

Michael Beißwenger
Lothar Lemnitzer
Carolin Müller-Spitzer

1. Einführung

Sprache war und ist ein hochspannender Forschungsgegenstand. Auf der einen Seite verwenden wir alle Sprache und fühlen uns dadurch mit ihr verbunden, auf der anderen Seite gibt es empirische Forschungen über bestimmte Aspekte von Sprache, die mit unseren eigenen subjektiven Eindrücken vielleicht gar nicht übereinstimmen. Die Möglichkeiten der empirisch-wissenschaftlichen Erforschung der deutschen Sprache, ihrer Verwendung und Entwicklung sind in den letzten Jahrzehnten enorm gestiegen. Mit den Möglichkeiten des Internets und der sozialen Medien tun sich Wege der sprachlichen Kommunikation und Partizipation auf, von denen noch vor 30 Jahren nur die Mutigsten zu träumen wagten. Zugleich entfaltet mit diesen neuen Medien die Sprache, oder genauer diejenigen, die sie verwenden, eine besondere Wirkmacht. Man denke nur an den Einfluss, der den sozialen Medien bei politischen Großereignissen wie der sog. Arabellion vor zehn Jahren oder den letzten beiden Präsidentschaftswahlen in den USA zugesprochen wird. Wörter und Texte, die Substanz der Sprache, können heute außerdem so leicht wie nie mit anderen Formen der Darstellung verbunden werden, z. B. mit Grafiken, Bildern und Videos, und so entstehen komplexe, multimodale, mitunter sehr suggestive Gebilde. Mit der Schnelligkeit, mit der heute Texte und deren Überarbeitungen publiziert werden können, ergeben sich Möglichkeiten, Nutzer*innen und Communities bei der Erschaffung von Texten zuzusehen (als Beispiel sei hier nur die Wikipedia genannt mit der Verfügbarkeit vieler Versionen der einzelnen Artikel und den artikelbegleitenden Diskussionsseiten). Nahezu jede*r mit einer Internetverbindung ist heute dazu in der Lage, am Unversum der Texte und damit am Gewebe der Sprache mitzuarbeiten. Der Einfluss, den diese Entwicklungen auf den Zustand der Sprache und die Vielfalt des sprachlichen Handelns haben, ist bislang kaum erforscht,

einfach aus dem Grund, dass diese Entwicklungen noch so neu sind.

Zugleich geraten Wissenschaft und Forschung unter einen stärkeren Druck. Wissenschaftler*innen müssen ihre Arbeit gegenüber einer breiteren Öffentlichkeit legitimieren, es wird von ihnen erwartet, dass sie harte und belastbare Fakten liefern. Dies wird besonders deutlich etwa beim Diskurs über den Klimawandel. Hier wird jede schlecht abgesicherte Aussage auf dem Operationstisch vor den Augen der Öffentlichkeit seziiert, von der Fachwelt, aber auch von den Medien und vor allem von denjenigen, denen eine bestimmte Erkenntnis gerade nicht in ihre Argumentation passt. Nun ist dies hier kein Buch über Klimaforschung. Doch auch die Sprache ist ein Gegenstand öffentlicher Diskussion. Wenn Sie die deutsche Sprache erforschen, haben Sie es mit über 100 Millionen Expert*innen zu tun, die diese Sprache sprechen und damit den Gegenstand, den Sie erforschen, kennen oder zu kennen glauben. Steilen Thesen wie „Das Deutsche verfällt zusehends unter dem Einfluss des Englischen, des Internets, der Migrant*innen usw.“, „Unsere Kinder können alle nicht mehr richtig schreiben“ oder „Gendern macht die Sprache unnötig kompliziert“ und Ähnlichem, was als Zeitungsschlagzeile oder zu einem Partygespräch taugt, wollen und müssen Sie harte Fakten entgegensetzen. Ihre Ausgangssituation als Linguist*in dafür ist aber sehr gut: Sie können zwar Sprache nicht in ihrer ganzen Bandbreite erforschen, aber auf mittlerweile sehr große Mengen digitalisierter und damit leicht durchsuchbarer Äußerungen in geschriebener und gesprochener Sprache (sog. *Korpora*) zurückgreifen. Sie finden spezialisierte Suchmaschinen und Analysewerkzeuge vor, die ihnen das systematische Durchsuchen dieser riesigen Datenmengen erleichtern. Sie können mit kostenfreien und relativ leicht zu handhabenden Werkzeugen arbeiten, also z. B. große Datenmengen statistisch aufberei-

ten. Bemerkungen wie „Der Konjunktiv wird heute kaum noch gebraucht“ oder „weil wird heute mehr und mehr als Konjunktion verwendet, die einen Hauptsatz einleitet und nicht einen Nebensatz“ können Sie mit statistisch gesicherten Erkenntnissen kontern, sei es ablehnend oder zustimmend oder, wie es aus linguistischer Perspektive bei solchen sprachlichen Zweifelsfällen meist der Fall ist, mit einer differenzierenden wissenschaftlichen Bewertung. Genauso können Sie empirisch untersuchen, ob z. B. geschlechtergerechte Texte wirklich weniger verständlich sind als Texte im sogenannten generischen Maskulinum, anstatt darüber Vermutungen auf Basis subjektiven Empfindens anzustellen. Die Möglichkeiten, die sich Ihnen bieten, sollten Sie daher auch als eine Verpflichtung zu methodisch solidem wissenschaftlichem Arbeiten verstehen: weg von der Spekulation, hin zu empirisch gesicherten, d. h. datengestützt gewonnenen und sorgfältig dokumentierten Erkenntnissen. Wenn Sie hier zustimmen, dann wurde das vorliegende Buch für Sie geschrieben.

