

Gohar Schnelle, Mathilde Hennig, Carolin Odebrecht und  
Anke Lüdeling

# Historische Korpora in sprachhistorisch orientierter germanistischer Hochschullehre

## Anregungen, Beispiele, Perspektiven

<https://doi.org/10.1515/bgsl-2023-0012>

**Abstract:** This paper argues for incorporating corpus data into the teaching of historical linguistics. While deeply annotated historical corpora are becoming available and corpus data is already widely used to answer various research questions, corpora are as yet rarely used in teaching. We believe they are ideally suited to make the variation in historical data transparent and help students to explore contexts and parameters. In our first study, we show how the KaJuK corpus and its more elaborated version, the GiesKaNe corpus, can be exploited to study adverbial sentences. Using the RIDGES corpus, the second study deals with phrasal and lexical development. Both studies focus on explaining the method and its extension to other corpora and research questions.

## I. Einleitung

Mit diesem Aufsatz möchten wir zeigen, wie historische Korpora zur Vermittlung von sprachhistorischen Inhalten Einsatz finden können. Die korpusbasierte Lehre sehen und schildern wir hierbei ausdrücklich als *Ergänzung* zu anderen gebräuchlichen Lehrmethoden. Im Gegensatz zu den modular aufgebauten Lehrbüchern bieten Korpora die Flexibilität, mehrere grammatische und außersprachliche Themenbereiche in dynamisch modifizierbarer Kombination zu betrachten. Sie eignen sich daher besonders, wenn mehrere Aspekte eines sprachlichen Phä-

---

**Gohar Schnelle:** Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik, SFB 1412, 416591334, Unter den Linden 6, 10099 Berlin, E-Mail: [schnellg@hu-berlin.de](mailto:schnellg@hu-berlin.de)

**Prof. Dr. Mathilde Hennig:** Justus-Liebig-Universität Gießen, Institut für Germanistik, Otto-Behaghel-Straße 10 B, 35394 Gießen, E-Mail: [mathilde.hennig@germanistik.uni-giessen.de](mailto:mathilde.hennig@germanistik.uni-giessen.de)

**Dr. Carolin Odebrecht:** Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik, Unter den Linden 6, 10099 Berlin, E-Mail: [carolin.odebrecht@hu-berlin.de](mailto:carolin.odebrecht@hu-berlin.de)

**Prof. Dr. Anke Lüdeling:** Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik, SFB 1412, 416591334, Unter den Linden 6, 10099 Berlin, E-Mail: [anke.luedeling@rz.hu-berlin.de](mailto:anke.luedeling@rz.hu-berlin.de)

nomens fokussiert oder ebenenübergreifende bzw. nicht-kanonische Phänomene thematisiert werden. Außerdem bieten Korpora einen besonderen Zugang zu textuellen Phänomenen (Biber 2012). Gleichzeitig ermöglicht ein korpusbasierter Ansatz feinkörnigen Zugriff auf sprachliche Entwicklungen in diesen Bereichen genauso wie eine Feststellung von erwarteten wie nicht erwarteten Kontinuitäten (Fischer 2018).

Historische Korpora finden bereits in vielen Disziplinen und in verschiedenen Kontexten Verwendung. In der Forschung zur historischen Germanistik wurden bereits viele Phänomene und Forschungsfragen korpusbasiert bearbeitet (bspw. Coniglio/Hinterhölzl 2021; Ágel/Sievers 2020; Schnelle 2020; Höllein 2019; Zeige 2019; Kempf 2018), was die Forschung quantitativ und qualitativ bereichert (vgl. bspw. Wegera 2013; Hilpert 2008) und die Diskussion um einen methodisch reflektierten Umgang mit linguistischen Daten hervorgebracht hat (vgl. Maitz 2012).<sup>1</sup> Profitiert hat hierbei vor allem die Variationsforschung, da durch die Digitalisierung und Annotation größerer Textmengen in kurzer Zeit viele Varianten in ihrer Distribution ermittelt und verschiedene Distributionsfaktoren mit Hilfe von im Korpus enthaltenen Metadaten geprüft werden können (s. bspw. Hilpert/Gries 2016). Die wissenschaftliche Nutzung von Korpora wird in vielen wissenschaftlichen Veranstaltungen adressiert.<sup>2</sup>

Im Bereich der synchronen Grammatikanalyse und im modernen Fremdsprachenunterricht werden Korpora bereits vielfältig auch in der Lehre eingesetzt (Hirschmann 2019; Bubenhofer 2011),<sup>3</sup> in der universitären Lehre zur historischen Linguistik gibt es hingegen bisher nur wenig systematische Angebote und Erfahrungsberichte zum Einsatz von Korpora. Eine wichtige Ausnahme bildet das Lehrbuch »Deutsche Sprachgeschichte. Grundzüge und Methoden« von Stefan Hartmann (2018), das einen Einstieg in die Nutzung digitaler historischer Korpora darstellt. Insgesamt ist das didaktische Potential historischer Korpora für die systematische Wissensvermittlung noch ausbaufähig.

Der mögliche didaktische Einsatz von Korpora in der universitären Lehre zur historischen Linguistik kann unter zwei Perspektiven thematisiert werden. Einer-

---

<sup>1</sup> Besonders hervorzuheben und zu würdigen ist hier die auf empirischer Grundlage erstellte Mittelhochdeutsche Grammatik (Klein [u. a.] 2018), die mit einer Revolution auf dem Gebiet der historischen Grammatikschreibung vergleichbar ist.

<sup>2</sup> Dies geschieht in einschlägigen Curricula, aber oft auch außerhalb in regelmäßig angebotenen Sommerschulen wie ›Culture & Technology: European Summer University in Digital Humanities‹ der Universität Leipzig, der ›Lancaster Corpus Linguistics Summer School‹ sowie in zahlreichen einmaligen Angeboten.

<sup>3</sup> Siehe auch die ›Teaching and Language Corpora‹-Konferenzen (<https://en.wikipedia.org/wiki/TaLC>, Aufrufdatum: 14.04.2021).

seits können Studierende den Umgang mit Korpora erlernen und beherrschen dann ein Forschungsinstrument für die eigene empirische Forschung. Andererseits können Lehrende Korpora als didaktisches Instrument einsetzen, um Inhalte zu vermitteln. Für beide Perspektiven mangelt es noch an konkreten korpusbasierten Lehrangeboten. Mit diesem Aufsatz zeigen wir anhand von zwei Fallbeispielen, wie Studierende sich anhand von korpusbasiert ermittelten Belegen Phänomene erschließen<sup>4</sup> und wie Lehrende mit Hilfe von Korpora flexibel Themengebiete gemeinsam mit den Studierenden erarbeiten können.

Die Vermittlung sprachhistorischer Inhalte basiert traditionell auf der Verwendung von Lehrbüchern, Nachschlagewerken, Forschungsliteratur und Originaltexten. Korpora können eine hilfreiche Ergänzung zu diesen Lehrmaterialien bieten. Durch ihren Einsatz können Vorkommen eines Phänomens systematisch gesucht werden, um so:

- 1.) Belege für ein Phänomen mit Informationen zu den Textmetadaten zu ermitteln,
- 2.) die Varianz eines Phänomens zu erfassen,
- 3.) die Frequenz von Varianten zu ermitteln.

Darüber hinaus ist je nach Korpus eine ebenenübergreifende und modellunabhängige Annäherung an linguistische Themenbereiche möglich. Durch die Filtrierung nach Metadaten können Belege bspw. diachron, diatopisch oder diaphasisch betrachtet werden, was sich gerade für die Vermittlung variationslinguistischer Themen anbietet, die möglicherweise infolge fehlender geeigneter Lehrmaterialien wenig Beachtung erfahren.

Die erste Fallstudie (s. Abschnitt IV.1) behandelt verschiedene Varianten des syntaktischen Phänomens *vorangestellter Adverbialsatz*. Das aus dem KaJuK-Korpus stammende Beispiel (1) zeigt eine dieser Varianten, und zwar den resumptiven Bezug auf den vorangestellten Konzessivsatz durch das Korrelat *so*.

- (1) *Obwol die Schweden und die Hesen eines Wesens war,  
so nam doch ein ider, was er bekommen konte.*  
(Bauernleben I, 1636)<sup>5</sup>

---

4 Der Einsatz von Korpora im Unterricht muss nicht zwangsläufig das Ziel verfolgen, Studierende als Korpusexpert\*innen auszubilden. Korpora können als Instrumente für die Anfertigung von Übungen und für die Demonstration von Variation ein integraler Bestandteil des Kurses sein (Grath 2009, S. 49).

5 Bei sämtlichen historischen Belegen handelt es sich um Korpusbelege aus dem Kasseler Junktionskorpus (KAJUK) (Ágel/Hennig 2020) oder dem RIDGES-Korpus (Odebrecht [u. a.] 2017). Siehe

Abschnitt IV.1 zeigt exemplarisch anhand der Korpora KAJUK und GiesKaNe, wie Vorkommen der einzelnen Varianten ermittelt werden können, d. h., wie Annotationsmodelle in einer Mehrebenenarchitektur die Suche nach Belegen für komplexe linguistische Phänomene ermöglichen. Da sowohl der Gegenstand des Fallbeispiels als auch die an diesem Beispiel illustrierten Möglichkeiten der Suche in den beiden Korpora solide linguistische Grundkenntnisse voraussetzen, handelt es sich um einen Anwendungsfall für die akademische Lehre mit fortgeschrittenen Studierenden.

Der zweite Themenbereich steht im Kontext einer linguistischen Deskription einer Textgruppe (Kräutertexte). Es wird gezeigt, wie die Quantität und Qualität des halboffenen funktionalen Bundles *ist gut*<sup>6</sup> im RIDGES-Korpus (Odebrecht [u. a.] 2017, s. Abschnitt III.2) Diskussionsgrundlage für syntaktischen Wandel und Varietätenwandel sein kann. Das dem RIDGES-Korpus entnommene Beispiel (2) enthält einen Beleg von *ist gut*, der hier durch eine Nominalgruppe im Dativ erweitert ist.

- (2) *das hasselwurtz ist gut den frawen zu ihrer krankheit*  
(Johannes Tallat, ArtzneyBuchleinDerKreutter, 1532)

In dieser Studie wird ein Fokus auf das Arbeiten mit Oberflächenformen und deren (korpusbezogener) Frequenz in Verbindung mit im Korpus enthaltenen Metadaten gelegt. Die Studie nutzt die Mehrebenenarchitektur des Korpus<sup>7</sup> (siehe Abb. 1), setzt aber den Schwerpunkt auf das Arbeiten mit Datenaggregaten, die bereits aus dem sprachlichen Kontext abstrahieren. Datenaggregate sind bspw. Frequenzlisten der Bundles, wie wir sie in diesem Beitrag verwenden.

---

hierzu genauer Abschnitt III. Im elektronischen Anhang werden für alle Beispiele die jeweiligen Korpusbelege referenziert.

<sup>6</sup> Der Begriff *lexical bundle* (Biber/Conrad 1999) bezeichnet eine wiederkehrende feste Reihung von lexikalischen Einheiten; mehr dazu siehe Abschnitt IV.2.1.

dipl	das	haffelwurtz	ift	gut	den	frawen	zu	ihrer	kranckheit
clean	das	hasselwurtz	ist	gut	den	frawen	zu	ihrer	kranckheit
norm	dass	Haselwurtzel	ist	gut	den	Frauen	zu	ihrer	Krankheit
comp_lex		lex							
comp_n		N_N							
comp_n_graph		nospace							
comp_n_mod		n							
infl_fuge		n							
lang	deu	deu	deu	deu	deu	deu	deu	deu	deu
lb	lb			lb					
lemma	dass	<unknown>	sein	gut	die	Frau	zu	ihr	Krankheit
ocr	tcttlce			gut	den	frawen	zu	ihrer	kranckheit
pb	pb								
pos	KOUS	NN	VAFIN	ADJD	ART	NN	APPR	PPOSAT	NN
quote	yes	yes	yes	yes	yes	yes	yes	yes	yes
script	blackletter	blackletter	blackletter	blackletter	blackletter	blackletter	blackletter	blackletter	blackletter

**Abb. 1:** Beispiel für eine Mehrebenenarchitektur im RIDGES-Korpus (Visualisierung auf ANNIS, s. Krause/Zeldes 2016). Die verschiedenen Annotationsebenen basieren auf unterschiedlichen Segmentierungsgrundlagen, so bildet die ›norm‹-Ebene die Segmentierungsgrundlage für alle linguistischen Annotationen, die Analysen zum morphologischen Wort enthalten (›lang‹, ›pos‹, ›lemma‹ etc.), während ›dipl‹ Segmentierungsgrundlage für markup- und textstrukturelle Annotationen ist (›script‹, ›pb‹ (= page), ›lb‹ (= linebreak)).

Im Folgenden schildern wir zunächst, wie unserer Erfahrung nach das Thema ›Variation‹ in der universitären Lehre unterrichtet wird (II),<sup>7</sup> wir stellen dann die genannten historischen Korpora vor, die sich aus unserer Sicht zum Einsatz im Unterricht für die Vermittlung der oben geschilderten Phänomene eignen (III), schildern zwei konkrete Einsatzszenarien dieser Korpora (IV) und diskutieren abschließend den konkreten Mehrwert, der sich aus der Anwendung von Korpora in Bezug auf die beiden Themenbereiche und in Bezug auf das Bewusstsein für sprachliche Variation ergeben kann (V).

<sup>7</sup> Da aufgrund der sehr unterschiedlichen Konzeptionen von Studiengängen zur Germanistik an deutschen und ausländischen Hochschulen kein allgemein verbindliches Curriculum zur Vermittlung sprachhistorischer Inhalte im Germanistikstudium vorliegt, basieren Aussagen über die Vorgehensweise in der Lehre dieses Fachs ausschließlich auf eigener Lehrerfahrung und dem Austausch mit Kolleginnen und Kollegen.

## II. Variation als Themenschwerpunkt in der Lehre

Unseren Vorschlag der korpusbasierten Vermittlung von sprachhistorischen Lehrinhalten möchten wir mit Bezug auf das Themenfeld der Variation ausarbeiten. Unter Variation verstehen wir die Eigenschaft aller natürlichen Sprachen, Funktionen (Variablen) durch eine Vielzahl scheinbar äquivalenter sprachlicher Mittel (Varianten) zu realisieren (Lüdeling 2017; Croft 2010; Rissanen 2008; Labov 2006; Labov 2004; Trudgill 2002; Barbour/Stevenson 1998; Labov 1991). Das Vorhandensein sprachlicher Varianten ist zugleich das Ergebnis von und die Ursache für Sprachwandel (Zeige 2011; Schmidt/Herrgen 2011; Croft 2010; Koch 2005; Koch 2004). Die Ausbildung eines sprachhistorischen Bewusstseins für Variation als Normalfall ist also ein entscheidender Schritt bei der Vermittlung historischer Sprachstufen.<sup>8</sup> Variation kann aus unterschiedlichen Perspektiven modelliert werden (bspw. Szmrecsanyi 2019; Schmidt/Herrgen 2011; Zeige 2011; Koch 2010; Coseriu 1992; Labov 1991), wobei Auslöser, Art der Beschreibung der einzelnen Varianten und die Wandelmodelle sich erheblich unterscheiden. Da dies in unserem Aufsatz nicht thematisiert wird, werden wir hier nur festhalten, dass die Möglichkeit, aber auch die Notwendigkeit besteht, Variation jeweils sowohl qualitativ als auch quantitativ zu analysieren. Da synchrone Variation auch im Gegenwartsdeutschen beobachtbar ist (zugänglich u. a. über das Themenfeld Zweifelsfälle, vgl. Hennig 2017; Klein 2013), kann hier möglicherweise an die entsprechenden Vorerfahrungen der Studierenden angeknüpft werden.

