

**Kratochvílová, Iva/Wolf, Norbert Richard (Hrsg.) (2010): Kompendium Korpuslinguistik. Eine Bestandsaufnahme aus deutsch-tschechischer Perspektive. Heidelberg: Winter. 315 S. und Register. ISBN 978-3-8253-5793-1.**

Die Korpuslinguistik hat im vergangenen Jahrzehnt dank der immer vollkommeneren Software und Technologien schnelle Entwicklung genommen und an Attraktivität gewonnen. Von der immer größeren Verbreitung der Korpuslinguistik zeugen auch ihre kontrastiv ausgerichteten Fragestellungen, der sich durchsetzende korpusbasierte Sprachvergleich und der wachsende Gebrauch von Parallelkorpora in der Übersetzungswissenschaft. Seinen Beitrag zur Diskussion über diese Problembereiche leistet auch der 2010 erschienene Band, der von der deutsch-tschechischen Perspektive ausgeht und es ermöglicht, die Korpuslinguistik als festen Bestandteil auch der tschechischen germanistischen Sprachwissenschaft wahrzunehmen. Diese vor allem deutsch-tschechische Perspektive lag einer Anfang Oktober 2009 im Bildungszentrum Sambachshof bei Bad Königshofen und an der Universität Würzburg veranstalteten internationalen Konferenz zugrunde, aus der das Kompendium hervorgegangen ist.

I. Kratochvílová und N. R. Wolf haben einen Band herausgegeben, der einschließlich ihres Vorworts 27 Texte umfasst. An diese sind fünf Poster angeschlossen (davon vier in Englisch und eines in Deutsch), die den Leser eben an die Konferenzwurzeln dieses Werkes erinnern. Bezüglich ihrer Textsortenzugehörigkeit repräsentieren die 'Poster-Artikel' aber etwas Anderes als erwartbare völlig kohärente Texte. Die globale thematische Orientierung im Band wird dafür von einem Register im Umfang von mehr als 150 Stichwörtern erleichtert.

Im Folgenden habe ich nicht vor, mich in ausgezogener Weise mit allen 27 Beiträgen zu beschäftigen. Ich möchte mich eher auf tragfähige allgemeinere Zusammenhänge konzentrieren, die für das rezensierte Kompendium einen gemeinsamen Nenner darstellen und die sonst ohne das Lesen dieses aufschlussreichen Werkes etwas unschärfer geblieben wären. Zu diesen Zusammenhängen komme ich aber erst nach einem kurzen Überblick über den ganzen Band zurück.

Die Aufsätze sind in drei Ganze gegliedert. Der erste Block – mit dem Titel ‚Grundsätzliches‘ – konzentriert sich vor allem auf methodologische Fragestellungen. Deutsche Sprachwissenschaftler setzen sich hier in ihren Texten mit sowohl theoretischen, als auch anwendungsbezogenen Fragestellungen auseinander. Elektronische Korpora sind Produkte moderner Technologien und wachsender Ansprüche auf die Datenerhebung und den immer bequemeren Umgang mit Daten. So werden in diesem

ersten Block außer den Entstehungsprozessen von Korpora geschriebener Sprache im IDS (W. Schütte) auch wünschenswerte Aspekte der gesprochen-sprachlichen Korpora des gegenwärtigen Deutsch (J. Schwitalla) ebenso wie z. B. der diachronen (R. U. Schmid) oder der dialektbezogenen Korpora (R. Zimmermann) diskutiert. Daran, wie unentbehrlich die elektronischen Korpora für moderne Lexikografie sind, zweifelt heutzutage niemand mehr. Die Überzeugungskraft der Texte von H. Wellmann, A. Klosa oder S. Krome ist in dieser Hinsicht klar. Das soll aber gleichzeitig nicht bedeuten, dass die korpusbasierte Lexikografie völlig problemlos wäre.

