

Korpuslinguistik als Innovation im DaF-Studium und Korpora im DaF-Unterricht. Fallbeispiel Trnava (Slowakei)

Korpusno jazykoslovje kot inovacija pri študiju nemščine kot tujega jezika in korpusi pri pouku nemščine kot tujega jezika. Študija primera (Slovaška)

Corpus linguistics as an innovation in study field DaF and corpora in teaching DaF. Case study Trnava (Slovakia)

PETER ĎURČO

Univerzita sv. Cyrila a Metoda v Trnave, Námestie J. Herdu 2,
SK 917 01 Trnava, durco@vtronk.net

DOI <https://doi.org/10.18690/978-961-286-019-6.3>
ISBN 978-961-286-810-9

In der universitären Lehre für Studierende der philologischen Fachrichtungen in der Slowakei besteht u. E. ein großes Manko in der Vermittlung der Informationen über die neuesten Technologien bei der Verarbeitung von sprachlichen Daten, was direkt mit dem Bereich der Korpuslinguistik zusammenhängt. Dies steht in krassem Gegensatz zu der Situation z. B. in Deutschland, wo an vielen Universitäten die Kurse zur Korpuslinguistik seit längerer Zeit angeboten werden. Man kann natürlich fragen: Brauchen DaF/DaZ-Lehrende und Studierende überhaupt Korpora? Wenn ja, wozu? Von welchen Korpora könnten sie am meisten Gebrauch machen? Welche Korpustools und Korpuseigenschaften sind am ehesten einsetzbar und hilfreich im Studium und DaF/DaZ-Unterricht? Solche und noch viele andere Fragen kann man im Zusammenhang mit der rasanten Entwicklung im Bereich der sprachlichen Ressourcen und Quellen und dem immensen technologischen Fortschritt im Bereich von verschiedenen Tools für NLP (Natural Language Processing) formulieren.

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. Die Korpuslinguistik zeigt, dass Sprache sinnvoll durch automatisierte Untersuchung großer Textmengen erforscht werden kann. Die

Korpuslinguistik kann relevante und zuverlässige Erkenntnisse über Sprache liefern. Anhand realer Anwendungen und mit Hilfe von mathematisch-statistischen, musterorientierten, induktiven und datengesteuerten Verfahren kann sie zu statistisch tragbaren Aussagen über komplexe Zusammenhänge, wie z. B. Worthäufigkeitsstatistiken und ihrer Kookkurrenzen, gelangen. Somit kann die Korpuslinguistik auch der Verifizierung und/oder Falsifizierung von linguistischen Theorien dienen. Die Korpuslinguistik, als eine induktive/empirische Methode zum Gewinn von Wissen über die Sprache, wo man eine Theorie nach der Beobachtung von möglichst vielen Einzelbeispielen ohne vorausgesetzte Annahmen über vorhandene sprachliche Gesetzmäßigkeiten aufstellt, steht in direktem Kontrast zu der traditionellen deduktiven Methode, die auf der Introspektion der Wissenschaftler basiert. Beide Herangehensweisen soll man aber als komplementär betrachten, weil beide grundlegend unterschiedliche Teilaspekte der Linguistik und der Sprache betrachten und verschiedene alternative Ansätze anwenden.

Sprachkorpora kann man jedoch auch für eine effektive Vermittlung jeder Sprache nutzen und heutzutage existieren viele Internetseiten, die sich mit der Thematik »Korpuslinguistik« aus der Sicht ihrer Vermittlung im Hochschulunterricht in Form von Online-Lernmodulen befassen (s. die Links im Anhang). Die schnelle Verbreitung der Computertechnik und der schnelle Zugang zu Informationen hat auch die immense Entwicklung im Bereich der Sprachverarbeitung stark beeinflusst und trug zudem zu einer rasanten Entwicklung der Korpuslinguistik bei. 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. Aus diesem Grunde haben wir am Lehrstuhl für Germanistik an der Universität des Hlg. Kyrill und Method in Trnava als einziges fremdsprachiges Institut in der Slowakei das Fach Korpuslinguistik bereits im 3. Semester für die Bakkalaureats-Studenten eingeführt, um diese Lücke im Lehrangebot zu füllen. Einen weiteren Aufbaukurs absolvieren dann auch Doktorats-Studenten.

Die primäre Motivation für diese Innovation lag in der Tatsache, dass viele Studierende, aber auch Wissenschaftler und Wissenschaftlerinnen und fast alle Lehrenden die Möglichkeiten der heutigen Instrumente für Forschung und Lehre nicht kennen oder kaum verwenden. Die Gründe dafür sehen wir im fehlenden Wissen von der Existenz dieser Instrumente, oder auch in der Angst vor Komplexität, vor der vermeintlichen oder auch tatsächlichen Unübersichtlichkeit und der Aufwendigkeit der Nutzung von Sprachkorpora, was sich zu Beginn jedes Kurses überraschenderweise immer wieder bestätigt.