Abgesehen von Modellen und Theorien, denen sprachliche Variation inhärent ist, hat Variation disziplingeschichtlich lange einen problematischen Stellenwert gehabt: »The main problem is the evaluation of variation as an incomplete status of a certain developing state« (Schlachter 2020, S. 20; vgl. auch Ágel 2003). Variation wird vor allem in Überblickswerken zu den einzelnen Sprachstufen (Hartweg/Wegera 2005) tendenziell im Abgleich mit einem synchron gedachten Idealzustand behandelt, der in der Regel durch eine schriftsprachliche Varietät repräsentiert wird. Das Vorhandensein von Variation in verschiedenen Dimensionen wird somit stark gewertet und als defizitär angesehen, was systematische Ausgleichsversuche und die Vernachlässigung ganzer Themenbereiche, wie zum Beispiel nächstsprach-

---

<sup>8</sup> Für die sprachhistorische Forschung und Lehre sind synchrone und diachrone Variation gleichermaßen interessant: Unter dem Stichwort *synchrone Variation* kann Variation zu einem eng gefassten Zeitraum (bspw. im 17. Jahrhundert oder im Gegenwartsdeutschen) betrachtet werden, unter dem Stichpunkt *diachrone Variation* Variation über eine umfassendere Zeitspanne (bspw. vom 17. bis zum 20. Jahrhundert). Die Begriffe und Möglichkeiten der Modellierung von evidenzbasierter Synchronie und Diachronie sind jedoch methodisch und theoretisch umstritten (Jäger 1998).

licher Texte, zur Folge haben kann. So ist die Einführung in das Frühneuhochdeutsche von Hartweg/Wegera (2005), in der Variabilität zwar als »Existenzform des Frühneuhochdeutschen« konstatiert wird, im Kern ausgerichtet auf »die wichtigsten Entwicklungsprozesse, die zum Neuhochdeutschen (= Nhd.) führen, wobei die Selektions- und Ersatzprozesse den zentralen Aspekt bilden«, (Hartweg/Wegera 2005, S. 3) und nivelliert somit eben diese Variation. Auch andere allgemeine Einführungen wie Schmid (2009) setzen ihr Augenmerk auf Ausgleich und Variantenreduktion (Sprachgeschichte als »Weichenstellung« Schmid 2009, S. 2). Von großer Bedeutung ist hierbei, dass viele Lehrwerke einen sehr viel älteren Stand der Wissenschaft wiedergeben. Allerdings haben auch einige sprachgeschichtliche Lehrwerke gar nicht den Anspruch, Variation zu thematisieren, da andere Themen im Fokus stehen (Jäger 1998; Sonderegger 1979). So werden im Kontext der Beschreibung von Sprachwandelmechanismen einzelne konkrete Phänomene nur beispielhaft adressiert, Sprachgeschichten des Deutschen (Schmidt <sup>11</sup>2013; Schmid 2009) belassen es bei der Aufzählung systematischer Wandelphänomene und der Nennung der Kontexte. Sprachstufengrammatiken sehen sich mit der Aufgabe konfrontiert, die Heterogenität einer Sprachstufe zu thematisieren, und sind sich der bestehenden Unmöglichkeit ihrer Umsetzung in einer statischen und begrenzten Buchform bewusst. Bedingt durch ihre Entstehungszeit werden für das Thema Sprachgeschichte in vielen Lehrwerken nur die systematische Abstraktion (»languebezogene Grammatik« Ebert [u. a.] 1993) und die grammatische Beschreibung einzelner Texte (»textbezogene Grammatik«) als Alternativen angesehen. Die Möglichkeit der korpusbasierten quantitativen Sprachstufenbeschreibung über größere heterogene Textmengen hinweg kann diese Ansätze ergänzen und bisher als gegeben hingestellte Aussagen der Grammatiken relativieren. Jüngere Lehrwerke (z. B. vor allem durch Hartmann 2018; Fleischer/Schallert 2011; Fleischer 2007 sowie Wegera/Waldenberger 2012) nehmen auf diese Möglichkeit Bezug.

Ein deduktiver korpusbasierter Ansatz bietet die Möglichkeit, Erfahrungen aus der Verwendung von Korpora als Lehrmittel auf das Themengebiet historische Variation zu übertragen, Methoden und Werkzeuge für die Auseinandersetzung mit sprachlicher Variation anhand von konkreten historischen Beispielen zu erlernen und so für die Studierenden Variation dynamisch erfahrbar zu machen. Studierende lernen so, das sprachhistorische Bewusstsein für sprachliche Variation zu schärfen. Varietäten, durch systematische Variation entstehende sprachliche Subsysteme, können durch eine Vielzahl externer Parameter (vgl. Abschnitt IV.2.2) charakterisiert sein, die in Form von Metadaten in einem Korpus operationalisiert sind. Innerhalb von Texten einer Varietät, die eine Ähnlichkeit hinsichtlich bestimmter externer Parameter aufweisen, kann man grammatische, pragmatische oder textexterne Unterschiede für die Verteilung von Varianten eines sprachlichen Phänomens fokussieren, wie dies anhand von nahe- und distanzsprachlichen Texten im

ersten Fallbeispiel gezeigt wird.<sup>9</sup> Eine andere Perspektive ist die Betrachtung von Gebrauchsvariation zwischen einzelnen Varietäten: in Fallbeispiel 2 werden die quantitativ abweichenden Vorkommen des Bundles *ist gut* in Kräuterbeschreibungen und in Vorworten der Texte hervorgehoben, wobei das kommunikative Handlungsziel als Parameter verglichen wird.

### III. Historische Korpora

Historische Korpora sind digitalisierte annotierte Sammlungen natürlicher Äußerungen vergangener Sprachstufen (McEnery/Hardie 2012; Claridge 2009). Motiviert durch den Forschungskontext unterscheiden sie sich hinsichtlich ihres Designs, ihrer Größe, Annotationsschemata und -konzepte sowie der verwendeten Formate (vgl. Lüdeling [u. a.] 2011). Die Korpuserstellung erfolgt damit funktional, d. h. immer durch eine wissenschaftliche Intention motiviert. Moderne Korpora werden in einer Mehrebenenarchitektur erstellt, die es erlaubt, mehrere ggf. auch konfligierende Annotationen und Annotationskonzepte in einem Korpus zu repräsentieren. Damit wird dem Fakt Rechnung getragen, dass Annotationen immer Interpretationen der Daten sind. Jede Art der Repräsentation von Sprache liegt in Form einer Transkription oder/und Normalisierung vor, die wiederum die Kriterien von Annotation erfüllen (vgl. Odebrecht 2018, S. 52–58). Eine solche offene Korpusarchitektur ermöglicht eine kollaborative Arbeit und Forschung in unterschiedlichen (Unterrichts-)Kontexten. Annotationen können mithilfe von verschiedenen Annotationskonzepten wie Token-, Spannen- und Konstituentenannotationen ausgewiesen werden (z. B. Krause 2019).

Gerade für historische Sprachstufen des Deutschen wurden und werden immer mehr Korpora öffentlich zugänglich gemacht. So stehen der Öffentlichkeit umfassende Referenzkorpora des Altdeutschen (Zeige [u. a.] 2023),<sup>10</sup> Mittelhochdeutschen (Klein/Dipper 2016), Frühneuhochdeutschen (Herbers [u. a.] 2021) und Mittelniederdeutschen/Niederrheinischen (ReN-Team 2021) zur Verfügung, ein Beitrag zu einem Referenzkorpus Neuhochdeutsch ist mit dem Deutschen Textarchiv (Berlin-Brandenburgische Akademie der Wissenschaften 2020) und dem GiesKaNe-Projekt

---

<sup>9</sup> Dies entspricht grob dem allgemeinen Untersuchungsschema bei Typ-A-Studien (Varianten innerhalb eines Subkorpus) und Typ-B-Studien (Vergleiche von Subkorpora) bei Korpusstudien, siehe dazu Biber/Jones (2009). In vielen wissenschaftlichen Korpusstudien werden beide Ansätze kombiniert.

<sup>10</sup> Die URLs zu den im folgenden genannten Korpora und Korpusprojekten werden unter Referenzen genannt.



(vgl. Abschnitt III.2) in Vorbereitung. Weiterhin gibt es viele in einem spezifischen Forschungskontext erstellte Korpora wie das Historische Predigtenkorpus zum Nachfeld (Coniglio [u. a.] 2014), das Hexenverhörprotokollkorpus (Szczepaniak/Barteld 2016), das Kasseler Junktionskorpus (Ágel/Hennig 2020), das Register in German Diachronic Science Corpus (Odebrecht [u. a.] 2017) oder das Fürstinnenkorrespondenzkorpus (Lühr [u. a.] 2013).

Um Korpora in Kontexten der Hochschullehre verwenden zu können, muss der Zugang zu und der Zugriff auf historische Korpora frei sein. Freie Zugänge können mit Hilfe von Forschungsdatenrepositorien<sup>11</sup> und freien Lizenzen<sup>12</sup> ermöglicht werden. Viele der genannten historischen Korpora sind bspw. im LAUDATIO-Repository (Guescini/Odebrecht 2020) nachhaltig publiziert. Einen Zugriff auf die Korpusdaten zum Zwecke der Analyse (in unserem Sinne bspw. die Wiederverwendung in Unterrichtskontexten) ermöglichen Such- und Visualisierungstools wie ANNIS (Krause 2019; Krause/Zeldes 2016).<sup>13</sup>

Die in diesem Artikel verwendeten Korpora KAJUK und RIDGES eignen sich gut für den Einsatz in der Lehre, da beide eine offene Mehrebenenarchitektur besitzen, frei zugänglich über das Such- und Visualisierungstool ANNIS zur Verfügung stehen und umfangreich dokumentiert sowie frei verfügbar im LAUDATIO-Repository publiziert sind. Die in den Korpora enthaltenen Texte decken einen ähnlichen Zeitraum (spätes Frühneuhochdeutsch (= Fnhd.)/Nhd. bis ca. 1900) ab. Allerdings unterscheiden sie sich in Bezug auf ihr Korpusdesign und ihr Annotationskonzept. KAJUK enthält nahe- und distanzsprachliche Texte mit u. a. umfangreichen Annotationen zu syntaktischen Phänomenen. Das gilt gleichermaßen für GiesKaNe, das bei der Besprechung des ersten Fallbeispiels (vorangestellte Adverbialsätze) ergänzend hinzugezogen wird. RIDGES enthält u. a. normalisierte Kräuterkundetexte, die diachrone Analysen ermöglichen. Die genannten Korpora werden im Folgenden kurz vorgestellt.

---

11 Forschungsdatenrepositorien sind Publikationsplattformen für Datensätze wie Korpora. Eine umfangreiche Übersicht über existierende Repositorien gibt das Projekt ›re3data‹, vgl. <https://www.re3data.org/> (Aufrufdatum: 23.10.2020).

12 Häufig werden hierfür Lizenzen der Creative Commons verwendet, vgl. <https://creativecommons.org/> (Aufrufdatum: 23.10.2020).

13 Für einen Überblick zu Suchwerkzeugen für Korpora vgl. Geyken/Lemnitzer (2018).

### III.1 Das Kasseler Junktionskorpus (KAJUK) sowie das Gießen-Kassel-Neuhochdeutsch-Korpus (GiesKaNe)

Das Kasseler Junktionskorpus (Version 1.1) ist im Rahmen des DFG-Projekts ›Explizite und elliptische Junktion in der Syntax des Neuhochdeutschen. Pilotprojekt zu einer Sprachstufengrammatik des Neuhochdeutschen‹ (2007–2009) entstanden (vgl. Ágel/Hennig 2006). In diesem Kontext entsteht derzeit auch das Gießen-Kassel-Neuhochdeutsch-Korpus (GiesKaNe). KAJUK ist ein diachrones Mehrebenenkorpus mit ca. 100.000 Wortformen und enthält acht Texte aus drei dialektalen Großräumen (Oberdeutsch, Mitteldeutsch, Niederdeutsch). Es wurden in je vier Texten aus dem 17. und 19. Jahrhundert junktionsrelevante grammatische Merkmale annotiert, durch deren Kombination verschiedene Techniken der expliziten und elliptischen Junktion rekonstruiert werden können (vgl. Hennig 2013). Da nicht Junktionstechniken annotiert wurden (vgl. Ágel 2012), sondern die für die Rekonstruktion der Techniken relevanten grammatischen Merkmale, ist das Korpus auch für weiterführende Fragestellungen nutzbar (vgl. bspw. zur Serialisierung im Verbalkomplex Hennig 2009). KAJUK wurde ausschließlich manuell mit syntaktischen und morphologischen Informationen sowie mit einem Fokus auf Sachverhaltsdarstellungen annotiert. Es enthält keine Normalisierungsannotation, folglich sind die Belege nur über die Ermittlung konkreter Oberflächenformen, also der im historischen Text vorliegenden Schreibvarianten suchbar.

GiesKaNe (= Gießen Kassel Neuhochdeutsch) entsteht derzeit im Rahmen eines DFG-Langfristvorhabens zu syntaktischen Grundstrukturen des Nhd. (Leitung: Vilmos Ágel, Kassel, und Mathilde Hennig, Gießen). Das Projekt versteht sich als ein Beitrag zu einem syntaktisch erschlossenen Referenzkorpus des Nhd. und ist als solches als Kooperationsprojekt mit dem Deutschen Textarchiv angelegt. Ein erster Zugriff auf annotierte syntaktische Grundstrukturen in zwei neuhochdeutschen Texten (GiesKaNe 0.1) ist seit Januar 2019 über ANNIS möglich. GiesKaNe ist als ein ebenfalls syntaktisch tiefenannotiertes Korpus angelegt, dessen Korpusdesign eine umfangreiche Integration von Texten aus dem 17. bis 19. Jahrhundert aus verschiedenen Registern vorsieht.<sup>14</sup> KAJUK (wie auch GiesKaNe) ist prädestiniert für die Arbeit an variationslinguistisch ausgerichteten Fragestellungen, da die Nähe-Distanz-Dimension zentraler Eckpfeiler der Korpuserstellung war (vgl. dazu Abschnitt IV.1.2) und somit der außersprachliche Kontext systematisch in das Korpusdesign integriert ist.

<sup>14</sup> URL: <https://gieskane.com/> (Aufrufdatum: 19.02.2023). Die Bereitstellung weiterer 24 Texte ist ebenfalls auf der Plattform ANNIS für Frühjahr 2023 vorgesehen.

## III.2 Das Register in Diachronic German Science Korpus (= RIDGES)

RIDGES (Version 8.0) ist ein kollaborativ erstelltes diachrones Mehrebenenkorpus, mit dessen Hilfe die Entstehung und Entwicklung der deutschen Wissenschaftssprache seit dem Ende des 15. Jahrhunderts untersucht werden kann. Es enthält 61 Auszüge aus Drucken von Kräuterkundetexten von 1482 bis 1914 mit insgesamt 257.537 historischen Wortformen (Odebrecht [u. a.] 2017). Das RIDGES-Korpus wird im Rahmen der universitären, forschungsorientierten Lehre seit 2011 in BA- und MA-Studiengängen erstellt und im Rahmen des LAUDATIO-Repository kuratiert. Das Korpus wächst kontinuierlich. Die Textauszüge sind zwischen wenigen hundert und mehreren tausend historischen Wortformen groß. Die kleinen Ausschnitte enthalten meist Vorworte der Bücher, die längeren Einheiten Kapitel oder Kapitelauszüge mit Beschreibungen von Kräutern, Anwendungsformen und Krankheiten. Die Texte sind opportunistisch gesammelt, stammen aus verschiedenen dialektalen Räumen (Oberdeutsch, Mitteldeutsch, Niederdeutsch) mit jeweils unterschiedlichen thematischen Schwerpunkten (heilkundlich, botanisch, medizinisch oder chemisch), die jedoch alle unter anderem das Thema ›Kräuterkunde‹ beinhalten. Damit wird eine möglichst große Varianz der Kräuterkundetexte abgebildet. Zu bedenken ist, dass bei einer diachronen korpuslinguistischen Untersuchung die Eigenschaften des Korpusdesigns (nicht repräsentativ, nicht balanciert) berücksichtigt werden müssen.

Für die Erstellung des Korpus werden die Inhalte der historischen Drucke nach dokumentierten Richtlinien (Odebrecht [u. a.] 2020) diplomatisch transkribiert und zusätzlich in zwei Ebenen normalisiert. Das Korpus ist manuell mit graphischen, (im weitesten Sinne auch) textstrukturellen und lexikalischen Annotationen versehen. Alle Korpusdokumente enthalten eine automatische Wortartenannotation (PoS-Tagging), ein Subset wurde auch automatisch syntaktisch annotiert (Abhängigkeiten, Konstituenten).

## IV. Fallbeispiele

### IV.1 Vorangestellte adverbiale Nebensätze als Variationsphänomen an der Syntax-Lexikon-Semantik-Schnittstelle

Als erstes Fallbeispiel dienen vorangestellte adverbiale Nebensätze im Neuhochdeutschen. Das Phänomen wurde ausgewählt, weil es dafür geeignet ist, alle in der

Einleitung aufgeführten Möglichkeiten der Korpusnutzung zu illustrieren. Es wird in der exemplarischen Korpusarbeit folglich darum gehen,

- das Phänomen als solches zu belegen;
- die verschiedenen Realisierungsvarianten adverbialer vorangestellter Nebensätze aufzuzeigen;
- exemplarische Angaben zur Verteilungshäufigkeit der Varianten vorzulegen.