Wir haben als Herausgebende zwei Typen von Beiträgen konzipiert, die verschiedene Funktionen erfüllen sollen: erstens *Fallstudien* und zweitens kompakte *Überblickskapitel zu einzelnen methodischen Aspekten, Forschungsressourcen und Werkzeugen*. Den Fallstudien und Überblickskapiteln vorangestellt ist ein Grundlagenkapitel, in dem die typischen Schritte einer empirischen Studie in der germanistischen Linguistik skizziert werden (→ Kapitel 2).

Die Fallstudien (→ die Beiträge in Teil II) sollen an konkreten Untersuchungen aufzeigen, wie empirisches Forschen in unterschiedlichen linguistischen Forschungsfeldern „funktioniert“, d. h. wie ausgehend von einer konkreten Forschungsfrage auf der Grundlage empirischer Daten und mit bestimmten Forschungsmethoden linguistische Erkenntnisse gewonnen werden. Dazu haben wir Kolleg*innen eingeladen, bereits durchgeführte und an anderer Stelle schon publizierte Untersuchungen nach einem vorgegebenen Muster aufzubereiten, das die jeweilige Untersuchung ausgehend von den Fragestellungen über die Gewinnung und Analyse der

Daten bis hin zu den Befunden für Studierenden nachvollziehbar macht. Dabei wird insbesondere das methodische Vorgehen bei der jeweiligen Untersuchung reflektiert und es werden die untersuchten Fragestellungen in den weiteren und engeren Fachkontext des zugehörigen linguistischen Forschungsfeldes eingebettet. Die Fallstudien wurden so ausgewählt, dass einerseits eine breite Palette linguistischer Forschungsfelder, andererseits eine große Bandbreite theoretischer und methodischer Zugänge abgedeckt sind. Wir haben diesen Fallstudien in diesem Band einen so breiten Raum gegeben, weil wir davon ausgehen, dass sich ein grundlegendes Methodenbewusstsein am besten an Fallbeispielen erwerben und schärfen lässt.

Mit der Einladung, zu diesem Buch mit einer Fallstudie beizutragen, haben wir, die Herausgeber*innen, den Autor*innen den folgenden Wunsch übermittelt:

Die einzelnen Fallstudien haben die Funktion, den Studierenden Beispiele von empirischen Untersuchungen aus verschiedenen Bereichen der Linguistik zu zeigen. Die Beiträge sollen keine üblichen Sammelbandbeiträge sein, sondern den Schwerpunkt entsprechend dem Ziel des Buches auf den methodischen Aufbau der Studien legen.

Damit verbunden war ein Vorschlag für die Gliederung der Fallstudien: 1. Einleitung, 2. Fragestellung, 3. Material, 4. Methode und Analyse, 5. Ergebnisse und Diskussion, 6. methodische Reflektion und 7. weiterführende Literatur und Literaturverzeichnis. Bei den zehn Fallstudien, die in diesen Band Eingang gefunden haben, handelt es sich also nicht um Berichte über neueste Forschungsergebnisse, sondern um didaktisch besonders aufbereitete Werkstattberichte. Wir hoffen, dass diese Blaupause Ihnen die Orientierung in den einzelnen Fallstudien erleichtert. Um selbst linguistische Untersuchungen durchzuführen, muss man, je nachdem, was man plant, aber nicht nur Fallbeispiele kennen, sondern sich zudem detaillierter mit einzelnen Methoden, Ressourcen und Werkzeugen auseinandersetzen. Eine grundlegende Orientierung dafür sollen die *Überblickskapitel* bieten, die in → den

Teilen III bis VII zusammengestellt sind und die zur weiteren Beschäftigung mit der jeweiligen Ressource bzw. Forschungsmethode einladen sollen. Auch für diese Kapitel haben wir ausgewiesene Spezialist*innen des jeweiligen Gebietes als Autor*innen eingeladen.

Die Fallstudien

Teil II umfasst insgesamt zehn Fallstudien, die von 18 Autor*innen verfasst wurden:

Angelika Storrer und Laura Herzberg stellen in → Kapitel 3 [Fallstudie „OKAY“]¹ zwei Studien zum Gebrauch des Internationalismus OKAY im Deutschen und Französischen vor. Für diese medienvergleichenden Studien verwenden sie ein Korpus der gesprochenen Sprache (das Forschungs- und Lehrkorpus Gesprochenes Deutsch FOLK) sowie zwei Korpora der geschriebenen Sprache (die deutsche und die französische Wikipedia, hier vor allem die Diskussionsseiten), um Unterschiede zwischen schriftlicher und mündlicher Verwendung von OKAY zu ermitteln. Im Mittelpunkt ihrer methodischen Reflexion steht das Vorgehen bei der Erhebung, Bereinigung und Analyse der Daten. Die beschriebenen Studien sind quantitativ ausgerichtet, korpusbasiert und verfolgen das Ziel, Hypothesen aus der Fachliteratur empirisch zu überprüfen.

Die Lektüre der Untersuchung von Storrer und Herzberg kann durch die Lektüre des → Kapitel 20 [Korpusdaten] sowie die → Beiträge in Teil VI, u.a die Überblicke über Korpora der geschriebenen und gesprochenen Sprache, vor allem aber des Kapitels zu Korpora der internetbasierten Kommunikation, vertieft werden.

Die von Evelyn Ziegler und Ulrich Schmitz in → Kapitel 4 [Fallstudie „Metropolenzeichen“] vorgestellte Fallstudie stellt das Unter-

suchungsdesign und die zentralen Ergebnisse des interdisziplinären Forschungsprojekts „Metropolenzeichen: Visuelle Mehrsprachigkeit in der Metropole Ruhr“ vor, ein Projekt, das von Linguist*innen, Integrationsforscher*innen und Stadtsoziolog*innen durchgeführt wurde. Den Analysen der Autor*innen bzw. der beteiligten Forscher*innen lagen Daten verschiedener Art zugrunde, nämlich Bilddaten und Interviewdaten. Diese wurden mit verschiedenen Verfahren analysiert und interpretiert. In ihrem Beitrag reflektieren Ziegler und Schmitz die mit den quantitativen und qualitativen methodischen Zugängen verbundenen Möglichkeiten und Grenzen der Analyse verschiedener Datentypen.