Der Kurs strebt mehrere Ziele an. Vor allem will er einen Überblick über die wichtigsten Begriffe der Korpuslinguistik, sowie der Typologie und Eigenschaften der größten Korpora geben. Bei der Auswahl der präsentierten Tools beschränken wir uns auf diejenigen Instrumente, die dem Bereich DaF/DaZ nützlich sein könnten, kontrastive Verfahren unterstützen und Studierende ansprechen können. Ein weiteres Ziel ist, den Umgang mit den Korpora einfach und transparent darzustellen und so die potentielle Angst vor technisch oft aufwendigen Instrumenten abzubauen. Die konkrete Korpuserarbeit wird anschaulich zuerst mit der Formulierung konkreter Suchanfragen für verschiedene linguistische und sprachliche Phänomene zugänglich gemacht und endet mit eigenen Fallstudien der Studenten zu ausgewählten Phänomenen der deutschen Sprache. In den Seminararbeiten konzentrieren sich Studenten auf die Analyse des kollokationellen Verhaltens der Wörter; ihre Aufgabe ist sodann die Eruierung der statischen und dynamischen Daten aus Lexika und Korpora und Erstellung sowie Interpretation der Kollokationsprofile von untersuchten Lexemen.

Nach einer kurzen Behandlung der wichtigsten korpuslinguistischen Begriffe werden Zugänge zur Sprache über verschiedene Suchmaschinen und verschiedene Korpustools in gängigen Programmen und Korpora erklärt. Im Weiteren werden die Studenten mit der Typologie der Korpora vertraut gemacht, die auf verschiedenen Kriterien basiert, 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 (monolinguale, bilinguale und multilinguale Korpora mit der wichtigen Differenzierung zwischen Vergleichs- und Parallelkorpora). Diese Unterrichtseinheit enthält auch kurze Erläuterung der technischen Eigenschaften zu externen Metadaten und internen Zusatzinformationen der linguistischen Aufbereitung (Tokenisierung, Lemmatisierung, Tagging, ev. Parsing und externe Annotation der Texte in Korpora). Außerdem steht den Studenten die relevante Literatur zum Thema Korpuslinguistik zur Verfügung (s. Literaturliste zu aktuellen Publikationen im Bereich der Korpuslinguistik).

In einem weiteren Modul wird die Problematik der Relevanz der Korpora für Fremdsprachenunterricht behandelt und es werden Fragen besprochen, von welchen Quellen sie am meisten Gebrauch machen können und welche Korpustools und Korpuseigenschaften am ehesten einsetzbar und hilfreich für den Fremdsprachenunterricht sind. In diesem Modul werden in Auswahl auch Informationen über das Nationale Korpus für Slowakisch vorgestellt, wobei

die Studenten auch den Umgang mit dem Corpus-Browser Bonito trainieren (s. <http://bonito.korpus.sk/>).

Die Korpora für Deutsch werden detailliert dargestellt. Wir konzentrieren uns auf die bekanntesten: *Wortschatzportal* der Universität Leipzig, *DWDS-Korpus* der Berlin-Brandenburgischen Akademie der Wissenschaften, *Deutsches Referenzkorpus* (DeReKo) des Instituts für deutsche Sprache in Mannheim. Neben der Präsentation der Funktionalitäten und Optionen werden konkrete Abfragemöglichkeiten in diesen Korpora mit der Präsentation der formalen Sprachen vorgestellt und anhand zahlreicher Beispiele für die Suchanfragen werden die Möglichkeiten der Nutzung dieser Tools erklärt.

Andere Korpora der geschriebenen Sprache für Deutsch werden nur kurz vorgestellt, wie z. B. *deWaC*, *AAC*, *Korpus Südtirol*, *Schweizer Textkorpus*, *TenTen-Korpora*, die erst im Aufbaumodul für Doktoranden näher behandelt werden.

Die im Bereich der Korpora mit gesprochenem Deutsch existierenden 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 werden nur als Quellenangaben vorgestellt, jedoch ohne die eigentliche Korpusarbeit zu trainieren.

Der Schwerpunkt der praktischen Arbeit der Studenten liegt im Erlernen verschiedener Suchstrategien und methodischer Verfahren zur Suche nach verschiedenen linguistischen Phänomenen, wie z. B. zur Suche nach dem »richtigen« Schriftbild, zur Suche nach Flexionsformen, zur Ermittlung von Wortfamilien und Wortbildungsarten, zu Analysen von einzelnen Wortarten und Wortkomplexen, zu Kollokationen und Phraseologismen, zu syntagmatischen Mustern, zu sinnverwandten Wörtern etc.

Ein wichtiger Bestandteil der Arbeit liegt in der Analyse, Auswertung und Interpretation der statistischen Berechnungen zur Frequenz einzelner Wortarten, Wortformen, Wortelemente und Kookkurrenzen auf Grund der gewonnenen Daten. Abschließend erstellen die Studenten selbständig Kollokationsprofile von ausgewählten Lexemen.