Ausgehend von der Überlegung, dass ein Einsatz von Korpora in der Lehre auch das Knowhow der Korpusnutzung voraussetzt, legt dieses Kapitel den Schwerpunkt auf eine Anleitung zur Suche nach den relevanten Phänomenen in den Korpora KAJUK und GiesKaNe. Dieser methodische Schwerpunkt hat zur Folge, dass hier keine vollständige, ergebnisorientierte Korpusstudie vorgelegt wird. Zu systematischen Ergebnissen zur Verteilung der Adverbialsatztypen im Neuhochdeutschen sei auf die Studie von Kappel (2016) verwiesen.

#### IV.1.1 Gegenstand

Die Fachliteratur unterscheidet im Allgemeinen in Anlehnung an König/van der Auwera (1988) drei Typen von vorangestellten Adverbialsätzen:

Integrativ:

- (3) *Wäre ich weniger intelligent, würde ich mehr Geld verdienen.*  
(Lötscher 2005, S. 347)
- (4) *Als ich zur Portten hinauß kam, fandt ich meine 2 Spießgesellen von Ambsterdam in eines Bauren Hauß auff dem Mist im Stro ligen, zu welchen ich mich schlaffen auff den Mist legte.*  
(Güntzer I)

Resumptiv:

- (5) *Wäre ich weniger intelligent, dann würde ich mehr Geld verdienen.*  
(Lötscher 2005, S. 347)
- (6) *wan den die schildtwach, auff den tuhrm, edtwas kommen sieht, so lest sie den korb nieder singken, da richten sich die leute nach, so leuff alles der kirch zu, vief, vndt leute, vndt lassen die dörffer lehr stehen*  
(Söldnerleben I)

Nicht-integrativ:

- (7) *Wäre ich weniger intelligent, ich würde mehr Geld verdienen.*  
(Lötscher 2005, S. 347)

- (8) *Wir arme Leuth meineten, weil die kayserische Arme alles sogar verwüstet und verderbet hatten, wir würden ja nun ein wenig Fridt haben, und machten uns aus dem Kirchain widerumb ein ider nach seinem Dorf, dan es ist ein Sprichwort »Uff geraubter Strassen ist sicher wandern«.*  
(Bauernleben I)

Die drei Typen sind hier jeweils durch ein konstruiertes Beispiel aus Lötscher (2005) sowie ein authentisches Beispiel aus KAJUK illustriert. Die konstruierten Beispiele veranschaulichen den Gebrauch, wie wir ihn auch im Gegenwartsdeutschen kennen; die Beispiele aus KAJUK dokumentieren die Belegbarkeit der Typen in historischen Korpora.

Die drei Typen unterscheiden sich wie folgt: Beim integrativen Typ ist der Adverbialsatz in die Felderstruktur des übergeordneten Hauptsatzes integriert. Das finite Verb des Hauptsatzes folgt unmittelbar auf den Adverbialsatz, der somit das Vorfeld besetzt. Der resumptive Typ zeichnet sich durch die Verwendung eines Korrelats aus (*dann* in 5, *so* in 6), das die Position vor dem finiten Verb im Hauptsatz einnimmt. Beim nicht-integrativen Typ steht an dieser Position ein anderes Satzglied (das Subjekt *ich* bzw. *wir* in 7 und 8).

#### IV.1.2 Eignung des Phänomens für die korpusbasierte Lehre

Das Themenfeld der vorangestellten adverbialen Nebensätze eignet sich aus den folgenden Gründen als Fallbeispiel für Überlegungen zur korpusbasierten Lehre:

- 1.) Es ist verhältnismäßig gut erforscht: Mit den Arbeiten von Axel (2002), Lötscher (2005) und Kappel (2016) liegt eine Reihe von Studien zur historischen Entwicklung des Verhältnisses vorangestellter Adverbialsätze zum Folgesatz vom Althochdeutschen (= Ahd.) bis zum Nhd. vor. Für die korpusbasierte Lehre bietet das die Möglichkeit, die Konturierung einer Fragestellung für die eigene Korpusarbeit systematisch zu untermauern und Ergebnisse eigener Korpusarbeit mit den Ergebnissen der bisherigen Forschung abzugleichen.
- 2.) Vorangestellte adverbiale Nebensätze als *Variationsphänomen*: Verschiedene Varianten der Verknüpfung vorangestellter adverbialer Nebensätze mit dem Folgesatz lassen sich sowohl synchron als auch diachron beobachten. Die unter 1.) genannten Arbeiten bieten sowohl Evidenz für *diachrone Variation* (dabei kann man zusammenfassend von einer Entwicklung hin zu mehr Integration sprechen, ganz im Sinne von Czicza/Hennig 2013), aber auch für die *synchrone Variation*, für die insbesondere Lötscher (2005) und Kappel (2016) pragmatische Erklärungsfaktoren anführen. So spricht Lötscher von einer »Rhetorik der Mündlichkeit« in Bezug auf das Mittelhochdeutsche

(= Mhd.) (Lötscher 2005, S. 363) sowie von einer »Tolerierbarkeit von komplexen Satzstrukturen« im Gegenwartsdeutschen (Lötscher 2005, S. 358). Kappel (2016) korreliert das Vorkommen der verschiedenen Adverbialsatztypen systematisch mit der Nähe-Distanz-Dimension. Ein weiterer Vorteil besteht darin, dass es die drei Varianten auch noch im Gegenwartsdeutschen gibt, sodass die Varianz für Studierende unmittelbar nachvollziehbar sein dürfte, wobei auch Unterschiede zum gegenwartssprachlichen Gebrauch rasch zutage treten. Als Variationsphänomen, das auf der Basis von syntaktischen Oberflächenmerkmalen gut zugänglich für Korpusanalysen ist, eignet sich das Phänomen u. E. sehr gut als Anwendungsbeispiel zur Illustration der korpuslinguistischen Möglichkeiten der Ermittlung von Varianz und Frequenz.

- 3.) **Semantik-Lexikon-Syntax-Schnittstelle:** In der Grammatiktradition werden Adverbialsätze wie auch sonstige Adverbiale seit jeher semantisch klassifiziert. Dem können wir entnehmen, dass es bei Adverbialsätzen nicht nur um eine syntaktische Relation zu einem Bezugssatz geht, sondern auch um eine semantische (zu aktuell gängigen Klassifikationen siehe Wöllstein (2016) sowie von Polenz (2008); zur Analyse adverbial-semantischer Relationen im Rahmen der Junktion siehe Ágel (2012)). Insbesondere Kappel (2016) hat nachweisen können, dass die semantische Relation einen Einfluss auf den Gebrauch der Adverbialsatztypen hat; Syntax und Semantik sind also interdependent. Schließlich spielt auch das Lexikon eine Rolle: Bekanntlich werden bestimmte Subjunkturen für bestimmte semantische Typen von Adverbialsätzen genutzt. Auch diese Lexikon-Semantik-Schnittstelle unterliegt dem Wandel (vgl. Ágel 2000, S. 1879–1888 zum Nhd.). Die moderne Korpuslinguistik bietet uns die Möglichkeit, Phänomene ebenenübergreifend und nicht modular zu betrachten, sodass die Korpusmethodik in der Lehre gewinnbringend genutzt werden könnte, um Grenzen einer streng modularen Sicht- und Herangehensweise aufzuzeigen.
- 4.) **Theorie und Wandel:** Bei der Ausbildung eines Bewusstseins für synchrone und diachrone Variation kommt es nicht nur darauf an, Unterschiede und Veränderungen sichtbar zu machen, sondern zentral ist auch die Frage, wie man diese deutet. Hier bietet die Fachliteratur eine eingängige Kontroverse. Dabei bilden u. a. die nicht-integrativen Adverbialsätze eine Herausforderung für die Grammatiktheorie: Axel (2002) geht von einem V3-Muster aus, also davon, dass das finite Verb hier an der dritten Stelle im Satz steht. Lötscher (2005) hingegen weist darauf hin, dass die Verbzweitstellung im Ahd. schon grammatikalisiert gewesen sei und begegnet dem Positionsproblem mit einer Verortung der nicht-integrierten Adverbialsätze im sogenannten linken Außenfeld (Lötscher 2005, S. 355). Aufschlussreich ist auch der Theoriestreit zur Verortung des Wandelphänomens. So nimmt Axel syntaktischen Wandel an und begründet das damit, »daß es anscheinend nicht mit der ›Schwere‹ des

AS zusammenhängt, ob V2- oder V3-Stellung gewählt wird« (Axel 2002, S. 13–14). Lötcher setzt dem die These eines pragmatischen Wandels entgegen, da er davon ausgeht, dass sich nicht

»die grammatisch-semantische Natur von Adverbialsätzen geändert [hat], sondern die Bewertung und das gegenseitige Verhältnis von grammatischer Integration von Nebensätzen, grammatischer Kohärenz und Tolerierbarkeit von komplexen Satzstrukturen.« (Lötcher 2005, S. 358)

Korpuslinguistische Methoden können genutzt werden, um Hypothesen zum Wandel zu generieren und zu überprüfen; die hier kurz angerissenen theoretischen Deutungsmöglichkeiten zeigen aber auch, dass sie eine Theoriediskussion nicht ersetzen.

#### IV.1.3 Umgang mit der Thematik in gängigen Lehrwerken zur Sprachgeschichte

Ein unmittelbarer Vergleich der Ergebnisse der zitierten Forschungsliteratur mit gängigen Lehrwerken zur Sprachgeschichte ist nicht möglich, da das Thema in den Lehrwerken kaum Niederschlag findet. So spielt syntaktischer Wandel in vielen Lehrwerken eine gegenüber lautlichem, morphologischem und lexikalischem Wandel nachrangige Rolle. Aber auch im Syntaxlehrbuch von Fleischer/Schallert (2011) liegt der Schwerpunkt auf anderen Themen (bspw. Ersatzinfinitiv und Negation). Bezeichnend ist die Einschätzung von Ebert [u. a.] (1993) in der ›Frühneuhochdeutschen Grammatik‹ im Kapitel ›Zum Satzgefüge‹:

»Über das frnhd. Satzgefüge ist man nicht gut unterrichtet. Es fehlt überhaupt eine Typologie von Verknüpfungsformen sowie genaue Studien zum Bestand der Formen.« (Ebert [u. a.] 1993, S. 483)

So enthält keines der hier zu Rate gezogenen, in der germanistischen Hochschullehre zur Sprachgeschichte gängigen Lehrwerke ein eigenes Kapitel zu vorangestellten Adverbialsätzen. Die meisten Werke enthalten Ausführungen zur Verbstellung, hier geht es aber vorrangig um die Verbstellung im Aussage Hauptsatz. Fleischer/Schallert (2011, S. 153 ff.) behandeln dabei auch die *Verb spä terstellung* im Ahd., jedoch ohne Rückgriff auf die oben erwähnte Diskussion zur grammatiktheoretischen Einordnung des Phänomens. Die Integration von Nebensätzen in die Felderstruktur wird nicht gesondert behandelt. Die ›Frühneuhochdeutsche Grammatik‹ widmet den einzelnen Adverbialsätzen ein ausführliches Kapitel. Hier geht es vorrangig um die Klassifikation der Subjunktorexeme. Das Thema wird auch in den meisten Lehrwerken angesprochen: Hartweg und Wegera (2005, S. 177–178)

sprechen im Lehrbuch zum Fnhd. von der »Profilierung der subordinierenden Konjunktionen«; Schmidt (<sup>11</sup>2013) spricht in seiner ›Geschichte der deutschen Sprache‹ in Bezug auf das Nhd. von »Variantenreduzierung und Monosemierung«. In Bezug auf das Ahd. weist Schmidt auf die Möglichkeit der Verwendung eines Korrelats im übergeordneten Satz hin (Schmidt <sup>11</sup>2013, S. 274).

Das bedeutet für unseren Vorschlag der Nutzung des Themenfelds vorangestellter Adverbialsätze für eine korpusbasierte Lehre: Die Möglichkeit eines unmittelbaren Abgleichs mit den in gängigen Lehrwerken zu findenden Formen der Darstellung ergibt sich nur für das in der Lexikon-Semantik-Schnittstelle angesiedelte Thema der Profilierung und Monosemierung der Semantik von Subjunktorexemen. Hier bieten einige Lehrwerke Detailangaben (bspw. Hartweg/Wegera 2005; Schmidt <sup>11</sup>2013); insbesondere die ›Frühneuhochdeutsche Grammatik‹ enthält umfangreiche Informationen zu systematischen Beziehungen zwischen Inhaltsrelationen und den entsprechenden Lexemen sowie zum Wandel des Systems. Hier liegt folglich eine gute Grundlage für einen Vergleich von Ergebnissen der Korpusrecherche in der Lehre mit den genannten Lehrwerken entnehmbaren Vorkenntnissen vor.

Das hier eigentlich im Fokus unseres Vorschlags stehende Themenfeld der syntaktischen Integration vorangestellter Nebensätze lässt sich hingegen nicht unmittelbar auf die durch die durchgesehenen Lehrwerke dokumentierte Praxis der Lehrgegenstände der auf die Sprachgeschichte ausgerichteten Hochschullehre beziehen. Bei einem Abgleich von Korpusarbeit mit Lehrwerken könnte es hingegen als Ergänzung zu den Angaben zur Verbstellung begriffen werden, bietet es doch Diskussionsraum für eine genauere Auseinandersetzung mit der Frage nach dem Stellenwert der Verbzweitstellung.

Wir haben uns hier dafür entschieden, das Themenfeld vorangestellte Adverbialsätze als mögliches Fallbeispiel für korpusbasierte Hochschullehre vorzuschlagen, weil wir meinen, dass es die Möglichkeit bietet, Themenfelder, die bisher eher getrennt behandelt wurden, zu integrieren: Mit Adverbialsätzen und Felderstruktur werden ein Klassiker der traditionellen Grammatik und ein in der jüngeren Grammatiktradition zentraler (und dabei bereits in vielen didaktischen Kontexten genutzter) Ansatz verbunden. Die Möglichkeit der integrativen Betrachtung von syntaktischem, semantischem und lexikalischem Wandel rechtfertigt es u. E., dem Themenfeld der vorangestellten Adverbialsätze einen Platz in der sprachhistorisch ausgerichteten germanistischen Hochschullehre zuzuweisen.

#### **IV.1.4 Möglichkeiten der Suche nach vorangestellten Adverbialsätzen in KAJUK**

In KAJUK sind die wichtigsten annotierten Phänomenbereiche für unser Themenfeld die Sachverhaltsdarstellungen, die Junktoren, die Inhaltsrelationen sowie



schließlich auch die Korrelate. Sachverhaltsdarstellungen sind deshalb ein zentrales Element des Annotationsmodells, weil KAJUK das auf Raible (1992) zurückführbare Verständnis von Junktion als Verknüpfung von Sachverhaltsdarstellungen zugrunde liegt (Ágel/Diegelmann 2010; Ágel 2012). Als Sachverhaltsdarstellungen (= SVDen) kommen vor allem einfache Sätze sowie Teilsätze in Satzgefügen in Frage; aber auch Präpositionalgruppen bieten die Möglichkeit der Inkorporation einer Sachverhaltsdarstellung (Ágel/Diegelmann 2010, S. 367–368). Im Korpus sind Sachverhaltsdarstellungen als ›line breaks‹ (lb) annotiert (Ágel/Hennig 2020). Als Junktoren sind vor allem Konjunktoren, Subjunktoren und Adverb-Partikeljunktoren erfasst. Die Verknüpfung von zwei Sachverhaltsdarstellungen durch einen Junktor wird in KAJUK mit IDs, genannt »Hausnummern«, angegeben, d. h., zwei verknüpfte Sachverhaltsdarstellungen erhalten die jeweils gleiche ID, damit auch bei komplexeren Beziehungsgeflechten klar ist, welche Sachverhaltsdarstellung Gegenstand der Relation ist. Die für unsere Thematik relevanten Subjunktoren haben eine ID zwischen 1 und 999. Der jeweiligen durch einen Junktor indizierten Relation wird auch eine semantische Relation zugewiesen. Das Inventar semantischer Relationen ist eng an von Polenz (2008) angelehnt.