Die Fallstudie von Ziegler und Schmitz kann durch die Lektüre des Kapitels → Kapitel 18 [Interviewerhebungen] über Interviewerhebungen vertieft werden. Da in → Kapitel 4 [Fallstudie „Metropolenzeichen“] auch rechtliche und ethische Aspekte im Umgang mit den verwendeten Daten thematisiert werden, bietet sich die Lektüre der beiden → Kapitel in Teil IV zu juristischen und ethischen Fragen im Umgang mit Daten an. Da vor allem die Bilddaten im Projekt selbst erhoben und archiviert wurden, spielt hier das Thema der Relevanz von Metadaten eine Rolle. Näheres dazu finden Sie in → Kapitel 13 [Daten und Metadaten].

Die von Pia Bergmann in → Kapitel 5 [Fallstudie „Sprachliche Variation“] präsentierte Untersuchung stammt aus dem Bereich der gesprochenen Sprache. Anhand der in der gesprochenen Sprache häufig und in unterschiedlicher Funktion verwendeten Phrase „keine Ahnung“ untersucht Bergmann die Frage, ob lautliche Reduktion in diesem Fall zufällig auftritt oder diese systematisch mit dem Verwendungskontext und mit ihren Funktionen im Gespräch zusammenhängt. Die Autorin verwendet Belegbeispiele aus

1 In den folgenden Abschnitten verweisen wir auf die einzelnen Kapitel des Bandes mit derselben Konvention, die im gesamten Band auch den Querverweisen zwischen den Kapiteln zugrundeliegt. Ein Verweis wird eingeleitet durch einen Verweispfeil (→) und besteht aus einer Kennzeichnung des Kapitels (im vorliegenden Fall: „Kapitel 3“) und der näheren Bezeichnung des Kapitels unter Verwendung eines Kurztitels (hier: „Fallstudie ‚OKAY‘“), die in eckigen Klammern steht. Die in den Verweisen gewählten, im Band einheitlich verwendeten Kurztitel finden sich auch in den Kopfzeilen wieder, so dass das betreffende Kapitel nicht nur anhand des Inhaltsverzeichnis, sondern auch beim Durchblättern im Band aufgefunden werden kann.

dem FOLK-Korpus des gesprochenen Deutsch und unterzieht diese Beispiele einer lautlichen und funktionalen Analyse. Dabei spielen sowohl die Frequenzspektren des Lautsignals als auch die Transkripte der Äußerungskontexte eine Rolle. Die Arbeit ist in dieser Hinsicht empirisch und qualitativ angelegt. Zur Methode der Datenauswertung bietet sich im Anschluss die Lektüre von → Kapitel 23 [Gesprächsanalytische Transkription] an.

Wolfgang Imo stellt in → Kapitel 6 [Fallstudie „Diskursmarker“] eine Fallstudie zu Diskursmarkern vor. Sein Beitrag ist so angelegt, dass er auch als Einführung in das Feld und die Methode der Interaktionalen Linguistik gelesen werden kann. Der Autor geht sowohl auf den ‚Kernbereich‘ der Interaktionalen Linguistik, die gesprochene Sprache, als auch auf neuere und noch wenig erschlossene Bereiche wie computervermittelte Schriftkommunikation oder verschriftete literarische Interaktionen ein. Dem Gegenstandsbereich entsprechend stehen Beispiele aus der gesprochenen Sprache im Vordergrund. Diese werden den in Münster und Hamburg aufgebauten Audiodatenbanken sowie dem FOLK-Korpus entnommen. Dazu kommen Exzerpte aus einem Korpus mit Kurznachrichtenkommunikation und aus Texten von Fontane und Gryphius, die die Modi der internetbasierten Kommunikation und der geschriebenen Sprache abdecken. Methodisch ist diese Studie quantitativ, korpusbasiert und umfasst mehrere Sprachmodi.

Einige der von Imo verwendeten Korpora werden in → den Kapiteln des Teils VI zu Korpora gesprochener Sprache und zu Korpora internetbasierter Kommunikation ausführlicher vorgestellt. Als weitere vertiefende Lektüre ist der Text von Thomas Schmidt zu Werkzeugen für die Transkription gesprochener Sprache (→ Kapitel 27 [Transkriptionswerkzeuge]) zu empfehlen.

Heidrun Kämper stellt in → Kapitel 7 [Fallstudie „Emotion“] eine Untersuchung vor, die sie hinsichtlich ihres Gegenstands dem Bereich ‚Sprache und Emotion‘ zuordnet. Die Fragestellung, die den Ausgangspunkt der Studie bildet, bezieht sich auf die Kodierung von Gefühlen und auf deontisch markierte Ausdrücke. Als Datengrundlage verwendet

sie ein Korpus mit Berichten von Nationalsozialist*innen, die ihren Weg zur NSDAP schildern, in die sie in der späten Weimarer Republik eintraten. Der Beitrag analysiert diese Texte, indem er danach fragt, welche Gefühlsbezeichnungen in den untersuchten Texten verwendet werden und worauf sie referieren. Die Studie ist primär qualitativ, diskurslinguistisch und auf die genaue Analyse markanter Beispiele aus dem Korpus angelegt und leistet einen linguistischen Beitrag zur Entstehungsgeschichte des Nationalsozialismus.

