

1 Einleitung

1.1 Was ist Korpuslinguistik?

Die Folklore der Sprachwissenschaft¹ kennt zwei Forschertypen:

- **Der Denker**² verbringt die meiste Zeit in seinem Sessel und denkt nach. Die Sprachtheorie, die er sich mit den Jahren in seinem Kopf zurechtgelegt hat, wird durch Beispiele, die unmittelbar seiner Sprachkompetenz entspringen, bestätigt oder widerlegt. Hin und wieder notiert sich der Denker besonders komplizierte und abwegige Beispiele, die durch eine Grammatik, die dieser Sprachtheorie entspricht, hergeleitet werden können. Diese Sätze legt er Sprechern der untersuchten Sprache mit der Frage vor, ob diese Sätze denn wohlgeformt seien. Daraus, ob die befragten kompetenten Sprecher seine Beispiele gutheißen oder ablehnen, zieht der Denker weit reichende Schlüsse über den Aufbau der Grammatik dieser Sprache und der zugrunde liegenden Sprachtheorie. Was für den Denker alleine zählt, ist das Urteil kompetenter Sprecher, das auf deren Sprachgefühl und sprachlichem Wissen fußt. Der Denker hält sich an den Rändern der Sprache auf, in Bereichen, die wenig mit dem alltäglichen Sprachgebrauch zu tun haben. Im Gegenteil, der Denker ist an den Äußerungen, die tagtäglich produziert werden, herzlich wenig interessiert. Sie sind wenig erleuchtend für seine Theorie.
- **Der Beobachter** ist an authentischen Sprachdaten interessiert: je mehr Daten, desto besser. Die Theorien, die er entwickelt, sind auf die Beobachtung dieser Daten gestützt. Seine Aussagen und Hypothesen werden durch immer neue Daten bestätigt oder verworfen. Mit seinen Kollegen spricht der Beobachter vor allem darüber, welche interessanten Beobachtungen er gemacht hat. Ansonsten hält er sich überwiegend an seinem Computer auf. Das Bild, das er durch diese Beobachtungen gewinnen möchte, sollte möglichst vollständig sein, deshalb ist er vor allem an den Phänomenen interessiert, die in unserem alltäglichen Sprachgebrauch vorkommen.

¹ Wer nicht glaubt, dass es eine Folklore der Sprachwissenschaft gibt, der möge sich einmal Pulium (1991) ansehen. Auch allen anderen Lesern möchten wir dieses vergnüglich zu lesende Buch empfehlen.

² Wir verwenden in diesem Buch das generische Maskulinum bei Bezeichnungen von Personen und schließen damit selbstverständlich alle weiblichen Personen mit ein. Die Wahl dieser Form hat einzig und allein den Grund, dass ihre Verwendung das Lesen des Textes etwas einfacher macht.

Der Denker erweist sich als scharfsinniger Theoretiker, der die Grundlagen des Sprachvermögens erforscht, das allen Menschen gemeinsam ist, und dies *Universalgrammatik* nennt. Für seine Forschungen muss er seinen Sessel nur äußerst selten verlassen. Den Beobachter hingegen findet man häufig dort, wo es um die möglichst umfassende Beschreibung einer Sprache in ihrer alltäglichen Verwendung und die Vermittlung dieses Sprachgebrauchs, z.B. in Lexikographie und Fremdsprachunterricht, geht.

Diese plastische Beschreibung zweier Typen von Forschern in der Linguistik ist nicht neu. Sie findet sich so ähnlich schon bei Charles Fillmore (Fillmore, 1992). Fillmore hat in den achtziger Jahren das Lager gewechselt und sich vom theoretisierenden Linguisten zum Beobachter gewandelt. Es ist jedoch keinesfalls so, dass die Entscheidung für eine Richtung die andere Richtung ausschließt: Wer sammelt, hat damit das Denken nicht aufgegeben, und auch der Denker profitiert hin und wieder von den Erkenntnissen der Beobachter. Wir werden Beispiele dafür noch kennen lernen.