Eingeleitete vorangestellte Adverbialsätze sind folglich als Kombination der Information ›Subjunktor‹ und Reihenfolge der aufeinander folgenden Sachverhaltsdarstellungen mit folgendem Ausdruck in der ANNIS-Query-Language<sup>15</sup> suchbar:

```
(9) lb & lb & J=>J & J_type=>SUB & #3 _ _ #4 & #1. #2 & #1_o_#3 &
#1 ->SubKon #2
```

Gesucht wird auf diese Weise nach zwei Sachverhaltsdarstellungen (lb), einem Junktor (J="J") und einem Junktortyp Subjunktor. Die folgenden Bedingungen für diese vier Parameter werden angeschlossen: Der Junktor muss ein Subjunktor sein (J\_type="SUB"). Die beiden Sachverhaltsdarstellungen müssen aufeinander folgen (#1.#2). Der Junktor (#3) soll in der ersten SVD (#1) enthalten sein (#1\_o\_#3). Zwischen den beiden Sachverhaltsdarstellungen soll eine Subjunktions- oder Konjunktionsrelation bestehen (#1 ->SubKon #2).

Diese Suchanfrage liefert für KAJUK 676 Treffer. Diese noch recht allgemeine Suchanfrage kann weiter eingegrenzt werden. Tab. 1 fasst die Möglichkeiten der Eingrenzung zusammen:

<sup>15</sup> Die ANNIS Query Language (AQL) ist die Abfragesprache, die im Such- und Visualisierungstool ANNIS verwendet wird. Sie hat die volle Mächtigkeit regulärer Ausdrücke und mit ihr kann man über alle annotierten Ebenen hinweg suchen. AQL ist ausführlich dokumentiert unter <http://korpling.github.io/ANNIS/3.6/user-guide/>.

Tab. 1: AQL-Suchanfragen, Beispiele und Trefferzahlen für verschiedene Adverbialsätze in KAJUK<sup>16</sup>

Bsp.	Gegenstand der Suchanfrage	Begründung der Suchanfrage	Suchanfrage	Beispiel	Anzahl Treffer
1	Integrativer vorangestellter Adverbialsatz	Gesucht wird nach zwei aufeinanderfolgenden SVDen, von denen die zweite mit einem finiten Verb beginnt.	lb & lb & J="J" & J_type="SUB" & Verbform=/Fin/ & #3 _= #4 & #1 . #2 & #1_o_#3 & #1 ->SubKon #2 & #2 _l_ #5	<i>Derohalber sie von mir kein Schiffflon fordern, gaben mir umsunst Essen und Trincken</i> (Güntzer I)	142
2	Resumptiver vorangestellter Adverbialsatz	Gesucht wird nach zwei aufeinanderfolgenden SVDen, von denen die zweite mit einem Korrelat beginnt.	lb & lb & J="J" & J_type="SUB" & Satzglied=/KOR/ & #3 _= #4 & #1 . #2 & #1_o_#3 & #1 ->SubKon #2 & #2_o_#5 & #2 _l_ #5	<i>Obwol die Schweden und die Hesen eines Wesens war, so nam doch ein ider, was er bekommen konte.</i> (Bauernleben I)	103
3	Temporaler vorangestellter Adverbialsatz	Gesucht wird nach einem Adverbialsatz mit einer bestimmten Inhaltsrelation (hier = temporal).	lb & lb & J=>» & J_type=>SUB & IR=->temp & #3 _= #4 & #4 _= #5 & #1 . #2 & #1_o_#3 & #1 ->SubKon #2	<i>als wier 22 meilen zurück gelegte hatten mußte halt gemacht werden</i> (Zimmer V)	236
4	Uneingeleiteter Adverbialsatz	Bei uneingeleiteten Nebensätzen ist die Inhaltsrelation der lb zugeordnet.	lb & lb & IR_lb & Verbform=/Fin/ & #1 . #2 & #1_o_#3 & #1 _l_ #4	<i>Hatten wir zuvor einen Freund, so bekamen wir da einen Feind.</i> (Briefwechsel V)	129
5	Integrativer uneingeleiteter vorangestellter Adverbialsatz	Kombination aus 1 und 4	lb & lb & IR_lb & Verbform=/Fin/ & Verbform=/Fin/ & #1 . #2 & #1_o_#3 & #1 _l_ #4 & #2 _l_ #5	<i>Kommt nicht rechtzeitig Brief, ist sie verstimmt.</i> (Koralek V)	8

<sup>16</sup> Die permanenten Korpusreferenzlinks zu den Suchen sind im elektronischen Anhang zu finden.

Tab. 1 (fortgesetzt)

Bsp.	Gegenstand der Suchanfrage	Begründung der Suchanfrage	Suchanfrage	Beispiel	Anzahl Treffer
6	Resumptiver uneingeleiteter vorangestellter Adverbialsatz	Kombination aus 2 und 4	lb & lb & IR_lb & Verbform=/Fin/ & Satzglied=/KOR/ & #1 . #2 & #1_o_#3 & #1 _#4 & #2 _#5	<i>Jst er Schreiber / so will er Secretarius , sein</i> (Thomasius I)	50
7	Adverbialsatz im Text ›Nietzsche‹	Kombination der Suche nach dem grammatischen Phänomen mit dem Metadatum für den Dokumentnamen	lb & lb & J=>» & J_type=>SUB» & #3 _#4 & #1 . #2 & #1_o_#3 & #1 ->SubKon #2 & meta::doc=/Nietzsche/	<i>Als er ihm endlich in die Hände gefallen ist, fragt der König (Nietzsche V)</i>	50
8	Subjunktorlexeme mit kausaler Relation	Kombination der Suche nach einem Subjunktor mit der Suche nach der Inhaltsrelation ›kausak‹	lb & lb & J="J" & J_type="SUB" & IR="kaus" & #3 _#5 & #3 _#4 & #1 . #2 & #1_o_#3 & #1 ->SubKon #2	<i>da die arme ist aufgebrochen , vndt da die arme , ist in die altemargk gegangen , bin ich alweil stil gelehgen</i> (Söldnerleben I)	110

Wie bei vielen Korpusanfragen bedarf es auch hier bei jeder Trefferzahl einer qualitativen Belegsichtung und ggf. einer iterativ modifizierten Suche. Bei jeder Arbeit mit Korpusbelegen kommt es immer auch darauf an, Fehlbelege (*false positives*) zu identifizieren, also Belege, die der Treffermenge angehören, weil sie die Suchkriterien erfüllen, die aber aus bestimmten Gründen nicht der gesuchten Kategorie entsprechen. So führt die oben in (9) genannte Suchanfrage zu vorangestellten Adverbialsätzen bspw. zu folgendem Fehlbeleg:

- (10) *Das Mal hatten wir ein Hauptman von den Heßen uf Omeneburg liegen, das mir das Mal in hesischer Conterbutio waren und wir also das Mal vor den Schweden bleiben konten.*  
(Bauernleben I)

Die beiden hier fett gedruckten Sachverhaltsdarstellungen sind als Treffer aufgeführt, weil tatsächlich die Bedingung gegeben ist, dass die erste Sachverhalts-

darstellung einen Subjunktor enthält. Dabei handelt es sich allerdings um einen Subjunktor, der in einer Beziehung zur vorausgehenden Sachverhaltsdarstellung steht und nicht zur folgenden. Der Fehlbeleg ist hier darauf zurückzuführen, dass zwischen den beiden fett gedruckten Sachverhaltsdarstellungen eine Relation des Typs SubKon besteht, allerdings ist es keine Subordinations-, sondern eine Koordinationsbeziehung.

Es wird hier bewusst darauf verzichtet, die mit den vorgeschlagenen Suchanfragen ermittelten Beleglisten auszuwerten und die entsprechenden Ergebnisse zu präsentieren, denn die Auswertung und Ergebnisdiskussion sollte u. E. im Sinne des eingangs beschriebenen Ziels der selbstständigen Erkenntnisgewinnung Gegenstand der Hochschullehre sein.

In der Korpusarbeit mit Studierenden hat man nun einerseits die Möglichkeit, Treffermengen mit Fehlbelegen bewusst für eine Sensibilisierung für subtile Unterschiede zwischen Phänomenen sowie zur Abhängigkeit der Beleglage von der Annotation zu nutzen. Man kann aber auch die Suchanfragen weiter einschränken, um a) bestimmte Typen von Fehlbelegen auszuschließen und b) gezielt nach Subtypen des gesuchten Phänomens zu suchen.

#### IV.1.5 Ausblick: Vorangestellte Adverbialsätze in GiesKaNe

In einem Ausblick seien nun die Möglichkeiten der Suche nach vorangestellten Adverbialsätzen im noch im Aufbau befindlichen Korpus GiesKaNe vorgestellt, das einen umfassenden Ansatz zu syntaktischen Grundstrukturen verfolgt und nicht wie das Junktionsprojekt nur einen eingeschränkten Phänomenbereich. Die bereits in KAJUK erarbeitete Grundstrategie der Annotation atomarer Kategorien, die in Suchanfragen dann beliebig kombinierbar sind, wurde hingegen beibehalten. GiesKaNe ist ein Mehrebenenkorpus, das neben einer Wortartebene vor allem textgrammatische und satzsyntaktische Analyseebenen beinhaltet. Die satzsyntaktische Analyseebene ist als Baumbank konzipiert.

Im Sinne des Vorhabens der Erschließung syntaktischer Grundstrukturen erfolgt in GiesKaNe eine systematische und vollständige Satzgliedannotation. Die wichtigste theoretische Grundlage bildet dabei die ›Grammatische Textanalyse‹ (Ágel 2017). Folglich müssen Adverbialsätze hier nicht wie in KAJUK indirekt, also als subordinierte Sachverhaltsdarstellungen mit Inhaltsrelation erschlossen werden, sondern sind als Realisierungsformen von Adverbialen als solche annotiert und können so direkt erfragt werden. Annotiert ist hier der Satzgliedwert Adverbial (mit der jeweiligen semantischen Subklasse) sowie die Realisierungsform (in unserem Fall: als Nebensatz). Die Beziehung des Adverbialsatzes zum übergeordneten Teilsatz ist über eine Konstituentenstruktur- und Hierarchieannotation erschlossen,

die in der Baumstruktur in Form einer Konstituentenannotation hinterlegt ist. Die folgende Tab. 2 listet einige Suchanfragen in GiesKaNe (Version 0.1) auf. Wir verzichten hier auf Beispiele; vielmehr werden wir im Anschluss an Tab. 2 allgemeine Erläuterungen zur Suche in der Baumbank vornehmen.

**Tab. 2:** AQL-Suchanfragen, Beispiele und Trefferzahlen für verschiedene Adverbialsätze in GiesKaNe

Bsp.	Gegenstand der Suchanfrage	Suchanfrage	Anzahl Treffer
1	Vorangestellter Nebensatz	<code>tree="s" &amp; tree="ns" &amp; #1 &gt;@l #2</code>	162
2	Vorangestellter Adverbialsatz	<code>tree="s" &amp; tree="ns" &amp; #1 &gt;@l #2 &amp; #1 &gt;[tree=/adv.+/] #2</code>	145
3	Vorangestellter Konditionalsatz	<code>tree="s" &amp; tree="ns" &amp; #1 &gt;@l #2 &amp; #1 &gt;[tree="adv_kond"] #2</code>	43
4	Integrativer vorangestellter Adverbialsatz	<code>tree="s" &amp; tree="ns" &amp; #1 &gt;@l #2 &amp; #1 &gt;[tree=/adv.+/] #2 &amp; #2 . " , " . finit="fin"</code>	58
5	Resumptiver vorangestellter Adverbialsatz	<code>tree="s" &amp; tree="ns" &amp; node &amp; #1 &gt;@l #2 &amp; #1 &gt;[tree=/adv.+/] #2 &amp; #1 &gt;[tree="kor"] #3 &amp; #2 . " , " . #3</code>	85
6	Resumptiver vorangestellter Adverbialsatz mit folgendem <i>so</i> als Korrelat	<code>tree="s" &amp; tree="ns" &amp; #1 &gt;@l #2 &amp; #1 &gt;[tree=/adv.+/] #2 &amp; #2 . " , " . „so“</code>	81
7	Vorangestellter Adverbialsatz in einem Wissenschaftstext	<code>tree="s" &amp; tree="ns" &amp; #1 &gt;@l #2 &amp; #1 &gt;[tree=/adv.+/] #2 &amp; meta::stil="wissenschaft"</code>	64

Die Gestaltung der Suchanfragen baut auf der Knoten-Kanten-Struktur der Baumbank auf, die systematisch Form-Funktions-Zusammenhänge abbildet. Die Grundstruktur der Suchanfragen basiert darauf, dass Funktionen als Relationen zwischen Knoten abgebildet werden. In Suchanfrage 1 bspw. werden mit *tree="s"* und *tree="ns"* zunächst die Knoten festgelegt, die Gegenstand der Relation sind. Dabei ist mit *tree* für Baum die Domäne festgelegt, innerhalb der nach einem solchen Knoten gesucht wird (im vorliegenden Fall einmal nach einem Satz, einmal nach einem Nebensatzknoten). In der Suchanfrage wird mit `& #1 >@l #2` dann festgelegt, welcher Art die Beziehung zwischen den beiden Knoten sein soll, in diesem Fall, dass der Knoten 2 (also der NS-Knoten) am weitesten links steht – auf diese Weise ist



bei KAJUK wiederum liegt der Schwerpunkt auf dem Zusammenspiel von Semantik und Syntax bei der Verknüpfung von Sachverhaltsdarstellungen.

## IV.2 Formelhafte Sprache in RIDGES als Ausgangspunkt zur Registererschließung

### IV.2.1 Gegenstand

Durch die Schilderung des folgenden Fallbeispiels zeigen wir, wie der Zusammenhang zwischen sprachlicher Variation und außersprachlichem Kontext korpusbasiert erschlossen und vermittelt werden kann. Gegenstand der Untersuchung ist hier eine Gruppe von Texten (repräsentiert durch das Korpus) und ihr sprachliches Profil einerseits sowie der Wandel möglicher syntaktischer Erweiterungen des untersuchten sprachlichen Phänomens andererseits. Hierbei dient eine Instanz formelhafter Sprache, das Bundle *ist gut*, der Erschließung eines Teilbereichs der Registerkonstellationen im Zeitraum zwischen dem 15. und 20. Jahrhundert.

Unter Registern verstehen wir Varietäten, die funktional mit einer kulturell konventionalisierten Kommunikationssituation korrelieren (Matthiesen 2019; Neumann 2014; Biber/Conrad 2009). Varietäten konstituieren sich als quantitative Sets kookkurrierender sprachlicher Merkmale und kommen situationsabhängig zum Einsatz. Kommunikationssituationen werden durch eine Konstellation außersprachlicher Komponenten wie Produzent\*in, Rezipient\*in, Thema, kommunikatives Ziel, Konzeption und Diskursmodus charakterisiert, wobei die außersprachlichen Komponenten einen nachvollziehbaren funktionalen Einfluss auf die Wahl eines konkreten sprachlichen Mittels haben. Die Kommunikationssituationen, die in einer Kultur vorkommen, beeinflussen das Registerinventar einer Sprache. Dementsprechend können sich auch neue Register entwickeln, sobald Kommunikationssituationen kumulieren, und verschwinden, sobald Kommunikationssituationen seltener werden.

Registerunterschiede findet man auf allen sprachlichen Ebenen. Als ein sprachliches Merkmal auf der lexikalischen Ebene (unter vielen), das ein Register konstituiert, gilt die Formelhaftigkeit, die sowohl quantitativ, also in Bezug auf die relative Häufigkeit formelhafter Wendungen, als auch qualitativ durch die Prominenz bestimmter formelhafter Wendungen ein Registermerkmal bilden kann.<sup>18</sup> Als Hin-

---

<sup>18</sup> In der historischen Linguistik ist formelhafte Sprache unter verschiedenen Terminologien (›Phrasem‹, ›Phraseologie‹) vor allem im lexikographischen Kontext behandelt worden, durch die Etablierung diachron ausgerichteter konstruktionsgrammatischer Ansätze weitet sich die Perspektive auf die Schnittstelle zur Syntax aus (Habermann 2010, S. 105–106).

weise auf ein Register können Instanzen formelhafter Sprache untersucht werden, die in Form von Mehrwortsequenzen, Kollokationen und Lexical Bundles an der Oberfläche ermittelbar sind. Formelhafte Sprache meint hier die konventionalisierte Verfestigung von Wortfolgen, die funktional bedingt sehr häufig in Kombination auftreten (Filatkina 2018, S. 1).<sup>19</sup> Es gibt unterschiedliche Ansätze zur Ermittlung und Modellierung von formelhafter Sprache – einige sind eher semantisch (mit dem Kriterium Nichtkompositionalität), bei vielen anderen ist die Auswahl frequenzbasiert (Evert 2009).

