

Marek Konopka 

# Einleitung

## 1 In aller Kürze

Eine deskriptive, konsequent korpuslinguistisch fundierte und an das Fachpublikum gerichtete Grammatik des Deutschen gibt es bisher nicht. Es wird sie auch noch nicht so schnell als abgeschlossenes Werk geben, denn die Aufgabe ist enorm, wenn man sich zum Ziel setzt, mit modernen korpuslinguistischen Methoden die Vielfalt und Variabilität der deutschen Grammatik in großer Detailschärfe zu erfassen, und gleichzeitig für die Validierbarkeit der Ergebnisse sorgen will. Zeitnah möglich ist aber die Erarbeitung einzelner Bausteine einer solchen Grammatik. Mit den „Bausteinen einer Korpusgrammatik des Deutschen“ werden nach und nach korpusgestützte Untersuchungen im Open-Access-Format vorgelegt, die als einzelne Kapitel einer neuen wissenschaftlichen Grammatik des Deutschen gestaltet sind. Die Idee der „Bausteine“ ist im Projekt „Korpusgrammatik – grammatische Variation im standardsprachlichen und standardnahen Deutsch“<sup>1</sup> des Leibniz-Instituts für Deutsche Sprache in Mannheim (IDS) entwickelt worden. Die „Bausteine“ entstehen am IDS und werden in Kooperation mit dem Germanistischen Seminar der Universität Heidelberg unter Trägerschaft des Europäischen Zentrums für Sprachwissenschaften (EZS) herausgegeben. Die Reihe setzt sich dabei das ausdrückliche Ziel, einen Beitrag zur offenen Wissenschaft im Sinne der Open-Science-Bewegung zu leisten, indem Bestandteile des wissenschaftlichen Prozesses frei zugänglich und auch nachnutzbar gemacht werden: Neben Texten in Form von Grammatikkapiteln werden Ergebnisse korpuslinguistisch-statistischer Analysen und bei ausgewählten Themen die zugrunde liegenden Sprachdaten veröffentlicht. Geplant sind jährlich bis zu zwei Ausgaben jeweils in einer Online- und einer Print-on-Demand-Version. Bei der Qualitätssicherung werden die Herausgebenden der

1 Vgl. <https://www.ids-mannheim.de/gra/projekte/korpusgrammatik.html>, abgerufen am 27.12.2019.

„Bausteine“ durch den [Wissenschaftlichen Beirat](#) unterstützt, der u. a. für die Begutachtung der Beiträge sorgt.

## 2 Warum braucht man eine neue wissenschaftliche Grammatik des Deutschen?

In der neueren wissenschaftlichen Grammatikschreibung zum Gegenwartsdeutschen hat sich ein besonderes Paradigma herauskristallisiert. Demnach sollen Grammatikwerke deskriptiv und gleichzeitig theoretisch fundiert sein, ohne jedoch dabei nur einer Theorie folgen zu müssen (vgl. z.B. Eisenberg 2013a: 7). Ein gewisser theoretischer Eklektizismus scheint sowieso notwendig zu sein, wenn man eine konsequente Deskription anstrebt, da Theorien nicht lückenlos sind (dazu Hoffmann 1997: 4). Im Zweifelsfall soll man versuchen, dem grammatischen Phänomen und nicht einer bestimmten Theorie gerecht zu werden. Diesen Ideen fühlen sich auch die Autor/inn/en der „Bausteine“ verpflichtet. Allerdings gibt es bereits einige Grammatiken, in denen eine weitgehend theorieneutrale Deskription beherzigt wird (dazu z.B. Paranhos Zitterbart 2010: 67–70, Eisenberg 2013a: 1–2). Warum soll also eine weitere wissenschaftliche Grammatik des Deutschen entstehen?

In den folgenden Ausführungen werden einige ausgewählte Grammatiken besonders häufig als Beispiele herangezogen. Es sind dies die Dudengrammatik (2016), der „Grundriss der deutschen Grammatik“ von Eisenberg (2013a, 2013b), die „Grammatik der deutschen Sprache“ (GDS) von Zifonun et al. (1997) und die „Grammatik des Deutschen im europäischen Vergleich“ (GDE) von Gunkel et al. (2017). Zu betrachten ist zunächst, an wen sich diese Werke richten. Im populärsten, der Dudengrammatik, die als einzige das Prädikat *wissenschaftlich* nicht für sich beansprucht, werden als Einsatzorte Schulen und Universitäten und als Zielgruppe beruflich mit Sprache Beschäftigte, persönlich an sprachlichen Fragen Interessierte sowie Lehrende und Lernende des Deutschen als Fremdsprache genannt (Dudengrammatik 2016: 5). Eisenberg (2013a: IX–X) empfiehlt seine Grammatik als Lehrbuch und hebt unter ihrer Leserschaft Studierende hervor. Hoffmann (1997: 2) sieht die Zielgruppe der GDS in Sprachinteressierten mit Vorkenntnissen: Linguisten, Akademikern aus Nachbardisziplinen, Lehrenden und Studierenden. In der GDE (2017: 1) schließlich werden sprachwissenschaftliche Vorbildung, aber keine speziellen grammatiktheoretischen Kenntnisse vorausgesetzt. Man könnte hier eine von Werk zu Werk zunehmende Spezialisierung der Kenntnisse herauslesen, die von den Adressierten erwartet werden. Der nächste Schritt wäre es, direkt die Sprachwissenschaftler/innen zum anvisierten Kernpublikum zu erklären.

Ein Bekenntnis zu einem solchen Schritt ist aus unserer Sicht im Sinne des wissenschaftlichen Fortschritts notwendig. So soll die neueste Auflage der Dudengrammatik (2016) dem Vorwort zufolge den heutigen Stand des Wissens über Formen und Funktionen der deutschen Standardsprache präsentieren, und sie erreicht tatsächlich einen Grad an deskriptiver Präzision und theoretischer Durchdringung, der auch einer deklariert wissenschaftlichen Grammatik gut zu Gesicht stünde. Doch als traditionelles Nachschlagewerk „für alle“ muss die Dudengrammatik den Spagat zwischen deskriptiver Adäquatheit und Allgemeinverständlichkeit, zwischen Deskription und Normierung versuchen. Autor/inn/en rein wissenschaftlicher Grammatiken dagegen brauchen sich um die Allgemeinverständlichkeit nicht so große Sorgen zu machen und dürfen die Normierung sogar – so der Konsens der allermeisten Linguist/inn/en – gar nicht anstreben (dazu differenziert Hoffmann 1997: 6, Habermann 2010: 11–12, Eisenberg 2013a: 8–10). Unter den Grammatiken mit wissenschaftlichem Anspruch gibt es einige, in denen explizit didaktische Zwecke verfolgt werden (z. B. Eisenberg 2013a, 2013b, tendenziell Hentschel & Weydt 2013 und Engel 2009). Es gibt aber auch welche, in denen die Grammatikbeschreibung mit weniger Rücksicht auf die Didaktik betrieben wird (z. B. GDS, GDE). Gäbe es darunter Grammatiken, die sich explizit an Sprachwissenschaftler/innen als Kernpublikum richteten, so hätten ihre Autor/inn/en endgültig die Berechtigung, die Grenzen der grammatikografischen Deskription auszureizen. Sie könnten es auch zu einem ihrer zentralen Ziele erklären, mit der tiefer gehenden Beschreibung direkte Anknüpfungspunkte für Spezialforschung zu bieten, die den wissenschaftlichen Fortschritt noch weiter vorantreibt.