Die Texte der Autorinnen im zweiten Block beruhen auf dem Troppauer DeuCze-Korpus und widmen sich primär konkreten morphologischen Variablen – den Übersetzungsmöglichkeiten von *lassen* ins Tschechische (V. Kotůlková), der *würde*-Form in narrativen Texten (I. Kratochvílová) oder den Komposita mit Partizipien als zweiter Konstituente (G. Rykalová). Methodologisch gut vergleichbar sind mit den gerade genannten Analysen auch die von H. Peloušková (Es-Konstruktionen im Deutschen und ihre Äquivalente im Tschechischen) und T. Káňa, der die synthetischen substantivischen Diminutive in beiden Sprachen vergleicht. Indem sie sich aber nicht auf das DeuCze-Korpus stützen, sondern auf das mehrsprachige Parallelkorpus InterCorp (<http://ucnk.ff.cuni.cz/intercorp/>) bzw. noch auf andere Korpora, sind die Texte von H. Peloušková und T. Káňa erst im nächsten Kapitel geraten.

Dieses dritte Kapitel besteht meistens aus Beiträgen tschechischer Forscher, die sowohl konkrete Fallstudien als auch Beschreibungen von Projektdokumentationen anbieten. Gleich drei Fallstudien sind dabei phraseologischer Orientierung: H. Bergerová zeigt an der Wendung *sein Mütchen kühlen* die Unentbehrlichkeit der korpusbasierten Forschung für die möglichst adäquate (d. h. auch semantische und morphosyntaktische) lexikografische Beschreibung von Phraseologismen. Zum gleichen Schluss kommen auch E. Ciešlarová (Phraseologismen mit dem Begriff ‚Gegenstand‘) und J. Malá (Phraseologismen, die zum Ausdruck von Emotionen dienen wie z. B. *sich die Haare raufen* oder *die Nase rümpfen*). Den Emotionen in ihrer metasprachlichen Erfassung in Romanen widmet sich weiter auch L. Vaňková. Die Herausforderungen im Bereich der Übersetzung thematisiert J. Kovář. Das Fragezeichen in der Überschrift seines Artikels ‚Korpora – endlich eine linguistische Fachhilfe für die literarischen Übersetzer?‘ lässt ahnen, dass die Antwort darauf nicht eindeutig sein muss.

Die Diskussion über die allgemeineren methodologischen Punkte des ganzen Bandes lässt sich von verschiedenen Ecken führen. Eine der Möglichkeiten bieten auf S. 9 in der Einleitung die Herausgeber

an, wo als Typen von sprachlichen Daten für sprachwissenschaftliche Zwecke vier Quellen identifiziert und bewertet sind: Introspektion, Informantenbefragung, Belegsammlung ('Zettelkasten') und Korpus. Als die relativ beste Quelle sind die Korpora ausgewertet.

Die Verlässlichkeit der Introspektion wird besonders in der Korpuslinguistik aus gut bekannten Gründen kritisiert. Diese Introspektion fällt aber mit der sog. systematischen Selbstbeobachtung nicht zusammen (Systematic Self-Observation, mehr dazu z. B. in Rodriguez/Ryave 2002). Eine der Schwierigkeiten der Fragebogenmethode besteht u. a. darin, dass die Befragten ohne die Assistenz der Forscher nicht adäquat verstehen müssen, wonach sie gefragt werden. Diese Gefahr kann durch Interviews vermieden werden (es muss jedoch nicht immer gelingen).

Was aber für die Sprachwissenschaft neben den Korpora als sehr wertvolle Datenquelle in Frage kommt, sind teilnehmende Beobachtungen. Wenn im Idealfall ein Forscher als *Insider* auftritt, ist die Qualität/Authentizität solcher Daten praktisch nicht zu übertreffen. In dieser Hinsicht hat die Ethnomethodologie – auch wegen der interaktiven Komponente – je nach der konkreten Fragestellung kaum Konkurrenz. Gute Ethnomethodologen bleiben mit ihren Interpretationen prinzipiell ziemlich vorsichtig und sprechen ihren Ergebnissen die Indexikalität nicht ab. Irgendwelches Herantragen von draußen entwickelten Kategorisierungen wird abgelehnt (vgl. Lamnek 2010).