Lexical Bundles können als eine Art erweiterte Kollokationen verstanden werden. Sie bestehen aus einer Sequenz von drei oder vier Wörtern, kookkurrieren hochfrequent und erfüllen eine bestimmte Text- oder Diskursfunktion (Cortes 2004, S. 400; vgl. ebenfalls Biber [u. a.] 1999, S. 990; Sinclair 1991).<sup>20</sup>

Lexical Bundles behandeln wir als Phänomen an der Grammatik-Lexikon-Schnittstelle (vgl. Filatkina 2018, S. 58), die mit dem Register des untersuchten Textes in unmittelbarer Wechselbeziehung steht (vgl. Filatkina 2018, S. 69). Die rein quantitative Ermittlung von Lexical Bundles ermöglicht es uns

»durch induktive Identifikationen von statistisch frequenten, konventionalisierten Wortverbindungen den Sprachgebrauch zu beschreiben, der für bestimmte Diskurse als typisch im Vergleich zu anderen Sprachauschnitten [...] und deshalb als musterhaft gelten muss.« (Filatkina 2018, S. 50)

Wir untersuchen Lexical Bundles als Strukturen von festen und offenen Elementen, so genannte hybride Bundles. Hybride Bundles sind Wortketten, in denen ein Teil der Wortsequenz festgelegt wird und ein anderer Teil wertoffen bleibt (vgl. Tab. 3).

**19** Dabei ist die reine Frequenz noch keine Bedingung für eine Konventionalisierung von sprachlichen Mustern (Filatkina 2018, S. 11). Mehrwortsequenzen sind ganz allgemein kookkurrierende Wortfolgen, deren grammatischer Status sowie deren Funktion nicht vorher festgelegt wird. Kollokationen stellen eine Abstraktion auf einer syntagmatischen Ebene dar, die nicht direkt auf einer konzeptuellen oder semantischen Wortbedeutungsebene der im Wortpaar enthaltenen Wortformen fußt (Firth 1957, S. 196). Die Entscheidung, ob die hier untersuchten Mehrwortsequenzen den Status einer Kollokation erlangen, nehmen wir hier nicht vor. Zu den Schwierigkeiten der Kollokationsermittlung in historischen Texten siehe Habermann (2010).

**20** »Lexical bundles are recurrent expressions, regardless of their idiomaticity, and regardless of their structural status. That is, lexical bundles are simply sequences of word forms that commonly go together in natural discourse« (Biber [u. a.] 1999, S. 990).



**Tab. 3:** Bundles-Struktur mit Kern *ist gut* und wertoffenen Stellen vor und hinter dem Bundle-Kern. Die wertoffenen Stellen können unterschiedlich syntaktisch eingebunden sein. Für die Beispiele verwenden wir die normalisierte Textrepräsentation aus dem RIDGES-Korpus.

Linksadjazente Erweiterung	Bundle-Kern	Rechtsadjazente Erweiterung
<i>Das Kraut</i>	<i>ist gut</i>	<i>wenn, so</i>
<i>Auf die Lider gestrichen</i>		<i>bei, für, wider etc. denen, dem</i>

Wir zeigen in unserer Fallstudie, wie man solche hybriden Bundles exemplarisch in Kräutertexten suchen und aus einer Registerperspektive diskutieren kann. Ein durch Texte gut belegter und historisch-außersprachlich nachvollziehbarer Registerwandel, genauer eine Registerspezifizierung, vollzieht sich im Bereich Medizin. Im 15. Jahrhundert kann man für die Vernakularsprache Deutsch<sup>21</sup> aufgrund von fehlender Spezialisierung und mangelnder Netzwerkbildung noch nicht von medizinisch-wissenschaftlichen Registern mit klarem Profil sprechen. Durch die Herausbildung einer wissenschaftlichen Community im Bereich der Medizin und die damit einhergehenden Kommunikationssituationen entstehen neue kommunikative Anforderungen, die sich in der Verwendung bestimmter konventionalisierter sprachlicher Merkmale manifestieren (vgl. Schäfer 2018; Zeige 2015; Gloning 2007; Habermann 2001).<sup>22</sup> Wissenschaftliche Register zeichnen sich im Gegensatz zu alltagssprachlichen Registern durch eine eigene konventionalisierte Formelhaftigkeit und die Verwendung von Fachtermini mit starker Informationsdichte aus, wobei eine spezialisierte, in höherem Maße eingeweihte Leserschaft adressiert wird (Roelcke 2018; Pörksen 1998; Klein 2011). Die Entstehung wissenschaftlicher Register ist an die Entstehung einer wissenschaftlichen Community gekoppelt und erstreckt sich im Bereich der Medizin für das Deutsche über einen Zeitraum von mehreren Jahrhunderten (Prinz/Schiewe 2018).

Die Entwicklung oder Spezialisierung wird durch eine sukzessive sprachliche Konventionalisierung rekurrenter Funktionen einer Kommunikationssituation gekennzeichnet. In diesem Kontext ist der uns interessierende Bundle-Kern *ist gut*

<sup>21</sup> Wissenschaftliche Kommunikation findet lange auf Latein statt (für einen Überblick zum Sprachenwandel siehe Prinz/Schiewe 2018).

<sup>22</sup> Im 21. Jahrhundert können wir die Existenz vieler medizinischer Subregister wie *Forschungsbericht*, *Vortrag*, *Fachartikel* oder *Arztbrief* annehmen, die jeweils durch verschiedene außersprachliche Komponenten wie Adressaten (z. B. Laien oder Fachpublikum), kommunikatives Ziel (z. B. Information oder Aufforderung) oder Diskursmodus (z. B. narrativ oder argumentativ) konstituiert sind.

zu sehen (vgl. Tab. 3), der eine Referenz auf Anwendungsbereiche eines Krautes signalisiert und somit im Sinne von Cherubim (1990) als indexikalisches Zeichen eine Kommunikationssituation einleiten könnte.

Untersuchungsobjekte des hier geschilderten Fallbeispiels sind drei Aspekte des hybriden Lexical Bundles *ist gut x*:

- 1.) Das Verteilungsmuster des Bundle-Kerns *ist gut* im Korpus;
- 2.) Die Vielfalt der rechtsadjazenten Erweiterungen;
- 3.) Das Verteilungsmuster ausgewählter rechtsadjazenter Erweiterungsvarianten.

#### IV.2.2 Eignung des Phänomens für die korpusbasierte Lehre

In fnhd. Zeit nehmen der Umfang und die Diversität der Textüberlieferung zu. Mit Beginn der nhd. Epoche bleiben Umfang und Diversität erhalten, werden jedoch zunehmend ab dem 17. Jahrhundert von Standardisierungsprozessen überlagert, die jedoch nicht für alle Varietäten hinsichtlich Intensität und Qualität identisch sind. Dieser Umstand schafft sowohl die Möglichkeit als auch die Notwendigkeit, Wandelphänomene im Produktionskontext der Texte, in denen sie vorkommen, zu untersuchen. Das Frühneuhochdeutsche und das Neuhochdeutsche bieten somit besonders viel Anschauungsmaterial für eine Diskussion des Verhältnisses von Variation und Standard in der universitären Lehre. Die Interdependenz zwischen Szenarien der Textproduktion und den ermittel- und analysierbaren Oberflächenformen kann durch eine korpusbasierte diachrone Betrachtung von Texten eines Registertyps illustriert werden. Hierbei ist die linguistische Analyse der untersuchten Form ebenso interessant wie die Funktion ihrer Verwendung und die außersprachlichen Veränderungen der Registertypen. Somit eignet sich die Untersuchung von Lexical Bundles für die Vermittlung historischer Varietätenkonstellationen.

Eine Annäherung an die diachrone Beschreibung des medizinischen Registers ist nur mit einschlägigem historischem Textmaterial möglich, wie es bspw. durch das diachrone Kräuterkundekorpus RIDGES vorliegt.

Speziell die Behandlung des Lexical Bundles *ist gut* und die Datengrundlage des RIDGES-Korpus eignen sich für die Vermittlung von Variation im Fnhd. und Nhd. aus folgenden Gründen:

- 1.) Das Bundle ist mit Hilfe des RIDGES-Korpus methodisch leicht ermittelbar.
- 2.) Das Bundle und seine Erweiterungen können auch ohne tieferegehende theoretische Kenntnisse ausgehend von am Nhd. orientierten Modellen beschrieben werden.
- 3.) Die im RIDGES-Korpus enthaltenen Metadaten erlauben einen systematischen und dynamisch erweiterbaren Bezug auf außersprachliche Kontexte.

- 4.) Da das RIDGES-Korpus eine teilweise opportunistische Textsammlung beinhaltet und die Texte in Bezug auf den Umfang nicht identisch sind, eignet sich die Datengrundlage, um die Aussagekraft der Ergebnisse zu problematisieren. Das heißt, die Studierenden werden angeleitet, über Datengrundlagen zu reflektieren.
- 5.) Ein Abgleich der Daten mit sprachwandeltheoretischen Modellen, die eine Interaktion von Varietätenkonstellation und sprachlichem Wandel beinhalten (Keller 2014; Zeige 2011; Labov 2006; Koch 2004), kann als Diskussionsgrundlage die Frage aufwerfen, inwiefern der Sprachvergleich des Standardneuhochdeutschen mit historischen Nicht-Standardvarietäten überhaupt gerechtfertigt ist.

#### IV.2.3 Umgang mit der Thematik in gängigen Lehrwerken

Die in dieser Studie verknüpften Themenbereiche Register und formelhafte Sprache wurden aus mehreren Gründen in der universitären Lehre bisher kaum systematisch thematisiert.<sup>23</sup> Die bereits angesprochene, in der Tradition der Junggrammatiker stehende weitestgehende Schwerpunktsetzung auf die grammatischen Ebenen Phonologie und Morphologie mag die starken Einschränkungen in der didaktischen Aufbereitung syntaktischer Themen, insbesondere aber ebenenübergreifender Themen wie Kollokationen, Bundles und Phraseologismen überhaupt erklären.<sup>24</sup> Darüber hinaus ist die Behandlung konkreter Wortkombinationen im Allgemeinen aufgrund ihrer starken Varietätengebundenheit in einem grammatischen Lehrwerk nicht zweckmäßig. Zweckmäßig ist aber die Vermittlung der Fähigkeit, sich über frequente Wortkombinationen ein Varietätenspektrum empirisch zu erarbeiten. Letzteres setzt den Umgang mit empirischem Material (Korpora) und methodischen Werkzeugen der Datenextraktion, -verarbeitung und -analyse voraus.<sup>25</sup>

---

<sup>23</sup> »Sie [Instanzen formelhafter Sprache; G. S.] gehören besonders in den Nachschlagewerken älteren Datums zu ›versteckten Informationen‹ mit zufälligem ›Beispielsatz‹-Charakter, die nicht eigens ausgewiesen werden, und vermitteln ein verzerrtes, wenn nicht unzureichendes Bild über die phraseologische Vielfalt früherer Sprachstufen« (Filatkina 2018, S. 83).

<sup>24</sup> Vgl. zur allgemeinen Problematik der modularen Behandlung des Themas ›Syntax‹ auch Riecke [u. a.] (2004).

<sup>25</sup> Das Aufkommen von historischen Korpora und Werkzeugen der Korpusanalyse hat der Forschung zu ebenenübergreifenden Wortkombinationen im Allgemeinen großen Aufschwung verschafft (vgl. Helbig 2006, S. 165).

Die korpusbasierte Forschung hat eine Hinwendung zum Themenbereich Wortsequenzen vor allem im angloamerikanischen Raum in Bezug auf moderne Gegenwartssprachen eingeleitet (Evert 2009; Sinclair 1991). In der Lexikographie und Phraseologie (Habermann 2010, S. 104), aber auch in der situativen und funktionalen Variationsforschung spielen Kollokationen, Lexical Bundles und formelhafte Wendungen zunehmend eine konstitutive Rolle (Filatkina 2018). Eine verstärkte Hinwendung zu Mehrwortphänomenen an der Grammatik-Lexikon-Schnittstelle ergibt sich in jüngerer Zeit auch durch die Konstruktionsgrammatik (bspw. Hilpert/Gries 2016; Goldberg 2006).

In Bezug auf die beiden Themen Formelhaftigkeit und Register besteht eine starke Divergenz zwischen dem Forschungsstand und seiner Rezeption in einschlägigen Grundlagen- und Lehrwerken zur Sprachgeschichte. Eine umfangreiche tabellarische Übersicht über die Thematisierung von formelhafter Sprache in Lehr- und einschlägigen Grundlagenwerken zur Sprachgeschichte bietet Filatkina (2018). Unter anderem weist sie nach, dass formelhafte Sprache weder in speziell syntaktischen Lehrwerken wie Fleischer/Schallert (2011), noch in sprachstufenspezifischen Lehr- und Nachschlagewerken wie der ›Frühneuhochdeutschen Grammatik‹ (Hartweg/Wegera 2005; Ebert [u. a.] 1993) oder allgemeinen Sprachgeschichten wie Schmidt (<sup>1</sup>2013) oder Schmid (<sup>2</sup>2013) thematisiert wird.

#### IV.2.4 Möglichkeiten der Suche von hybriden Lexical Bundles in RIDGES

In diesem Abschnitt demonstrieren wir, wie Instanzen des Bundles *ist gut* im RIDGES-Korpus ermittelt werden können. Ziel ist das Aufzeigen datenbasierter Hypothesenbildung zu variationslinguistischen Themenbereichen unter Einbezug text- und sprachstufenspezifischen sowie außersprachlichen Wissens. Zunächst wird die Korpusuche des Lexical Bundles *ist gut* erläutert. In einer exemplarischen korpusbasierten Datenvisualisierung bilden wir anschließend die quantitative Distribution der rechtsadjazenten Erweiterungsvarianten des Bundles auf den Faktor ›Zeit‹ ab und diskutieren aus der Visualisierung ableitbare Hypothesen.

Der erste Schritt einer Lexical-Bundles-Studie dient somit der Ermittlung einer potentiellen Instanz von formelhafter Sprache bezogen auf ein Korpus. Die Grundlage für die Ermittlung bilden Textrepräsentationen in Form von Transkription und Normalisierung (Odebrecht [u. a.] 2017). Eine Bundle-Ermittlung ist vollständig korpusgetrieben. Wir verwenden eine Mischform, bei der die Analyse der Länge und der Werte potenzieller Bundles vor dem Hintergrund einer basalen Struktur, im weitesten Sinne formelhafter Konstruktion, festgelegt wird.

Zunächst erfolgt eine quantitative Ermittlung von Wortformsequenzen. Lexical Bundles werden anhand ihrer Frequenz *ad hoc* und korpusbezogen identifiziert.

Hierfür werden die Zeichenketten *ist* und *gut* in direkter Präzedenz auf der Annotationsebene ›norm‹ gesucht.<sup>26</sup> Die auf dieser Annotationsebene enthaltene Normalisierungsannotation erlaubt eine gebündelte Suche nach sämtlichen belegten Schreibvarianten, die einer am Nhd. orientierten normalisierten Form zugeordnet sind. Die Durchsicht der Kombinationslisten wie in Tab. 3 erfordert eine Formulierung von Kriterien für die Bewertung eines Bundles als *Lexical Bundle*, ohne dass dieses zuvor als theoretische kategoriale Einheit bekannt sein muss. Kriterien für die Bewertung einer Sequenz als *Lexical Bundle* sind die ihr beigemessenen kommunikativen Funktionen. Tab. 4 enthält einen Ausschnitt der Vorkommensliste von Sequenzen mit einer absoluten Häufigkeit von 15 und 16, was die höchste absolute Vorkommenszahl von Kollokationen im untersuchten Korpus darstellt. Während die absoluten Vorkommenszahlen der Sequenzen ähnlich sind, unterscheiden sie sich hinsichtlich ihrer Funktionen, die wir aufgrund unserer Kenntnis der im Korpus enthaltenen Texte ableiten können. Während die Sequenzen *Die Kraft und Wirkung, /Natur/ Zubereitung, trocken im dritten Grad* und *im Juli und August* Hinweise auf die Häufung konkreter Inhalte geben, indiziert das Bundle *ist gut wider die* das Vorkommen einer bestimmten Funktion.<sup>27</sup> Die Texte des RIDGES-Korpus haben die registerrelevanten Komponenten *Thema* (›Kräuterkunde‹) und *Handlungsziel* (›Information‹, ›Instruktion‹) gemein. Es scheint also plausibel, dass der Bundle-Kern *ist gut* die kommunikative Funktion ›Anzeigen eines Anwendungsbereiches/einer Indikation für den Einsatz eines Krautes‹ hat, die in allen diesen Texten bedingt durch die inhaltliche Komponente gehäuft vorkommt. Die Feststellung dieser mit einer Sequenz assoziierten kommunikativen Funktion liefert das notwendige Kriterium für die Konstatierung eines *Lexical Bundles*.