Eine Steigerung der Deskriptionstiefe, die in größerer Detailschärfe resultiert, passt also nicht nur zum Konzept einer deskriptiven Grammatik, sie ist auch dem wissenschaftlichen Fortschritt zuträglich. Sie lässt aber den Umfang einer Grammatik anwachsen und kann in der Praxis der Werkerstellung auf Kosten der Vollständigkeit gehen, die oft von Grammatiken erwartet wird. Allerdings ist das Vollständigkeitspostulat in seiner Absolutheit ohnehin utopisch, sodass sich wissenschaftliche Grammatiken mit mehr oder weniger fragmentarischen Lösungen begnügen müssen (dazu Hoffmann 1997: 4). Unter den weiter oben genannten Grammatiken sind die GDS und die GDE besonders umfangreich. Sie bestechen auch durch eine besondere Detailschärfe (zur GDS vgl. z. B. Paranhos Zitterbart 2010: 70) und eine beeindruckende theoretische Fundiertheit. Sie weisen aber zwangsläufig Lücken auf oder behandeln nur Teilbereiche.<sup>2</sup>

2 Hoffmann (1997: 11) weist in der Einführung der GDS u. a. auf das Fehlen „eine[r] wirklich ausgearbeitete[n] Morphosyntax“ und einer Behandlung der deutschen Wortbildung hin. Die GDE (2017: 1) ist zunächst nur als Grammatik des Nominals angelegt.

In beiden Grammatiken wird die Zielgruppe zwar noch nicht entsprechend eingeschränkt, jedoch kommen sie in der Praxis unserem Postulat einer Grammatik für Sprachwissenschaftler/innen schon sehr nahe. Dabei ist die GDE 2017 erschienen und damit ziemlich aktuell. Die Autor/inn/en nehmen darin allerdings eine spezielle Perspektive ein, die auf die Grammatik des Deutschen im Vergleich mit europäischen Kontrastsprachen ausgerichtet ist, und beschreiben dabei nur nominale Kategorien und Konstruktionen. Die GDS wiederum konzentriert sich zwar auf das Deutsche allein und bietet dabei eine überaus breit angelegte Beschreibung, sie ist aber bereits vor 23 Jahren erschienen. Was also offensichtlich fehlt, ist eine Grammatik, die einerseits dadurch, dass sie nur das Deutsche in den Fokus nimmt, ein entsprechend höheres Maß an Deskriptionsbreite und -tiefe erreichen kann und die andererseits einen Beitrag im aktuellen Forschungsdiskurs darstellt.

Warum ist die Aktualität so wichtig? Es liegt zunächst auf der Hand, dass die Forschung nicht stillsteht und immer wieder neue Erkenntnisse hervorbringt, die auch in wissenschaftliche Grammatiken einfließen sollen. Es ist außerdem zu prüfen, ob in neuester Zeit nicht sprachliche Entwicklungen stattgefunden haben, die zu Phänomenen geführt haben, die in wissenschaftlichen Grammatiken noch nicht beschrieben sind. Und schließlich sind in der grammatischen Beschreibung neue Forschungsparadigmen zu berücksichtigen, die bisher in Grammatiken noch nicht angewendete Methoden, noch nicht eingenommene Blickwinkel und noch nicht erörterte Fragestellungen relevant machen. So wurde in der grammatischen Forschung der letzten Jahrzehnte die empirische Fundierung immer wichtiger. Wesentlich dazu beigetragen hat die rasante Weiterentwicklung von Sprachkorpora, die computerbasiert aufgebaut, aufbereitet, bereitgestellt, durchsucht sowie analysiert werden und heute zum Teil Größen von Milliarden von Wörtern erreichen können. Was die Arbeit mit Korpora dabei anderen Empiriearten wie Befragung oder Experiment voraushat, ist neben der potenziell sehr großen Menge an Sprachdaten deren natürlichere Entstehungssituation (vgl. Konopka 2018: 152–153). Übrigens verstehen sich schon alle weiter oben aufgezählten Grammatiken als empirisch orientiert. Die Dudengrammatik (2016), die GDS und die GDE nutzen auch in größerem Maße Korpora.<sup>3</sup> Sie ziehen sie vor allem als Quelle für Belege heran, die theoretische Feststellungen exemplifizieren. In der neueren grammatischen Forschung werden Korpora hingegen immer häufiger dazu benutzt, theoretische Feststellungen systematisch zu evaluieren (vgl. Meurers & Müller 2009). Sie werden außerdem zunehmend als Quelle für erklärungsbedürftige Phänomene und neue

3 Eisenberg (2013a: 9) dagegen äußert sich kritisch zum konsequenten „Datensammeln“, das sich schnell verselbständige und damit das Entstehen einer Grammatik verhindere.

Fragestellungen exploriert.<sup>4</sup> Korpusorientierte Grammatikforschung heißt also heute Untersuchung großer Mengen authentischen, elektronisch durchsuchbaren und mit Metadaten angereicherten Textmaterials. Sie hat in den letzten Jahren über Einzelphilologien hinweg zu einer regelrechten Welle interessanter Studien geführt (vgl. z. B. die Beiträge in Konopka et al. 2011 und Fuß et al. 2018). In der Grammatikografie des Englischen hat sie auch bereits vor einiger Zeit ihren Niederschlag in korpusgestützten Grammatiken gefunden, etwa der „Cambridge Grammar of English“ von Carter & McCarthy (2006) und vor allem der „Longman Grammar of Spoken and Written English“ von Biber et al. (2006 [1999]).<sup>5</sup> In der Germanistik jedoch wurde dieser Ansatz bisher nur im Rahmen von kleineren Studien und Monografien bzw. von Untersuchungen zu überschaubaren Grammatikausschnitten verfolgt. Eine breit angelegte Grammatik des Deutschen auf korpuslinguistischer Basis stellt immer noch ein Desiderat dar.

Korpuslinguistik ermöglicht genauere, detailliertere Deskription. Besiegelt ist damit nicht nur die endgültige Abwendung von der normativen Grammatik oder von einer Grammatik, die einen einheitlichen Sprachgebrauch vorgaukelt. Unausweichlich ist jetzt konsequenterweise auch die Hinwendung zu einem Grammatikkonzept, gemäß dem die Sprache in ihrem ganzen Facettenreichtum, ihrer Variabilität und in ihrem tatsächlichen Gebrauch beschrieben werden soll (vgl. Conrad 2010: 228). Variation ist mittlerweile zu einem der zentralen Themen der Grammatikschreibung geworden, während das Konzept der Standardsprache an Bedeutung eingebüßt hat bzw. jetzt anders gedacht werden muss (dazu Eichinger 2017: v. a. 17–19, Felder 2016: 23–26). Von einer neuen, korpuslinguistisch fundierten Grammatik des Deutschen ist in diesem Kontext zu erwarten, dass sie die Variation nicht nur festhält, sondern auch im Hinblick auf ihre Abhängigkeit von verschiedenen sprachinternen und/oder -externen Einflussfaktoren erklärt (vgl. Konopka 2018: 166).<sup>6</sup> Dies stellt besondere Anforderungen an die Methodik der Untersuchungen, die der Grammatikbeschreibung zugrunde liegen. Die Bedeutung korpuslinguistisch-statistischer Verfahren, mit denen man die Datenfülle und -vielfalt zu fassen bekommen kann, wächst. Die Anwendung statistischer Verfahren verpflichtet dabei schon dazu, die Datenerhebung, -aufbereitung und -analyse für die Leserschaft zu dokumentieren, was sich sehr gut in das Konzept einer wissenschaftlichen Grammatik einfügt, die die Validierbarkeit der zugrunde liegenden Untersuchungen sichert. Damit wird die direkte Anschlussforschung ermöglicht, die unter Umständen sogar

4 Zur Evidenz aus dem Korpus, Validierung am Korpus und Induktion mithilfe des Korpus vgl. Konopka (2018: 151–164).

5 Zur kritischen Diskussion vgl. Mukherjee (2006).

6 Felder (2016: v. a. 7–19) zeigt instruktiv die Möglichkeiten, sprachliche Variation im Allgemeinen zu beschreiben und zu erfassen.

mit vergleichbaren Methoden und denselben sprachlichen Daten betrieben werden kann.

Die „Bausteine“ sollen es möglich machen, nach und nach eine wissenschaftliche Grammatik „aufzubauen“, die dem skizzierten Bedarf gut begegnen kann. Sie wird im Weiteren als *Korpusgrammatik* bezeichnet.

