

**Tomáš Káňa: *Sprachkorpora in Unterricht und Forschung*  
 DaF/DaZ. Brno: Masarykova univerzita, Pedagogická fakulta  
 2014, 244 S. ISBN 978-80-210-6994-7**

*Peter Ďurčo*

Sprachkorpora sind heute Informationsquellen, die Sprachwissenschaftlern für die Erforschung der Sprachen und für den Vergleich von Sprachen als ein unentbehrliches Instrument zur Verfügung stehen und man kann bei einer effizienten Erforschung von Sprachen sowie bei der Datenerhebung um Korpora gar nicht herumkommen. Sprachkorpora kann man jedoch auch für eine effektive Vermittlung jeder Sprache nutzen. Die schnelle Verbreitung der Computertechnik und der schnelle Zugang zu Informationen hat auch stark die immense Entwicklung im Bereich der Sprachverarbeitung beeinflusst und bedeutete zudem eine rasante Entwicklung der Korpuslinguistik. Deswegen erscheinen als große Defizite die ungenügende Vermittlung der Kenntnisse und das Fehlen der praktischen Fähigkeiten im Umgang mit korpuslinguistischen Tools bei Studierenden der philologischen Fachrichtungen. Die als Lehrbuch konzipierte Publikation hat eine gute Chance, diese Lücke im Lehrwerkangebot zu füllen.

Der Autor der vorliegenden Publikation erklärt seine Motivation zur Arbeit an diesem Buch durch die Tatsache, dass viele Studierende aber auch Wissenschaftler/-innen und fast alle Lehrenden die Möglichkeiten der heutigen Instrumente für Forschung und Lehre nicht kennen oder kaum verwenden. Die Gründe dafür sieht der Autor im fehlenden Wissen von der Existenz dieser Instrumente oder auch in der Angst vor ihrer Komplexität, ihrer vermeintlichen oder auch tatsächlichen Unübersichtlichkeit und der Aufwendigkeit ihrer Nutzung.

Sein Buch strebt mehrere Ziele an. Vor allem will es einen Überblick über die wichtigsten Begriffe der Korpuslinguistik, sowie der Typologie und Eigenschaften der größten Korpora geben. Bei der Auswahl der präsentierten Tools beschränke sich der Autor auf diejenigen Instrumente, die dem Bereich DaF/DaZ nützlich sein könnten, kontrastive Verfahren unterstützen und ermöglichen und Schüler/-innen, Studierende und Lehrkräfte ansprechen können. Ein weiteres Ziel war, den Umgang mit Korpora einfach und transparent darzustellen und so die potentielle Angst vor den technisch oft aufwendigen Instrumenten abzubauen. Zu innovativem Ansatz in der Publikation gehört die Herstellung der Verbindung zwischen der Korpuslinguistik und DaF/DaZ. Die konkrete Korpusarbeit wird anschaulich mit Fallstudien zu einigen Phänomenen der deutschen Sprache illustriert und die Monographie trägt damit auch zur kontrastiven quantitativen Forschung der deutschen Sprache bei. Der Autor bezeichnet selbst seine Publikation auch als ein „Kochbuch“ für die Korpusarbeit.

Die Publikation ist in acht Kapitel eingeteilt. Nach einer kurzen Behandlung der Zugänge zur Sprache über verschiedene Suchmaschinen und einem Vergleich der Korpustools in gängigen Programmen und Korpora erklärt der Autor im zweiten Kapitel die wichtigsten korpuslinguistischen Begriffe.

Die Typologie der Korpora von Tomáš Káňa basiert auf verschiedenen Kriterien, wie Stadium der gespeicherten Sprache (synchrone, diachrone und historische Korpora), Medium (Korpora der geschriebenen und gesprochenen Sprache), Repräsentativität (allgemeine, spezifische Korpora und Varietätenkorpora; ausgewogene und opportunistische Korpora), Sprache (monoguale, bilinguale und multilinguale Korpora mit der wichtigen Differenzierung zwischen Vergleichs- und Parallelkorpora). Dieses Kapitel enthält auch kurze Erläuterung der technischen Eigenschaften zu externen Metadaten und internen Zusatzinformationen der linguistischen Aufbereitung (Tokenisierung, Lemmatisierung, Tagging, Parsing und Markierung der prosodischen Eigenschaften in Korpora der gesprochenen Sprache).

