Michael Beißwenger et al. (Hg.). 2023. *Korpusgestützte Sprachanalyse. Grundlagen, Anwendungen und Analysen* (Studien zur deutschen Sprache 88). Tübingen: Narr Francke Attempto. 438 S.

Besprochen von **Barbara Aehnlich:** Universität Bremen, FB 10, Universitäts-Boulevard 13, D-28359 Bremen, E-Mail: ba_ae@uni-bremen.de

https://doi.org/10.1515/zrs-2025-2034

Der vorliegende Band umfasst 25 Beiträge von 49 Autor:innen und eine Einleitung, die ich allesamt in meiner Besprechung berücksichtige, da sie die vielfältigen Möglichkeiten korpusgestützter Sprachanalyse widerspiegeln. Das ist durchaus nicht üblich und resultiert in einer umfangreicheren Rezension des Buches, das nicht nur Einblicke in grundlegende Fragen der Korpuslinguistik bietet, sondern auch aktuelle Forschungsarbeiten und Entwicklungen vorstellt und den Einsatz von Sprachkorpora im Schulunterricht und in der universitären Lehre thematisiert.

Die Einleitung der Herausgeber:innen Michael Beißwenger, Eva Gredel, Lothar Lemnitzer und Roman Schneider (S. 11–24) würdigt die wissenschaftlichen Arbeiten Angelika Storrers – ihr wurde der Band zum 65. Geburtstag gewidmet. Die Beiträge des Bandes sollen den aktuellen Stand korpuslinguistischer Forschung darlegen und knüpfen an thematische Forschungsschwerpunkte der Jubilarin an. Erklärtes Ziel ist es, nicht nur Fachwissenschaftler:innen für die Lektüre zu gewinnen, sondern auch fortgeschrittene Linguistik-Studierende. Dieses Buch soll laut den Herausgeber:innen dazu dienen,

> „i) Sprachressourcen für das Deutsche und ii) Werkzeuge für deren Nutzung kennenzulernen, iii) aktuelle Forschungsfelder und Forschungsfragen zu identifizieren sowie Wissenschaftler:innen bei der Beantwortung einschlägiger Forschungsfragen und Korpusentwickler:innen bei aktuellen Problemstellungen im Zusammenhang mit dem Aufbau und der Aufbereitung von Korpusressourcen zu folgen und iv) den Umgang mit empirischen Korpusdaten sowie dessen theoretische und methodische Implikationen besser zu verstehen.“ (S. 20)

Die sechs Unterkapitel des Bandes werden auch im Folgenden nacheinander besprochen.

Empirie, Korpora und linguistische Theoriebildung

Der erste Abschnitt umfasst zwei reflektierende Aufsätze und wird eröffnet mit dem Beitrag von Ludger Hoffmann zum Thema „Sprachwissenschaft – Theorien und empirische Zugänge“ (S. 27–44). Der Autor skizziert in seinem Beitrag sprach-

wissenschaftliche Grundlagen und beschreibt den Zusammenhang zwischen Theorie und Gegenstand. Dabei zeigt er eindrücklich auf, dass eine „untheoretische Empirie“ (S. 41) keine tragfähigen und interessanten Ergebnisse liefern kann. Er macht deutlich, dass die Linguistik dringend eine Form der Lehre benötigt, die neue Datenzugänge und ihr Verhältnis zu Theorien umfasst.

Der Beitrag von Ulrich Schmitz „Infinity Corpus – Linguistischer Größenwahn einmal durchgespielt“ (S. 45–57) ist ein unterhaltsamer theoretischer Aufsatz zu Infinity Corpora. Schmitz beschreibt den Traum der Korpuslinguistik, auf der Grundlage unendlicher Datenmengen bzw. eines unendlich großen Korpus mit allen sprachlichen Äußerungen von Menschen Sprache umfassend beschreiben zu können, und zeigt auf, dass diese Idee nicht nur unrealistisch, sondern sogar weltfremd, überflüssig, ineffizient und dumm ist. (S. 52) Wichtig ist vielmehr, passende Daten zu vielversprechenden Forschungsfragen zur Verfügung zu stellen; alles andere sei zum Scheitern verurteilt.