Eine Einführung in die Korpuslinguistik wendet sich in erster Linie an die Beobachter unter den Sprachwissenschaftlern. Wer Korpuslinguistik betreibt, dem geht es in erster Linie um das Beobachten und Beschreiben sprachlicher Phänomene. Wir wenden uns aber auch an die Denker und werden zeigen, dass und wie sie von den Beobachtungen und Erkenntnissen der Korpuslinguisten profitieren können. Eine enge Zusammenarbeit zwischen Denkern und Beobachtern, also zwischen theoretischen Linguisten und empirisch arbeitenden Linguisten, erscheint uns fruchtbar für beide Seiten. Eine solche Haltung ist in der Zukunft aber keinesfalls selbstverständlich. Randy Allen Harris hat sein Buch über die Sprachwissenschaft in den sechziger und siebziger Jahren des letzten Jahrhunderts „Linguistic Wars“ genannt, und dies ist sicher nicht allzu stark übertrieben. Charles Hockett, ein Vertreter der empirischen Arbeitsweise, bezeichnete die Methode, Selbstauskünfte von Sprechern über ihr sprachliches Wissen heranzuziehen, als im günstigsten Fall überflüssig (*superfluous*) und im ungünstigsten Fall als widerwärtig (*obnoxious*)³. Viele theoretische Sprachwissenschaftler im Umfeld der generativen Sprachtheorie, allen voran Noam Chomsky, bezeichnen das Werk der Korpuslinguistik als irrelevant und nutzlos⁴. Es gibt, wie gesagt, Berichte von „Lagerwechseln“⁵, was auch nicht gerade für ein friedliches Zusammenleben spricht.

Wir werden im zweiten Kapitel zeigen, dass mindestens ein Teil der Kritik, die von Sprachtheoretikern gegenüber empirisch arbeitenden Linguisten geäußert wurde, berechtigt ist. Sie betrifft Annahmen, die von der Korpuslinguistik in der Zeit vor dem Entstehen der generativen Grammatik in den fünfziger Jahren getroffen wurden. Die moderne Korpuslinguistik hat daraus gelernt. Es ist aber auch heute noch so, dass jeder, der korpuslinguistisch arbeitet, eine Antwort auf die Kritik aus dem sprachtheoretischen Lager haben sollte. Wir werden auf diese Antworten ausführlicher im dritten Kapitel eingehen.

Zunächst jedoch wollen wir eine Antwort auf die Frage geben, was Korpuslinguistik eigentlich ist. Das Wort ist ein Kompositum, es setzt sich aus den Bestandteilen *Korpus*

³ Vgl. Hockett (1964), zitiert nach McEnery und Wilson (1996). Wir werden in Abschnitt 2.2 auf die Probleme eingehen, die Selbstauskünfte von Sprechern tatsächlich mit sich bringen.

⁴ Z.B. Chomsky (1986), S. 27.

⁵ Vgl. zum Beispiel Fillmore (1992) und Sampson (1996).

und *Linguistik* zusammen. Eine Antwort auf die Frage führt also zunächst über diese beiden Begriffe.

Definition 1 (Korpus⁶). *Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen. Die Daten des Korpus sind typischerweise digitalisiert, d.h. auf Rechnern gespeichert und maschinenlesbar. Die Bestandteile des Korpus bestehen aus den Daten selber sowie möglicherweise aus Metadaten, die diese Daten beschreiben, und aus linguistischen Annotationen, die diesen Daten zugeordnet sind⁷.*

Die Sammlung von Äußerungen ist meist das Ergebnis sorgfältiger Planung⁸, was nicht ausschließt, dass auch ad hoc oder zu anderen als linguistischen Zwecken entstandene Textsammlungen einen Wert als Datenbasis haben können. Je besser ein Korpus geplant ist, um so nützlicher ist es für die spätere Forschung.

Heutzutage liegen Korpusdaten in maschinenlesbarer Form vor. Es gibt auch heute noch nicht digitalisierte Textsammlungen bzw. Recherchen, die sich auf solche beziehen. Wir werden in Kapitel 8 solche Untersuchungen vorstellen. Die Verwendung nicht digitalisierter Texte führt jedoch zu methodischen Problemen. Auch dies werden wir in Kapitel 8 zeigen. „Ältere Texte werden heute in vielen Projekten nachträglich digitalisiert. Das Gleiche gilt für Tonaufzeichnungen von Interviews, Gesprächen usw. Man tut gut daran, sich Gedanken zu machen, ob es digitalisierte Daten für die eigenen Untersuchungen gibt bzw. ob und wie man die eigenen Daten digitalisieren kann. Wir betrachten hier das digitale Korpus als die Norm.