Der Beitrag von Sandra Schwinning und Miriam Morek in → Kapitel 8 [Fallstudie „Sprachliche Förderung“] beschreibt eine Fallstudie aus dem Bereich der Sprachdidaktik. Am Beispiel einer Studie zu Satzbildungsfähigkeiten von Schüler*innen werden Aufbau und Probleme von Interventionsstudien erläutert, mit denen die Wirksamkeit bestimmter didaktischer Maßnahmen überprüft wird. Der Artikel skizziert die Entwicklung des Untersuchungsdesigns, den Entwurf und die Überprüfung eines Messinstruments (Test) und zeigt zudem anhand einer Beispielhypothese, wie Forschungshypothesen formuliert und statistisch überprüft werden. Methodisch zeigt diese Fallstudie, wie eine Interventionsstudie im Rahmen einer linguistischen Forschungsfrage ausgestaltet werden kann. In der Studie wird dazu ein selbst erstelltes Korpus von Schülertexten als ein Teil der Datenbasis verwendet.

Carolin Müller-Spitzer und Sascha Wolfer präsentieren in → Kapitel 9 [Fallstudie „Wörterbücher“] eine Studie, die der Frage nachgeht, ob und in welchem Ausmaß Wörterbücher und andere lexikographische Ressourcen die Ergebnisse von Textüberarbeitungen verbessern. Das Design der Studie ist experimentell: Studierende wurden gebeten, zwei Texte zu optimieren und waren dabei zufällig in drei unterschiedlichen Gruppen mit Versuchsbedingungen eingeteilt: 1. ein Ausgangstext ohne Hinweise auf potenzielle Fehler im Text, 2. ein Ausgangstext, bei dem problematische Stellen im Text hervorgehoben waren und 3. ein Ausgangstext mit hervorgehobenen Problemstellen zusammen mit lexikographischen Ressourcen, die zur Lö-

sung der spezifischen Probleme verwendet werden konnten. Gemessen werden die Bearbeitungszeiten für die Aufgaben und die Qualität der Lösungen. Mithilfe teststatistischer Mittel wird gezeigt, dass die Bereitstellung lexikographischer Hilfsmittel einen Effekt auf die Qualität der Lösungen hatte. Die Studie verwendet verschiedene statistische Methoden für die Hypothesenprüfung.

Begleitend oder vertiefend zu den beiden letztgenannten Kapiteln empfehlen sich → Kapitel 21 [Aufbereitung Untersuchungsergebnisse] und → Kapitel 28 [Werkzeuge statistische Analyse].

Michael Beißwenger und Steffen Pappert präsentieren in → Kapitel 10 [Fallstudie „Emojis“] eine Fallstudie zum höflichen kommunikativen Handeln in einem Lehr-Lern-Kontext. Vor dem Hintergrund einerseits der medienlinguistischen Erforschung internetbasierter Kommunikation und andererseits der linguistischen Höflichkeitsforschung werden Funktionen der Emoji-Verwendung bei der Bearbeitung einer Peer-Feedback-Aufgabe untersucht. Die Autoren betten die Arbeit in den weiteren Kontext der angewandten Linguistik ein und verstehen die Fallstudie als ein typisches Beispiel für dieses Forschungsfeld. Das Datenset, auf das sich die Studie bezieht, umfasst eine Sammlung von Postings aus einer digitalen Lernumgebung, mit denen Studierende im Rahmen eines sprachdidaktischen Seminars Arbeitsergebnisse ihrer Kommiliton*innen bewertet und kommentiert haben. Beim Verfassen der Postings konnten die Studierenden auf eine Auswahl an Emojis zurückgreifen. Die Studie vereint qualitative und quantitative Elemente und zeigt vor dem Hintergrund der linguistischen Theorie höflichen Handelns, dass die Studierenden Emojis systematisch als Mittel zur Abfederung gesichtsbedrohender sprachlicher Handlungen eingesetzt haben.

Vertiefend zu dieser Studie bietet sich v. a. → Kapitel 26 [Korpora internetbasierter Kommunikation] an.

Gegenstand der Untersuchung von Sandra Hansen, Felix Bildhauer und Marek Konopka in → Kapitel 11 [Fallstudie „Fugenelemente“] sind korpuslinguistische Zugänge zur Varia-

tion im Auftreten des Fugenelements in Komposita aus zwei Nomen (*Arbeit | s | weg*). Die Studie fokussiert dann die sehr variable Verfügung nach Erstglied auf Konsonant (*Arbeit | s | weg* vs. *Heimat | ort*) und modelliert statistisch den Einfluss von Größen, deren Bedeutung in der bisherigen Forschung nur angenommen, aber nicht überprüft werden konnte. Die Studie ist damit quantitativ angelegt und bezieht – für Studierende der Linguistik – sehr weitreichende statistische Methoden und Analysen ein. Diese Fallstudie zeigt somit auch, wie weit das Feld der statistischen Analysen von geschriebener Sprache methodisch ist und welche Kompetenzen es zu erwerben gilt, wenn tieferegehende statistische Methoden angewandt werden sollen.

Vertiefend zu dieser Studie legen wir Ihnen das Kapitel → Kapitel 21 [Aufbereitung Untersuchungsergebnisse] und Kapitel → Kapitel 28 [Werkzeuge statistische Analyse] ans Herz.

Der Beitrag von Annelen Brunner und Fotis Jannidis in → Kapitel 12 [Fallstudie „Redewiedergabe“] stellt eine Fallstudie vor, in der die quantitative Verteilung von direkten und nicht-direkten Formen von Redewiedergabe im Vergleich zwischen den Literaturtypen Hochliteratur und Heftrromanen untersucht wurden. Die Studie ist ein Beitrag zum Feld der (literaturwissenschaftlichen) Erzählforschung. Sie basiert auf manuell annotierten Daten aus einem selbst zusammengestellten Korpus und überprüft daran die Verlässlichkeit automatischer Annotationswerkzeuge. Die Studie ist damit korpusbasiert, quantitativ ausgerichtet und bezieht dabei auch das Thema ‚maschinelles Lernen‘ ein. In ihr konnte nachgewiesen werden, dass sich die Literaturtypen sowie auch unterschiedliche Genres von Heftrromanen hinsichtlich der verwendeten Wiedergabeformen unterscheiden.