Erhebung und Aufbereitung von Sprachkorpora

Die vier Aufsätze des zweiten Abschnitts beleuchten die Erstellung, Pflege und Organisation von Sprachkorpora. Marc Kupietz, Harald Lüngen und Andreas Witt führen in ihrem Beitrag „DeReKo im Kontext deutschsprachiger Gegenwartskorpora: Perspektiven – Ziele – Visionen“ (S. 61–78) in die Arbeit des Deutschen Referenzkorpus (DeReKo) am Leibniz-Institut für Deutsche Sprache ein und stellen Visionen zu dessen Weiterentwicklung vor. Zu diesen gehören etwa die Verteilung von ausgewählten Korpusressourcen auf mehrere Standorte, die Weiterentwicklung multilingual vergleichbarer Korpora, pragmatische Aspekte wie „Standardkomponenten für Lizenzvereinbarungen zwischen Autor*innen und Verwertern“ (S. 70) oder der Ausbau von Citizen Science. Als Herausforderung beschreiben die Autoren den Umgang mit KI-generierten Texten und deren Erkennbarkeit, da das DeReKo ausschließlich den menschlichen Sprachgebrauch abbilden soll. Hier formulieren die Autoren eine weitere Vision – ein Mechanismus, der unterscheiden kann, ob ein Text maschinell oder von Menschen verfasst wurde.

Der kurze Beitrag von Henning Lobin „Was bieten heutige Korpora? Über die Herausforderungen der Erfassung besonderer Textsorten bei der Erforschung gegenwärtigen Sprachgebrauchs“ (S. 79–86) beschreibt das Potential von Sprachsammlungen verschiedener Art und die damit verbundenen Probleme. Die Entstehung von Korpora ist eng an die Verfügbarkeit von Texten gebunden, weshalb einige Textsorten wie Zeitungen in größerem Maße verfügbar sind als etwa Texte aus der privaten Kommunikation. Die Nutzung solcher persönlicher Daten für linguistische Zwecke ist durch rechtliche Vorgaben und mangelndes Vertrauen der jeweili-

gen Verfasser:innen in die Datensicherheit eingeschränkt. Literarische Texte wiederum sind mit materiellen Interessen verknüpft, so dass auch sie nur bedingt verfügbar sind, ebenso wie Fachtexte oder weitere besondere Arten von Texten. Um auch diese Daten korpuslinguistisch erschließen zu können, muss Vertrauen geschaffen und erhalten werden. Eine Möglichkeit, Vertrauen zu gewinnen und dadurch Textspenden zu generieren, ist die bürgerwissenschaftliche Beteiligung.

Aleksandra Pushkina und Erhard Hinrichs stellen in ihrem Beitrag „The IVK-Ler Corpus of Adolescent Foreign-Language Learners of German“ (S. 87–104) ein kleines Lerner-Korpus mit annotierten Arbeiten einer Internationalen Vorbereitungsklasse (IVK) von erwachsenen DaF-Lernenden ohne jegliche Vorkenntnisse vor. Besonders interessant ist Kap. 3 des Beitrags, in welchem die Datensammlung, die Annotationsprozesse und die Darstellung des IVK-Ler Corpus genau beschrieben werden. Der Aufsatz ist ein wunderbares Beispiel für ein konkretes Lerner-Korpus, von denen es für die deutsche Sprache zu wenige gibt – zum Zeitpunkt der Entstehung des Aufsatzes hatten nur 19 von 199 verfügbaren Lerner-Korpora Deutsch als Zielsprache. Digitalisierte und annotierte Lerner-Korpora mit Deutsch als Zielsprache sind überhaupt eine relativ neue Entwicklung und die meisten von ihnen sind nicht öffentlich zugänglich. Auch das vorgestellte IVK-Ler Corpus ist leider nur über die Autorin und den Autor des Beitrags zugänglich.

„Natürlichkeit vs. Reichhaltigkeit vs. Vergleichbarkeit: Wie Widerstreitendes bei der Erhebung von Gesprächskorpora versöhnt werden kann“ diskutiert Uta Quasthoff in ihrem Aufsatz (S. 105–120). Während in der Konversationsanalyse ursprünglich rein deskriptiv-rekonstruktive Verfahren angewendet wurden, müssen diese mit der Erweiterung auf interdisziplinäre oder anwendungsorientierte Forschungsziele methodisch angepasst werden. Quasthoff stellt sich dabei die forschungspraktische Frage, wie diese verschiedenen, sich teilweise widersprechenden Interessen beim Korpusaufbau miteinander in Einklang gebracht werden können. Sie bezieht sich dabei auf den Zusammenhang zwischen der jeweiligen Fragestellung, dem theoretischen Kontext und der projektbezogenen Datenerhebung. Als beispielhafte Ressource beschreibt die Autorin kurz die „chronologisch verschränkten Datenerhebungen der DASS- und OLDER-Studien zur längsschnittlichen Analyse von Diskurserwerbsprozessen bei Grundschulkindern“ (S. 112).