Der Wert eines Korpus wächst, wenn seine *Primärdaten* mit beschreibenden Daten versehen werden, die z.B. Auskunft geben über die Autoren von Texten oder die Sprecher von Tonaufnahmen, über den Zeitpunkt der Entstehung usw. Man spricht hierbei auch von *Metadaten*. Von diesen Daten, die ganze Texte oder zusammenhängende Äußerungsfolgen beschreiben, unterscheiden wir die *Annotationen*, die sich auf Teile von Texten bezieht, also auf Wörter, Sätze, Paragraphen usw. Annotationen markieren und klassifizieren bestimmte Einheiten, zum Beispiel Wörter mit ihrer Wortart.

Von anderen Medien außer Text oder Ton sehen wir ab, wollen aber darauf hinweisen, dass es interessante Korpora gibt, in denen Text und Ton mit stehenden oder bewegten Bildern verbunden werden. Man spricht dann von *multimedialen* oder *multimodalen* Korpora⁹.

Der zweite konstituierende Begriff ist *Linguistik*. Diese Disziplin wird im deutschen Sprachraum meistens als *Sprachwissenschaft* bezeichnet. Damit ist der Gegenstand dieser Disziplin im weitesten Sinn umschrieben. Das Wort *Sprache* ist aber mehrdeutig, wie die folgenden Beispiele zeigen:

⁶ Im Deutschen wird das Neutrum verwendet, es heißt also *das Korpus*, wenn von einer Sammlung von Äußerungen die Rede ist. In allen anderen Bedeutungen wird das Wort im Maskulinum verwendet.

⁷ In diesem Buch wird es überwiegend um Korpora geschriebener Texte gehen. Eine gute Einführung in Korpora gesprochener Sprache liegt nun mit Draxler (2008) vor.

⁸ Vgl. hierzu ausführlich Hunston (2008).

⁹ Einen guten Überblick über multimodale Korpora gibt Jens Allwood (2008).

- (1) ... weil Deutsch die Sprache ist, in der ich meine Gedanken am schönsten darlegen kann. (taz, 25.6.1993)
- (2) ... als ich die ersten Bilder sah, verschlug es mir die Sprache. (taz, 15.11.1996)
- (3) Aber auch der Kosovo, Afghanistan und der Kaukasus kamen zur Sprache. (taz, 5.2.1999)
- (4) Sie verzichten darauf, Hölderlins Sprache mit Bedeutung aufzuladen. (taz, 6.8.1990)

In Beispiel (1) ist mit *Sprache* eine konkrete natürliche Sprache, zum Beispiel das Deutsche, gemeint. In Beispiel (2) geht es allgemeiner um das Sprachvermögen und den Zugang zu diesem, welcher bei dem entgeisterten Betrachter momentan blockiert ist. Er wäre weder in der Lage sich in Deutsch, noch in irgendeiner anderen Sprache zu äußern. In Beispiel (3) ist mit *zur Sprache kommen* ein konkretes sprachliches Ereignis gemeint. In Beispiel (4) schließlich bezieht sich der Autor auf die Eigensprache einer einzelnen Person.

Dass mit *Sprache* Unterschiedliches bezeichnet werden kann, hat Auswirkungen auf die Wissenschaft von der Sprache bzw. den Sprachen. All die in diesen Beispielen dargestellten Aspekte können Gegenstand der wissenschaftlichen Betrachtung sein. Ein Grund für den Streit zwischen den verschiedenen sprachwissenschaftlichen Lagern ist, dass der Gegenstand der eigenen wissenschaftlichen Betrachtung verabsolutiert wird und die anderen Gegenstände nicht der wissenschaftlichen Untersuchung wert befunden werden.

Korpuslinguisten haben es mit Sprache in dem Sinn zu tun, der in Beispiel (3) zum Ausdruck kommt. Die Korpora, die untersucht werden, stellen Sammlungen konkreter sprachlicher Äußerungen dar. Natürlich werden diese in einer bestimmten Sprache getätigt, z.B. im Deutschen, Spanischen oder Chinesischen. Wir werden uns in diesem Buch auf deutsche Korpora und die korpuslinguistische Untersuchung der deutschen Sprache konzentrieren¹⁰. Inwieweit von Äußerungen als Gegenstand der Untersuchung auf das Sprachvermögen der Sprecher geschlossen werden kann, ist umstritten. Es ist sogar umstritten, ob dies ein wissenschaftliches Ziel der Korpuslinguistik sein sollte¹¹.