---

<sup>26</sup> Zum genauen Aufbau der Suchanfrage sowie zur Referenzierung der Suchergebnisse siehe elektronischer Anhang.

<sup>27</sup> Typischerweise werden die Sequenzen funktionalen Kategorien wie *Bundle der Meinungsäußerung*, der *Themeneinführung* oder der *Zeit/Raum-Deixis* zugeordnet und deren Distribution wird über verschiedene sprachliche Varietäten verglichen (vgl. bspw. Biber/Barbieri 2007).

**Tab. 4:** Ausschnitt aus einer anhand des RIDGES-Korpus erstellten Liste an Vier-Wort-Bundles aus der Frequenzanalyse in ANNIS.<sup>28</sup>

Normierte Wortform	Normierte Wortform	Normierte Wortform	Normierte Wortform	Vorkommen
Die	Kraft	und	Wirkung	16
ist	gut	wider	die	15
/	Natur	/	Zubereitung	15
trocken	im	dritten	Grad	15
im	Juli	und	August	15

Die Behandlung der Wortkombination in den historischen Wörterbüchern von Adelung und der Gebrüder Grimm stützen den Status von *ist gut* als formelhafte Wendung. So verweist Adelung (Adelung 1811, S. 852–861, Punkt 4) in einem gesonderten Punkt auf die Verwendung von *ist gut* in medizinischen Kontexten. Auch Grimm konstatiert den »teleologischen charakter« der formelhafte Wendung: sie

»ist gut weil sie im Hinblick auf einen im beziehungswort mehr oder minder deutlich erkennbaren zweck als »passend, zweckdienlich, tauglich, brauchbar, »tüchtig, »günstig« usw. erscheinen«

(Grimm [u. a.] 1984, S. 1225–1375, Punkt IA, 1b)

Auf dieser Grundlage schränken wir unsere Suchanfrage auf Strukturen ein, die die Wortabfolge *ist gut* beinhalten, um gezielter weitere Belege untersuchen zu können.

Aus den in der Liste vorkommenden Instanzen (*ist gut wider; ist gut denen; ist gut, wenn*) abstrahieren wir das hybride Lexical Bundel *ist gut X*. Nachdem die Instanz *ist gut X* explorativ ermittelt worden ist, erfolgt die linguistische Deskription unter zwei Aspekten (siehe Tab. 5):

- 1.) Struktur der festen Sequenz
- 2.) Struktur der Erweiterungen

Das untersuchte Bundle stellt syntaktisch eine Kopulakonstruktion bestehend aus der 3. Person Singular von *sein* und dem prädikativen Adjektiv *gut* dar. Da wir einen oberflächenbasierten quantitativen Ansatz verfolgen, sehen wir von der

<sup>28</sup> N-Gramme sind Frequenzanalysen von Sequenzen. Gesucht wurden Sequenzen aus 4 normierten Wortformen. Der Ausschnitt zeigt die Sequenzen und ihre Vorkommen. Bei der Suche nach Sequenzen wurden keine Stopplisten (u. a. für Interpunktion) berücksichtigt.

Untersuchung weiterer Varianten der kommunikativen Funktion ›Anzeigen eines Anwendungsbereiches/einer Indikation für den Einsatz eines Krautes‹, wie zum Beispiel auf formaler Ebene der Pluralform *sind gut* sowie auf funktionaler Ebene von Äquivalenten zu *gut* wie *geeignet*, *nutz* und *verwendbar* ab. Semantisch weist das Bundle einer Entität oder einem bestimmten Bearbeitungszustand die Eigenschaft ›gut‹ im Sinne von ›brauchbar‹ zu.

**Tab. 5:** Schematische Darstellung des Bundles *ist gut* mit Beispielen für die zwei offenen Positionen für Subjekte vor und Adverbiale sowie Objekte nach dem festen Kern.

Linksadjazente Erweiterung	Bundle-Kern	Rechtsadjazente Erweiterung
<i>damit das Haupthaar gewaschen</i>	<i>ist gut</i>	<i>wider das Haarausfallen</i>
<i>ab Stabwurtzel getrunken</i>		<i>dem der gebissen ist</i>
<i>Geißklauen gebrant zu pulver</i>		<i>wenn man das haupt damit salbt</i>

Die linksadjazente Erweiterung wird, wenn vorhanden, typischerweise durch das Subjekt der festen Sequenz gefüllt. Das Subjekt kann durch ein Lexem für ein bestimmtes Kraut oder eine Beschreibung einer Zubereitungs- oder Darreichungsform realisiert werden.<sup>29</sup>

Die zweite offene Stelle nach *ist gut* referiert auf Zustände, die die Bedingung für das Zutreffen der Konstatierung darstellen. Syntaktisch kann man die Belege in zwei Gruppen einteilen: Komplemente von *gut* oder adverbiale Bestimmungen der Kopulakonstruktion. Die Komplemente von *gut* werden durch Dativ- und Präpositionalobjekte realisiert (Beispiel 11 und 12: *dem*, *den Frauen*; Beispiele 13, 14 und 15: *wider das Haarausfallen*, *für das Hauptweh*, *in Skorbut*). Die adverbialen Bestimmungen machen die konditionale Bedeutungskomponente durch die Füllung mit einem konditionalen oder konditional interpretierbaren temporalen Nebensatz explizit (Beispiel 16: *wenn man das Haupt damit salbet*).

- (11) *ist gut dem, der gebissen ist von wilden Tieren*  
(J. Tallat, 1532, Artzney Büchlein der Kreutter)

<sup>29</sup> Die linksadjazenten Elemente referieren sehr häufig auf bestimmte Zubereitungsarten oder Zustände der Heilkräuter, die syntaktisch die Funktion des Subjekts zum finiten Verb *gut* bilden. Die Tatsache, dass die syntaktischen Subjekte durch ungewöhnliche Phrasentypen (adverbiale Nebensätze, Partizipialgruppen) ausgedrückt werden und regelmäßig in Bezug auch Numerus nicht mit dem finiten Verb kongruieren, ist eine interessante Beobachtung, die wir in diesem Rahmen nicht weiter vertiefen können.

- (12) *ist gut den Frauen zu ihrer Krankheit*  
(J. Tallat, 1532, Artzney Büchlein der Kreutter)
- (13) *ist gut wider das Haarausfallen Allopitia*  
(J. Tallat, 1532, Artzney Büchlein der Kreutter)
- (14) *ist gut für das Hauptweh*  
(J. Tallat, 1532, Artzney Büchlein der Kreutter)
- (15) *der Same ist gut in Skorbut, Wassersucht und dem Schwellen des Milzes*  
(E. Blackwell, 1750, Blackwellisches Kraeuterbuch)
- (16) *ist gut wenn man das Haupt damit salbt*  
(J. Tallat, 1532, Artzney Büchlein der Kreutter)

Die rechtsadjazente Erweiterung zeigt hinsichtlich der syntaktischen Funktionen (Objekt, Adverbial), der formalen Realisierung auf Satzgliedebene (Nominalgruppe im Dativ, Präpositionalgruppe, Nebensatz), der lexikalischen Füllung (Präpositionen *wider*, *für*, *in* 11–16) und in Bezug auf ihre Semantik Variation, die wir im Folgenden untersuchen.

Weiter untersuchbare kategoriale Variablen der rechtsadjazenten Erweiterungen sind demnach die Komplemente (realisiert durch Präpositionalobjekte und Dativobjekte) und Adjunkte (realisiert durch Adverbialsätze). Aufgrund der besseren Beleglage und der oberflächenbasierten Suchbarkeit wählen wir die Komplemente für unsere weitere Untersuchung aus.

Innerhalb der Komplemente gibt es die Erweiterungen durch eine Nominalgruppe im Dativ (*denen*, *den Frauen*) und eine Erweiterung durch eine Präpositionalgruppe, die durch verschiedene Präpositionen eingeleitet werden kann. Wir zählen ausschließlich die ersten Elemente der rechtsadjazenten Komplemente, also die Präpositionen der angeschlossenen Präpositionalgruppen sowie die Determinierer der Nominalgruppen, oberflächenbasiert als gleichberechtigte Varianten und sehen von einer weiteren Kategorisierung ab. Tab. 6 zeigt das Inventar der quantitativ auszuwertenden Varianten.

Ein Durchgehen der Belegliste lässt erkennen, dass (1) aus gegenwartssprachlicher Sicht ungewöhnliche Präpositionen (z. B. *in*, *wider*) für einen Anschluss sorgen und (2) sowohl commodiale (*für*, *bei*) als auch incommodiale Präpositionen (*wider*, *gegen*, *in*) für die Angabe der Indikation verwendet werden. Mit Hilfe der Belegliste (Suche nach Fünf-Gram-Bundles mit dem Anfang *ist gut* und der Angabe der Wortartenkategorie für die dritte Stelle) lässt sich ein genaues Bild zeichnen, welche Präpositionen auf den Bundle-Kern *ist gut* folgen. Damit wird die Fallstudie flexibel auf die Untersuchung von Bundles erweitert bzw. die Anfrage konkretisiert. Mit den Ergebnissen dieser Suchanfrage kann dann die Verteilung der unterschiedlichen Bundles im Korpus visualisiert werden (Abb. 3).



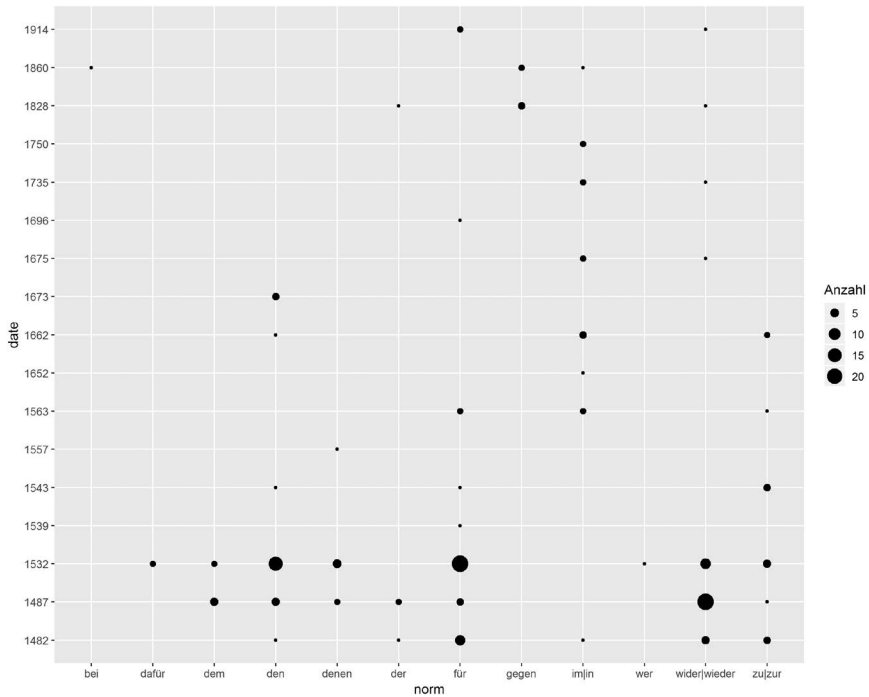
Tab. 6: Im Korpus ermittelte Varianten für die kategoriale Variable ›Komplementkopf‹

Bundle-Kern	Varianten für Köpfe der rechtsadjazenten Komplementerweiterungen
<i>ist gut</i>	<i>den</i>
	<i>denen</i>
	<i>dem</i>
	<i>für</i>
	<i>gegen</i>
	<i>in</i> (beinhaltet <i>in</i> und <i>im</i> )
	<i>wider</i> (beinhaltet auch Belege mit der fehlerhaften Normalisierung <i>wieder</i> , die als <i>wider</i> interpretiert wurden)
<i>zu</i> (beinhaltet <i>zu</i> und <i>zur</i> )	

Im Korpus sind 184 Instanzen von *ist gut* über den gesamten Textzeitraum ermittelbar, davon liegen 160 Belege mit Komplementerweiterung vor.

Abbildung 3<sup>30</sup> zeigt, dass die Verwendung von *ist gut* mit Komplementerweiterung diachron sowohl in Bezug auf die Quantität der Belege (Größe der Punkte), als auch in Bezug auf die Vielzahl der Komplementköpfe (Anzahl der besetzten Koordinaten) im Korpus mit dem Faktor ›Zeit‹ stark abnimmt. So enthalten die ersten drei Textüberlieferungsjahre 1482, 1487 und 1532 jeweils 17, 47 und 75 Belege, während es im 17. Jahrhundert schon nur noch insgesamt 15 und im jüngsten Text von 1914 lediglich drei Belege sind. Damit können wir eine erste Tendenz ableiten: Die 9 über den gesamten Zeitraum vorkommenden Varianten finden sich in einem Spektrum von 7 Varianten im Jahr 1532 bis zwei Varianten im Jahr 1914. Diese Beobachtung muss im Kontext des spezifischen Korpusdesigns gesehen werden, das mit der Vorgabe der gemeinsamen Komponente ›Kräuterkunde‹ ein sehr breites Kriterium zur Aufnahme von Texten bereitstellt. So bedingt der Wandel in der textuellen Repräsentation des Themas ›Kräuterkunde‹ auch die Wahrscheinlichkeit des Vorkommens bestimmter Funktionen: Während in der Frühen Neuzeit das Thema vorwiegend in einem praktischen Zusammenhang behandelt wurde, der sowohl die

<sup>30</sup> Die Visualisierung kann je nach Vorkenntnissen der Studierenden von der Lehrperson ausgeführt werden, sodass in der Lehrsituation lediglich die Diskussion der Visualisierung durchgeführt wird. An vielen Universitäten ist die Vermittlung des Umgangs mit R und anderen Datenvisualisierungsprogrammen jedoch mittlerweile lohnenswerter Bestandteil der grundlegenden Methodenausbildung, die eine wichtige Kompetenz für die Verfassung empirischer Haus-, Bachelor- und Masterarbeiten vorbereitet.



**Abb. 3:** Visualisierung der Vorkommen der Komplementköpfe des Bundles *ist gut*. Die Größe der Punkte gibt einen ungefähren Hinweis auf das absolute Vorkommen der Belege. In einem Zeitraum von 1482 bis 1914 sind nach Verfügbarkeit in zeitlichen Intervallen von ca. 30 Jahren 61 Texte enthalten, von denen 28 das Bundle *ist gut* enthalten. Die Visualisierung wurde mit Hilfe der R-Pakete `ggplot2` (Wickham 2016) und `plyr` (Wickham 2020) und auf Grundlage des reduzierten Dataframes aus der N-gram-Suche mit dem Beginn *ist gut* und der Frequenzanalyse in ANNIS erstellt. Betrachtet werden hier nur die Anschlüsse, gefiltert wird nach dem Publikationsdatum des Textauszugs, wobei mehrere Textauszüge auf ein Datum fallen können.

Bestimmung als auch die Zubereitung, Indikation und Anwendung eines Krautes enthielt, spezialisiert sich das Thema ›Kräuterkunde‹ mit Beginn des 19. Jahrhunderts in jeweils gesonderte Texte zur Bestimmung und Beschreibung (Botanik), Zubereitung (Pharmazie) sowie Indikation und Anwendung (Medizin). Da von den genannten spezialisierten Texten nur in der Botanik die Kräuterkunde prominentes Thema ist, weisen die im Korpus enthaltenen Texte des 19. Jahrhunderts eine klare Präferenz für die Beschreibung der Kräuter, nicht aber für deren Anwendung auf, womit die funktionale Verwendung des Bundles *ist gut* und insgesamt die Funktion einer Angabe der Indikation weniger prominent wird.