Methodologisch wichtig im Sinne der Trennlinie zwischen *korpus-basiert* vs. *korpus-gesteuert* ist im rezensierten Band aber nicht nur die gemeinsame Einleitung von I. Kratochvilová und N. R. Wolf (neben der diskutierten Passage auf S. 9 auch S. 10 oder der selbständige Aufsatz von N. R. Wolf). Auch I. Mindt geht in den grundsätzlichen Bereichen *Methodologie* und *Anwendung* auf die Möglichkeiten der beiden Ansätze ein. (In diesem Kontext ist dabei nicht entscheidend, dass sie als Anwendungsbereich die Variable der englischen *dass*-Sätze (*that*-clauses) wählt). Zu welchen Ergebnissen kann die Forschung gelangen, wenn auf elektronisch verarbeitete Mengen von Daten bestimmte Kategorien von außen herangetragen werden, von denen die Analysen so wie so abhängen (müssen)? Komplizierter ist es dagegen, sich z. B. die oben aufgeworfene Kategorisierung als Ziel zu setzen und sie erst aufgrund gut durchdachter Suchanfragen zu entwickeln. Beides lässt sich jedoch in Abhängigkeit von konkreten Forschungsfragen verteidigen: Wenn die Frage lautet *Sind die und die konkreten Varianten belegbar?* oder *Wie wird die und die Struktur in einem Parallelkorpus übersetzt?*, so kann man natürlich dank der Korpora einiges Nützliche erfahren. Eine solche Analyse liefert *eine Basis für Datenauswertungen*. Wie diese

*interpretiert* werden, stellt allerdings noch eine andere Geschichte dar ebenso wie der eigentliche Umgang mit den Interpretationen in Praxis.

Das Kompendium erlaubt es also nicht, weitreichende methodologische Fragen der Sprachwissenschaft in Vergessenheit geraten zu lassen: Welches Verhältnis gibt es zwischen einem *Korpus* und einer *Einzelsprache*? Neigt man – besonders im Falle der sehr umfangreichen Korpora – bereits dazu, beide Konstrukte als identisch wahrzunehmen, oder wird das potenzielle Gleichheitszeichen zwischen *Korpus* und *Einzelsprache* doch noch kritisch reflektiert? Wird z. B. selbst mit den korpus-gesteuerten Ergebnissen – geschweige denn von den korpusbasierten – so umgegangen, als ob sich die Schlüsse ohne Weiteres auf ganze Einzelsprachen beziehen ließen? Wie oft kommen solche Analysen vor, die ihre Schlüsse vorsichtig „nur“ auf die Korpora beziehen, ohne sie stillschweigend auf die Sprachen zu extrapolieren? Diese Gefahr der nicht immer angemessenen Ausdehnung der Aussagewerte der korpus-basierten bzw. auch -gesteuerten Schlüsse ist im rezensierten Band nämlich nicht immer deutlich artikuliert.

Das Kompendium Korpuslinguistik bringt aber außer wertvollen methodologischen Anregungen und ganz praktischen Fallstudien noch einen wesentlichen Wert mit sich, der gleich beim Lesen der ersten Beiträge deutlich zu werden beginnt. Da es nicht besonders umfangreiche, nicht mit Theorie überladene und gleichzeitig doch thematisch abwechslungsreiche Artikel umfasst, bekommen die Leser einen Band, der die Attraktivität der Korpuslinguistik bzw. das Interesse an Linguistik überhaupt stärken kann – besonders wenn die Schritte auf dem Wege zur Linguistik über die Computer-Technik und Informationstechnologien führen. Auch der attraktive Titel des ganzen Bandes wird dazu beitragen.

Vit DOVALIL

Literatur:

- LAMNEK, Siegfried (2010): *Qualitative Sozialforschung*. Basel; Weinheim.  
 RODRIGUEZ, Noelle/Ryave, Alan (2002): *Systematic Self-Observation* (Qualitative research methods series – Volume 49). London.