Im vierten Kapitel behandelt der Autor die Problematik der Relevanz der Korpora für Fremdsprachenunterricht und stellt sich die Fragen, ob Lehrer und Schüler überhaupt Korpora brauchen und wenn ja, von welchen sie am meisten Gebrauch machen können und welche Korpustools und Korpuseigenschaften am ehesten einsetzbar und hilfreich für den Fremdsprachenunterricht sind. In diesem Kapitel finden wir in Auswahl kurze Informationen über einige nationale Korpora für Englisch, Französisch, Griechisch, Italienisch, Kroatisch, Polnisch, Russisch, Slowenisch, Slowakisch, Spanisch, Tschechisch, Türkisch und Ungarisch.

Die Korpora für Deutsch werden detailliert dargestellt. Der Autor konzentriert sich auf die bekanntesten Korpora für Deutsch: Wortschatzportal der Universität Leipzig, DWDS-Korpus der Berlin-Brandenburgischen Akademie der Wissenschaften, Deutsches Referenzkorpus (DeReKo) des Instituts für deutsche Sprache in Mannheim und C4-Korpus (gemeinsames Korpus des Digitalen Wörterbuchs der deutschen Sprache des 20. Jahrhunderts (DWDS), des Austrian Academy Corpus (AAC), des Korpus Südtirol und des Schweizer Textkorpus (CHTK). Neben der Präsentation der Funktionalitäten und Optionen präsentiert der Autor auch konkrete Abfragemöglichkeiten in diesen Korpora mit der Präsentation der formalen Sprachen und auf zahlreichen Beispielen für die Suchanfragen erklärt er die Möglichkeiten der Nutzung dieser Tools.

Andere Korpora der geschriebenen Sprache für Deutsch werden nicht näher erwähnt. Dies erklärt der Autor dadurch, dass die meisten von ihnen (deWaC, AAC, Korpus Südtirol, Schweizer Textkorpus) in andere Projekte bzw. Großkorpora integriert sind, weitere sind schwer zugänglich (TenTen corpora) oder er hält sie für DaF/DaZ nur wenig relevant.

Auch im Bereich der Korpora mit gesprochenem Deutsch existieren mehrere Quellen, z. B. Bayerisches Archiv für Sprachsignale (BAS), Österreichisches Aussprachewörterbuch/Österreichische Aussprachedatenbank (adaba), Korpus Gesprochene Sprache beim DWDS, die Datenbank für gesprochenes Deutsch (DGD) oder das GeWiss-Korpus der gesprochenen Wissenschaftssprache und verschiedene Kleinkorpora mit unterschiedlichen Zielsetzungen. Der Autor widmet sich näher der Datenbank für gesprochenes Deutsch (DGD) (<http://dgd.ids-mannheim.de/>) und dem DWDS-Korpus Gesprochene Sprache (<http://retro.dwds.de/>), weil die anderen Korpora seiner Meinung nach klein sowie stilistisch und regional nicht ausgewogen und für die gesprochene deutsche Sprache nicht genug repräsentativ sind.

Das Bild über die Korpuslandschaft ergänzen Informationen über Korpora mit historischen deutschen Texten. Die historischen und diachronen Korpora können den Germanistikstudenten/-innen nicht nur als Demonstration der Vielfalt deutscher Sprache dienen, sondern auch im Studium behilflich sein. In kurzen Skizzen werden die historischen Korpora vorgestellt, die praktisch alle Perioden der Entwicklung der deutschen Sprache abdecken, beginnend mit dem Korpus des Althochdeutschen (DDD) (<http://www.deutschdiachron-digital.de/>), des Mittelhochdeutschen (Mittelhochdeutsche Begriffsdatenbank MHDBDB) (<http://mhdbdb.sbg.ac.at/>), des Frühmittelhochdeutschen (Bonner Frühneuhochdeutschkorpus (FNHD) (<http://www.korpora.org/Fnhd/>) und mit dem diachronen Archiv HIST im Deutschen Referenzkorpus DeReKo (<https://cosmas2.ids-mannheim.de/cosmas2-web/>), das Texte aus der zweiten Hälfte des 17. Jahrhunderts bis 1972 enthält.