### 3 Eine *Korpusgrammatik* in spe

Die ersten konzeptionellen Überlegungen zu einer *Korpusgrammatik* wurden in Bubenhofer et al. (2014) vorgestellt. Von 2011 bis 2019 entstand am IDS außerdem eine Reihe von Pilotstudien, in denen anhand detaillierter Untersuchungen von Variationsphänomenen aus verschiedenen Grammatikbereichen exemplarische Fragestellungstypen und dazu passende korpuslinguistische Herangehensweisen herausgearbeitet wurden. Einen Überblick über diese Arbeiten geben Fuß et al. (2019). Nun werden die ersten Versionen von Kapiteln der künftigen *Korpusgrammatik* als „Bausteine“ vorgestellt. Es handelt sich dabei gewissermaßen um Probekapitel – die Anlage der Kapitel und die Darstellungsformen werden sich mit weiteren „Bausteinen“ weiterentwickeln: Es muss ja einiges ausprobiert werden, denn die *Korpusgrammatik* hat auch Neues vor. Zu beginnen ist aber mit den eher traditionellen Rahmeninformationen.

Gegenstand der *Korpusgrammatik* ist die Grammatik des geschriebenen Deutsch, wie es sich in Texttypen manifestiert, die Standardsprachliches erwarten lassen (vgl. Bubenhofer et al. 2014: 23–32). Der besondere Fokus liegt auf der Variation. Den grammatischen Gegenstandsbereich bilden Morphologie und Syntax, zuzüglich der Ausschnitte benachbarter Gebiete (Phonologie und Graphematik einerseits, „Textgrammatik“ andererseits), die morphologische und syntaktische Variation direkt begründen oder beeinflussen. Die *Korpusgrammatik* strebt eine empirisch fundierte Deskription an und stützt sich dabei hauptsächlich auf das sogenannte Untersuchungskorpus, das ein speziell aufbereitetes Teilkorpus des Deutschen Referenzkorpus (DeReKo, vgl. Kupietz et al. 2018)<sup>7</sup> darstellt (Genaueres in Abschnitt 5 weiter unten). Das primär anvisierte Publikum der *Korpusgrammatik* sind Sprachwissenschaftler/innen.

Da es offenkundig utopisch ist, die gesamte Grammatik einer Einzelsprache in ihrer Breite und Tiefe zu erfassen, wurden in der obigen Kurzcharakteristik für die *Korpusgrammatik* von vorn herein gewisse Einschränkungen vorgenommen. In der Tradition neuerer Grammatiken beschränken wir uns auf das Standarddeutsche. Zwei weitere Einschränkungen sind durch die korpuslinguistische

<sup>7</sup> Siehe auch <https://www.ids-mannheim.de/kl/projekte/korpora.html> (abgerufen am 29.11.2019).

Herangehensweise mitbedingt. Die eine ist medialer Natur: Die empirisch fundierte Deskription beschränkt sich auf das geschriebene Deutsch. Ein wichtiger Grund dafür ist, dass es sehr umfangreiche, inzwischen auch ausdifferenzierte und gut aufbereitete Korpora des geschriebenen Deutsch gibt, aber noch keine vergleichbaren Sammlungen des gesprochenen Deutsch.<sup>8</sup> Die andere Einschränkung resultiert aus der Umgewichtung der grammatischen Themen und Perspektiven, wie sie in Abschnitt 2 beschrieben wurde: Besondere Aufmerksamkeit gilt grammatischer Variation. Selbstverständlich wird dabei davon ausgegangen, dass die Standardsprache Variation zulässt. So sollen sich die behandelten Phänomene auf das Standarddeutsche beziehen, da aber die Abgrenzung der Standardsprache ein Problem an sich darstellt (vgl. Bubenhofer et al. 2014: 21–32), werden zuweilen auch Übergänge zum Substandard ausgeleuchtet. Um solche Übergänge zu erfassen, werden ergänzend zum Untersuchungskorpus diejenigen Teilkorpora des Web-Korpus DECOW16 (Schäfer & Bildhauer 2012; Schäfer 2015) herangezogen, die „Grenzwertiges“ erwarten lassen. Im Weiteren soll sich die *Korpusgrammatik* weitgehend auf die Morphologie (Wortbildung und Flexion) und Syntax beschränken. Ein zentraler Grund für diese Einschränkung sind Machbarkeitsüberlegungen. Ausgewählt wurden mit Morphologie und Syntax aber Gebiete, die traditionell als Kernbereiche der Grammatik aufgefasst werden. Zuallerletzt führt die Tatsache, dass sich die *Korpusgrammatik* primär an Sprachwissenschaftler/innen richtet, zusammen mit der Priorisierung der Variation dazu, dass Phänomene, die schwer zu erklären bzw. in der Forschung umstritten sind, „einfachere“ und durch wissenschaftlichen Konsens geprägte Themen in den Hintergrund drängen bzw. im Falle von unstrittigen „Basics“ manchmal ganz verschwinden lassen.

Wie von einer empirisch fundierten Grammatik dieser Art zu erwarten, wird sich die *Korpusgrammatik* zunächst an Oberflächenphänomenen orientieren. Die zugrunde liegenden Untersuchungen von Variationsphänomenen werden unter Suspendierung des Konzepts „freie Variation“ durchgeführt, wobei versucht wird, inner- und außergrammatische Einflussgrößen der Variation zu ermitteln. Die Ermittlung ersterer soll grundlegend durch Korpusannotationen nach grammatischen Merkmalen ermöglicht werden, die Erfassung letzterer durch Dokumentmetadaten, die die Korpora etwa nach zentralen Textklassen (Registern), geografisch oder zeitlich strukturieren. Trotz der primären Orientierung an Oberflächenphänomenen werden durchaus theoretische Erklärungen anvisiert,

8 Zum geschriebenen Deutsch (insbesondere in den Druckmedien) sind außer DeReKo vor allem die Korpora des Digitalen Wörterbuchs der deutschen Sprache (DWDS) zu nennen (vgl. <https://www.dwds.de/d/korpora>, abgerufen am 20.6.2020). Ein Beispiel für ein im Aufbau begriffenes ausdifferenziertes Korpus der gesprochenen Sprache ist Folk (vgl. <http://agd.ids-mannheim.de/folk.shtml>, abgerufen am 26.5.2020).

und es wird theoretische Fundiertheit angestrebt, allerdings ohne die Theorieutralität aus den Augen zu verlieren. Letztere ist überhaupt die Voraussetzung dafür, dass die *Korpusgrammatik* von einem breiteren Fachpublikum umfassend benutzt werden kann. Einem solchen Publikum soll sie nämlich nicht nur Interpretationen sprachlicher Daten liefern. Sie soll auch quantitative Übersichten und differenzierte Datensammlungen bereitstellen, auf denen theoretische Forschung verschiedener Prägung aufsetzen kann. Außerdem hat sie so beschaffene Voruntersuchungen von Variationskontexten zu liefern, dass die empirische Spezialforschung direkt daran anschließen kann, auch wenn sie in spezielleren Frameworks oder mit anderen Methoden arbeitet.

Die *Korpusgrammatik* zeigt sich in noch einem Bereich anders als die bisherigen Grammatiken. So beeinflusst die Benutzung korpuslinguistischer Methoden die Inhalte und die Anlage der Grammatikkapitel. Diesen liegen oft statistische Untersuchungen zugrunde, die transparent dargestellt sein müssen, um methodisch nachvollziehbar zu sein, eine Bewertung der statistischen Befunde zuzulassen und idealerweise Replikation zu erlauben. Eine adäquate Darstellung von Ergebnissen solcher Untersuchungen führt dazu, dass viele Kapitel der *Korpusgrammatik* stärker durch methodische Argumentation geprägt sind und einen weniger deklarativen Stil aufweisen als für andere Grammatiken üblich. Dafür sind sie auch stärker problemlösungsorientiert und zwar in dem speziellen Sinne, dass sie nicht nur das Resultat zeigen, sondern auch den Weg zur Erkenntnis nachzeichnen, was dem forschenden Publikum nur entgegenkommen kann.

