploration*. In: grammis. <https://grammis.ids-mannheim.de/korpusgrammatik/6302> (Stand: 5.3.2018).
- Konopka, Marek/Waßner, Ulrich H., unter Mitwirkung von Sandra Hansen (2013): *Standarddeutsch messen? Frequenz und Varianz negativ-konditionaler Konnektoren*. In: *Korpus – grammatika – axiologie* 8, S. 12–35.
- Konopka, Marek/Wöllstein, Angelika (Hg.) (2017): *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*. (= Jahrbuch des Instituts für Deutsche Sprache 2016). Berlin/Boston: De Gruyter.

- Kroch, Anthony (1989): Reflexes of grammar in patterns of language change. *Journal of Language Variation and Change* 1, 3, S. 199–244.
- Labov, William (1966): The linguistic variable as a structural unit. In: *Washington Linguistic Review* 3, S. 4–22.
- Labov, William (2008): Quantitative analysis of linguistic variation. *Quantitative Analyse sprachlicher Variation*. In: Ammon, Ulrich et al. (Hg.): *Sociolinguistics. An international handbook of the science of language and society. Soziolinguistik. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*. Bd. 1. 2., vollst. überarb. u. erw. Aufl. (= Handbücher zur Sprach- und Kommunikationswissenschaft 3). Berlin/New York: De Gruyter, S. 6–21.
- Münzberg, Franziska/Hansen-Morath, Sandra (2018): *Die Wucht und Strömung war immens* – wie stark ist der Ellipseneffekt? In: Fuß, Eric/Konopka, Marek/Trawiński, Beata/Waßner, Ulrich H. (Hg.): *Grammar and Corpora 2016*. Heidelberg: Heidelberg University Publishing, S. 73–95.
- Pasch, Renate et al. (2003): *Handbuch der deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfers (Konjunktionen, Satzadverbien und Partikeln)*. (= Schriften des Instituts für Deutsche Sprache 9). Berlin/New York: De Gruyter.
- Rabanus, Stefan (1996): *Die Sprache der Internetkommunikation*. (= *Internet im Gardez!* 1). Mainz: Gardez!
- Rapp, Irene et al. (2017): Lexikalisch-semantische Passung und argumentstrukturelle Trägheit – eine korpusbasierte Analyse zur Alternation zwischen *dass*-Sätzen und *zu*-Infinitiven in Objektfunktion. In: *Deutsche Sprache* 45, S. 193–221.
- Schneider, Roman (2014): GenitivDB – a corpus-generated database for German genitive classification. In: Calzolari, Nicoletta et al. (Hg.): *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík: European Language Resources Association (ELRA).
- Viana, Vander/Zyngier, Sonia/Barnbrook, Geoff (Hg.) (2011): *Perspectives on corpus linguistics*. (= *Studies in Corpus Linguistics* 48). Amsterdam/Philadelphia: Benjamins.
- Volodina, Anna (2018): *angenommen* ist nicht *vorausgesetzt* – eine korpuslinguistische Analyse. In: *Deutsche Sprache* 46, Themenheft: *Über den Tellerrand hinaus: Konnektorenforschung aus unterschiedlichen Blickwinkeln*, S. 67–81.
- Waßner, Ulrich Hermann (2018): Wortbildung und Variation bei Konnektoren. In: *Deutsche Sprache* 46, Themenheft: *Über den Tellerrand hinaus: Konnektorenforschung aus unterschiedlichen Blickwinkeln*, S. 52–66.
- Wiese, Bernd (1999): Unterspezifizierte Paradigmen. Form und Funktion in der pronominalen Deklination. In: *Linguistik Online* 4, 3. [https://www.linguistik-online.net/3\\_99/wiese.html](https://www.linguistik-online.net/3_99/wiese.html) (Stand: 5.12.2018).
- Wöllstein, Angelika (2015): Grammatik – explorativ. Hypothesengeleitete und -generierende Exploration variierender Satzkomplementationsmuster im standardnahen Deutsch. In: Eichinger, Ludwig M. (Hg.): *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*. (= *Jahrbuch des Instituts für Deutsche Sprache* 2014). Berlin/Boston: De Gruyter, S. 93–120.

Wurzel, Wolfgang U. (1984). Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung. (= *Studia grammatica* 21). Berlin: Akademie Verlag.

Zifonun, Gisela/Hoffmann, Ludger/Strecker, Bruno (1997): Grammatik der deutschen Sprache. (= *Schriften des Instituts für Deutsche Sprache* 7). Berlin/New York: De Gruyter.