

**Joybrato Mukherjee.** 2009. *Anglistische Korpuslinguistik. Eine Einführung* (Grundlagen der Anglistik und Amerikanistik 33). Berlin: Erich Schmidt. 223 S.

Carolin Biewer

Universität Zürich  
Englisches Seminar  
Plattenstraße 47  
CH-8032 Zürich  
carolin.biewer@es.uzh.ch

Neben klassischen Einführungen in die Grundbegriffe und Methodik der anglistischen Korpuslinguistik wie Sinclair (1991), McEnery & Wilson (1996), Kennedy (1998) und Meyer (2002) liegt mit dem neuen Buch von Mukherjee nun die erste deutschsprachige Einführung vor,<sup>1</sup> die nicht nur dadurch besticht, dass sie aktuelle Studien und neue methodische Ansätze aus dem sich schnell fortentwickelnden Gebiet vorstellt. Sie demonstriert auch eindrücklich, dass die Korpuslinguistik unser Erforschen und Bewerten von sprachlichen Merkmalen maßgeblich verändert hat und dabei nicht nur als eine neue Methode, sondern auch als eigenständiges Teilgebiet der Linguistik gesehen werden kann. Mukherjee – selbst Autor von zahlreichen korpuslinguistischen Studien (z. B. Mukherjee 2001, 2002) – weiß die zahlreichen Facetten dieses Gebietes didaktisch so aufzubereiten, dass Studienanfänger wie auch fortgeschrittene Studenten mit Vorkenntnissen dazu animiert werden, sich mit der Korpuslinguistik vertieft zu beschäftigen. Nebenbei werden sie mit der Terminologie und Grundsätzen und Ergebnissen verschiedenster linguistischer Bereiche vertraut gemacht, in denen korpuslinguistische Methoden angewendet werden können, während auch Dozierende Anregungen für die Gestaltung ihres eigenen Unterrichts finden.

---

<sup>1</sup> Deutschsprachige Einführungen wie Scherer (2006) und Lemnitzer & Zinsmeister (2006) beziehen sich auf die germanistische und nicht auf die anglistische Korpuslinguistik.

Die vorliegende Einführung ist daher in vielerlei Hinsicht für alle, die ein wissenschaftliches Interesse an der Korpuslinguistik haben, eine lohnende Lektüre.

Das Buch ist in sieben Kapitel aufgeteilt, die nacheinander in die Grundsätze, Methodik und Anwendungsmöglichkeiten der Korpuslinguistik einführen. Im ersten Kapitel wird definiert, was man in der modernen Korpuslinguistik unter einem Korpus versteht, welche allgemeinen Grundsätze es bei der Kompilierung und der Annotation eines Korpus zu beachten gilt, und wie man sich eine Korpusanalyse vorzustellen hat. Der Leser erfährt insbesondere einige interessante Fakten zur Entstehung und Bedeutung der modernen Korpuslinguistik. Zwar wurde die Beschreibung von Sprache auf Basis von Textsammlungen selbstverständlich auch schon lange vor der Entwicklung von Computern praktiziert. Erst aber durch die Computertechnik gelingt es, große maschinenlesbare Textsammlungen anzulegen (z. B. das BNC mit 100 Millionen Wörtern), die sich in kurzer Zeit vollständig und systematisch nach linguistischen Merkmalen durchsuchen und analysieren lassen. So wird es möglich, mit so genannten *general corpora*, die den alltäglichen Sprachgebrauch einer Sprachgemeinschaft repräsentieren, allgemeine Aussagen über den Sprachgebrauch zu machen. Als das erste maschinenlesbare Korpus in den 60er Jahren in Form des *Brown Corpus* kompiliert wurde (eine Sammlung von amerikanischen Texten verschiedener Genres), stieß dieser neue Ansatz in der Linguistik, der die Untersuchung authentischer Daten des alltäglichen Sprachgebrauchs in den Vordergrund rückte, zuerst auf heftige Kritik. Es herrschte ein generativistischer Zeitgeist, der sich nur für die muttersprachliche Kompetenz, für das Mögliche und nach dem Regelsystem der Sprache Erlaubte, aber nicht für das tatsächlich Verwendete interessierte. Die Bedeutung von Korpora gerade auch für die Erstellung von Wörterbüchern wurde aber frühzeitig auch auf kommerzieller Ebene schnell erkannt (COBUILD-Projekt, 80er Jahre).

Ein weiteres wichtiges Fazit dieses ersten Kapitels ist, dass der korpuslinguistische Ansatz in der Linguistik das Wesen der anglistischen Sprachwissenschaft mit der neuen Datengrundlage und Methodik entscheidend beeinflusst hat. Durch eine Neuordnung von Sätzen eines Textes durch Konkordanzen wird es möglich, den vertrauten Sprachgebrauch auf neue Weise unter die Lupe zu nehmen (Hunston 2002: 2). Forschungsdesiderata zur Frequenz von Formen und Strukturen, wie sie lange vor Beginn der korpuslinguistischen Analyse formuliert wurden, können nun ausgeführt werden, und dabei wird deutlich, dass auf Intuition basierende Annahmen zur Sprachverwendung oft nicht mit dem eigentlichen Sprachgebrauch übereinstimmen. Didaktisch besonders gelungen an diesem Kapitel ist der Abschnitt 1.4., in dem beispielhaft anhand von verschiedenen Möglichkeiten der Komplementation des Verbs *provide* Schritt für Schritt nachvollzogen wird, wie eine korpuslinguistische Analyse funktioniert, so dass Studierende eine gute Vorstellung davon bekommen, wie sie selbst eine solche Analyse durchführen können.