Textkorpora: Untersuchungen und Anwendungen

Mit acht Aufsätzen ist dieser Abschnitt der umfangreichste des Bandes. Er widmet sich konkreten Untersuchungen und Anwendungen, die auf unterschiedlichen Korpora der geschriebenen Sprache beruhen. Der Abschnitt startet mit dem Beitrag von Ludwig M. Eichinger „Anpassungsfähigkeit und Akzentuierung. Von moder-

nen Dingen und den vielfältigen Möglichkeiten adjektivischer Wortbildung“ (S. 123–138), der eine korpusgestützte Untersuchung komplexer Adjektive mit dem Erstglied „gender-“ mit DeReKoVecs vorstellt. Damit zeigt er, welche Wortbildungsmethoden im Deutschen Verwendung finden, um neue Elemente wie „gender-“ in bestehende Strukturen zu integrieren. So finden sich beispielsweise Determinativkomposita wie *genderkonform* oder *gendersensibel*, Partizipialkomposita wie *genderorientiert* und *genderverwirrt* oder auch Konstruktionen, die sich an einem fließenden Übergang zwischen den Wortbildungsarten befinden, wie *gendergerecht* (klassifiziert als zwischen Komposition und Derivation stehend) oder *gendermäßig* (zwischen Komposition und Suffigierung).

Anschließend untersucht Stefan Engelberg („Argumentstrukturen in expressionistischer Lyrik“, S. 139–154) in seinem Korpus expressionistischer Gedichte Auffälligkeiten im Stil und verbindet damit Literatur- und Sprachwissenschaft. Die in Gedichten verwendete Sprache soll sich gemäß der vom Autor vorgestellten lyrischen Abweichungstheorie von der außerhalb der Lyrik benutzten Sprache unterscheiden (S. 139). Die von Engelberg in diesem Beitrag betrachteten Abweichungen sind phonologischer, morphologischer, syntaktischer, semantischer und pragmatischer Art; sein Korpus umfasst etwa 320 expressionistische Gedichte mit rund 450 annotierten Valenzauffälligkeiten, die gruppiert und typisiert werden. Er macht deutlich, dass zumindest auf der grammatiktheoretischen Basis die allgemeine Gültigkeit der Abweichungstheorie anzuzweifeln ist.

Thomas Gloning beschreibt in seinem Beitrag „Wissensräume von Zeitschriften in Beiträgen, Heften und Heft-Serien. Textorganisation, Multimodalität, Wortgebrauch“ (S. 155–170) und stellt dabei einen Zusammenhang von sprachlich-textueller Organisation mit Fragen der Wissensorganisation und -vermittlung in thematisch ausgerichteten Zeitschriften her. In diesen themen- und wissensorientierten Publikationen untersucht der Autor die Organisation der Texte, wissensvermittelnde Strategien, die Koordination von Texten und Bildern, Multimodalität und den Wortgebrauch. Er betrachtet diese Aspekte sowohl hinsichtlich einzelner Beiträge und ganzer Zeitschriften, aber auch bezogen auf Serien von Heften im historischen Längsschnitt. Anhand einiger Beispiele zeigt Gloning auf, dass sich Textsorten im herkömmlichen Sinne nur schwer identifizieren lassen und eine „flexiblere Problemlösekonzeption zu Produktion und Analyse der Machart von Texten, Textclustern und besonders von multimodalen Angeboten“ (S. 163) nötig ist. Er betrachtet im Weiteren die thematischen Räume und Wissensräume, die durch Zeitschriften konstituiert werden, und deren Angebotsstruktur. Ausgehend von seinen ersten Analyseergebnissen leitet der Autor weitere Aufträge für künftige Untersuchungen ab, etwa die weitergehende Erforschung des Zusammenhangs von sprachlicher Form und der Organisation von Wissen oder breiter angelegte Wortschatzanalysen.