So hat die *Korpusgrammatik* zwar einige Unterschiede zu den in Abschnitt 2 genannten Grammatiken des Deutschen aufzuweisen, dennoch sind diese für die *Korpusgrammatik* einerseits Fundament und andererseits eine große Inspirationsquelle. Sie sind eine Fundgrube des grammatischen Wissens, das als weitgehend gesichert gilt. Sie liefern mit ihren Beschreibungen auch Ausgangspunkte für Vertiefungen und laden zur Evaluation und Weiterentwicklung von Generalisierungen ein. Sie bleiben in vielerlei Hinsicht – etwa bei den Darstellungsformen, der präzisen Sprache und dem grammatikografischen Duktus – ein Vorbild. Letztlich machen sie es möglich, dass die *Korpusgrammatik* als Handbuch für Sprachwissenschaftler/innen mehr in die Tiefe gehen und sich nicht nur der Präsentation von bereits Bekanntem, sondern auch verstärkt der Problemlösung widmen kann.

#### 4 Warum „Bausteine“?

Vom Beginn der Arbeiten an war allen Beteiligten klar, dass sehr viel Zeit vergehen würde, bis eine Grammatik mit diesen Ansprüchen als umfassendes Buch vorgelegt werden könnte, und es sprachen mehrere gewichtige Argumente

dagegen, mit der Publikation der Teilergebnisse zu warten, bis die gesamte *Korpusgrammatik* auf einmal erscheinen kann.

Wie andere Grammatiken kann auch die *Korpusgrammatik* nie wirklich vollständig werden – besonders offensichtlich wird dies angesichts der Ziele, die Methodik nachvollziehbar zu gestalten und auf die Deskriptionstiefe zu setzen (die im Prinzip vor keiner natürlichen Grenze haltmachen muss). Es ist sinnvoll, sich nicht nur damit abzufinden, sondern sich an den Umstand auch anzupassen: Da eine Grammatik mehr oder weniger fragmentarisch bleiben muss und sowieso immer ergänzt werden kann, spricht wenig dagegen, die *Korpusgrammatik* von vornherein in nach und nach erscheinenden Einzelteilen, sprich Bausteinen, zu veröffentlichen, die sich mit der Zeit zu einem immer vollständigeren (aber nie gänzlich vollständigen) Bild der deutschen Grammatik zusammensetzen.

Die zeitnahe Veröffentlichung von Grammatikkapiteln sichert deren Aktualität. Die ihnen zugrunde liegenden Untersuchungen orientieren sich am aktuellen Forschungsstand, und ihre Ergebnisse sind eigentlich Beiträge im laufenden Forschungsdiskurs. In einigen Jahren allerdings werden die Untersuchungen ihre innovative Wirkung verloren und die Ergebnisse an Aktualität eingebüßt haben. Im ungünstigsten Fall werden die Ergebnisse schon überholt sein, wenn man mit ihrer Veröffentlichung darauf wartet, bis alle vorgesehenen Kapitel der Grammatik vorliegen. Das spricht für die „Zwischenveröffentlichung“ bereits erstellter Kapitel.

Wird im Team gearbeitet, übernehmen einzelne Autor/inn/en typischerweise verschiedene thematische Bereiche. Sie bearbeiten sie dann oft parallel. Die Grammatik entsteht also nicht unbedingt linear entlang einer angestrebten Gliederung. Die einzelnen Kapitel können außerdem – auch wenn man eine formale und inhaltliche Einheitlichkeit anstrebt – spürbar durch die Persönlichkeit der jeweiligen Autor/inn/en geprägt sein.<sup>9</sup> Beides begründet oft eine gewisse Eigenständigkeit der Kapitel, die zu ihrer separaten Publikation in den „Bausteinen“ gut passt. Die Kapitel, die für eine Ausgabe der „Bausteine“ ausgewählt werden, müssen nicht thematisch zusammenhängen. Auch eignen sich nicht alle Kapitel dafür, in die „Bausteine“ aufgenommen zu werden. Die Kriterien, nach denen Kapitel für die „Bausteine“ ausgesucht werden, sind ein ausreichender linguistischer Informationsgehalt und eben gerade ein gewisses Maß an Eigenständigkeit.<sup>10</sup>

9 Vgl. die Erfahrungen der Autor/inn/en der GDE (2017: 8).

10 So dienen manche Überblickskapitel (vgl. Abschnitt 5) nicht so sehr der linguistischen Beschreibung wie einer Vorstrukturierung der Themen, die in den darauf folgenden Kapiteln beschrieben werden.

## 5 Sprachliche Daten, Untersuchungsmethoden, Präsentationsformen und Qualitätssicherung

Die den „Bausteinen“ zugrunde liegenden Untersuchungen basieren vor allem auf dem Untersuchungskorpus des IDS-Projekts „Korpusgrammatik – grammatische Variation im standardsprachlichen und standardnahen Deutsch.“ Das Untersuchungskorpus umfasst zurzeit 23 Mio. Dokumente, 5,8 Mrd. Wörter sowie 7 Mrd. Tokens und stellt einen Ausschnitt des Deutschen Referenzkorpus (DeReKo, vgl. Kupietz et al. 2018), Release 2017-II, dar. Es enthält vor allem klassische Druckerzeugnisse und wurde im Rahmen des Projekts mit Dokumentmetadaten wie ‚Land‘, ‚Region‘, ‚Register‘, ‚Medium‘, ‚(inhaltliche) Domäne‘ und ‚Entstehungszeit‘ angereichert und stratifiziert (Bubenhofer et al. 2014: 54–76). Seine aktuelle Zusammensetzung ist unter <https://grammis.ids-mannheim.de/korpusgrammatik/6616> (abgerufen am 23.7.2020, DOI: 10.14618/korpusgrammatik) beschrieben. Mitunter werden Teilkorpora des Untersuchungskorpus gebildet, die in den einzelnen Bausteinen an einschlägigen Stellen spezifiziert werden. Um Internettexte besser in den Blick zu bekommen, werden ergänzend Teilkorpora des DECOW16 (Schäfer & Bildhauer 2012; Schäfer 2015) benutzt.

Das Untersuchungskorpus verfügt über zahlreiche Annotationsebenen. In dieser Ausgabe der „Bausteine“ wurden insbesondere Annotationen nach Lemma und Wortart (v. a. TreeTagger, vgl. Schmid 1994, und SMOR, vgl. Schmid et al. 2004), morphologischen Merkmalen (MarMoT, vgl. Müller et al. 2014), Wortbildung (Canoo)<sup>11</sup> sowie Eigennamen (Stanford NER, Finkel et al. 2005) verwendet (vgl. <https://grammis.ids-mannheim.de/korpusgrammatik/6625>, abgerufen am 23.7.2020, DOI: 10.14618/korpusgrammatik). Für die einzelnen Untersuchungsthemen werden die sprachlichen Daten auch spezifisch nachannotiert, was an entsprechenden Stellen beschrieben wird.

In den statistischen Analysen werden in dieser Ausgabe der „Bausteine“ verschiedene Maße und Verfahren benutzt, darunter Chi-Quadrat-Tests und Pearson-Residuen mit entsprechenden Visualisierungen, Baayens (2009: 902) *potenzielle Produktivität* als morphologisches Produktivitätsmaß, Modellierungen von Worthäufigkeitsverteilungen mithilfe des Softwarepakets zipfR (Evert & Baroni 2007) sowie binäre logistische Regressionen. Bei der statistischen Modellierung und Visualisierung kommt insbesondere die freie Open-Source-Software R (R Core Team 2020) zum Einsatz. Neben den statistischen Analysen werden umfangreiche und vielschichtige qualitative Sichtungen der Daten vorgenommen, die öfter auch explorative Vorstufen für quantitative Analysen bilden. Die

11 Siehe <https://web.archive.org/web/20200326230620/http://canoonet.eu/> (abgerufen am 10.5.2020).

Kenndaten der Analysen sowie Hintergrunderläuterungen befinden sich an einschlägigen Stellen im Haupttext oder sind bei logistischen Regressionen in die Anhänge ausgelagert.