Im zweiten Kapitel werden die wichtigsten Korpora des Englischen vorgestellt und nach ihrer Größe, ihrem Inhalt und ihrem Design kategorisiert. Neben den so genannten Standardkorpora der ersten Generation, die den Aufbau und die Größe des *Brown Corpus* für andere Varietäten oder Zeitabschnitte komplett oder in leichter Abwandlung reproduzieren (FLOB, Frown, aber auch die *ICE family*), werden größere Korpora wie das BNC oder die 500 Millionen Wörter umfassende *Bank of English* kurz vorgestellt. Es werden die Unterschiede zwischen statischen Korpora und dynamischen *monitor corpora* aufgeführt und neben den genannten synchronen auch diachrone Korpora vorgestellt. Gerade dadurch, dass der Autor auch viele eher unbekannte, spezialisierte Korpora aufführt, erhält der Leser einen guten Eindruck von der Vielfalt und den Möglichkeiten der thematischen Erstellung von Korpora.

Datensammlungen, die nicht eigens für die korpuslinguistische Analyse erstellt wurden, werden aber ebenso für die korpuslinguistische Forschung genutzt, weshalb in diesem Kapitel auch korpusähnliche Ressourcen wie Zeitungsarchive besprochen werden. Insbesondere geht der Autor auf neuere Forschung zur Nutzung des Internets als Korpus oder als Quelle zur Erstellung eines Korpus (*web-derived corpora*) ein. Unter anderem wird hier, basierend auf Hoffmann (2007), eine Anleitung gegeben, wie man sich aus Internetdaten ein eigenes Korpus aufbauen kann.

Im dritten Kapitel stehen die methodischen Ansätze der Korpusanalyse im Vordergrund. Es wird Software wie *WordSmith Tools* vorgestellt, mit der man Korpora nach Morphemen, Worten oder Wortkombinationen durchsuchen kann. Dabei wird insbesondere erklärt, wie man Wortlisten für lexikalische Analysen und *keyword*-Analysen für die Beschreibung von Textsorten bzw. Schreibstilen verwenden kann. In Hinblick auf die Arbeit mit Konkordanzen wird darauf eingegangen, wie man eine Suche mit diesem Tool sinnvoll gestaltet, so dass man wirklich alle relevanten Belege findet und dabei nicht zu viele nicht relevante Belege von Hand wieder aus-sortieren muss. Diese Frage nach *precision* und *recall*, die sich im Kern um eine akkurate Definition der Variablen dreht, ist ein zentraler Punkt, der Studierenden am Anfang Schwierigkeiten bereitet. Auch wird beschrieben, welche Formen von Korpusannotation es gibt und wie diese eine Analyse der Daten unterstützen kann. Besonders wichtig in diesem Kapitel scheint mir der Abriss zu quantitativen Methoden, mittels derer man Häufigkeitsanalysen vornehmen und die statistische Validität der gefundenen Häufigkeiten überprüfen kann. Gerade dieser Teil ist gut verständlich und mit eindrucklichen Beispielen gestaltet, wobei auch die Grenzen und die Nichtanwendbarkeit bestimmter statistischer Tests nicht verschwiegen werden. Allerdings haben sich in den Tabellen zum *chi-square test* einige Fehler eingeschlichen, die in einer Neuauflage berücksichtigt werden sollten. Weitere Themen des Kapitels sind die gewinnbringende Verbindung von korpuslinguistischen Methoden mit anderen Methoden der Linguistik und das Erstellen eigener Korpora.

Kapitel 4 ist der *deskriptiven Korpuslinguistik* gewidmet. Hier wird anhand einschlägiger Studien demonstriert, wie korpuslinguistische Methoden in den verschiedenen Gebieten der Linguistik eingesetzt werden können. Es werden nacheinander korpuslinguistische Studien zur Wortschatzanalyse, Lexikogrammatik, Semantik, Syntax und Sprachwandelforschung vorgestellt, ebenso, wie sich korpuslinguistische Ansätze in der Soziolinguistik, Textlinguistik und Varietätenforschung gewinnbringend einsetzen lassen und wie diese sogar zu neuen Einsichten in der Prosodieforschung und in der linguistischen Theoriebildung führen können. In allen Teilabschnitten zu den einzelnen linguistischen Disziplinen werden die Grundbegriffe der jeweiligen Disziplin kurz vorgestellt. Es wird dargelegt, wie man eine korpuslinguistische Analyse in diesem Gebiet durchführen kann, z. B., was Kollokationen sind und wie man Kollokationen korpuslinguistisch bestimmen kann. Es folgen Beispielstudien, die bestehende Hypothesen in den Gebieten überprüfen und neue Erkenntnisse ans Licht bringen.