Im Beitrag „20 Jahre Wortwarte. Wie alles anfing (und endete)“ (S. 171–180) berichtet Lothar Lemnitzer von dem von 2000–2020 laufenden Projekt und Vorläufer des DeReKo zur Wortschatzentwicklung von Tylman Ule und Lothar Lemnitzer. Der Aufsatz beschreibt das Vorgehen beim Erstellen dieser lexikalischen Ressource und zieht Bilanz. Die sprachlichen Beobachtungen, die die Wortwarte bot, machen Erkenntnisse zur Lexik und ihrem Wandel möglich. Ein Fazit des Aufsatzes: Das Konzept „Neologismus“ ist nach wie vor nur vage zu umreißen.

An diesen Aufsatz schließt der Beitrag „Filtern, Explorieren, Vergleichen: Neue Zugriffsstrukturen und instruktive Potenziale von OWID^{plus}“ (S. 181–196) von Frank Michaelis, Carolin Müller-Spitzer, Jan-Oliver Rüdiger und Sascha Wolfer hervorragend an. OWID^{plus} ist ein Zusatzangebot zur Wörterbuchplattform OWID und damit ein Ort, an dem sämtliche wortschatzbezogenen Projekte gesammelt sind; eine Plattform, die lexikalische Datenbanken, Korpuswerkzeuge und Analysen vereint. Hier wird Sichtbarkeit gewährleistet – ohne diesen Ort wären die Ressourcen unter verschiedenen URLs im Internet verteilt zu finden. Der Aufsatz zeigt, was getan wird, damit unterschiedliche Korpora besser auffindbar sind, und stellt zwei Ressourcen (OWID^{plus}LIVE und Wortschatzwandel im Spiegel) und deren Einsatz in der universitären Lehre genauer vor. Beide Tools haben einen ähnlichen Einsatzbereich, zeigen aber Unterschiede im Zugriff (Abfrage über bestimmte sprachliche Formen vs. Angabe von Frequenzparametern).

Der Beitrag von Bernhard Schröder „Induktiv oder intuitiv? Die Gewinnung von Frames aus mathematischen Beweistexten“ (S. 197–216) kommt anhand der Untersuchung von 20 mathematischen Beweistexten ab Mitte des 20. Jh. aus unterschiedlichen Bereichen und Publikationen (von Lehrbüchern für Studierende bis hin zu wissenschaftlichen Veröffentlichungen) zur Erkenntnis, dass man bei der empirisch gestützten Spezifikation von Frames innerhalb einer Strategie folgendermaßen vorgehen sollte: Bei prototypischen Frames sollten im frühen didaktischen Erwerb induktive Verfahren eingesetzt werden, wobei diese jedoch bei weniger prototypischen Subtypen von Frames durch hypothesenbildende Verfahren zu ergänzen sind. Außerdem sollte die Plausibilität der Hypothesen an Texten getestet werden. Schröder prognostiziert eine Übertragbarkeit auf konstruktionsgrammatische und textlinguistische Typologien, weshalb sich die Lektüre auch für Nicht-Mathematiker:innen lohnt.

Über die Entwicklung und Implementierung eines Online-Glossars für „Klimakomposita“ berichten Manfred Stede, Anna-Janina Goecke, Noël Simmel und Birgit Schneider in: „Der reine Klimawahnsinn! Zur Konzeption eines Diskursglossars von Klimakomposita“ (S. 217–230). Darin untersuchen sie die strategische Verwendung von Komposita mit „Klima-“ im politischen Diskurs und deren Konnotationen. Die Datengrundlage bilden zunächst zwei Websites – einerseits die des klimawandelskeptischen Europäischen Instituts für Klima und Energie e.V. und andererseits

die Seite der Klimaaktivist:innen von Fridays for Future. Demzufolge gibt es zwei Subkorpora mit zwei dazugehörigen Subdiskursen, die die Pole der Positionen der Diskursteilnehmenden bilden. Beide Korpora wurden im Laufe der Arbeit mit der Extraktion weiterer Websites angereichert und hinsichtlich der Gebrauchshäufigkeiten ausgewertet. Im Rahmen einer Bachelorarbeit wurde eine App entwickelt (www.klimadiskurs.info), über welche das im Projekt entstandene Klimaglossar abrufbar ist. Ein wunderbares Beispiel, um Studierenden Bedeutungswandel und korpuslinguistische Basics nahezubringen!