Die Arbeit mit der korpuslinguistischen Methodik und die Notwendigkeit, die statistischen Methoden genau darzustellen, werden die Kapiteltypen in der *Korpusgrammatik* und damit die Präsentationsformen der „Bausteine“ beeinflussen. Wie in den meisten anderen Grammatiken sollen auch in der *Korpusgrammatik* größere Themenbereiche soweit möglich in der Abfolge vom Allgemeinen zum Besonderen behandelt werden. Dies wird sich in verschiedenen Typen von Kapiteln spiegeln. An einem Ende der Skala steht dann ein Überblickskapitel, in dem ein ganzer Phänomenbereich unter Verwendung korpuslinguistischer Perspektiven gesichtet wird und in dem auf interessante Variationsbereiche hingewiesen wird, am anderen Ende ein Kapitel, in dem ein spezielles Variationsphänomen eingehend analysiert wird. Ein solches Kapitel kommt schon einer Fallstudie nahe: Darin werden oft die Einflussgrößen des konkreten Variationsfalls ermittelt, statistisch geprüft und in ihrer Effektstärke bestimmt. Zwischen diesen beiden Kapiteltypen stehen einerseits Überblickskapitel, die auf niedrigeren Abstraktionsstufen angesiedelt sind, und andererseits solche Kapitel, die die Variabilität der Grammatik auf abstrakteren Stufen untersuchen als der eines konkreten Variationsphänomens. Tendenziell werden in Überblickskapiteln mehr deskriptiv-statistische und explorative Methoden angewendet. Die inferenzstatistischen Methoden nehmen zu, je näher das Kapitel an die Fallstudie herankommt. Hier wächst auch der Bedarf an Dokumentation der Methoden. In dieser Ausgabe der „Bausteine“ erscheinen zwei Überblickskapitel unterschiedlichen Abstraktionsgrades und zwei Kapitel, die einer Fallstudie nahekommen (vgl. Abschnitt 7 weiter unten).

Mit den „Bausteinen“ wird eine Veröffentlichungsstrategie verfolgt, die den Mehrwert einer Online-Publikation ausnutzt, um die zugrunde liegenden Untersuchungen im intendierten Sinne nachvollziehbar und damit anschlussfähig zu machen. Anders als es bei bisherigen Grammatiken des Deutschen der Fall ist, werden bei ausgewählten Themen die ausgewerteten Sprachdaten online veröffentlicht und mit den „Bausteine“-Texten verlinkt. Ergebnisse komplexerer korpuslinguistisch-statistischer Analysen werden direkt in den Bausteinen dokumentiert oder – bei größerem Umfang – ebenfalls online zugänglich gemacht. Mit der Wahl eines Open-Access-Formats sowie mit der Veröffentlichung von Datensätzen (Open Data) und Dokumentationen statistischer Analysen (Open Methodology) trägt die Reihe zur offenen Wissenschaft bei.

Durch die geschilderten Maßnahmen wird die Entstehung der *Korpusgrammatik* weitgehend transparent gehalten, sodass den Autor/inn/en bei ihren grammatikografischen Bemühungen gewissermaßen zugesehen werden kann. So können auch die Reaktionen aus der Forschungsgemeinde auf das Vorhaben

*Korpusgrammatik* zurückwirken. Eine besondere Rolle im Entstehungsprozess der „Bausteine“ kommt dem [Wissenschaftlichen Beirat](#) zu. Er setzt sich aus namhaften Kolleg/inn/en zusammen, die über besondere Expertise in der Grammatikografie des Deutschen einerseits und/oder in der Korpus- und Computerlinguistik andererseits verfügen und sich bereit erklärt haben, uns auf dem Weg zu einer *Korpusgrammatik* zu begleiten.

## 6 Geplanter Aufbau der *Korpusgrammatik*

Die Bausteine müssen modular in eine Makrostruktur einpassbar sein, die der anvisierten Gliederung der künftigen *Korpusgrammatik* entspricht. Geplant ist eine in ihren Grundzügen traditionelle Gliederung mit zwei Hauptteilen. Im ersten wird die Grammatik von den herkömmlichen grammatisch bestimmbareren Ausdrucksklassen her betrachtet, die unterhalb der Satzebene liegen, wie Wort, Substantiv oder Nominalphrase. Im zweiten Hauptteil bildet der Satz als Realisierung der kommunikativen Minimaleinheit den Ausgangspunkt, und es werden dort syntaktische Funktionen und Relationen untersucht, die typischerweise auf der Satzebene zu beobachten sind. Wenn man umschließende Gliederungseinheiten wie Vorwort, eine allgemeine Einführung, ein einleitendes Kapitel zur korpuslinguistischen Basis und Methodik, ein Literaturverzeichnis und mögliche Register erst einmal beiseitelässt, sieht die Gliederung für den ersten Hauptteil der *Korpusgrammatik* folgende Großkapitel vor:

1. Wortbildung
2. Wörter und Wortgruppen
3. Substantiv, Nomen und Nominalphrase
4. Adjektiv und Adjektivphrase
5. Adverb und Adverbphrase
6. Andere nicht flektierbare Wortarten
7. Verb, Verbalkomplex, Verbgruppe (Verbalphrase)

Der erste Hauptteil ist somit als eine „Wort- und Phrasengrammatik“ konzipiert (zu einer in Grundzügen ähnlichen Anlage vgl. z.B. die korpusgestützte Grammatik des Englischen von Biber et al. 2006 [1999]: v.a. 44). Im Hinblick auf die grammatisch bestimmten Ausdruckseinheiten folgt der Aufbau im Kern einer aszendenten Betrachtungsweise, die die Reihenfolge Morphologie – Syntax vorgibt. Dabei wird zuerst in einem getrennten Großkapitel die Wortbildung behandelt. In jedem der darauf folgenden Großkapitel zu flektierbaren Wörtern und darauf basierenden Wortgruppen sollen dann die Bereiche Flexion und Syntax nach einer einführenden Übersicht über Flexionskategorien und ihre Realisierungsformen zu einer Morphosyntax zusammenfließen, sobald es um die

Betrachtung konkreterer Variationsfälle geht. Damit steht im ersten Hauptteil eine eher an formalen Aspekten orientierte Perspektive im Vordergrund.

Die Gliederung des zweiten Hauptteils soll durch Themen wie die folgenden bestimmt sein:

1. Kommunikative Minimaleinheit, Satz, Satzkomponenten
2. Komplementation
3. Supplementation
4. Kongruenz und Rektion
5. Wortstellung
6. Informationsstruktur
7. Komplexere Strukturen (Asyndese, Syndese, Koordination, Subordination, komplexe Sätze)

Zunächst wird mit dem Satz als einer auf dem finiten Verb basierenden, prototypischen Realisierung der kommunikativen Minimaleinheit (dazu Zifonun 1997a: 86; 1997b: 608) die zentrale Bezugsgröße der Beschreibung etabliert. Man kann somit beim zweiten Hauptteil tendenziell von einer „Satzgrammatik“ sprechen. Allerdings werden darin im Weiteren syntaktische Funktionen und Relationen thematisiert, die zwar typischerweise, aber nicht nur in Sätzen eine Rolle spielen. Der Aufbau ist bis in die Großkapitel Komplementation und Supplementation hinein deszendend. In diesen beiden Großkapiteln rückt dann langsam die funktionale Perspektive als Ordnungsprinzip in den Vordergrund, denn syntaktische Funktionen werden nicht nur für die Satzebene postuliert, sondern auch für Ausdruckseinheiten, die unterhalb der Satzebene liegen (vgl. z. B. Verbkomplemente einerseits und Komplemente in der Nominal- oder Adjektivphrase andererseits). So sind auch die Relationen der Kongruenz und der Rektion für Ausdruckseinheiten unterschiedlichen Ranges relevant.<sup>12</sup> Die folgenden, ebenfalls relational bzw. funktional interpretierbaren Themen werden (wie „Wortstellung“ und „Informationsstruktur“) üblicherweise im Kontext von Sätzen behandelt oder betreffen (wie „Komplexere Strukturen“) ein Spektrum von Ausdruckseinheiten, die die Satzebene sowohl deutlich unter- als auch überschreiten können. Der Schwerpunkt im zweiten Hauptteil liegt somit insgesamt auf dem Vergleich der verschiedenen Realisierungsformen für syntaktische Funktionen und Relationen, die unterschiedlichen grammatischen Ebenen angehören können.