Kapitel 5 ist als ein Appell zu lesen, Korpuslinguistik verstärkt im Fremdsprachenunterricht einzusetzen, d. h. auch an der Universität Lehramtskandidaten als zukünftige Lehrer mit der Korpuslinguistik und ihrem sinnvollen didaktischen Einsatz beim Erlernen des Englischen vertraut zu machen. Das sechste Kapitel führt vor, wie korpuslinguistische Methoden darüber hinaus in den Übersetzungswissenschaften und der Computerlinguistik ihre Anwendung finden, aber auch in thematisch entfernten Gebieten wie den Kulturwissenschaften oder der forensischen Linguistik gewinnbringend eingesetzt werden können. Einerseits kann das Identifizieren von häufigen sprachlichen Routinen in einer neuen Textsorte oder in einer repräsentativen Datenmenge einer Sprachgemeinschaft helfen, kulturelle Identifikationsmarker zu finden, die der Sprachgemeinschaft selbst unter Umständen nicht bewusst sind – wie der Ausdruck von *Britishness* in Form von Kollokationen, die für das britische Englisch typisch sind (Mair 2007). Andererseits kann im Vergleich mit herkömmlichem Sprachgebrauch, wie er in einem *general corpus* manifestiert ist, und mittels eines Dokumentes zweifelsfreier Autorschaft überprüft werden, ob andere Dokumente vom selben Autor stammen, in dem Sinne auch, ob ein Täter zweifelsfrei überführt oder ein Angeklagter zweifelsfrei entlastet werden kann (Blackwell 2000). Im siebten und letzten Kapitel wird noch einmal auf die Bedeutung der Korpuslinguistik nicht nur als linguistische Methode, sondern auch als eigenständige Disziplin hingewiesen und ein Ausblick gegeben, wie sich die Korpuslinguistik in den nächsten Jahren weiterentwickeln wird.

Insgesamt lässt sich feststellen, dass der Autor mit seinem neuen Buch zur anglistischen Korpuslinguistik eine inhaltlich wie strukturell gelungene Einführung vorgelegt hat, die kompakt und übersichtlich die Wesenszüge der Korpuslinguistik skizziert und mit zahlreichen Beispielen den Gegenstand nicht nur gut erklärt, sondern nebenbei auch in bekannte Studien des Gebietes einführt. Literaturhinweise zur weiterführenden Lektüre, eine Liste von

Internetseiten, die von korpuslinguistischem Interesse sind, sowie Angaben, wie man sein eigenes Korpus aufbaut und wie man Schritt für Schritt eine eigene korpuslinguistische Analyse durchführt, leiten die Studierenden zum kreativen und eigenständigen Arbeiten an, und regen auch die Dozierenden zur Erstellung interessanter korpuslinguistischer Übungen für den Unterricht an. Dazu sind Fachbegriffe im laufenden Text fettgedruckt und auch über Querverweise und ein Sachregister jederzeit auffindbar, so dass die Studierenden jederzeit überprüfen können, ob sie den dargestellten Sachverhalt verstanden haben. Wünschenswert für den Einsatz im Englischunterricht wäre eine englischsprachige Ausgabe oder zumindest ein Register aller in diesem Gebiet gebräuchlichen englischen Termini im Anhang. Auch scheint mir das Kapitel zur Computerlinguistik etwas zu kurz geraten. Gerade durch die Weiterentwicklung von Parsern und Softwareprogrammen zur Erstellung von Konkordanzen in der Computerlinguistik ist auch eine Weiterentwicklung korpuslinguistischer Analyseverfahren zu erwarten, weshalb die Wechselwirkung dieser beiden Gebiete sicherlich eine genauere Untersuchung verdient.

#### Literatur

- Blackwell, Susan. 2000. Looking up 'look'. Discourse Markers in the *Bank of English*. In: John Kirk (Hg.). *Corpora Galore. Analyses and Techniques in Describing English*. Amsterdam: Rodopi. 3-16.
- Hoffmann, Sebastian. 2007. From Webpage to Mega-Corpus: the CNN Transcripts. In: Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (Hg.). *Corpus Linguistics and the Web*. Amsterdam: Rodopi. 69-85.
- Hunston, Susan. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Lemnitzer, Lothar & Heike Zinsmeister. 2006. *Einführung in die Korpuslinguistik*. Tübingen: Gunter Narr.
- Mair, Christian. 2007. Varieties of English around the World: Collocational and Cultural Profiles. In: Paul Skandera (Hg.). *Phraseology and Culture in English*. Berlin, New York: Mouton de Gruyter. 437-468.
- McEnery, Tony & Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Meyer, Charles F. 2002. *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.
- Mukherjee, Joybrato. 2001. Principles of Pattern Selection: a Corpus-based Case Study. *Journal of English Linguistics* 29, 295-314.
- Mukherjee, Joybrato. 2002. *Korpuslinguistik und Englischunterricht. Eine Einführung*. Frankfurt am Main u. a.: Peter Lang.
- Scherer, Carmen. 2006. *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.