

Zeitschrift für Interkulturellen Fremdsprachenunterricht

Lemnitzer, Lothar & Zinsmeister, Heike (2006), *Korpuslinguistik: Eine Einführung*. Tübingen: Narr. ISBN 3-8233-6210-0. 220 Seiten.

Rezensiert von **Maryse Nsangou-Njikam** (Université de Yaoundé/Kamerun)

Erschienen online: 1. April 2008

© *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 2008

In der linguistischen Germanistik hat die Arbeit mit authentischen Daten in den letzten Jahren zugenommen; jedoch fehlt(e) es an theoretischen Grundlagen, um diese Arbeit zu unterstützen. Das Buch von Lemnitzer/Zinsmeister ist der Versuch, diese Forschungslücke zu schließen. Es richtet sich in erster Linie an „Forscher der Linguistik, die empirisch die deutsche Sprache untersuchen wollen“ (4). Dieses Buch ist die erste Publikation zur Korpuslinguistik im deutschen Sprachraum, in der die „Chancen und Grenzen der Arbeit mit Korpora“ (4) grundlegend und einführend dargestellt werden. Dank des Aufbaus der Kapitel, die alle mit einer ausführlichen Literaturliste und ausgewählten Aufgaben versehen sind, eignet es sich sowohl für die Forschungsarbeit in der empirischen Linguistik als auch für das Selbststudium.

„Korpuslinguistik“ wird definiert als „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“ (9). Das Buch fängt mit einer Diskussion über Sinn und Zweck der Korpuslinguistik an (Kapitel 2), wobei die erkenntnistheoretischen Positionen der beiden großen Strömungen der Linguistik, der generativen Grammatik und des Kontextualismus, ausführlich dargestellt werden; die erstgenannte Richtung sei eher eine korpusbasierte Spracharbeit, bei der zweiten seien Korpusdaten ein notwendiger Bestandteil linguistischer Erkenntnis. Auch die zweifellos vorhandenen Mängel korpusbasierter Sprachdaten werden nicht verschwiegen (vgl. 24 ff), zugleich aber auch ihre Bedeutung als materielle Basis der linguistischen Untersuchungen hervorgehoben (z.B. im Bereich der Lexikographie und der Lexikologie).

Im dritten Kapitel werden die wichtigsten Merkmale linguistischer Korpora aufgelistet und der Unterschied zu anderen, nicht-linguistischen Datensammlungen klar gemacht. Der Begriff ‚Korpus‘ im Sinne der Korpuslinguistik wird definiert als „Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehrerer Sprachen. [...] Die Bestandteile des Korpus, die Texte oder Äußerungsfolgen, bestehen aus den Daten selbst sowie möglicherweise aus Metadaten, die diese Daten beschreiben, und aus linguistischen Annotationen, die diesen Daten zugeordnet sind“ (40).

Besonders interessant ist die Aufzählung der Funktionen der unterschiedlichen Datentypen (44 ff.), der methodischen Probleme, die auf verschiedenen Datenebenen vorkommen können, sowie die Antworten, die auf methodologische Fragen zum Umgang mit Korpora und zu Untersuchungsgegenstand und -ziel gegeben werden (57 ff.).

Linguistische Korpusdaten werden in der Regel auf Rechnern gespeichert und sind maschinenlesbar. Sie sind unterschiedlicher Natur und dienen unterschiedlichen Zwecken (auch nicht-linguistischen) (vgl. 41 ff.). Das vierte Kapitel des Buches zeigt die Wichtigkeit und den Nutzen der gängigsten Annotationsebenen an. Es wird auf die Grenzen und Probleme linguistischer Annotation hingewiesen; Tipps werden bei der Annotation eigener Korpora gegeben. Bei ambigen Korpora helfe die automatische Sprachverarbeitung besonders, vor allem an Korpusdaten mit linguistischen Interpretationen wie z.B. Wortarten oder syntaktische Phrasen (vgl. 62 und 79 ff.). Man kann am Ende nachvollziehen, warum Sprachwissenschaftler und Computerlinguisten die zeitaufwändige Arbeit der Annotation auf sich nehmen, denn diese dient u.a. der Wiederverwendung gespeicherter Korpora, die dadurch gut dokumentiert und an Standards orientiert sind. Außerdem: „Je stärker ein Korpus annotiert ist, desto kleiner ist normalerweise die Datenmenge“ (105). Bedauernd ist jedoch, dass hier nur schriftliche Daten in Anspruch genommen werden (z.B. das Stuttgart-Tübingen-Tagset u.a.); gespeicherte und annotierte Korpora gesprochener Sprache werden an dieser Stelle nicht erwähnt.

