Eric Fuß, Marek Konopka & Angelika Wöllstein (Hg.). 2019. *Grammatik im Korpus. Korpuslinguistisch-statistische Analysen morphosyntaktischer Variationsphänomene*. Tübingen: narr. 357 S.

Besprochen von **Anna D. Havinga**: University of Bristol, 21 Woodland Road, UK-BS8 1TE Bristol, E-Mail: anna.havinga@bristol.ac.uk

<https://doi.org/10.1515/zrs-2020-2056>

Die Korpuslinguistik ist in der Sprachwissenschaft nicht mehr wegzudenken. Nach der ersten automatischen Untersuchung maschinenlesbarer Korpora im Jahr 1951 fand das computerunterstützte Analysieren von Textsammlungen ab den 1960er Jahren Verbreitung (McEnery & Hardie 2012: 37). Ein Anstieg korpuslinguistischer Studien ist vor allem in den 1990er Jahren zu verzeichnen, als die Software WordSmith (Scott 1996) auf den Markt kam. Ähnliche, frei verfügbare Softwareprogramme führten danach zu einem weiteren Zuwachs an korpuslinguistischen Analysen. Diese Entwicklung ging hauptsächlich vom englischsprachigen Raum aus. Lüdeling (2006) bemerkt, dass es vor dem 21. Jahrhundert fast so schien, „als wolle die germanistische Linguistik in Deutschland die Möglichkeiten der Korpuslinguistik verschlafen“. Als Ausnahme nennt Lüdeling das Institut für Deutsche Sprache in Mannheim (IDS), das auch heute noch als korpuslinguistisches Zentrum im deutschsprachigen Raum gilt.

Der hier besprochene Sammelband entstand aus am IDS durchgeführten korpuslinguistischen Pilotstudien. Er besteht aus einer Einleitung von ERIC FUß, MAREK KONOPKA und ANGELIKA WÖLLSTEIN, die den Schwerpunkt der korpusorientierten Untersuchung grammatischer Variation skizziert und dabei sowohl bisherige Untersuchungen als auch die in dem vorliegenden Band präsentierten Pilotstudien vorstellt (S. 9–25), sowie aus drei Pilotstudien zu bestimmten morphosyntaktischen Phänomenen (S. 29–297) und einem Kapitel zur statistischen Auswertung von Korpusrecherchen durch das Analysetool KoGra-R (=Korpusgrammatik-R) (S. 299–357). Die vier Kapitel des Sammelbandes erfüllen das von den HerausgeberInnen vorgegebene Ziel: Sie stellen anschaulich dar, „wie ein korpuslinguistischer Zugang genutzt werden kann, um die Vielfalt und Variabilität des Sprachgebrauchs in standardnahen Texten aufzuzeigen und in einer größeren Detailtiefe zu beschreiben“ (S. 20).

ANNA VOLODINA untersucht in ihrem Beitrag die Variation zwischen *angenommen* und *vorausgesetzt* als einbettende Prädikatsausdrücke aus syntaktischer, semantischer und metasprachlicher Perspektive anhand von Daten aus dem Deutschen Referenzkorpus (S. 29–90, 41). Basierend auf einer teils maschinellen, teils annotationsbasierten „manuellen“ korpuslinguistischen Analyse mit statischer

Auswertung von 6.949 Belegen für *angenommen* und 32.878 Belegen für *vorausgesetzt* zieht Volodina die folgenden Schlüsse: Beide Prädikatsausdrücke lizensieren identische syntaktische Strukturen, nämlich eine Präferenz für V2-Strukturen, wobei sich diese Präferenz bei *angenommen*-Strukturen deutlicher zeigt (98,7 %) als bei *vorausgesetzt*-Strukturen (90,3 %) (S. 47). Volodinas Analyse ergab aber auch Unterschiede im Hinblick auf topologische Präferenzen: *vorausgesetzt*-Konnekte werden präferiert nachgestellt, während *angenommen*-Konnekte häufiger in der Anteposition (v.a. im Vor-Vorfeld) vorkommen (S. 53). Außerdem lassen sich Korrelationen zwischen der Wahl der topologischen Position und der Wahl des Verbmodus (S. 56–59) sowie zwischen Verbstellung bzw. Prädikatsausdruck und Moduswahl (S. 60–64) feststellen. So kommt *angenommen* fünfmal häufiger mit Konjunktiv vor als *vorausgesetzt* (S. 57). Volodina schließt aus diesen Analysen, dass die Befunde semantisch-funktional in dieselbe Richtung deuten, nämlich dass *angenommen* (vor allem *angenommen*-V2) „eine stärkere Tendenz in Richtung des Hypothetischen, des Nicht-Faktischen“ hat (S. 63f.). In Bezug auf die Semantik stellt sie außerdem fest, dass sich die Wahl bedeutungsmodifizierender Ausdrücke zwischen *angenommen* und *vorausgesetzt* unterscheidet: Modifikatoren, die häufig vor *angenommen* vorkommen, wie *mal* oder *aber*, werden zum Beispiel nur selten mit *vorausgesetzt* verwendet und umgekehrt (*immer angenommen* kommt fast nie vor, *immer vorausgesetzt* jedoch häufig) (S. 64–66). Dies zeige eine stärkere semantische Unterscheidung dieser Prädikatsausdrücke als erwartet (S. 71), die jedoch nicht weiter generalisiert wird.