Nach diesen Begriffsbestimmungen wollen wir nun versuchen, eine Antwort auf die Eingangsfrage zu geben: Was ist Korpuslinguistik?

Definition 2 (Korpuslinguistik). *Man bezeichnet als Korpuslinguistik die Beschreibung von Äußerungen natürlicher Sprachen, ihrer Elemente und Strukturen, und die darauf aufbauende Theoriebildung auf der Grundlage von Analysen authentischer Texte, die in Korpora zusammengefasst sind. Korpuslinguistik ist eine wissenschaftliche Disziplin, d.h. sie muss wissenschaftlichen Prinzipien folgen und wissenschaftlichen Ansprüchen genügen. Korpusbasierte Sprachbeschreibung kann verschiedenen Zwecken dienen, zum Beispiel dem*

¹⁰ Natürlich ist der Begriff *deutsche Sprache* selbst eine Abstraktion, die von Dialekten wie dem Schwäbischen, nationalen Varianten wie dem Österreichischen oder Fachsprachen wie der Sprache der Informatik abstrahiert. Von diesen Varietäten kann man zu Recht fragen, in wie weit diese noch *deutsche Sprache* sind. Das Konstrukt *deutsche Sprache* ist jedoch den meisten Sprechern vertraut und hat sich als übergeordneter Begriff auch in der Sprachwissenschaft bewährt.

¹¹ „... the task of corpus linguists is to exemplify the dominant structural patterns of the language without recourse to abstraction, or indeed to generalization“ (Sinclair, 1991, S. 103).

Fremdsprachunterricht, der Sprachdokumentation, der Lexikographie oder der maschinellen Sprachverarbeitung bzw. Computerlinguistik.

Gegenstand von Korpora und damit der Korpuslinguistik sind natürliche Sprachen, nicht formale Sprachen wie z.B. Programmiersprachen. Das schließt die Untersuchung von älteren Sprachstadien natürlicher Sprachen, wie etwa des Althochdeutschen oder des Mittelhochdeutschen, ein. Eine Vorbedingung ist allerdings, dass die überlieferten Texte dieser Sprachdenkmäler in digitalisierter Form vorliegen. In den letzten Jahren werden solche Texte in verstärktem Maße digitalisiert, man spricht dabei von *Retrodigitalisierung*¹². Eine Stärke der Korpuslinguistik ist es, dass auf Grund der Datenbasis nicht nur die Struktur einer Sprache, sondern auch deren Verwendung untersucht werden kann.

Die Einhaltung gewisser Prinzipien ist die Grundvoraussetzung jeder wissenschaftlichen Tätigkeit. Dazu gehört, dass die Ergebnisse von Untersuchungen nachprüfbar oder sogar reproduzierbar sein müssen. Im Falle der Korpuslinguistik bedeutet dies, dass die Ergebnisse von Untersuchungen durch andere Personen am selben Korpus nachvollziehbar sein sollten. Idealerweise sollte es zusätzlich möglich sein, die Untersuchungen auch an vergleichbaren, anderen Korpora als denen, auf die sie sich ursprünglich stützen, nachzuprüfen. Die gemeinsame Nutzung eines Korpus für verschiedene Untersuchungen gewährleistet, dass Forschungsergebnisse miteinander verglichen werden können. Die Methoden der Untersuchung sollten den anerkannten wissenschaftlichen Standards entsprechen, und es muss Klarheit bestehen über die Reichweite und Sicherheit von Aussagen, die auf Grund von Beobachtungen getroffen werden. Dies trifft gleichermaßen für statistische über Regularitäten wie für Gesetzesaussagen zu. Statistische Aussagen benennen Tendenzen in den Daten, die durch einzelne Gegenbeispiele nicht widerlegt werden können. Bei dieser Art von Aussagen sollte aber die Sicherheit angegeben werden können, mit der die Aussage zutrifft. Hierfür gibt es in der Statistik etablierte Verfahren. Gesetzesaussagen hingegen sind absoluter – sie bezeichnen Regeln und Zusammenhänge, die immer zutreffen. Deshalb sind sie leichter, nämlich bereits durch ein einziges Gegenbeispiel, widerlegbar.

Korpuslinguistik ist stärker als andere Richtungen der Sprachwissenschaft zweckorientiert. Die Erkenntnisse der Korpuslinguistik beeinflussen u.a. die Übersetzungswissenschaft, die Lexikografie und den Sprachunterricht.