Der Beitrag von Gisela Zifonun stellt aus ihrer Sicht als Grammatikerin „Korpusbefunde und Grammatik am Beispiel des Genitivs im Deutschen“ (S. 231–242) vor. Dazu betrachtet sie zwei korpusgrammatische Studien am IDS genauer, die sich mit der Wahl des Markers *-s*/*-es* und Sonderfällen des Genitivattributs befassen. Zifonun sieht einen positiven Effekt der korpusgrammatischen Studien: Übergreifende Regularitäten werden sichtbar. Allerdings konstatiert sie auch ein Dilemma: „Die Forschungsgrammatik rät zum Verzicht auf grammatische Vorannahmen; ein adäquates Recherchedesign erfordert jedoch ein grammatisch informiertes Vorgehen.“ Zifonun rät dazu, diesen Konflikt im jeweiligen Einzelfall neu zu lösen. Sie schließt ihren Beitrag mit der Erkenntnis, dass die korpusgrammatische Herangehensweise zwar durchaus wissenschaftliche Fortschritte hervorbringen kann, aber weder das Sprachverständnis grundlegend ändert noch die Sicht auf Grammatik überflüssig macht.

Korpusgestützte Analyse gesprochener Sprache

Der Abschnitt zur korpusgestützten Analyse gesprochener Sprache umfasst zwei Aufsätze. Zunächst untersuchen Arnulf Deppermann und Silke Reineke die Metadatenverwendung bei der Interaktionsanalyse anhand des FOLK-Korpus in ihrem Beitrag „Zur Verwendung von Metadaten in der interaktionsanalytischen Arbeit mit Korpora – am Beispiel einer Untersuchung anhand des Korpus FOLK“ (S. 245–259). Sie konstatieren, dass Metadaten (Eigenschaften der Sprechenden, des Sprachereignisses oder der Gesprächsaufnahme) bisher kaum genutzt werden, obwohl sie der Identifizierung von Regularitäten von Gesprächspraktiken, Aktivitäten und Sprecherrollen dienlich sein können. Die seltene Verwendung von Metadaten resultiert aus einem reflexiven Kontextmodell: Man befürchtet, dass ihre Nutzung zu einer „verkürzten subsumptiven und deduktiven Analyse“ (S. 246) führt. Deppermann und Reineke zeigen mittels einer exemplarischen Analyse der Verwendung des Formats *was heißt X*, das zur Adäquatheitsreparatur und als Spezifikationsaufforderung verwendet wird, welcher Erkenntnisgewinn durch die Nutzung von Metadaten in der Interaktionsanalyse erwartet werden kann.

Rosemarie Tracy und Dafydd Gibbon stellen in ihrem Beitrag „The Beat Goes On: A Case Study of Timing in Heritage German Prosody“ (S. 261–283) eine Fallstudie vor, mit deren Hilfe sie herausfinden wollen, wie Sprechrhythmen die narrative Kohäsion fördern. Auch wenn die kleine Analyse zur Sprachproduktion einer bilingualen Sprecherin des Deutschen als Minderheitensprache in einer englischsprachigen Umgebung keine Generalisierung zulässt und es größere Datenmengen und ausgefeiltere Verfahren braucht, um das methodologische Potential auszuschöpfen, können Tracy und Gibbon zeigen, wie erkenntnisträchtig derartige Untersuchungen sein können.

Korpusgestützte Analyse internetbasierter Kommunikation

Die vier Aufsätze im nächsten Abschnitt befassen sich mit internetbasierter Kommunikation und deren Auswertung. Zunächst betrachten Michael Beißwenger und Sarah Steinsiek „Interpunktion als interaktionale Ressource. Eine korpusgestützte Untersuchung zur Funktion von Auslassungspunkten in der internetbasierten Kommunikation“ (S. 287–310). Mittels einer Stichprobe aus einem WhatsApp-Korpus (MoCoDa2) ergründen sie die Verwendung von Auslassungspunkten und deren Funktionen in Chats. Bei der Untersuchung von 98 Postings mit 108 Auslassungspunkt-Belegen gehen sie von vier Grundtypen aus: Auslassen, Andeuten, Organisieren und Segmentieren. Diese Funktionsbeschreibungen verdeutlichen sie an Beispielen und entwickeln sie für das interaktionsorientierte Schreiben in der internetbasierten Kommunikation weiter.