12 Zur Kongruenz vgl. z. B. die Numeruskongruenz zwischen Subjekt und finitem Verb auf der Satzebene und die Numerus- und Kasuskongruenz zwischen Determinierer und Nomen in einer Nominalphrase, zur Rektion etwa die Rektion der Verbkomplemente durch das Verb auf der Satzebene oder die Rektion des Kasus einer Nominalphrase durch die Präposition in einer Präpositionalphrase.

Die vorgestellte Gliederung ist tentativ. Sie ist modifizierbar und vor allem in Bezug auf den zweiten Hauptteil ergänzbar. Die Formulierungen der Kapitelüberschriften sind vorläufig. Die Gliederung muss einigermaßen flexibel sein, damit die Autor/inn/en mit Änderungen und Erweiterungen auf sich eventuell durch die Forschungsdynamik ergebende Entwicklungen reagieren können, die in ihrer Arbeit an der *Korpusgrammatik* oder in der aktuellen Forschung im Allgemeinen entstehen.

Die Zweiteilung der *Korpusgrammatik* spiegelt die Doppelperspektivik, nach der sich die Darstellung richten soll: Zum einen wird bei der Grammatikbeschreibung von formbezogen bestimmten Einheiten ausgegangen (so tendenziell im ersten Hauptteil), zum anderen von funktional bestimmten Einheiten bzw. bestimmten Funktionen (so tendenziell im zweiten Hauptteil). Die eine Perspektive involviert, dass formbezogene Einheiten in Bezug auf ihre verschiedenen Funktionen untersucht werden, d.h. eine Art semasiologische Vorgehensweise verfolgt wird, die andere Perspektive, dass für bestimmte Funktionen verschiedene Realisierungsformen ermittelt werden, d.h. quasi onomasiologisch vorgegangen wird.<sup>13</sup> Diese Doppelperspektivik bekommt unter dem Aspekt der Variation einen besonderen Sinn: Beide Male wird nach Varianten geforscht, nur sind diese in ersterem Fall funktional und in letzterem Fall formal ausdifferenziert (vgl. Fuß et al. 2017: 229–233). Interessanterweise hat sich in der modernen Linguistik unter dem Begriff Variation nur die zweite Perspektive richtig etablieren können, wozu entscheidend die soziolinguistischen Arbeiten Labovs (1960: 15–17) zur „linguistischen Variante“ beigetragen haben. Wir allerdings wollen auf das innovative Potenzial einer systematischen „reziproken“ Variationsbetrachtung setzen und den bewussten Perspektivenwechsel zum festen Bestandteil unserer grammatikografischen Herangehensweise machen. Die beiden Perspektiven lassen sich auf der Ebene der Hauptteile und der Großkapitel bis zu einem gewissen Grad auseinanderhalten. In kleineren Gliederungseinheiten soll demgegenüber ein zyklischer Perspektivenwechsel es möglich machen, in der grammatischen Erkenntnis immer weiter vorzudringen (dazu Fuß et al. 2017: 235).

## 7 Bausteine dieser Ausgabe

Geplant sind jährlich bis zu zwei Ausgaben der „Bausteine“ in einer Online- und einer Print-on-Demand-Version. Nur diese vorliegende Ausgabe enthält eine längere Einleitung – sie führt in die gesamte Reihe ein. Die Bausteine der

13 Zur Übertragung der semasiologischen und onomasiologischen Perspektive auf die Betrachtung der Variation vgl. Ammon (2005: 30) und schon von der Gabelentz (1995 [1901]: 84–86).

vorliegenden Ausgabe sind als Kapitel verschiedener Hierarchieebenen aus den Großkapiteln „Wortbildung“ und „Substantiv, Nomen und Nominalphrase“ angelegt. Wie sie sich in die bisher geplante Struktur der Großkapitel einfügen, zeigen die folgenden Übersichten. Auf der oberen, nummerierten Ebene erscheinen die Titel aller bisher vorgesehenen Kapitel dieser Ebene. Auf der unteren, nicht nummerierten Ebene hingegen sind nur diejenigen Kapitel aufgeführt, die als Bausteine publiziert sind. Die Titel aller vier Kapitel, die den Bausteinen dieser Ausgabe entsprechen, sind hervorgehoben:

### Wortbildung

1. **Grundlegende Aspekte der Wortbildung**
2. Wortbildungsarten
3. Spezialbereiche der Wortbildung
  - **Bau von und Umbau zu Adverbien: Präpositionen, Vergleiche und Flexion**

### Substantiv, Nomen und Nominalphrase

1. Flexion des Nomens
2. Aufbau der Nominalphrase
3. Determination und Quantifikation
4. Attribution
  - **Starke vs. schwache Flexion aufeinanderfolgender attributiver Adjektive: mit hohem technischen/technischem Aufwand**
  - **Reihenfolge attributiver Adjektive**
5. Syntaktische Funktionen von Nominalphrasen

Der Baustein **Grundlegende Aspekte der Wortbildung** eröffnet das Kapitel Wortbildung. Er besitzt dementsprechend einen expositorischen Charakter und ist als Überblickskapitel angelegt. Er zeigt grundlegende Mechanismen der Wortbildung im Deutschen, erkundet vor allem deskriptiv-statistisch Relationen (u. a. die Produktivitätsverhältnisse betreffend) zwischen Wortbildungen und primären Wörtern und präsentiert allgemeine Wortbildungsprofile der Hauptwortarten. Der Baustein **Bau von und Umbau zu Adverbien: Präpositionen, Vergleiche und Flexion** ist als Überblickskapitel einer etwas tieferen Ebene angelegt. Er behandelt ein für die empirische wie auch die theoretische Forschung besonders herausforderndes Thema. Die Wortbildung des Adverbs ist nicht allzu häufig Forschungsgegenstand und ist daher besonders geeignet, in die „Spezialbereiche der Wortbildung“ aufgenommen zu werden. Der Baustein gibt einen deskriptiv-statistisch fundierten Überblick über die Bildungsmuster des komplexen Adverbs. Besonders fokussiert werden Adverbien, die mit Formativen enden, die in der Flexion vorkommen, und solche mit präpositionalen

Bestandteilen (vgl. für beide Adverbklassen z.B. *zunächst*). Auf quantitativen und qualitativen Korpusauswertungen bauen theoretische Überlegungen auf, die grundlegende Bedingungen für die Entstehung komplexer Adverbien zum Gegenstand haben und Prozesse aufdecken, die komplexe Adverbien mit Präpositionalphrasen korrelieren. Mit **Starke vs. schwache Flexion aufeinanderfolgender attributiver Adjektive: mit hohem technischen/technischem Aufwand** wird ein Baustein aus dem Bereich Nominalphrase vorgelegt. Darin wird ein konkreter Variationsfall behandelt, der es auch zu einem notorischen Zweifelsfall des Sprachgebrauchs gebracht hat. Mithilfe logistischer Regressionen wird nach Einflussgrößen für die Wahl zwischen starker und schwacher Flexion des zweiten Adjektivs gesucht. Berücksichtigt werden sprachsystematische Bedingungen wie die Möglichkeit einer Koordination der beiden Adjektive sowie extralinguistische Einflüsse wie das Herkunftsland des Belegs. Im Anhang erscheint die Dokumentation der statistischen Modellierung, die die Analysen nachvollziehbar macht, aber aus dem Haupttext ausgelagert werden musste, um ihn besser lesbar zu machen. Ein Teil der Sprachdaten, die der Untersuchung zugrunde liegen, ist – angereichert um die untersuchungsrelevanten Annotationen – online veröffentlicht (vgl. Abschnitt 8 unten). Die Sprachdaten wurden im Rahmen dieser Untersuchung bereits so weit aufbereitet, dass sich auch der nächste Baustein **Reihenfolge attributiver Adjektive** darauf stützen konnte. In diesem Baustein wird geprüft, wie bei Abfolgen aus zwei Adjektiven wie in *mit begeistertem, minutenlangem Beifall* die Reihenfolge gesteuert wird. Es zeigt sich, dass dabei semantische Aspekte mit anderen Einflussgrößen interagieren wie etwa der Silbenanzahl oder der Häufigkeit der involvierten Adjektive. Die eingesetzte Regressionsanalyse wird auch bei diesem Baustein in einem Anhang genauer beschrieben.