1.2 Wer sollte dieses Buch lesen?

Diese Einführung wendet sich an Studierende und Forscher der Sprachwissenschaft, die empirisch die deutsche Sprache untersuchen wollen. Wir wollen ihnen mit diesem Buch das Wissen und die Mittel an die Hand geben, die für die Planung und Durchführung korpuslinguistischer Untersuchungen benötigt werden. Sie sollen mit diesem Buch in die Lage versetzt werden, ein für ihre Fragestellung geeignetes Korpus auszuwählen oder ein eigenes Korpus zu erstellen. Das Buch ist auch zum Selbststudium geeignet. Wir werden lediglich die Kenntnisse voraussetzen, die in einer allgemeinen Einführung in die (germanistische) Linguistik erworben werden können¹³.

¹² Vgl. hierzu Altrichter (2001) und Claridge (2008).

¹³ Zum Beispiel die allgemeine Einführung herausgegeben von Jacob Ossner und Heike Zinsmeister (2014) oder – für die syntaktische Analyse – das bewährte Arbeitsbuch von Karin Pittner und Judith Berman (2013).

1.3 Aufbau des Buchs

Im zweiten Kapitel werden wir ausführlicher auf die Kritik, die von sprachtheoretischer Seite gegen die Korpuslinguistik vorgebracht wurde, eingehen. Der Gegensatz zwischen Generativer Grammatik und Korpuslinguistik ist grundsätzlich. Er wurzelt in einer unterschiedlichen Auffassung von Gegenstand und Methode der Linguistik, wie wir darstellen werden. Wir stellen die im positiven wie negativen Sinne für die Korpuslinguistik einflussreichen linguistischen Strömungen der Generativen Grammatik und des Kontextualismus vor. Am Ende dieses Kapitels werden wir drei Ansätze korpuslinguistischer Forschung gegenüberstellen: einen korpusbasierten, rein quantitativen Ansatz, einen korpusbasierten, quantitativ wie auch qualitativ ausgerichteten Ansatz und einen korpusgestützten, qualitativen Ansatz.

Im dritten Kapitel werden wir ausführlicher darstellen, was linguistische Korpora sind, in Abgrenzung zu anderen Arten linguistischer Datensammlungen. Wir werden drei für linguistische Korpora relevante Datenebenen unterscheiden: die Primärdaten, die Metadaten und die linguistische Annotation. Für die Beschreibung linguistischer Korpora haben sich auf internationaler Ebene Standards durchgesetzt. Diese Standards werden wir vorstellen. Der abschließende Teil ist methodischen Problemen gewidmet, die man lösen sollte, bevor man Korpora für eine linguistische Untersuchung heranzieht. Wir werden die folgenden Fragen beantworten: Können Korpora repräsentativ sein? Wie findet man sprachliche Phänomene in großen Mengen von Sprachdaten? Was macht man, wenn ein zu untersuchendes Phänomen nicht im Korpus gefunden wird und was, wenn man etwas findet, das auf Grund einer entwickelten Theorie eigentlich nicht vorkommen dürfte?

Linguistische Annotationen helfen, relevante Vorkommnisse in größeren Datenmengen (wieder) zu finden. Hierzu werden die Daten linguistisch voranalysiert und mit Annotationen wie zum Beispiel Wortarten oder grammatischen Funktionen versehen. Wir werden im vierten Kapitel Mittel und Methoden der Annotation darstellen und dabei unter anderem syntaktisch annotierte Korpora vorstellen. Um die Leser an die eigene Nutzung solcher Ressourcen heranzuführen, werden wir im fünften Kapitel anschließend auf die linguistische Abfrage von Korpora eingehen, darüber hinaus Methoden für das eigene Annotieren und eine Reihe von Abfrage- und Annotierwerkzeuge vorstellen.

Im sechsten Kapitel werden wir die Leser an die quantitative Auswertung auf der Basis von Korpora heranführen. Dort, wo wir grundlegende Konzepte von Mathematik und Statistik benötigen, werden wir diese informell einführen und im Übrigen auf vertiefende Literatur zu diesem Thema hinweisen. Wir, die Autoren dieses Buches, haben die Erfahrung gemacht, dass es durchaus auch Nicht-Mathematikern gelingen kann, sich das Handwerkszeug quantitativer Forschung anzueignen.