Da in diesem Kapitel Annotationen und Annotationswerkzeuge thematisiert werden, bieten sich als ergänzende Lektüre das → Kapitel 14 [Annotation] zu linguistischer Annotation sowie das → Kapitel 30 [Werkzeuge für die automatische Sprachanalyse] (und damit auch zur linguistischen Annotation) an.

Überblickskapitel zu einzelnen methodischen Aspekten, Ressourcen und Werkzeugen

Das Feld der empirischen Linguistik ist sehr breit. Dementsprechend ist es nicht möglich, alle relevanten grundlegenden methodischen Aspekte, Ressourcen und Werkzeuge in einem Buch zu versammeln. Wir haben aber versucht, in → den Teilen III bis VII die grundlegenden Fragen im Zusammenhang mit Sprachdaten, wichtige Vorgehensweisen bei der Erstellung und Analyse von Daten sowie zentrale heute verfügbare Ressourcen und Werkzeuge anzusprechen.

Die beiden → Kapitel in Teil III behandeln zentrale Aspekte des Korpusaufbaus und -managements, die für Sie als Nutzer*innen solcher Ressourcen, eventuell auch als jemand, der oder die vergleichbare eigene Sammlungen aufbauen möchte, von hoher Relevanz sind. Ein Sprachkorpus besteht typischerweise aus drei Schichten: den eigentlichen Äußerungen als Primärdaten, weiteren, diesen Daten zugeordneten Annotationen und Metadaten, die die Primärdaten hinsichtlich verschiedener Eigenschaften beschreiben.

Thomas Schmidt thematisiert in → Kapitel 13 [Daten und Metadaten] verschiedene Funktionen von Metadaten, d. h. von Daten, die der Dokumentation bzw. Beschreibung empirischer Sprachdaten dienen. Er beschreibt ihre Rolle im Forschungsprozess und ihre Bedeutung für die Konzepte der Ausgewogenheit und Repräsentativität von Korpora. Anhand des *Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK)* stellt er sodann Metadaten beispielhaft für ein konkretes Korpus vor und zeigt, wie diese bei Korpusanalysen zum Einsatz kommen.

In → Kapitel 14 [Annotation] führen Lother Lemnitzer und Michael Beißwenger die Funktion und Praxis des Annotierens von Korpusdaten ein. Sie stellen verschiedene für das Deutsche verwendete Annotationsschemata (sog. *Tagsets*) vor und veranschaulichen an Datenbeispielen, wie diese in die Primärdaten eingebracht werden. Das Kapitel legt die Grundlagen dafür, dass Sie informiert mit annotierten Sprachdaten umgehen, den

Nutzwert von Annotationen bei der Abfrage und Analyse linguistischer Korpora einschätzen und grundsätzliche Überlegungen für die Planung eigener Annotationsprojekte anstellen können.

Sollten Sie selbst eigene Korpusdaten annotieren wollen, dann sollten Sie im Anschluss das Kapitel → Kapitel 30 [Werkzeuge für die automatische Sprachanalyse] zur technischen, genauer computerlinguistischen, Verarbeitung von Sprachdaten lesen. Die dort vorgestellten Methoden bilden die Basis für ein vertieftes Verständnis der technischen Aspekte von Annotation. Im Anschluss haben Sie die Wahl zwischen einfach zu bedienenden Korpusplattformen, in die Annotationswerkzeuge integriert sind (→ Kapitel 29 [Werkzeuge Korpusanalyse]) und einer Annotationsplattform, die auch den Ansprüchen eines größeren Projekts mit einer Vielzahl von Annotator*innen gerecht wird (→ Kapitel 31 [INCEPTION]).

In → Teil IV des Bandes haben wir zwei Kapitel zu Themen zusammengestellt, die man beim Forschungsdesign gerne verdrängt und verschiebt, denen man sich aber so früh wie möglich stellen sollte. Es geht um juristische und ethische Fragen beim Umgang mit Sprachdaten für die Zwecke empirischer linguistischer Untersuchungen. Die Problematik klingt in einigen der Fallstudien an, etwa im Zusammenhang mit Probandenbefragungen in → Kapitel 4 [Fallstudie „Metropolenzeichen“] und hinsichtlich der Verwendung von Korpora, die für die Öffentlichkeit gar nicht oder nur mit einiger Mühe zugänglich sind, wie in den Kapiteln → Kapitel 6 [Fallstudie „Diskursmarker“] und → Kapitel 12 [Fallstudie „Redewiedergabe“].

Pawel Kamocki und Andreas Witt führen in → Kapitel 15 [Juristische Fragen] in verschiedenen rechtliche Problematiken im Zusammenhang mit der Akquirierung und Nutzung von Sprachdaten ein. Da in den Sprachwissenschaften empirisch gearbeitet wird und Sprachdaten – insbesondere Texte, Ton- und Videoaufnahmen sowie Transkripte gesprochener Sprache, in den letzten Jahren aber auch verstärkt Sprachdaten internetbasierter Kommunikation – als Basis für die linguistische Forschung dienen, müssen

rechtliche Rahmenbedingungen für jede Art von Datennutzung beachtet werden. Sprachdaten basieren auf schriftlichen, mündlichen oder gebärdeten Äußerungen von Menschen, wodurch sich juristisch begründete Beschränkungen ihrer Nutzung ergeben. Diesen Fragen müssen Sie sich vor allem dann stellen, wenn Sie für ihr Forschungsprojekt selbstständig Daten erheben und die Ergebnisse Ihrer Forschung früher oder später veröffentlichen oder für Dritte zugänglich machen wollen.