Ausgehend von den Ergebnissen der Datenvisualisierung sind verschiedene Ansätze zum Aufzeigen unterschiedlicher variationslinguistischer Aspekte vor-

stellbar, die eine weitergehende Auseinandersetzung mit den Texten und ihren Charakteristika erfordern.

Beispielsweise kann nun eine Hypothese darüber aufgestellt werden, dass die Verwendung von *ist gut* rückläufig ist und überlegt werden, ob der außersprachliche historische Kontext Anhaltspunkte für die Begründung dieser Hypothese liefern könnte. Diese Hypothese muss nun in einschlägigen Korpora überprüft werden und müsste in Form einer variationistischen Studie umgesetzt werden, in der auch andere Varianten betrachtet werden. Die insgesamt zu beobachtende Variantenreduktion der Komplementköpfe würde außerdem einen allgemein konstatierten zeitgleichen Standardisierungsprozess (Reichmann 1990) stützen, sie muss aber durch weiteres Datenmaterial ergänzt werden.

In Bezug auf die Distribution einzelner Varianten gibt es sowohl erwartbare als auch überraschende Ergebnisse. Der Rückgang der Nominalgruppen im Dativ (Dativus commodi) und der commodialen Verwendung der Präposition *zu* entspricht den gegenwartssprachlichen Verhältnissen. Ausgehend von dieser Beobachtung können sowohl die historische Existenz der Phänomene (Lateinabhängigkeit, Registerangemessenheit), als auch der Schwund- bzw. Wandelprozess (begünstigende grammatische Kontexte, Registerwandel, Fachsprachenspezifizierung) auch unter Hinzunahme einschlägiger Literatur untersucht werden. Dagegen überrascht die Kontinuität von incommodiale Gebrauch des eigentlich commodialen *für* und die durchgehenden und bis 1914 reichenden Belege für das im gegenwartssprachlichen Standarddeutschen kontextuell sehr restringierten *wider*. Aus Sicht des modernen Standarddeutschen ungewöhnlich erscheint die zeitlich kontinuierliche Verwendung von incommodiale *in*. Auch in diesem Fall kann das Wandelphänomen unter den oben genannten Perspektiven untersucht werden. Ebenso hätte man im Kontext von Empfehlungen im Krankheitsfall eher ein gehäuftes Vorkommen eines Anschlusses mit der Präposition *gegen* erwartet, was sich anhand der Daten jedoch nicht bestätigt.

Da nur in 28 von 61 im Korpus enthaltenen Texten die Konstruktion enthalten ist, wäre es mit Hilfe der im Korpus zur Verfügung stehenden Metadaten interessant, mögliche Faktoren zu untersuchen, die ein Vorkommen von *ist gut* ausschließen, was sich explorativ und unter Anwendung von außersprachlichem Kontextwissen über die Metadaten vollziehen lässt. So kommen 179 der 184 Belege von *ist gut* in Texten vor, die thematisch nach einzelnen Kräutern sortiert sind, in den ebenfalls im Korpus enthaltenen Vorreden zu den Kräuterbüchern gibt es nicht einen Beleg für *ist gut*. Auf dieser Basis kann ein direkter funktionaler Zusammenhang zwischen Registerkomponenten und Vorkommen sprachlicher Merkmale diskutiert werden.

## V. Fazit

Mit den beiden vorgestellten exemplarischen Fallstudien haben wir illustriert, inwieweit mit Hilfe von Korpora in der akademischen sprachhistorischen Lehre

- der Grundgedanke sprachlicher Variabilität und Wandelbarkeit durch eine Auseinandersetzung mit einschlägigen Phänomenbereichen greifbar gemacht werden kann;
- der Zugriff auf Variations- und Wandelphänomene durch eine Beobachtung von Sprachdaten in ihrem jeweiligen Kontext erweitert werden kann;
- Sprachgeschichte in der Lehre als ein auch durch Studierende erforschbarer Gegenstand erfahrbar gemacht werden kann.

Dabei ist eine Nutzung von Korpora kein empirischer Selbstzweck. So können diese Methoden nur in Zusammenhang mit den entsprechenden theoretischen Grundlagen für das Gewinnen linguistischer Erkenntnisse sinnvoll genutzt werden. Für die Arbeit mit Korpora in der Lehre wie für die Arbeit mit Korpora in der Forschung gilt: Die empirische Korpusarbeit ersetzt keine Theoriearbeit, es kommt vielmehr auf eine sinnvolle Verzahnung theoretischer Grundlagen und methodischer Ansätze an. Die Interpretation von Daten setzt stets theoretische Vorarbeit bzw. die Einbettung in Forschungszusammenhänge voraus.

Insofern ist der hier erfolgte Vorschlag der Nutzung von Korpora in der sprachhistorischen Lehre als Vorschlag zur Ergänzung etablierter Methoden der akademischen Lehre gedacht. Die Arbeit mit Korpora kann diese nie ersetzen, weil Korpora zunächst objektsprachliche Daten bereitstellen und dabei nicht über die metasprachliche Systematik verfügen, die bspw. Überblicksdarstellungen zur Sprachgeschichte oder zu einzelnen Sprachstufen bieten. Außerdem können Korpora mit ihren Sprachdaten immer nur punktuelle Einblicke in Wandelphänomene bieten und ersetzen als solche keine Wandeltheorien.

Wir plädieren vor diesem Hintergrund folglich ausdrücklich für eine sinnvolle Integration korpusanalytischer und sonstiger, gut etablierter Methoden. Der zentrale Mehrwert, der u. E. durch den Einsatz von Korpora erzielt werden kann, liegt in ihrer Dynamik: Korpora können mit sehr unterschiedlichen Forschungsfragen adressiert werden, sie ermöglichen die Bearbeitung eines ebenenübergreifenden Themenfeldes und sind in vielen Fällen sogar durch weitere Annotationsebenen, Metadaten und Formate erweiterbar.

Als Sprachdaten sind Korpora offen in Bezug auf ihre Nutzung für verschiedene Fragestellungen, Analysen und Interpretationen. Zwar sind – wie die Beispiele in diesem Text gezeigt haben – die Nutzungsmöglichkeiten durch die jeweilige Konzeption eines Korpus begrenzt. Jedoch können Korpora durchaus im Rahmen der Möglichkeiten ihres jeweiligen Korpusdesigns als empirische Grundlage für Fra-

gestellungen genutzt werden, die über das ihrer Erstellung zugrundeliegende Anliegen hinausgehen. Korpora bieten ein Datenpaket, das objektsprachliche Daten (d. h. Primärtexte) mit metasprachlichen Erschließungsdaten verzahnt. Damit erlauben Korpora das Entdecken von Belegen in ihren Kontexten. Studierende und Lehrende, die ein Korpus selbst untersuchen können, werden in die Lage versetzt, Belege für viele unterschiedliche Phänomene zu finden und miteinander in Beziehung zu setzen.

Korpora bieten heutzutage häufig mehrere Ebenen der Texterschließung bzw. der Erschließung linguistisch relevanter Phänomene. Durch die Mehrebenenarchitektur ist mit einem Korpus folglich ein ebenenübergreifender Ansatz möglich, wie unser Fallbeispiel zu den verschiedenen Varianten von Adverbialsätzen anhand der dem Phänomen inhärenten Interdependenz von Syntax und Semantik illustriert hat. Die oberflächenbasierte Suche nach dem Lexical Bundle *ist gut* zeigt die Möglichkeit der Thematisierung nicht-kanonischer Phänomene am Beispiel der formelhaften Sprache. Auch hier haben wir nach der ersten Oberflächenanalyse verschiedene Annotationsebenen genutzt, um Erkenntnisse abzubilden, die über die reine Lexik (vgl. Biber [u. a.] 2004; Beyer/Schulz 2022) hinausgehen. Ein wesentlicher Vorteil ist hier, dass eine Interpretation der Instanzen nicht vorgegeben sein muss, sodass auch Induktion und Deskription innerhalb eines Beschreibungsmodells oder modellunabhängig als Lehransatz eingesetzt werden können.

Korpora und Korpusauswertungsumgebungen sind nicht statisch – sie können erweitert werden (sowohl im Umfang als auch in der Annotationstiefe) und viele Projekte arbeiten an besseren Annotations-, Such- und Visualisierungswerkzeugen.

Eine Dynamik besteht darüber hinaus auch in den möglichen Abstufungen der Intensität der Korpusnutzung in der Lehre. So sind (mindestens) die folgenden Szenarien der Korpusnutzung in der akademischen Lehre denkbar (die Übergänge sind dabei natürlich fließend):

- **Szenario 1:** Geringer Korpusanteil: Die Lehrenden greifen vereinzelt auf veröffentlichte Korpora zurück, um Belege oder ggf. Texte/Textausschnitte herauszusuchen.
- **Szenario 2:** Mittlerer Korpusanteil: Die Lehrenden beziehen die Studierenden in die Korpusnutzung ein. Sie führen in die Korpusarchitektur und -nutzung ausgewählter Korpora ein und leiten die Studierenden zur Arbeit mit dem Korpus an. Diesem Szenario sind die beiden Fallstudien in unserem Text zuzuordnen.
- **Szenario 3:** Hoher Korpusanteil: Die Lehrenden beteiligen die Studierenden am Korpusaufbau (entweder in Ergänzungsarbeit zu bestehenden Korpora oder im Sinne des Aufbaus eines neuen Korpus). So entsteht bspw. das Korpus RIDGES in dem über mehrere Jahre hinweg unterrichteten Seminar ›Historische Korpuslinguistik‹ an der Humboldt-Universität zu Berlin, in dem das

Korpus, sein Design, seine Architektur, sein Sampling, seine Erstellung und Anwendung selbst unterrichteter Gegenstand waren bzw. sind.

Es versteht sich von selbst, dass die Wahl und Ausgestaltung eines der genannten Szenarien vor dem Hintergrund der jeweiligen Lehrziele und Studienordnungen zu treffen ist. Wir hoffen, gezeigt zu haben, dass die Nutzung von Korpora viele lohnende Perspektiven für die akademische sprachhistorische Lehre bietet.

## Literaturverzeichnis

- Adelung, Johann Christoph 1811: Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart, Bd. 2, Wien.
- Ágel, Vilmos 2000: Syntax des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts, in: Werner Besch [u. a.] (Hgg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung, Berlin, S. 1855–1903.
- 2003: Prinzipien der Grammatik, in: Anja Lobenstein-Reichmann u. Oskar Reichmann (Hgg.): Neue historische Grammatiken. Zum Stand der Grammatikschreibung historischer Sprachstufen des Deutschen und anderer Sprachen, Tübingen, S. 1–46.
  - 2012: Junktionsprofile aus Nähe und Distanz. Ein Beitrag zur Vertikalisierung der neuhochdeutschen Grammatik, in: Jochen Bär u. Marcus Müller (Hgg.): Geschichte der Sprache – Sprache der Geschichte. Probleme und Perspektiven der historischen Sprachwissenschaft des Deutschen. Oskar Reichmann zum 75. Geburtstag, Berlin, S. 181–206.
  - 2017: Grammatische Textanalyse. Textglieder, Satzglieder, Wortgruppenglieder, Berlin u. Boston.
- Ágel, Vilmos u. Carmen Diegelmann 2010: Theorie und Praxis der expliziten Junktion, in: Vilmos Ágel u. Mathilde Hennig (Hgg.): Nähe und Distanz im Kontext variationslinguistischer Forschung, Berlin u. Boston, S. 345–396.
- Ágel, Vilmos u. Mathilde Hennig 2006: Grammatik aus Nähe und Distanz. Theorie und Praxis am Beispiel von Nähertexten 1650–2000, Tübingen.
- 2020: Kasseler Junktionskorpus (Version 1.1), Justus-Liebig-Universität Gießen, DOI: <https://doi.org/10.34644/laudatio-dev-nCQsCnMB7CArCQ9CDmun>.
- Ágel, Vilmos u. Laura Sievers 2020: Markierte Vorfeldbesetzung im Neuhochdeutschen. Zur Grammatikalisierung einer neuen Vorfeldstruktur, in: Delphine Pasques u. Claudia Wich-Reif (Hgg.): Textkohärenz und Gesamtsatzstrukturen in der Geschichte der deutschen und französischen Sprache vom 8. bis zum 18. Jahrhundert. Akten zum Internationalen Kongress an der Universität Paris-Sorbonne vom 15. bis 17. November 2018, Berlin, S. 461–491.
- Axel, Katrin 2002: Zur diachronen Entwicklung der syntaktischen Integration linksperipherer Adverbialsätze im Deutschen. Ein Beispiel für syntaktischen Wandel?, in: PBB 124, S. 1–43.
- Barbour, Stephen u. Patrick Stevenson 1998: Variation im Deutschen. Soziolinguistische Perspektiven, Berlin u. Boston.
- Behaghel, Otto 1928: Deutsche Syntax. Eine geschichtliche Darstellung, Heidelberg.
- Berlin-Brandenburgische Akademie der Wissenschaften 2020: Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache, Berlin.

- Beyer, Andrea u. Konstantin Schulz 2022: New Insights and methods of vocabulary acquisition in Latin classes, in: *Forma y Función* 35(2). DOI: 10.15446/fyf.v35n2.91129.
- Biber, Douglas 2012: Register as a predictor of linguistic variation, in: *Corpus Linguistics and Linguistic Theory* 8, S. 9–37.
- Biber, Douglas u. Frederica Barbieri 2007: Lexical Bundles in university spoken and written registers, in: *English for Specific Purposes* 26, S. 263–286.
- Biber, Douglas u. Susan Conrad 1999: Lexical Bundles in Conversation and Academic Prose, in: Hilde Hasselgar u. Signe Oksefjell (Hgg.): *Out of Corpora. Studies in Honor of Stig Johansson*, Amsterdam, S. 181–190.
- 2009: *Register, Genre, and Style*, Cambridge.
- Biber, Douglas u. James K. Jones 2009: Quantitative methods in corpus linguistics, in: Anke Lüdeling u. Merja Kytö (Hgg.): *Corpus Linguistics. An International Handbook*, Berlin, S. 1286–1304.
- Biber, Douglas [u. a.] 1999: *Longman Grammar of Spoken and Written English*, Harlow.
- Biber, Douglas [u. a.] 2004: A corpus linguistic investigation of vocabulary-based discourse units in university registers, in: Ulla Connor u. Thomas Albin Upton (Hgg.): *Applied Corpus Linguistics. A Multidimensional Perspective*, Amsterdam u. New York, S. 53–72.
- Bubenhöfer, Noah 2011: Korpuslinguistik in der linguistischen Lehre. Erfolge und Misserfolge, in: *JLCL* 26(1), S. 141–156.
- Cherubim, Dieter 1990: Rituell formalisierte Syntax in Texten des 16. und 19. Jahrhunderts, in: Anne Betten u. Claudia M. Riehl (Hgg.): *Neuere Forschungen zur historischen Syntax des Deutschen*, Berlin u. New York, S. 269–285.
- Claridge, Claudia 2009: Historical corpora, in: Anke Lüdeling u. Merja Kytö (Hgg.): *Corpus Linguistics. An International Handbook*, Berlin, S. 246–259.
- Coniglio, Marco [u. a.] 2014: HIPKON: Historisches Predigtenkorpus zum Nachfeld (Version 1.0), Humboldt-Universität zu Berlin, DOI: <https://doi.org/10.34644/laudatio-dev-yiTKcNB7CArCQ9CxLh>.
- Coniglio, Marco u. Roland Hinterhölzl (2021): Discourse conditions on relative clauses: A crosslinguistic and diachronic study on the interaction between mood, verb position and information structure, in: Anke Holler [u. a.] (Hgg.): *Information Structuring in Discourse*, Leiden u. Boston (Current Research in the Semantics/Pragmatics Interface 40), S. 201–233.
- Cortes, Viviana 2004: Lexical Bundles in published and student disciplinary writing. Examples from history and biology, in: *English for Specific Purposes* 23, S. 397–423.
- Coseriu, Eugenio 1992: *Einführung in die allgemeine Sprachwissenschaft*, Tübingen.
- Croft, William 2010: Language Evolution. Contact, Competition and Change by Salikoko S. Mufwene, in: *World Englishes* 29, S. 306–311.
- Czicza, Dániel u. Mathilde Hennig 2013: Aggregation, Integration und Sprachwandel, in: *Jahrbuch für Germanistische Sprachgeschichte* 4, S. 1–33.
- Donhauser, Karin 2015: Das Referenzkorpus Altdeutsch, in: Jost Gippert u. Ralph Gehrke (Hgg.): *Historical Corpora. Challenges and Perspectives*, Tübingen, S. 35–49.
- Ebert, Robert Peter [u. a.] 1993: *Frühneuhochdeutsche Grammatik*, Tübingen.
- Evert, Stefan 2009: Corpora and Collocations, in: Anke Lüdeling u. Merja Kytö (Hgg.): *Corpus Linguistics. An International Handbook*, Berlin, S. 1212–1248.
- Filatkina, Natalia 2018: Historische formelhafte Wendungen als Konstruktionen. Möglichkeiten und Grenzen der diachronen Konstruktionsgrammatik, in: *Linguistik Online* 90(3).
- Firth, John Rupert 1957: *Papers in Linguistics 1934–1951*, Oxford.
- Fischer, Hanna 2018: *Präteritumschwund im Deutschen*, Berlin u. Boston.
- Fleischer, Jürg 2007: Das prädikative Adjektiv und Partizip im Althochdeutschen und Altniederdeutschen, in: *Sprachwissenschaft* 32, S. 279–348.