Im Aufbaumodul für Doktoranden werden Informationen über Korpora mit historischen deutschen Texten ergänzt. Die historischen und diachronen

Korpora können den Germanistikstudierenden nicht nur als Demonstration der Vielfalt deutscher Sprache dienen, sondern auch im Studium behilflich sein. Die historischen Korpora werden in kurzen Skizzen 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 enthalten.

Im Aufbaukurs für Doktoranden wird vor allem die Arbeit mit *TenTen-Corpora* und dem Corpus-Tool *SketchEngine* trainiert. Von der Vielfalt der Parallelkorpora und der vergleichbaren Korpora, wo auch die deutsche Sprache in unterschiedlichem Umfang vertreten ist, bekommen die Studenten detaillierte Informationen auch erst im Aufbaumodul für Doktoranden, und zwar über das *OPUS-Korpus* (<http://opus.lingfil.uu.se/>), *InterCorp* (<http://ucnk.ff.cuni.cz/intercorp/>) und *Aranea-Corpora* (http://sketch.juls.savba.sk/aranea_about/), 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* und *Aranea-Web-Corpora* und Möglichkeiten ihrer Nutzung besonders bei der kontrastiven Analyse behandelt. Auf Grund von zahlreichen Beispielen wird sehr anschaulich die Suchanfragesprache erklärt, die das System von Regulären Ausdrücken (<http://www.regular-expressions.info/>) verwendet.

Die Studenten im Aufbaukurs werden auch über Konkordanzprogramme und korpusähnliche Instrumente, wie *TextSTAT*, *Linguee* informiert, die unabhängig von Korpora existieren und viele von ihnen als Free- oder Shareware ins Internet gestellt werden. Diese Softwaretools können in einem Text oder in einer Textsammlung nach unterschiedlichen sprachlichen Elementen suchen und diese in Form von Konkordanzzeilen anzeigen. Ihre Nutzung ist jedoch auf kleine Experimente mit eigenen Kleinkorpora beschränkt, denn ihnen fehlt eine substanzielle Korpuseigenschaft: entweder basieren sie auf einer zu kleinen Textsammlung oder sie ermöglichen keine Suche nach sprachlichen Elementen, bzw. sie können keine statistischen Angaben liefern.

Die bereits zehnjährigen Erfahrungen haben uns fest überzeugt, dass die Einführung dieses Faches sehr nützlich und sinnvoll gewesen ist. Einerseits hat sich bestätigt, dass die vermeintliche Furcht vor korpuslinguistischen

Technologien bei einer durchdachten Methodik der Präsentation einzelner Produkte und ihrer Anwendung sehr schnell verschwindet und andererseits sind die Studenten in relativ kurzer Zeit dazu vorbereitet, selbständig ihre eigenen Analysen durchzuführen und diese Kenntnisse und Fähigkeiten in ihren Seminar-, Bachelor- und Magisterarbeiten auch effektiv einzusetzen. Natürlich, die permanente und schnelle Entwicklung in diesem Bereich erfordert auch vom Lehrer permanente Verfolgung der Neuigkeiten und Änderungen der gängigen Programme und Ressourcen und permanente Aktualisierung der Inhalte in Curricula zu diesem Fach. Auf Grund unserer Erfahrungen und Erkenntnisse sind wir jedoch fest überzeugt, dass man das Manko in der universitären Lehre im Bereich der Korpuslinguistik für Studierende der philologischen Fachrichtungen beheben kann und beheben muss.

Literatur zum Fach Korpuslinguistik für Studierende

Noah BUBENHOFER: *Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge*. <http://www.bubenhofer.com/korpuslinguistik/>

Noah BUBENHOFER, 2009: *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin, New York: de Gruyter.

Noah BUBENHOFER, 2011: Korpuslinguistik in der linguistischen Lehre: Erfolge und Misserfolge. http://www.jlcl.org/2011_Heft1/10.pdf

Wolf DIETRICH (Hg.), 2006: *Lexikalische Semantik und Korpuslinguistik*. Tübingen: Narr Verlag.

Tomáš KÁŇA, 2014: *Sprachkorpora in Unterricht und Forschung DaF/DaZ*. Brno: Masarykova univerzita.

Lothar LEMNITZER, Heike ZINSMEISTER, 2010: *Korpuslinguistik. Eine Einführung*. Tübingen: Narr Verlag.

Rainer PERKUHN, Holger KEIBEL, Marc KUPIETZ, 2012: *Korpuslinguistik*. Paderborn: Fink / UTB.

Unterrichtsfach Korpuslinguistik im Internet

<http://www.bubenhofer.com/korpuslinguistik/kurs/>

<https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik>

<http://www.computerlinguistik.org/portal/portal.html?s=Korpuslinguistik>

<http://www.spitzmueller.org/docs/ws2004-05/praes-korpusling-2005-11-23.pdf>

http://www.uni-regensburg.de/Fakultaeten/phil_Fak_IV/Korpuslinguistik/ressourcen/parallelkorpus.php