Von der Vielfalt der Parallelkorpora, wo auch die deutsche Sprache in unterschiedlichem Umfang vertreten ist, hat der Autor nur zwei gewählt, und zwar das OPUS-Korpus (<http://opus.lingfil.uu.se/>) und InterCorp (<http://ucnk.ff.cuni.cz/intercorp/>), die für eine seriöse kontrastive Arbeit viele Belege aus unterschiedlichen Texten zumindest die grundlegenden statistischen Werkzeuge und Recherchemöglichkeiten anbieten. Detailliert wird vor allem die Arbeit mit dem InterCorp beschrieben. Auf Grund von zahlreichen Beispielen wird sehr anschaulich die Suchanfragesprache erklärt, die das System von Regulären Ausdrücken verwendet.

Informationen über Konkordanzprogramme und korpusähnliche Instrumente, wie TextSTAT, Linguee, ADABA und ParZu, sind im fünften Kapitel enthalten. Sie existieren auch

unabhängig von Korpora, viele von ihnen werden als Free- oder Shareware ins Internet gestellt. Es handelt sich um Softwaretools, die in einem Text oder in einer Textsammlung nach unterschiedlichen sprachlichen Elementen suchen können und diese in Form von Konkordanzzeilen anzeigen. Nach der Meinung des Autors ist ihre Nutzung jedoch beschränkt auf kleine Experimente mit eigenen Kleinkorpora, denn ihnen fehlt eine substanzielle Korpuseigenschaft: entweder basieren sie auf einer zu kleinen Textsammlung oder sie ermöglichen keine Suche nach sprachlichen Elementen, bzw. können keine statistischen Angaben liefern. Die Ausnahme bildet das Portal Linguae.de, das als eine Kombination aus einem Wörterbuch und einer Suchmaschine mit Textsammlungen aus mehreren Millionen zwei- und mehrsprachigen (fast ausschließlich Internet-)Texten besteht und als Translatortool benutzt werden kann.

Der besondere didaktische Mehrwert dieser Publikation liegt im sechsten Kapitel mit Fallstudien zur Arbeit mit Korpora. Es werden Suchstrategien und methodische Verfahren zur Suche nach verschiedenen linguistischen Phänomenen präsentiert, wie z. B. Suche nach der „richtigen“ Aussprache, Suche nach dem „richtigen“ Schriftbild, Suche nach Flexionsformen, Ermittlung von Wortfamilien und Wortbildungsarten, Analysen von Verben und Verbalkomplexen, Präpositionen, Kollokationen, syntagmatischen Mustern, Anglizismen, etc.

Das siebte Kapitel enthält eigene statische Berechnungen des Autors zur Frequenz einzelner Wortarten, Wortformen und Wortelemente auf Grund der Daten des Korpus InterCorp\_de (Gesamtkorpus) sowie auf Daten des deTenTen-Korpus.

Das letzte achte Kapitel enthält Korpusabfragehilfen - einen sehr nützlichen praktischen Überblick über die Einstellungen der Suchoptionen in Korpora, über Tagsets, Tag-Kürzel, Tastenkürzel, Shortcuts und CQL-Abfragen im InterCorp, die für die Studierenden bei der praktischen Arbeit sehr hilfreich sein können.

Den Schluss dieser nützlichen Publikation bildet die Literaturliste zu aktuellsten Publikationen im Bereich der Korpuslinguistik und Index der korpuslinguistischen Terminologie.

Das Manko in der universitären Lehre im Bereich der Korpuslinguistik ist für Studierende der philologischen Fachrichtungen das schon eingangs erwähnte Fehlen einführender Lehrwerke zur Korpuslinguistik. Mit der Publikation von Tomáš Kaňa wird dieses große Manko in einem sehr wichtigen Aspekt der Korpuslinguistik behoben.