Die Beiträge der Reihe „Bausteine einer Korpusgrammatik des Deutschen“ werden weitgehend wie Kapitel einer korpusgestützten, variationssensitiven Grammatik gestaltet, aber dennoch nicht ganz so, wie sie in einem bereits fertiggestellten Werk erscheinen würden. Vor allem bei dieser ersten Ausgabe muss vieles an Ort und Stelle erklärt werden, was ansonsten in ein anderes Kapitel hätte ausgelagert werden können. Dies betrifft in besonderer Weise Ausführungen zur Methodik, deren Explizitmachung uns so wichtig erscheint. Ein Teil dieser Ausführungen wird sich in einem Kapitel zur korpuslinguistischen Basis und Methodik der ganzen *Korpusgrammatik* wiederfinden, da einige methodische Aspekte konstant bleiben und einige Verfahren immer wieder angewendet werden. Jetzt schon einen Baustein zu schreiben, der einem solchen Kapitel entspricht, ist jedoch verfrüht, da sich das Spektrum der eingesetzten Methoden mit weiteren Themen der *Korpusgrammatik* und damit einhergehenden neuen Problemstellungen um einiges erweitern wird. Dadurch, dass die *Korpusgrammatik* mit der Publikation der „Bausteine“ portionsweise

entsteht, sind also gewisse Redundanzen vorprogrammiert. Die Leser/innen mögen durch die Aktualität der präsentierten Forschungsergebnisse, die relative Eigenständigkeit einzelner Bausteine und die Möglichkeit, das Entstehen der *Korpusgrammatik* mit frühzeitigen Reaktionen zu begleiten, entschädigt sein.

Die einzelnen Bausteine haben verschiedene Autor/inn/en, die sich auf ein gemeinsames inhaltliches Programm und die wichtigsten Aspekte der formalen Gestaltung geeinigt haben. Auf eine stilistische Vereinheitlichung der Texte wurde bewusst verzichtet. Wir hoffen, dass Stilunterschiede die Leser/innen von den Inhalten nicht ablenken werden.

## 8 Veröffentlichte Forschungsdaten zu dieser Ausgabe und Verlinkung mit Internetquellen

Die Bausteine **Starke vs. schwache Flexion aufeinanderfolgender attributiver Adjektive: mit hohem technischen/technischem Aufwand** und **Reihenfolge attributiver Adjektive** basieren in wichtigen Teilen auf dem aufbereiteten Datensatz `ATTRIBUTIVE_ADJEKTIVE_1`. Dieser ist die Grundlage für die „Datenbank attributive Adjektive“, die unter der folgenden Adresse öffentlich zugänglich ist:

<https://grammis.ids-mannheim.de/attributiveadjektivedb>  
(abgerufen am 11.7.2020)  
DOI: 10.14618/attributiveAdjektiveDB

Die Datenbank besteht aus 1.598 Belegen zu artikellosen Nominalphrasen mit je zwei attributiven Adjektiven im Dativ Singular Maskulinum oder Neutrum, angereichert durch zahlreiche Annotationen. Die komplette Datensammlung kann auch heruntergeladen und für eigene Anschlussuntersuchungen verwendet werden.

In den in einem Open-Access-Format erscheinenden „Bausteinen“ soll der Leserschaft nach Möglichkeit der direkte Zugang zu allen zitierten Internetquellen ermöglicht werden. Alle Internetadressen sind mit einem Abrufdatum versehen und in den online erscheinenden Texten als Links aktiv. Sofern den Quellen ein Permalink oder ein DOI zugeordnet sind, bemühen wir uns, diese anzugeben. Zusätzlich werden alle zitierten Internetseiten in [Internet Archive](#)<sup>14</sup> gespeichert. Dort sollten alle Versionen mit dem angegebenen Abrufdatum einsehbar sein.

14 Vgl. <https://web.archive.org/>, abgerufen am 27.7.2020.

## 9 Danksagung

Die erste Ausgabe der „Bausteine“ hätte ohne die tatkräftige und inhaltlich sehr anregende Hilfe vieler Personen nicht und vor allem nicht in der vorliegenden Form zustande kommen können. Der Dank der Autor/inn/en und Herausgebenden gilt den Mitgliedern des wissenschaftlichen Beirats der Reihe für die gesamte Unterstützung der Unternehmung und insbesondere für die Begutachtung der Bausteine dieser Ausgabe, Kristin Kopf und Ulrich H. Waßner für die aufwendige Redaktion der Beiträge, Thilo Weber für die Organisation der reibungslosen Zusammenarbeit mit Heidelberg University Publishing, der Leitung und dem gesamten Team von Heidelberg University Publishing für die Ermöglichung und die Umsetzung der Open-Access-Publikation, einschließlich des Satzes, des Designs und Aufbaus des Web-Auftritts sowie der Realisierung der Print-Ausgabe, Stefan Falke für die Programmierung der „Datenbank attributive Adjektive“ und unseren Hilfskräften Nagehan Cetin, Frederik Dumele, Oskar Jabs, Anna Konovalova, Dang Hoang Nguyen und Viktoria Arendt für die Unterstützung in allen Phasen der Untersuchungen und der Beitragserstellung.

ORCID®

Marek Konopka  <https://orcid.org/0000-0001-9627-0918>

## Literatur

- Ammon, Ulrich. 2005. Standard und Variation. Norm, Autorität, Legitimation. In Ludwig M. Eichinger & Werner Kallmeyer (Hgg.), *Standardvariation. Wie viel Variation verträgt die deutsche Sprache?* (Jahrbuch des Instituts für Deutsche Sprache 2004), 28–40. Berlin & New York: De Gruyter
- Baayen, Harald R. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (Hgg.), *Corpus Linguistics: An International Handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 29), Bd. 2, 899–919. Berlin: De Gruyter.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 2006 [1999]: *Longman grammar of spoken and written English*. Harlow: Longman.
- Bubenhofer, Noah, Marek Konopka & Roman Schneider. 2014. *Präliminarien einer Korpusgrammatik*, unter Mitwirkung von Caren Brinckmann, Katrin Hein & Bruno Strecker (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache, CLIP 4). Tübingen: Narr.