- Fleischer, Jürg u. Oliver Schallert 2011: Historische Syntax des Deutschen. Eine Einführung, Tübingen.
- Gass, Susan M. u. Larry Selinker 2008: Second language acquisition. An introductory course, New York.
- Geyken, Alexander u. Lothar Lemnitzer 2018: Automatische Gewinnung von lexikografischen Angaben, in: Annette Klosa u. Carolin Müller-Spitzer (Hgg.): Internetlexikografie. Ein Compendium, Berlin u. Boston, S. 195–237.
- Gloning, Thomas 2007: Deutsche Kräuterbücher des 12. bis 18. Jahrhunderts. Textorganisation, Wortgebrauch, funktionale Syntax, in: Andreas Meyer u. Jürgen Schulz-Grobert (Hgg.): Gesund und krank im Mittelalter. Marburger Beiträge zur Kulturgeschichte der Medizin, Leipzig, S. 9–88.
- Goldberg, Adele 2006: Constructions at Work. The Nature of Generalizations in Language, Oxford.
- Granath, Solveig 2009: Who benefits from learning how to use corpora?, in: Karin Aijmer (Hg.): Corpora and Language Teaching, Amsterdam, S. 47–65.
- Grimm, Jacob u. Wilhelm Grimm 1984: Deutsches Wörterbuch, München (Bd. 9 Greander-Gymnastik).
- Guescini, Rolf u. Odebrecht, Carolin 2020: Laudatio Repository – Long-term Access and Usage of Deeply Annotated Information. URL: <https://www.laudatio-repository.org/> (Aufrufdatum: 13.12.2021).
- Habermann, Mechthild 2001: Deutsche Fachtexte der frühen Neuzeit: naturkundlich-medizinische Wissensvermittlung im Spannungsfeld von Latein und Volkssprache, Berlin.
- 2010: Kollokationen und ihre Funktion in der mittelhochdeutschen Syntax, in: Jahrbuch für Germanistische Sprachgeschichte 1, S. 104–122.
- Hartmann, Stefan 2018: Deutsche Sprachgeschichte. Grundzüge und Methoden, Tübingen.
- Hartweg, Frédéric u. Klaus-Peter Wegera 2005: Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit, Tübingen.
- Helbig, Gerhard 2006: Funktionsverbgefüge – Kollokationen – Phraseologismen, in: Ulrich Breuer u. Irma Hyvärinen (Hgg.): Wörter – Verbindungen. Festschrift für Jarmo Korhonen zum 60. Geburtstag, Frankfurt/Main, S. 165–174.
- Hennig, Mathilde 2009: Syntaktische Relationen in Nichtsätzen, in: Andrea Bachmann-Stein u. Stephan Stein (Hgg.): Mediale Varietäten. Gesprochene und geschriebene Sprache und ihre fremdsprachendidaktischen Potenziale, Landau, S. 211–238.
- 2013: The Kassel Corpus of Clause Linking, in: Paul Bennett [u. a.] (Hgg.): New Methods in Historical Corpora, Tübingen, S. 207–219.
- 2017: Grammatik und Variation im Spannungsfeld von Sprachwissenschaft und öffentlicher Sprachreflexion, in: Marek Konopka u. Angelika Wöllstein (Hgg.): Grammatische Variation. Empirische Zugänge und theoretische Modellierung, Berlin.
- Herbers, Birgit [u. a.] 2021: Dokumentation zum Referenzkorpus Frühneuhochdeutsch und Referenzkorpus Deutsche Inschriften, in: Bochumer Linguistische Arbeitsberichte 24.
- Hilpert, Martin 2008: Germanic future constructions. A usage-based approach to language change, Amsterdam u. Philadelphia.
- Hilpert, Martin u. Stefan Gries 2016: Quantitative approaches to diachronic corpus linguistics, in: Merja Kytö u. Päivi Pahta (Hgg.): The Cambridge Handbook of English Historical Linguistics, Cambridge, S. 36–53.
- Hirschmann, Hagen 2019: Korpuslinguistik. Eine Einführung, Stuttgart.
- Hölllein, Dagobert 2019: Präpositionalobjekt vs. Adverbial. Die semantischen Rollen der Präpositionalobjekte, Berlin u. Boston.
- Jäger, Ludwig 1998: Das Verhältnis von Synchronie und Diachronie in der Sprachgeschichtsforschung, in: Besch, Werner [u. a.] (Hgg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung, Berlin (HSK 2.1), S. 816–824.



- Kappel, Péter 2016: Vorangestellte Temporalsätze und ihre Integration in den Hauptsatz. Eine empirische Analyse von Nähe- und Distanztexten (1650–2000), in: *Temeswarer Beiträge zur Germanistik* 13, S. 7–21.
- Keller, Rudi 2014: Sprachwandel. Von der unsichtbaren Hand in der Sprache, Tübingen.
- Kempf, Luise 2018: Genre influence on word formation (change): A case study on German adjectival derivation, in: Richard Jason Whitt (Hg.): *Diachronic Corpora, Genre and Language Change*, Amsterdam, S. 302–331.
- Klein, Wolf Peter 2013: Sprachliche Zweifelsfälle als linguistischer Gegenstand. Zur Einführung in ein vergessenes Thema der Sprachwissenschaft, in: *Linguistik Online* 16(4).
- Klein, Thomas u. Stefanie Dipper 2016: Handbuch Referenzkorpus Mittelhochdeutsch, in: *Bochumer Linguistische Arbeitsberichte* 19.
- Klein, Thomas [u. a.] (2018): *Mittelhochdeutsche Grammatik. Tl. II: Flexionsmorphologie*, 2 Bde., Berlin u. Boston.
- Koch, Peter 2004: Sprachwandel, Mündlichkeit und Schriftlichkeit, in: *Zeitschrift für romanische Philologie* 120, S. 605–630.
- 2010: Sprachgeschichte zwischen Nähe und Distanz: Latein – Französisch – Deutsch, in: Vilmos Ágel u. Mathilde Hennig (Hgg.): *Nähe und Distanz im Kontext variationslinguistischer Forschung*, Berlin u. Boston, S. 155–206.
- König, Ekkehard u. Johan van der Auwera 1988: Clause integration in German and Dutch conditionals, concessive conditionals, and concessives, in: John Haiman u. Sandra A. Thompson (Hgg.): *Combining in Grammar and Discourse*, Amsterdam, S. 101.
- Krause, Thomas 2019: ANNIS: A Graph-based Query System for Deeply Annotated Text Corpora. Dissertation, Humboldt-Universität zu Berlin. DOI: 10.18452/19659.
- Krause, Thomas u. Amir Zeldes 2016: ANNIS 3. A new architecture for generic corpus query and visualization, in: *Literary and Linguistic Computing* 31, S. 118–139. DOI: 10.1093/llic/fqu057.
- Labov, William 1991: *Sociolinguistic patterns*, Philadelphia.
- 2004: Quantitative analysis of linguistic variation, in: Ulrich Ammon [u. a.] (Hgg.): *Sociolinguistics. An International Handbook of the Science of Language and Society*, Berlin, S. 6–21.
- 2006: *The social stratification of English in New York City*, Cambridge.
- Lötscher, Andreas 2005: Linkspanthetische Adverbialsätze in der Geschichte des Deutschen. Pragmatische Aspekte eines grammatischen Wandels, in: *PBB* 127, S. 347–376.
- Lüdeling, Anke 2017: Variationistische Korpusstudien, in: Marek Konopka u. Angelika Wöllstein (Hgg.): *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*, Berlin, S. 129–144.
- Lüdeling, Anke [u. a.] 2011: Variationism and underuse statistics in the analysis of the development of relative clauses in German, in: Yuji Kawaguchi (Hg.): *Corpus Analysis and Diachronic Linguistics*, Amsterdam.
- Lühr, Rosemarie [u. a.] 2013: *Fuerstinnenkorrespondenz 1.1 (Version 1.1)*, Universität Jena. DOI: <https://doi.org/10.34644/laudatio-dev-ZCSVC3MB7CARCQ9CVed>.
- Maitz, Péter 2012 (Hg.): *Historische Sprachwissenschaft. Erkenntnisinteressen, Grundlagenprobleme, Desiderate*, Berlin.
- Matthiesen, Christian M. I. M. 2019: Register in systemic functional linguistics, in: *Register Studies* 1, S. 10–41.
- McEnery, Tony u. Andrew Hardie 2012: *Corpus linguistics. Method, theory and practice*, Cambridge.
- Neumann, Stella 2014: *Contrastive Register Variation. A Quantitative Approach to the Comparison of English and German*, Berlin.
- Odebrecht, Carolin 2018: *MKM – ein Metamodell für Korpusmetadaten. Dokumentation und Wiederverwendung historischer Korpora*, Diss. Berlin.

- Odebrecht, Carolin [u. a.] 2017: RIDGES Herbiology: designing a diachronic multi-layer corpus, in: *Language Resources and Evaluation* 51, S. 695–725.
- Odebrecht, Carolin [u. a.] 2020: Dokumentation und Annotationsrichtlinien für das Korpus Ridges Herbiology Version 9.0 (ANNIS- und PAULA-Format).
- Petrova, Svetlana 2015: Synchronic variation and diachronic change in the expression of indefinite reference. Evidence from historical German, in: *Zeitschrift für Sprachwissenschaft* 34(2), S. 213–246.
- Polenz, Peter von 2008: *Deutsche Satzsemantik*, Berlin.
- Pörksen, Uwe 1998: *Deutsche Sprachgeschichte und die Entwicklung der Naturwissenschaften. Aspekte einer Geschichte der Naturwissenschaftssprache und ihrer Wechselbeziehung zur Gemeinsprache*, in: Werner Besch [u. a.] (Hgg.): *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, Berlin, S. 193–210.
- Prinz, Michael u. Jürgen Schiewe 2018: *Vernakuläre Wissenschaftskommunikation. Beiträge zur Entstehung und Frühgeschichte der modernen deutschen Wissenschaftssprachen*, Berlin u. Boston.
- Raible, Wolfgang 1992: *Junktion. Eine Dimension der Sprache und ihre Realisierungsformen zwischen Aggregation und Integration*, vorgetragen am 4. Juli 1987, Heidelberg.
- Reichmann, Oskar 1990: *Sprache ohne Leitvarietät vs. Sprache mit Leitvarietät*, in: Werner Besch (Hgg.): *Deutsche Sprachgeschichte. Festschrift für Johannes Erben zum 65. Geburtstag. Grundlagen, Methoden, Perspektiven*, Frankfurt/Main, S. 141–158.
- ReN-Team 2021: *Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650). Version 1.1: Archiviert im Hamburger Zentrum für Sprachkorpora*. Publikationsdatum 06.01.2021. Online verfügbar unter <https://doi.org/10.25592/uhhfdm.9195>.
- Riecke, Jörg [u. a.] 2004: *Einführung in die historische Textanalyse*, Göttingen.
- Rissanen, Matti 2008: From ›quickly‹ to ›fairly‹: on the history of ›rather‹, in: *English Language and Linguistics* 12, S. 345–359.
- Roelcke, Thorsten 2018: *Von der wissenschaftlichen Eignung der deutschen Sprache*, in: Michael Prinz u. Jürgen Schiewe (Hgg.): *Vernakuläre Wissenschaftskommunikation. Beiträge zur Entstehung und Frühgeschichte der modernen deutschen Wissenschaftssprachen*, Berlin u. Boston, S. 361–374.
- Römer, Ute 2005: *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*, Amsterdam.
- Schäfer, Pavla 2018: *Formelhafter Sprachgebrauch in Fachdiskursen der Schulmedizin, Naturheilkunde und Homöopathie. Erste Überlegungen zu einem Forschungsvorhaben*, in: Sören Stumpf u. Natalia Filatkina (Hgg.): *Formelhafte Sprache in Text und Diskurs*, Berlin u. Boston, S. 311–350.
- Schlachter, Eva 2020: *A complex grammaticalization scenario for the definite article*, in: Massimo Cerruti u. Stavroula Tsiplakou (Hgg.): *Intermediate Language Varieties*, Amsterdam, S. 18–41.
- Schmid, Hans Ulrich 2009: *Einführung in die deutsche Sprachgeschichte*, Weimar.  
– 2013: *Einführung in die deutsche Sprachgeschichte*, Stuttgart u. Weimar.
- Schmidt, Jürgen Erich u. Joachim Herrgen 2011: *Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung*, Berlin.
- Schmidt, Wilhelm <sup>1</sup>2013: *Geschichte der deutschen Sprache. Ein Lehrbuch für das germanistische Studium*, Stuttgart.
- Schnelle, Gohar 2020: *Verbstellungsvarianten als Indikator für Narrativität im Deutschen?*, in: Delphine Pasques u. Claudia Wich-Reif (Hgg.): *Textkohärenz und Gesamtsatzstrukturen in der Geschichte der deutschen und französischen Sprache vom 8. bis zum 18. Jahrhundert*. Akten zum Internationalen Kongress an der Universität Paris-Sorbonne vom 15. bis 17. November 2018, Berlin, S. 11–48.
- Schulz, Konstantin 2020: *CALLIDUS Project-Learner Data from a Study on Latin Language Learning*.
- Sinclair, John 1991: *Corpus, Concordance, Collocation*, Oxford.

- Sonderegger, Stefan 1979: Grundzüge deutscher Sprachgeschichte. Diachronie des Sprachsystems, Berlin.
- Szczepaniak, Renata u. Fabian Barteld 2016: Hexenverhörprotokolle als sprachhistorisches Korpus, in: Sarah Kwekkeboom [u. a.] (Hgg.): Perspektivwechsel oder: Die Wiederentdeckung der Philologie, Berlin, S. 43–70.
- Szmrecsanyi, Benedikt 2019: Register in variationist linguistics, in: Register Studies 1, S. 76–99.
- Trudgill, Peter 2002: Sociolinguistic variation and change, Edinburgh.
- Wegera, Klaus-Peter 2013: Language data exploitation: design and analysis of historical language corpora, in: Paul Bennett [u. a.] (Hgg.): New Methods in Historical Corpora, Tübingen, S. 55–76.
- Wegera, Klaus-Peter u. Sandra Waldenberger 2012: Deutsch diachron. Eine Einführung in den Sprachwandel des Deutschen, Berlin.
- Weiss, Zarah u. Gohar Schnelle 2017: Annotation of an Early New High German Corpus. The LangBank Pipeline. AG 4: Encoding language and linguistic information in historical corpora. 39. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Saarbrücken.
- Wickham, Hadley 2016: ggplot2: Elegant Graphics for Data Analysis, New York.  
– 2020: Package ›plyr‹. Tools for Splitting, Applying and Combining Data, New York.
- Wöllstein, Angelika (Hg.) 2016: Duden – Die Grammatik. Unentbehrlich für richtiges Deutsch, Berlin.
- Zeige, Lars Erik 2011: Sprachwandel und soziale Systeme, Hildesheim.  
– 2015: From Saussure to sociology and back to linguistics. Niklas Luhmann's reception of signifiant/signifié and languae/parole as the basis for a model of language change, in: Semiotica 207, S. 327–368.  
– 2019: Die Präpositionalgruppe im ältesten Deutsch. Korpus, Grammatik, Darstellung, 2 Bde., Habilitationsschrift, Humboldt-Universität zu Berlin.
- Zeige, Lars Erik [u. a.] 2023: Deutsch Diachron Digital – Referenzkorpus Altdeutsch (1.2). Humboldt-Universität zu Berlin. Homepage: <http://www.deutschdiachrondigital.de/> (Aufrufdatum 19.02.2023).