Bei der Erhebung von Daten, vor allem aus den sog. sozialen Medien, stehen Ethik und Erkenntnisinteresse in einem Spannungsfeld, das Konstanze Marx in → Kapitel 16 [Ethische Fragen] darstellt. Soziale Medien bilden zwischenmenschliche Interaktion unter quasi-öffentlichen Bedingungen ab. Für die sprachwissenschaftliche Forschung legen forschungspraktische Erwägungen, vor allem hinsichtlich der Verfügbarkeit von Daten, eine Zuwendung zum digitalen Interaktionsraum nahe. Die hiermit verbundenen ethischen Herausforderungen haben u. a. mit den folgenden Fragen zu tun: Was ist zu tun, wenn die Urheber*innen der Daten nichts von der Untersuchung wissen, was, wenn sie eigentlich Autor*innenstatus haben? Welche Daten sind öffentlich, welche privat und darf in Seminararbeiten darauf zurückgegriffen werden?

Die gute Nachricht ist, dass es eine Vielzahl von Korpora der internetbasierten Kommunikation gibt, bei deren Zusammenstellung diese Fragen bereits geklärt wurden. Es empfiehlt sich daher, → Kapitel 26 [Korpora internetbasierter Kommunikation] als Überblick zu diesen Korpora zu konsultieren, bevor Sie sich den rechtlichen Fragen und ethischen Herausforderung einer eigenen Sammlung von Online-Daten stellen (diese ist aber immer dann die Mühe wert, wenn die existierenden Bestände partout nicht zu Ihrem Forschungsinteresse passen).

Die Beiträge in → Teil V führen in verschiedene empirische Arbeitsmethoden ein, die der Erhebung und Aufbereitung von Sprachdaten dienen. In → Kapitel 17 [Audio- und Videografie] stellt Anja Stukenbrock methodische Grundlagen der Audio- und Videogra-

fie von verbaler Interaktion aus der Sicht von multimodaler Konversationsanalyse und verwandten Forschungsrichtungen vor. Der Fokus liegt auf der Erhebung von Videodaten für eigene Forschungszwecke. Den Ausgangspunkt bildet die Reflexion des Konstruktcharakters von Daten, aus dem sich ein enger Zusammenhang zwischen Gütekriterien für Aufnahmen und für darauf basierende Analysen ergibt. → Kapitel 18 [Interviewerhebungen] von Katharina König vermittelt einen Überblick über den Einsatz von Interviews als Erhebungsinstrument in der Linguistik. Neben der Vorstellung verschiedener Interviewformen werden anhand zweier Studien mögliche Fragestellungen und Analysebefunde interviewbasierter linguistischer Forschung skizziert. Abschließend werden prototypische Phasen einer Interview-Studie dargestellt und forschungspraktische und -ethische Herausforderungen diskutiert.

In → Kapitel 19 [Lautes Denken] führen Caroline Schuttkowski und Katharina Staubach in die Methode des Lauten Denkens ein. Dabei werden Proband*innen bei der Bearbeitung einer Aufgabe aufgefordert, alle kognitiven Prozesse zu verbalisieren, um diese einer Analyse zugänglich zu machen. In dem Kapitel werden zunächst Methodik und Durchführung des Lauten Denkens präsentiert und eine Möglichkeit zur Auswertung der nach dieser Methode erhobenen Sprachdaten anhand der qualitativen Inhaltsanalyse vorgestellt.

Die Erhebung, Aufbereitung und Auswertung von Korpusdaten steht im Zentrum des Kapitels → Kapitel 20 [Korpusdaten] von Lothar Lemnitzer. In diesem Kapitel werden die Entscheidungen und Prozesse dargestellt, die beim Aufbau und vor allem der Aufbereitung von Korpora, besonders von großen Referenzkorpora, getroffen bzw. ausgeführt werden müssen. Die Kenntnis dieser Entscheidungen und Prozesse stellt ein wichtiges Hintergrundwissen dar, um Korpora reflektiert für eigene linguistische Untersuchungen nutzen zu können.

Die Aufbereitung von linguistischen Analysen ist Gegenstand des Kapitels → Kapitel 21 [Aufbereitung Untersuchungsergebnisse] von Sandra Hansen und Sascha Wolfer.

Sobald eine statistische Datenanalyse abgeschlossen ist, müssen in einem weiteren Schritt die Untersuchungsergebnisse aufbereitet und dargestellt werden. Hierzu gibt es verschiedene Möglichkeiten, die davon abhängig sind, welche Art von Analyse man durchgeführt hat. Der Beitrag geht dabei im Besonderen auf die Aufbereitung der Daten in Tabellenform ein und zeigt an einem Beispiel, wie man die Ergebnisse von statistischen Tests darstellen bzw. visualisieren kann.

Lothar Lemnitzer und Nils Diewald befassen sich in ihrem Beitrag in → Kapitel 22 [Korpusabfragen] mit der Abfrage und Analyse von Korpusbelegen. Es werden grundlegende Konzepte von Abfragesystemen und Abfragesprachen für die Suche in Korpora vorgestellt. Diese Konzepte sollen helfen, die einzelnen Abfragesprachen besser verstehen und vergleichen zu können. Die Details und die Möglichkeiten und Grenzen zweier als Beispiel gewählter Abfragesprachen werden im zweiten Teil anhand vieler Beispiele durchgeführt.

Miriam Morek führt in → Kapitel 23 [Gesprächsanalytische Transkription] am Beispiel des *Gesprächsanalytischen Transkriptionssystems (GAT 2)* in die gesprächsanalytische Transkription ein. Transkription bezeichnet den Prozess des Verschriftens gesprochener Sprache und stellt einen zentralen Schritt der Untersuchung mündlicher Kommunikation dar. In diesem Kapitel wird erläutert, welchen Zweck Transkripte im Forschungsprozess erfüllen und wie beim Transkribieren von Gesprächsaufzeichnungen vorgegangen wird. Skizziert werden darüber hinaus auch Möglichkeiten der graphischen Repräsentation von Multimodalität. So zeigen diese verschiedenen Kapitel nicht alle, sondern ausgewählte Methoden, die für die Erhebung und Aufbereitung von Sprachdaten für viele Studienarbeiten in der germanistischen Linguistik wichtig sind.