Der Beitrag von Leonie Bröcher, Eva Gredel, Laura Herzberg, Maja Linthe und Ziko van Dijk „Linguistische Wikipedistik und Wikipedaktik Revisited (2018–2023)“ (S. 311–328) gibt einen Überblick über zwei Forschungsfelder: die Linguistische Wikipedistik, deren Untersuchungsgegenstände Wikipedia und Wikis im Allgemeinen darstellen, und die Wikipedaktik, bei der das didaktische Potential der genannten Untersuchungsgegenstände in Vermittlungskontexten im Mittelpunkt steht. Für die linguistische Wikipedistik werden interessante Bereiche aufgeführt: (Korpus-)Ressourcen, text-, interaktions- und diskursanalytische Studien, Sprach- und Kulturvergleich, Genderlinguistik. Die Wikipedia-Daten bilden ein Archiv des Deutschen Referenzkorpus (DeReKo) und stellen dort mit 53 Mrd. Wörtern (Stand März 2022) die umfangreichste Sammlung geschriebener Gegenwartssprache dar – die Korpora sind über die IDS-Recherchesysteme COSMAS II und KorAP verfügbar. Der Beitrag zeigt in anschaulicher Weise das Potential der Wikipedaktik in Schule und Hochschule: So können Wikipedia und andere Wikis als Wissensressourcen

genutzt werden. Zudem werden verschiedene kindgerechte digitale Plattformen aufgeführt. Auch in der akademischen Lehre oder bei älteren Schüler:innen bieten sich verschiedene Einsatzmöglichkeiten an. Wikis können verschieden genutzt werden: als Lehr-Lern-Plattformen, als „Orte digitaler Partizipation und Emanzipation" (S. 318) oder als Reflexionsgegenstände. Ihre Vorteile sind die vollständige Transparenz des Schreibprozesses (Versionsgeschichte), Feedbackmöglichkeiten und die Nachhaltigkeit über das Projektende hinaus. Ein Beitrag insbesondere für angehende Lehrkräfte.

Wolfgang Imo legt in seinem Aufsatz „‚Ich glaub mein Schwein pfeifft' – ein Fall für die *Mobile Communication Datatabase*. Oder: Das Possessivpronomen *mein* aus korpusbasierter Perspektive" (S. 329–340) eine Fallstudie zur Verwendung des Possessivpronomens *mein* in Messengerchats aus der Mobile Communication Database MoCoDa 2 vor, die auch bei der oben besprochenen Analyse von Beißwenger und Steinsiek eingesetzt wurde. Der Autor beschreibt die verschiedenen Verwendungsweisen des Possessivpronomens *mein* in der informellen Chatkommunikation und kann auf semantisch-funktionaler Basis vier Kategorien ausmachen, die hier nach ihrer Häufigkeit geordnet werden: 1) Realia im Besitz einer Person, 2) Ausdruck einer Affiliation (Verwandtschaft, Beruf, Studium u.ä.), 3) Abstrakta im (metaphorischen) Besitz einer Person, 4) floskelhafte Ausdrücke. Imos Beitrag zeigt neben den Ergebnissen seiner Studie das Potential von MoCoDa2 für Untersuchungen zu sprachlichen Strukturen in Messengerchats.

Konstanze Marx unterbreitet mit ihrem Beitrag „Die INSTAB-Formel. Ein Vorschlag für die Erstellung von Instagram-Datensammlungen für studentische Arbeiten" (S. 341–356) eine Empfehlung, wie in Studien- oder Abschlussarbeiten schrittweise eine Datensammlung angelegt werden kann, ohne dass Studierende über Programmierkenntnisse oder besondere technische Voraussetzungen verfügen. Die anzulegende Datenbasis soll es den Studierenden im Rahmen internetlinguistischer Fragestellungen ermöglichen, qualitative Analysen, etwa zur Plattform Instagram, durchzuführen. Die Auswahl der Social-Media-Plattform Instagram beruht auf deren Nähe zur Lebenswelt der Studierenden und schafft so Interesse – eine hervorragende Voraussetzung für forschungsorientierte akademische Lehre. Die Autorin schildert zunächst die Problemstellung: Welche Prozesse und Vorgänge sind konkret mit dem meist ungenau beschriebenen manuellen Zusammentragen, Sammeln, Erheben und Sichern verbunden? Wie kann man diese Vorgänge für Studierende so gestalten, dass sich der Betreuungsaufwand in Grenzen hält und nicht in Einzelberatungen zerfällt? Marx formuliert deshalb die INSTAB-Formel als „Vorschlag für ein sachbezogenes, vereinheitlichendes Vorgehen, das Studierende in der Phase, in der noch keine Standards entwickelt worden sind, beim Verfolgen ihrer Forschungsinteressen unterstützen soll" (S. 344). Die einzelnen Bestandteile dieser Formel sind: IN – In Time (frühzeitiger Beginn der Datensammlung), S – Speichern, T – Transkribieren, A – Annotie-