Darüber hinaus zeigen sich auf der metasprachlichen Ebene Unterschiede. In Bezug auf den Faktor Medium, zum Beispiel, sind *vorausgesetzt*-Strukturen in der Publikumspresse signifikant überrepräsentiert, während *angenommen*-Strukturen signifikant unterrepräsentiert sind, was Volodina darauf zurückführt, dass Presstexte „eher über Fakten berichten als Texte in den Kategorien Literatur, Gesprochenes und Internet“ (S. 74). Genaueres zur Funktion der Strukturen in Presstexten oder in den anderen Medien wird nicht ausgeführt. Insgesamt kommt Volodina zum Ergebnis, dass sich die zwei untersuchten Prädikatsausdrücke „in der Art, wie sie die Aufmerksamkeit der Adressaten lenken, radikal unterscheiden“ (S. 86). Im Gegensatz zu *vorausgesetzt* wird mit dem *angenommen*-Konnekt die Aufmerksamkeit auf die Möglichkeit gelenkt, dass der Sachverhalt des *angenommen*-Konnekts wahr ist oder wahr sein könnte (S. 86). Aufmerksamkeitssteuerung könnte laut Volodina somit „zur grammatikalisierten Bedeutung dieser prädikativen Ausdrücke gehören“ (S. 86). Diese Argumentation ist anhand Volodinas Analysen gut nachvollziehbar. Allerdings erscheinen Erklärungsansätze auf metasprachlicher Ebene weniger gut ausgebaut. Außerdem hätte der Aufbau des Korpus (zum Beispiel, dass dem Medium „Internet“ hauptsächlich Wikipedia-Quellen zugrunde liegen) stärker problematisiert werden können.

PATRICK BRANDT & ERIC FUß untersuchen in ihrem Beitrag die Relativpronomenselektion und grammatische Variation von *was* und *das* in attributiven Relativsätzen (S. 91–209). Sie stellen fest, „dass die Alternation zwischen *das* und *was* systematisch von der Natur der syntaktischen Umgebung abgänglich ist, in der das Relativpronomen auftritt“ (S. 94). Aber auch sprachexterne Faktoren (Region) sind relevant (S. 95, 181–193). 18.307 Belegsätze von Relativpronomina wurden aus der Korpusgrammatik-Datenbank (KoGra-DB) extrahiert, weiterbearbeitet und untersucht (S. 106f.). Darüber hinaus wurden ergänzende Analysen im gesamten Deutschen Referenzkorpus bzw. in Teilkorpora durchgeführt. Nach sorgfältigen Analysen kommen die Autoren zum Schluss, dass die Wahl zwischen *d*- und *w*-Relativum von der Präsenz bzw. Absenz eines lexikalischen Bezugsnomens beeinflusst wird (S. 118). So wird normalerweise nach Antezedentien mit Nomen *das* als Relativum ausgelöst, während auf Antezedentien ohne Nomen fast ausschließlich die Relativierung mittels *was* folgt (S. 121). Konträr zur vorwiegenden Darstellung des *d*-Pronomens als Default-Relativum in der Literatur argumentieren Brandt und Fuß, dass *w*-Pronomina als (Default-) Relativierer verwendet werden, „wenn kein geeignetes substantivisches Antezedens vorliegt“ (S. 127). Im Gegensatz zu *d*-Formen verlangt *was* nämlich keine bestimmte syntaktische Kategorie seines Bezugselements. *W*-Pronomen werden also „generell zur Einleitung von Relativsätzen verwendet“ (S. 129). Die Autoren fügen jedoch hinzu, dass auch andere Faktoren, wie die Semantik des Bezugselements sowie interne Eigenschaften des Relativsatzes (z.B. die Satzgliedfunktion des Relativpronomens), die Verteilung von *das* und *was* beeinflussen können. Im Vergleich zu der sehr sorgfältigen Analyse grammatischer und semantischer Faktoren werden außergrammatische Aspekte (Medium, Land, Region) nur sehr kurz beleuchtet und liefern zum Teil vorhersehbare Ergebnisse. So zeigen die Autoren zum Beispiel, dass es „in Texten, die dem Bereich der Mündlichkeit näherstehen, eine deutliche Tendenz zum Relativsatzeinleiter *was* gibt“ (S. 186). Darüber hinaus zeigt sich „in Texten österreichischer oder schweizerischer (also südlicher und vor allem südöstlicher) Provenienz eine Tendenz zu vergleichsweise stärkerem Gebrauch des Relativsatzeinleiters *das*“ (S. 190f.). Allerdings wurde die Untersuchung sprachexterner Faktoren auf *das/was* nach den Quantoren *alles/manches/eines/keines* beschränkt. Obwohl diese Einschränkung gerechtfertigt wird, wirkt die Analyse sprachexterner Faktoren nicht so systematisch wie die der sprachinternen Aspekte. Auch in dem letzten Unterkapitel, das Lücken und nicht ausreichend behandelte Fragen aufzeigt und den Beitrag von Brandt und Fuß recht abrupt abschließt (S. 203–205), werden sprachexterne Gesichtspunkte nicht erwähnt, obwohl offene Fragen durchaus bestehen bleiben. Dies mag mit dem Fazit der Autoren, dass „die Selektion von *d*-Relativa in erster Linie strukturell, d.h. formal bedingt ist“ (S. 202), zusammenhängen. Um diesen Schluss über-