Das fünfte Kapitel ist der Typologie deutschsprachiger Korpora gewidmet. Eine systematische Klassifizierung von Korpora erlaubt es Forschern im Bereich der Korpuslinguistik einen wichtigen und interessanten Wegweiser in die Korpuslandschaft zu bekommen. Die Klassifizierung dieser Korpora wurde nach bestimmten systematisch beschriebenen und erklärten Kriterien durchgeführt, und zwar nach Funktionalität, Sprachenauswahl, Medium, Annotation, Größe, Persistenz, Sprachbezug und Verfügbarkeit (vgl. 103 ff.). Die aufgestellte Liste ist zwar nicht vollständig, aber breit und hilfreich genug, um eigene korpuslinguistische Projekte durchführen zu können. Hinweise auf existierende Korpora, die der vorliegenden Typologie zugrunde liegen, wurden durch eine Umfrage, durch Anschreiben geeigneter Personen und eine Internetanfrage im Forum ‚Gesprächsforschung‘ gewonnen. Interessant ist auch die Angabe einer speziellen Webseite (<http://www.lemnitzer.de/lothar/KoLi>), auf der die Dokumentation zu den Korpora zu finden ist.

Der praktische Nutzen korpuslinguistischer Arbeit wird im sechsten Kapitel am Beispiel der Darstellung korpusbasierter Forschungsarbeiten aus den Bereichen Orthographie, Wortbildung, Computerlinguistik, Syntax, Lexikologie/Lexikographie, Partikelforschung und Sprachlehre gezeigt. Grenzen und Vorzüge dieser Studien werden diskutiert. Ein besonderes Interesse korpuslinguistischer Forschung gelte dem Sprachgebrauch in den neuen Medien (u.a. Internet, E-Mail, Foren, Chatträume, Gästebücher), wobei sich die Frage stellt, ob es sich hier um überwiegend schriftliche oder mündliche Daten handelt. Am Ende dieses Kapitels, bei dem gute sowie schlechtere korpuslinguistische Arbeiten dargestellt wurden, werden die methodischen Tendenzen, die diesen Arbeiten zugrunde liegen, offenbart. Daraus leiten die Autoren „Empfehlungen für ein gutes methodisches Arbeiten“ ab (170). Sie schließen ihr Buch mit Interviews von ausgewählten Linguisten über ihre Erfahrungen mit Korpora. Festzuhalten bleibt: „Um Korpora sinnvoll benutzen und ausnutzen zu können, bedarf es deshalb einer gründlichen Einarbeitung in die jeweiligen Recherche- und Analysewerkzeuge“ (193). Außerdem sollte man sich „vor der Korpusrecherche darüber klar sein, ob man das Korpus befragt, um die Bestätigung einer Vorannahme zu erhalten oder auch um geeignete Belege zu finden“ (192), wobei jede Korpusanalyse von ziemlich gutem linguistischen Vorwissen geleitet werden sollte, „damit man überhaupt weiß, wonach man guckt“ (194). Die Autoren schließen das gesamte Studienbuch mit einem Glossar ab.

Insgesamt eignet sich das Buch von Lemnitzer/Zinsmeister nicht nur vorzüglich, um sich mit den großen Linien korpuslinguistischer Forschung vertraut zu machen, sondern es zeigt auch an einem eindrucksvollen Spektrum von Theorien und praktischen empirischen Forschungsarbeiten, wie mit sprachlichen Korpora umgegangen werden sollte und wie unumgänglich sie für die Sprachwissenschaft geworden sind. Es hält auf eindrucksvolle Weise, was es verspricht, denn es bietet einen wirklich breit gefächerten Abriss des aktuellen Forschungsstandes im Bereich Korpuslinguistik.