ren, B – Bereitstellen. Mit diesen Hinweisen für ein adäquates methodisches Vorgehen schafft sie eine Orientierung für die Erstellung von Datensammlungen im Rahmen von universitärer Lehre, in der natürlich auch bisher derartige Untersuchungen erfolgreich stattfanden – ohne Formel. Durch den klaren Zeitplan und die überschaubare Arbeitsbelastung soll diese jedoch einer Überforderung der Studierenden vorbeugen. Damit handelt es sich bei diesem Beitrag um eine hilfreiche didaktische Anregung, die hoffentlich weithin aufgegriffen wird.

Korpusgestützte Analyse und Förderung sprachlicher Kompetenzen

Die fünf Aufsätze dieses Abschnitts beschäftigen sich mit dem Einsatz digitaler Sprachkorpora in der Lehre und der Förderung von sprachlicher Kompetenz. Thomas Bartz und Nadja Radtke starten mit der „Nutzung digitaler Textkorpora und Analysewerkzeuge beim materialgestützten Schreiben im Deutschunterricht“ (S. 359–376) und zeigen, wie die Sprachreflexion mit der Untersuchung authentischer Sprachdaten verknüpft und dadurch das materialgestützte Schreiben gefördert werden kann. Ein interessanter Beitrag insbesondere für alle angehenden Lehrkräfte und diejenigen, die sie ausbilden!

Im Aufsatz „Koordination – (k)ein Lernproblem für DaF?“ (S. 377–392) nimmt Eva Breindl die Koordination als Lerngegenstand in Lehrmaterialien für DaF unter die Lupe und zeigt, dass die Behandlung koordinativer Verknüpfungen darin unterrepräsentiert ist. In ihrer Pseudolongitudinalstudie am Lernerkorpus MERLIN (die Vergleichsdaten stammen aus DeReKo) wertet sie drei Subkorpora mit Texten der Sprachniveaus B1, B2 und C1 nach Koordinationstypen der Koordination mit *und* bzw. Konnektkategorien aus. Sie zeigt auf, dass auch Lernende auf C1-Niveau noch erhebliche Probleme bei der Verwendung komplexer Koordinationsstrukturen haben.

Carolina Flinz, Ruth M. Mell, Christine Möhrs und Tassja Weber beschreiben in ihrem Beitrag „Korpora für Deutsch als Fremdsprache – Potenziale und Perspektiven“ (S. 393–408) verschiedene Ressourcen und deren Anwendungsbereiche im DaF-Unterricht: So können Lernerkorpora mit Deutsch als Zielsprache, Spezial- und Fachkorpora, Vergleichskorpora, Korpora gesprochener Sprache oder Wörterbuchressourcen eingesetzt werden. Die Autorinnen identifizieren und benennen das hohe Potential, das der Einsatz von Korpora und der genannten Ressourcen bietet: Korpora können bestehende Lehr- und Lernmaterialien sinnvoll ergänzen und eine große Menge an Daten bereithalten, die Lernende mit der Sprachverwendung vertraut machen und dabei helfen, Deutsch zu verstehen.