Korpora unterscheiden sich in vielfältiger Weise. Im siebten Kapitel werden wir anhand von konkreten Korpusbeispielen eine Typologie einführen, in der verschiedene Korpus Typen systematisch unterschieden werden. Am Schluss dieses Kapitels stellen wir einige vielversprechende neue Korpusinitiativen vor.

Korpora sind die Materialgrundlage vielfältiger qualitativer und quantitativer sprachwissenschaftlicher Untersuchungen. Im achten und letzten Kapitel werden wir einige ausgewählte Untersuchungen präsentieren und damit die Vielfalt der Fragen sichtbar machen, die mit Hilfe von Korpora beantwortet werden können.

Glossar und Index im Anhang werden sicherlich auch denen helfen, die das Buch zum Nachschlagen oder zum Lernen auf eine Prüfung verwenden wollen. Begleitet wird dieses Buch von einer Webseite, die unter www.narr-studienbuecher.de/9783823378860 erreichbar ist. Hier finden Sie:

- Eine tabellarische Liste von Korpusprojekten. Diese Einträge werden nach den in Kapitel 7 eingeführten Kriterien beschrieben;
- Hinweise auf Werkzeuge, die die Arbeit mit Korpora erleichtern;
- Handreichungen zu einigen der gebräuchlicheren Korpuswerkzeuge;
- Lösungsansätze für die Übungsaufgaben;
- weitere nützliche Links;
- weitere Informationen zu den Autoren des Buchs.

Diese Einleitung ist ein guter Ort, um über weitere Einleitungen und Handbücher zu informieren, die unsere Leser auch interessieren könnten. Eine weitere deutsche Einführung in das Thema hat Carmen Scherer verfasst (Scherer, 2005). Dieser Text kann als eine etwas leichtgewichtige, an Germanisten gerichtete Alternative zu diesem Buch betrachtet werden. Drei Mitarbeiter des Instituts für Deutsche Sprache (IDS) haben eine methodisch ausgerichtete Einführung erarbeitet, die den Umgang mit großen Datenmengen und die speziellen Abfragemöglichkeiten der IDS-Korpora in den Mittelpunkt stellt (Perkuhn et al., 2012). Eine in Deutsch verfasste, aber an Anglisten gerichtete Einführung ist die von Joybrato Mukherjee (Mukherjee, 2009). Korpora gesprochener Sprache, die hier nur am Rande behandelt werden, stehen im Mittelpunkt einer gut lesbaren Einführung von Christoph Draxler (Draxler, 2008). Aus dem angelsächsischen Raum ist unbedingt das Buch von Tony McEnery, Richard Xiao und Yukio Tono zu erwähnen (McEnery et al., 2006). Es ist zum einen die Fortschreibung von McEnery und Wilson (2001), zum anderen enthält sie, über das ältere Werk hinausgehend, eine Dokumentation über wichtige methodische Diskussionen innerhalb der Korpuslinguistik (Teil B) und dreizehn Fallstudien, in denen beispielhaft Schritt für Schritt korpuslinguistische Projekte entwickelt werden (Teil C). Eine weiterführende Auseinandersetzung mit den verschiedenen Strömungen der Korpuslinguistik findet sich in McEnery und Hardie (2012). An Anfänger hingegen richtet sich die Einführung von Kübler und Zinsmeister (2015), die sich auf linguistisch annotierte Korpora konzentriert.

Neben diesen Einführungen sind auch zwei Handbücher erschienen. Bei de Gruyter wurden 2008 und 2009 zwei Bände des internationalen Handbuchs „Corpus Linguistics“ veröffentlicht (Lüdeling und Kytö, 2008, 2009). Auf einige Aufsätze aus diesem Handbuch werden wir im Laufe dieses Buches noch zurückkommen. 2010 erschien außerdem das „Routledge Handbook of Corpus Linguistics“. Ebenfalls aus Großbritannien kommt ein Werk, in dem Grundbegriffe (*Terms*) der Korpuslinguistik erläutert werden (Mahlberg und Brook O'Donnell, 2010). Schließlich möchten wir noch auf den sehr schönen Kurs hinweisen, den Noah Bubenhofer zusammengestellt und über das Web verfügbar gemacht hat (Bubenhofer, 2001).

Zunächst und vor allem wünschen wir Ihnen aber viel Spaß bei der Arbeit mit diesem Buch!