zeugender darzustellen, ist eine detailliertere Untersuchung der außergrammatischen Faktoren allerdings unerlässlich.

Auch die Untersuchung der Alternation von *zu*- und *dass*-Komplementen von PATRICK BRANDT unter Mitwirkung von FELIX BILDHAUER (S. 211–297) fokussiert auf grammatische und semantische Aspekte. Außergrammatische Faktoren (z.B. Mündlichkeit) werden nur kurz in einem Absatz angesprochen (S. 282). Dass diese Faktoren genauer untersucht werden müssten, wird allerdings expliziert. Trotz dieser Lücke liefern die Autoren interessante Erkenntnisse zur Variation zwischen *zu*- und *dass*-Komplementen. Zur Untersuchung dieser Alternation wurden Belege von *zu*-Infinitiven im Nachfeld sowie im Mittelfeld und *dass*-Sätzen (im Nachfeld) aus dem mit dem TreeTagger-Tagset annotierten Teil der Korpusgrammatik-Datenbank (KoGra-DB) extrahiert und daraus zwei Datensätze erstellt (S. 232–237). Datensatz I „umfasst Belege mit Akkusativobjektkontrollverben (*biten*, *zwingen*), inhärent reflexiven Verben (*sich anstrengen*, *sich weigern*) und Antikontrollverben (*anordnen*, *missbilligen*)“, während Datensatz II „Belege mit Subjektkontrollverben (*versprechen*, *drohen*), Dativobjektkontrollverben (*empfehlen*, *raten*) und wiederum Antikontrollverben“ enthält (S. 235). Von über 70.000 gültigen Belege wurden 7.685 Belege für sechs Variablen manuell annotiert und von Bildhauer statistisch ausgewertet. Diese Auswertung legt dar, dass das Bestehen einer Koreferenzbeziehung zwischen einem Element im Matrixsatz und dem Subjekt des eingebetteten Satzes „die entscheidende Rolle für die Wahl zwischen *zu*- und *dass*-Komplementierung spielt“ (S. 263). Die Analysen von Brandt und Bildhauer zeigen außerdem, dass *zu* infinitivisch komplementiert wird, wenn Kontrollbedingungen vorliegen (S. 289, siehe dazu v.a. Tab. 9 auf S. 249). Genauer gilt laut Brandt: „Je besser die Qualität der Kontrolle, desto eher wird *zu* infinit komplementiert“, wobei „besser“ hier

„im Sinne von Köpcke als eine Hierarchie der Qualität im Hinblick auf das Ideal der Identifizierung eines verantwortlichen Matrixargumentreferenten mit einem verantwortlichen Referenten des eingebetteten (logischen PRO-) Subjekts“

verstanden wird (S. 289). Neben Kontrolle kann aber auch „der Wunsch nach Ausdruck von Modalität die *zu*-Komplementation motivieren“ (S. 291, siehe dazu Abschnitt 4.4.1). Zum Abschluss werden methodische Aspekte angesprochen, die bei korpuslinguistischen Untersuchungen und statistischen Auswertungen generell berücksichtigt werden sollten (S. 292–293). Außerdem wird auf das Potenzial manueller Datensichtung und Annotationsarbeit hingewiesen. Diese Arbeit ist zwar zeitaufwendig, kann aber zu neuen Hypothesen führen. In der Korpuslinguistik geht es eben nicht nur darum, bestimmte Belege zu zählen und einander gegenüberzustellen, sondern Sprachgebrauch möglichst akkurat zu beschreiben. Dies gelingt den Autoren des Beitrags auf grammatischer und semantischer Ebe-

ne. Um die Alternation von *zu*- und *dass*-Komplementen wirklich zu verstehen, müssten aber auch, wie es die Autoren selbst ansprechen, außergrammatische Faktoren untersucht werden.