Aivars Glaznieks, Jennifer-Carmen Frey und Andrea Abel beobachten in ihrer Studie Phänomene der syntaktischen Variation in *weil*-Sätzen bei Deutschlernenden in der mehrsprachigen Provinz Südtirol und beschreiben in ihrem Aufsatz „*Weil*-Sätze bei Lernenden des Deutschen. Vergleich zwischen immersiv und nicht immersiv Deutschlernenden in Südtirol“ (S. 409–424), dass im Immersionsfall VL- und V2-Sätze verwendet werden (ähnlich bei L1), während bei nicht-immersiv Lernenden auch Systemfehler auftreten. Ihre Daten basieren auf den Korpora LEONIDE, Kolipsi und KoKo, die die mehrsprachige Situation der Schüler:innen mit verschiedenen sprachlichen Hintergründen repräsentieren. Die Autor:innen bestätigen mit ihrer Studie, dass die Wortstellung im Nebensatz für Lernende eine große Herausforderung darstellt.

Im letzten Beitrag des Abschnitts und des Bandes fragen Christian Lang, Roman Schneider und Angelika Wöllstein „Was ist, was soll sein – und warum? Sprachanfragen aus empirisch-linguistischer Perspektive“ (S. 425–438). Darin skizzieren sie Erkenntnispotentiale von Sprachanfragen für linguistische und transferwissenschaftliche Forschungsfragen und Methoden; der Fokus liegt auf wissenschaftskommunikativen Prozessen im Austausch zwischen linguistischen Laien und Experten. Um herauszufinden, welche Erkenntnisse aus Sprachanfragedaten für die Vermittlung von Sprachwissen gewonnen werden können, wurden ca. 50.000 Sprachnutzungsanfragen korpusgestützt analysiert. Das Material aus den Anfragen bietet einen riesigen Pool an Daten aus einem großen Nutzungskreis. Wichtig für den Transfer von Sprachwissen ist es, Hinweise zu gewinnen, welche linguistischen Informationenquellen bei der Laiensuche genutzt werden, und sich einen Überblick darüber zu verschaffen, wie diese dabei vorgehen. Die Sprachanfragen liefern spannende Einblicke in die Kommunikation zwischen linguistischen Laien und Experten. Sie zeigen aber auch, wo es inhaltliche Lücken in Informationsangeboten wie *grammis* gibt und welche Themen für die Wissenschaftskommunikation von Interesse sind.

Fazit

Der vorliegende Band könnte thematisch zusammengefasst werden als „Was Sie schon immer über korpusbasierte und -unterstützte Sprachanalyse wissen wollten ...“. Seine Anwendungsszenarien sind vielfältig: Einige Beiträge wie die von Gloning, Bröcher et al., Michaelis et al., Stede et al. oder Marx sind besonders gut für die akademische Lehre geeignet, andere stellen wichtige Werkzeuge und eigene interessante Anwendungsszenarien vor. Wieder andere Aufsätze unterstützen die Lehramtsausbildung in herausragender Weise und zeigen, wie man Korpusarbeit in die Schule bringen kann (bspw. der Aufsatz von Bartz und Radtke). Die eingangs

dargelegten Ziele des Bandes sind damit erreicht. Als roter Faden dienen die in der Einleitung aufgeführten Forschungsschwerpunkte von Angelika Storrer, die in den Beiträgen in unterschiedlicher Intensität aufgegriffen werden und dabei wie zu Beginn des Buches angekündigt den aktuellen „Stand der linguistischen, korpuslinguistischen und korpustechnologischen Forschung“ (S. 16) reflektieren. Die Themen, Zugänge und Methoden sind daher naturgemäß heterogen und spiegeln das weite Feld der auf Korpora beruhenden Analyse von Sprache. Wenn etwas vermisst werden könnte, dann wären es Projektvorstellungen mit korpuslinguistischem Bezug aus der historischen Sprachwissenschaft, wie sie sich beispielsweise im von Renata Szczepaniak, Stefan Hartmann und Lisa Dücker herausgegebenen Band „Historische Korpuslinguistik“ finden.

Das Buch zeigt die enorme Vielfalt korpusgestützter Sprachanalyse und erhöht damit die Corpus Literacy, also die Fähigkeit, mit Korpora umzugehen (Mukherjee 2002, S. 179f.), in ganz besonderem Maße.

Literatur

Mukherjee, Joybrato. 2002. *Korpuslinguistik und Englischunterricht. Eine Einführung.* Frankfurt a.M.: Lang.
Szczepaniak, Renata, Stefan Hartmann & Lisa Dücker (Hg.). 2019. *Historische Korpuslinguistik* (Jahrbuch für Germanistische Sprachgeschichte 10). Berlin, Boston: De Gruyter.