Teil VI und VII des Bandes geben Ihnen schlussendlich einen Überblick darüber, welche Korpusressourcen es bereits für das Deutsche gibt und welche Werkzeuge für die empirische Sprachanalyse zur Verfügung stehen.

In Teil VI werden bestehende Korpora und vor allem Korpusansammlungen vorgestellt,

die von großen Institutionen wie dem Leibniz-Institut für Deutsche Sprache, der Berlin-Brandenburgischen Akademie der Wissenschaften oder von anderen Anbietern für die Forschung bereitgestellt werden, meist im Paket mit Abfragesprachen und Rechercheumgebungen – mehr zu diesen erfahren Sie in → Kapitel 22 [Korpusabfragen].

Die unkompliziertesten und deshalb wohl am häufigsten verwendeten Korpora sind die Korpora geschriebener Sprache. In → Kapitel 24 [Korpora geschriebener Sprache] stellt Lothar Lemnitzer die wichtigsten Vertreter aus der germanistischen Linguistik vor. Computerlinguistische Verfahren (→ Kapitel 30 [Werkzeuge automatische Sprachanalyse]) und Annotationen bzw. Tagsets (→ Kapitel 14 [Annotationen]) sind vor allem anhand dieser Daten entwickelt worden und sind auf diese zugeschnitten. Bei den Korpora gesprochener Sprache bildet das Transkript des Audiosignals (dazu → Kapitel 23 [Gesprächsanalytische Transkription]) die Brücke für einen einfacheren, nämlich zeichenkettenbasierten Zugang zu den Sprachdaten. Bei avancierteren Systemen sind Transkript und Audiosignal aligniert und gelegentlich mit Videoaufnahmen der Äußerungssituation verbunden, sodass man mehrere Modi beobachten und erforschen kann. Die wichtigsten Korpora dieser Art stellt Thomas Schmidt in → Kapitel 25 [Korpora gesprochener Sprache] vor.

Wir betrachten Korpora der internetbasierten Kombination als einen eigenen Korpusstyp, aus dem einfachen Grund, dass die internetbasierte Kommunikation eine Sprachgebrauchsdomäne mit einer charakteristisch eigenständigen sprachlichen und strukturellen Merkmalhaftigkeit bildet. Ein Großteil der internetbasierten Kommunikation wird im Medium der Schrift organisiert. Die dahinterstehenden Kommunikationstechnologien und Kommunikationsumgebungen ermöglichen aber kommunikative Sequenzen, die interaktional strukturiert sind und daher (zumindest zu einem gewissen Grad) mehr mit den Strukturen eines Gesprächs gemeinsam haben als mit denen monologischer Texte. Durch ihre Schriftlichkeit weisen sie zugleich materiale und mediale Qualitäten von Textformen auf, die sich von

den Qualitäten gesprochener Sprache unterscheiden. Dies schlägt sich in charakteristischen sprachlichen Ausformungen nieder. In einigen der in Teil II vorgestellten Fallstudien stehen deshalb Daten der internetbasierten Kommunikation im Mittelpunkt der Untersuchung (→ Kapitel 3 [Fallstudie „OKAY“] und → Kapitel 10 [Fallstudie „Emojis“]) oder werden zumindest mitbedacht (→ Kapitel 6 [Fallstudie „Diskursmarker“]). Michael Beißwenger und Harald Längen stellen in → Kapitel 26 [Korpora internetbasierter Kommunikation] die wichtigsten Vertreter von Korpora internetbasierter Kommunikation vor und skizzieren die Herausforderungen, die sich beim Aufbau und der Nutzung solcher Sprachdatensammlungen in spezifischer Weise stellen. Der letzte Teil des Bandes, → Teil VII, ist Werkzeugen gewidmet, die für Sie hilfreich sein können, wenn Sie 1. selbst Sprachdaten zusammenstellen und diese Daten verwalten und nutzen möchten oder 2. aus größeren Datensammlungen spezifisch auf Ihre Forschungsfrage zugeschnittene Samples von Sprachdaten auswerten und im Hinblick auf eine Publikation angemessen visualisieren wollen. Thomas Schmidt stellt in → Kapitel 27 [Transkriptionswerkzeuge] Werkzeuge für die Transkription gesprochener Sprache vor und schließt dadurch an das methodisch ausgerichtete Kapitel → Kapitel 23 [Gesprächsanalytische Transkription] an. Transkriptionswerkzeuge sind spezialisierte Softwaretools für die Transkription und Annotation von Audio- oder Videoaufzeichnungen gesprochener Sprache. Thomas Schmidt erklärt einleitend, worin der Mehrwert solcher Werkzeuge gegenüber einfacher Textverarbeitungssoftware liegt, und gibt dann einen Überblick über grundlegende Prinzipien und einige weitverbreitete Tools. Am Beispiel der Editoren *FOLKER* und *OrthoNormal* demonstriert er den praktischen Einsatz zweier Werkzeuge im Zusammenhang mit typischen Arbeitsabläufen eines Korpusprojekts. Ebenfalls an ein konzeptuelles Kapitel, → Kapitel 21 [Aufbereitung Untersuchungsergebnisse] von Sandra Hansen und Sascha Wolfer, schließt das gleichermaßen von Hansen und Wolfer verantwortete Kapitel → 28 [Werkzeuge statistische Analy-