- Carter, Ronald & Michael McCarthy. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Conrad, Susan. 2010. What can a corpus tell us about grammar? In Anne O’Keeffe & Michael McCarthy (Hgg.), *The Routledge handbook of corpus linguistics* (Routledge Handbooks in Applied Linguistics), 227–240. London & New York: Routledge.
- Dudengrammatik 2016 = Wöllstein, Angelika & Dudenredaktion (Hgg.). 2016. *Dudenband 4 – Die Grammatik*, 9. Aufl. Berlin: Dudenverlag.
- Eichinger, Ludwig M. 2017. Standarddeutsch – die beste aller möglichen Sprachen. In Marek Konopka & Angelika Wöllstein (Hgg.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung* (Jahrbuch des Instituts für Deutsche Sprache 2016), 3–22. Berlin & Boston: De Gruyter.
- Eisenberg, Peter. 2013a. *Grundriss der deutschen Grammatik. Bd. 1: Das Wort*, 4. Aufl. Stuttgart & Weimar: Metzler.
- Eisenberg, Peter. 2013b. *Grundriss der deutschen Grammatik. Bd. 2: Der Satz*, 4. Aufl. Stuttgart & Weimar: Metzler.
- Engel, Ulrich. 2009. *Deutsche Grammatik*, Neubearbeitung, 2. Aufl. München: Iudicium.
- Evert, Stefan & Marco Baroni. 2007. zipfR: Wordfrequency distributions in R. In Sophia Ananiadou (Hg.), *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*. 29–32. Prag: ACL. <https://www.aclweb.org/anthology/P07-2008.pdf> (abgerufen am 7.7.2020).
- Felder, Ekkehard. 2016. *Einführung in die Varietätenlinguistik* (Germanistik kompakt). Darmstadt: WBG.
- Finkel, Jenny Rose, Trond Grenager & Christopher Manning. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In Kevin Knight, Hwee Tou Ng & Kemal Oflazer (Hgg.), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, 363–370. Ann Arbor: Association for Computational Linguistics. <http://nlp.stanford.edu/~manning/papers/gibbscrf3.pdf> (abgerufen am 18.7.2020).
- Fuß, Eric, Marek Konopka & Angelika Wöllstein. 2017. Perspektiven auf syntaktische Variation. In Marek Konopka & Angelika Wöllstein (Hgg.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung* (Jahrbuch des Instituts für Deutsche Sprache 2016), 229–254. Berlin & Boston: De Gruyter.
- Fuß, Eric, Marek Konopka & Angelika Wöllstein (Hgg.). 2019. Grammatik im Korpus. Korpuslinguistisch-statistische Analysen morphosyntaktischer Variationsphänomene (Studien zur deutschen Sprache 80). Tübingen: Narr.

- Fuß, Eric, Marek Konopka, Beata Trawiński & Ulrich Hermann Waßner (Hgg.). 2018. *Grammar and corpora 2016*. Heidelberg: Heidelberg University Publishing. <https://heiu.uni-heidelberg.de/catalog/book/361> (abgerufen am 20.7.2020).
- GDE = Gunkel Lutz, Adriano Murelli, Susan Schlotthauer, Bernd Wiese & Gisela Zifonun (Hgg.). 2017. *Grammatik des Deutschen im europäischen Vergleich. Das Nominal* (Schriften des Instituts für Deutsche Sprache 14). 2 Bände. Berlin: De Gruyter.
- GDS = Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker (Hgg.). 1997. *Grammatik der deutschen Sprache* (Schriften des Instituts für Deutsche Sprache 7). 3 Bände. Berlin & New York: De Gruyter.
- Habermann, Mechthild. 2010. Was ist eigentlich „Grammatik“? – Eine Einführung. In Mechthild Habermann (Hg.), *Grammatik wozu? Vom Nutzen des Grammatikwissens in Alltag und Schule* (Thema Deutsch 11), 9–14, Mannheim & Zürich: Dudenverlag.
- Hentschel, Elke & Harald Weydt. 2013. *Handbuch der deutschen Grammatik*, 4. Aufl. Berlin & Boston: De Gruyter.
- Hoffmann, Ludger. 1997. Einführung. In Gisela Zifonun, Ludger Hoffmann & Bruno Strecker (Hgg.), *Grammatik der deutschen Sprache* (Schriften des Instituts für Deutsche Sprache 7), 3 Bände. 1–16. Berlin & New York: De Gruyter.
- Institut für Deutsche Sprache (2017): *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2017-II* (Release vom 1.10.2017). Mannheim: Institut für Deutsche Sprache. <https://www.ids-mannheim.de/DeReKo> (abgerufen am 20.7.2020).
- Konopka, Marek. 2018. Korpuslinguistik, Grammatiktheorie, Grammatikographie. In Angelika Wöllstein, Peter Gallmann, Mechthild Habermann & Manfred Krifka (Hgg.), *Grammatiktheorie und Empirie in der germanistischen Linguistik* (Germanistische Sprachwissenschaft um 2020 1), 151–184. Berlin & Boston: De Gruyter.
- Konopka, Marek, Jacqueline Kubczak, Christian Mair, František Štícha & Ulrich Hermann Waßner (Hgg.). 2011. *Grammatik und Korpora 2009. Dritte internationale Konferenz. Mannheim, 22.–24.09.2009* (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache, CLIP 1). Tübingen: Narr.
- Kupietz, Marc, Harald Lünen, Paweł Kamocki & Andreas Witt. 2018. The German Reference Corpus DeReKo: New Developments – New Opportunities. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (Hgg.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 4353–4360. Miyazaki:

- European Language Resources Association (ELRA). <https://www.aclweb.org/anthology/L18-1689.pdf> (abgerufen am 20.7.2020).
- Labov, William. 1966. The linguistic variable as a structural unit. *Washington Linguistic Review* 3, 4–22.
- Meurers, W. Detmar & Stefan Müller. 2009. Corpora and syntax. In Anke Lüdeling & Merja Kytö (Hgg.), *Corpus Linguistics: An International Handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft 29), Bd. 2, 920–933, Berlin: De Gruyter.
- Mukherjee, Joybrato. 2006. Corpus linguistics and English reference grammars. In Antoinette Renouf & Andrew Kehoe (Hgg.), *The changing face of corpus linguistics. Papers from the 24th International Conference on English Language Research on Computerized Corpora (ICAME 24)*, 337–354. Amsterdam: Rodopi.
- Müller, Thomas, Helmut Schmid & Hinrich Schütze. 2013. Higher-order CRFs for morphological tagging. In David Yarowsky, Timothy Baldwin, Anna Korhonen, Karen Livescu & Steven Bethard (Hgg.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 322–332. Seattle: Association for Computational Linguistics.
- Paranhos Zitterbart, Jussara. 2010. Grammatiken des Deutschen im Vergleich. In Mechthild Habermann (Hg.), *Grammatik wozu? Vom Nutzen des Grammatikwissens in Alltag und Schule* (Thema Deutsch 11), 66–81, Mannheim & Zürich: Dudenverlag.
- R Core Team. 2020. *R: A language and environment for statistical computing*, Wien: R Foundation for Statistical Computing. <http://www.R-project.org/> (abgerufen am 26.6.2020).
- Schäfer, Roland. 2015. Processing and Querying Large Web Corpora with the COW<sub>14</sub> Architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lungen, & Andreas Witt (Hgg.), *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3)*, 28–34. Mannheim: Institut für Deutsche Sprache.
- Schäfer, Roland & Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (Hgg.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. Istanbul: European Language Resources Association (ELRA).
- Schmid Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing*. Manchester: o. V. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf> (abgerufen am 11.7.2020).
- Schmid, Helmut, Arne Fitschen & Ulrich Heid. 2004. SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. In

- Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (Hgg.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, 1263–1266. Lissabon: European Language Resources Association (ELRA). <https://www.cis.uni-muenchen.de/~schmid/papers/smor.pdf> (abgerufen am 18.6.2020).
- von der Gabelentz, Georg. 1995 [1901], Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse, 2. Aufl. In: Christopher Hutton (Hg.): *History of linguistics. 18th and 19th Century German linguistics*, Bd. 7. London: Routledge/Thoemmes Press.
- Zifonun, Gisela. 1997a. Kommunikative Minimaleinheit und Satz. In Gisela Zifonun, Ludger Hoffmann & Bruno Strecker (Hgg.), *Grammatik der deutschen Sprache* (Schriften des Instituts für Deutsche Sprache 7), 3 Bde., 85–97. Berlin & New York: De Gruyter.
- Zifonun, Gisela. 1997b. Der Modus kommunikativer Minimaleinheiten. In Gisela Zifonun, Ludger Hoffmann & Bruno Strecker (Hgg.), *Grammatik der deutschen Sprache* (Schriften des Instituts für Deutsche Sprache 7), 3 Bde., 605–675. Berlin & New York: De Gruyter.