Der letzte Beitrag dieses Sammelbandes (S. 299–357) stellt das webbasierte Analysetool KoGra-R vor, das in den besprochenen Studien verwendet wurde, um statische Auswertungen durchzuführen. Von den Autoren (HANSEN-MORATH, SCHMITZ, SCHNEIDER & WOLFER) wird es als „Werkzeugkasten“ für korpusgestützte grammatische Untersuchungen“ bezeichnet. Die Analysearten dieses ‚Werkzeugkastens‘ können erweitert werden, umfassen aber bereits die wichtigsten Funktionen, u.a. die Darstellung von Tabellen und Diagrammen für Rohdaten und normierte Werte, den Chi-Quadrat-Test und Assoziationsplots (S. 299). KoGra-R, dessen Funktionen auf Häufigkeitsangaben basieren (S. 306), ist öffentlich über die Homepage des IDS Mannheim zugänglich. Die Benutzeroberfläche des Tools ist klar und einfach gestaltet. Programmierkenntnisse werden für die statistische Auswertung nicht vorausgesetzt, der verwendete R-Code wird aber zur Verfügung gestellt (S. 306).

Im Beitrag wird zunächst mithilfe von Screenshots erklärt, wie Daten eingegeben werden können (Abschnitt 2, S. 301–306). ForscherInnen können nicht nur Cosmas-Exportdateien laden und miteinander vergleichen, sondern auch eigene Kontingenztabellen mit Häufigkeitsangaben eingeben oder hochladen. Diese Tabellen werden auf ihre Wohlgeformtheit geprüft, bevor sie ausgewertet werden (S. 305). Abschnitt 3 des Beitrags (S. 306–324) erläutert die einzelnen Analysearten anhand der Variation zwischen *Werkes* und *Werks*. Abschnitt 4 (S. 324–328) behandelt die technische Implementation. Danach werden Ergebnisse von zwei Evaluationen von KoGra-R besprochen (S. 328–332). Die TeilnehmerInnen dieser Evaluation gaben an, mit Korpusrecherchen gut, aber mit statischen Analysen weniger gut vertraut zu sein (S. 330f.). Insgesamt wurde KoGra-R von ihnen als nützlich und einfach zu bedienen eingestuft (S. 331). Dieser Beitrag gibt somit einen guten Überblick über die Möglichkeiten statistischer Auswertungen durch KoGra-R. Hat man vor, KoGra-R selbst zu verwenden, empfiehlt es sich aber auch, das auf der IDS Webseite frei verfügbare Online-Tutorial zu konsultieren, das detailliertere Anleitungen bietet.

Wäre dieser letzte Beitrag schon als zweiter Beitrag in dem Sammelband gereiht worden, wären detaillierte Beschreibungen der Funktionen von KoGra-R nicht in jedem Beitrag nötig gewesen. Andererseits erlauben diese Beschreibungen, die einzelnen Beiträge unabhängig voneinander zu lesen. Auch der Fokus auf sprachinterne Faktoren von Variation in allen Beiträgen ist verständlich und wird im Titel sowie in der Einleitung angekündigt. Zwar geht keiner der Beiträge in genügender Tiefe auf außergrammatische Aspekte ein, aber es wird aufgezeigt, dass diese genauer untersucht werden sollten. Diaphasische, diastrati-

sche und diatopische Variation könnte für alle behandelten Variablen in weiteren Forschungsprojekten genauer herausgearbeitet werden, um den Gebrauch von bestimmten Varianten noch akkurater beschreiben zu können. Wie von den HerausgeberInnen versprochen, liefert der Sammelband ein „besseres empirisches Fundament“ (S. 9) für die deutsche Sprachwissenschaft. Dieses könnte aber durch zusätzliche Analysen weiter gestärkt werden. Der Sammelband schafft somit Anreize für weitere Studien, zeigt aber auch sehr anschaulich den Wert der Korpuslinguistik auf und kann somit als Lektüre für SprachwissenschaftlerInnen empfohlen werden.

Literatur

- Lüdeling, Anke. 2006. Zum Geleit. In: Lothar Lemnitzer & Heike Zinsmeister. *Korpuslinguistik. Eine Einführung*. Tübingen: Narr, 9.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Scott, Mike. 1996. *WordSmith Tools*. Oxford: Oxford University Press.