se] zu Werkzeugen für die statistische Analyse an. In diesem Kapitel werden Werkzeuge vorgestellt, die Sie bei der quantitativen Analyse der von Ihnen erhobenen Daten unterstützen können. Im Zentrum der Darstellung steht dabei die in der Linguistik weitverbreitete Programmiersprache R sowie damit verbundene Entwicklungsumgebungen. Susanne Haaf stellt in → Kapitel 29 [Werkzeuge Korpusanalyse] Werkzeuge und Plattformen für die Verwaltung, Analyse und Auswertung von Korpusdaten vor. Die vorgestellten Tools sind das Mittel der Wahl, wenn die von großen Plattformen und Anbietern bereitgestellten Daten und Rechercheumgebungen für Ihre Untersuchungen nicht hinreichen. Das Spektrum reicht von einfachen Werkzeugen, die für genau eine Aufgabe etwa im Bereich der Annotation entwickelt wurden, über einfache, aber leicht zu bedienende Korpusplattformen bis hin zu ausgereiften, umfangreichen, von der Bedienung her aber etwas anspruchsvolleren Plattformen, die man sich auch als Werkzeugkisten vorstellen kann. Am Anfang aber steht dabei die Wahl zwischen verschiedenen Werkzeugkisten, aus denen man sich dann die geeigneten Werkzeuge heraussuchen kann. Diese wird dadurch erleichtert, dass die Möglichkeiten der jeweiligen Software am Beispiel einer durchgängigen Fallstudie demonstriert werden. Auch in → Kapitel 30 [Werkzeuge automatische Sprachanalyse] von Andrea Horbach werden Werkzeuge für die Verarbeitung von Sprachdaten vorgestellt. Insofern ergibt sich eine Überschneidung zu → Kapitel 29 [Werkzeuge Korpusanalyse]. Die gemeinsame Lektüre beider Kapitel ist daher sehr zu empfehlen. Das Kapitel → Kapitel 30 [Werkzeuge automatische Sprachanalyse] stattet Sie mit mehr Hintergrundwissen aus der Welt der Sprachtechnologie bzw. Computerlinguistik aus und befähigt Sie zu einer besseren Einschätzung zur Qualität der Ergebnisse, die Sie von einer automatischen Analyse und Annotation von Korpusdaten erwarten können. Diese Qualität ist sowohl von der Art der Daten, die analysiert werden sollen, abhängig als auch von der Komplexität der Analyse- bzw. Annotationsaufgabe. In → Kapitel 31 [INCEpTION] beschreibt Marcel Fladrich die

Software INCEpTION mit ihren wesentlichen Funktionen und stellt anschließend ein forschungsbezogenes Nutzungsszenario am Beispiel der Annotation literarischer Texte im DFG-Projekt „Interaktionale Sprache bei Andreas Gryphius – datenbankbasiertes Arbeiten zum Dramenwerk aus linguistisch-literaturwissenschaftlicher Perspektive“ vor. Es handelt sich um eine reichhaltige Software, die als Plattform vor allem für große, ein Team umfassende Annotationsprojekte geeignet ist und für diese kollaborative Form des Arbeitens eine Vielzahl von nützlichen Hilfsmitteln bereitstellt. Diese Hilfsmittel erleichtern und beschleunigen nicht nur viele Arbeitsschritte, sondern vereinfachen auch deren Überwachung. Die Lektüre der Kapitel 29 bis 31 kann mit der Lektüre des eher konzeptuellen Kapitels → Kapitel 14 [Annotation] zu Annotationen kombiniert werden.

Zu guter Letzt sei noch erwähnt, dass wir die Autor*innen des Bandes gebeten haben, eine Form geschlechtergerechter Sprache zu verwenden und auf die durchgängige Verwendung des sogenannten generischen Maskulinums zu verzichten. Welche Mittel ihnen dafür sinnvoll erschienen, war den Autor*innen überlassen. Innerhalb einzelner Beiträge sollten die Formen des Genderns einheitlich sein, über den Band hinweg sind sie es nicht. Dies scheint uns aber auch am besten den Sprachwandelprozess, in dem wir uns gerade befinden, abzubilden und den Präferenzen der einzelnen Autor*innen entgegenzukommen. Alle Beiträger*innen waren mit diesem Vorgehen einverstanden.

Vernetzung der Kapitel des Buches

Die Autor*innen der Beiträge und wir als Herausgeber*innen haben zahlreiche Querverweise in die einzelnen Kapitel gesetzt, die es Ihnen ermöglichen sollen, nicht-lineare

und interessengeleitete Pfade durch dieses Buch zu finden. Einige mögliche Lesepfade haben wir bereits in dieser Einleitung skizziert. Im Zentrum sollten Ihre Forschungsfrage und Ihr Projekt stehen. Eine empirische Studie durchzuführen verlangt immer, dass man sich sowohl mit den Daten, die man erforschen möchte, als auch mit den geeigneten Methoden sorgfältig auseinandersetzt und sich in verschiedene Teilaspekte einarbeitet. Trotzdem lohnt sich diese Investition, weil man im besten Fall mit interessanten, intersubjektiv nachvollziehbaren Forschungsergebnissen belohnt wird. Deshalb möchten wir Sie ausdrücklich dazu ermuntern, diese Mühe auf sich zu nehmen. Wir freuen uns, wenn es uns mit diesem Buch gelingt, Sie dabei in nützlicher Weise zu begleiten. Wir hoffen, dass Ihnen die Lektüre der einzelnen Kapitel ebenso viel Vergnügen bereitet wie uns das Verfassen und Zusammenstellen der Texte.

Die Entstehung des Buches wäre ohne die Mithilfe vieler Hände und Köpfe nicht möglich gewesen. In erster Linie gilt deshalb unser Dank den Autor*innen der Kapitel, die unsere Fragen und Anmerkungen mit viel Geduld und Verständnis beantwortet haben. Die studentischen Hilfskräfte Hanna Hoebink, Kristin Katschak und Cathrin Pichler in Essen sowie Dora Hinderer in Mannheim haben die fertigen Manuskripte noch einmal auf Punkt und Komma geprüft. Auch ihnen gilt unser Dank.

Schließlich wollen wir von Herzen dem Lektorat des Verlags und besonders Frau Nadine Albert für die ebenso sachkundige wie verständnisvolle Begleitung des Manuskripts auf dem Wege zur Veröffentlichung danken.

Essen, Berlin, Mannheim, 1. Dezember 2021
Michael Beißwenger, Lothar Lemnitzer und
Carolin Müller-Spitzer