

# Korpora in der Lexikographie und Phraseologie

Die mit Korpora ermittelten Daten sind im lexikographischen Prozess nicht mehr wegzudenken (vgl. Krishnamurthy, 2006)<sup>1</sup>. Ihr Einsatz in der Lexikographie reicht von der Ermittlung von Lemmata für konzipierte Wörterbücher, die mit Hilfe von Frequenzlisten erfolgen kann, über reiche syntagmatische Relationen, die sowohl die Valenz als auch weit verstandene syntagmatische Muster umfassen, bis hin zu der Ermittlung von Äquivalenzen in bi- und multilingualen Wörterbüchern.

Wie Korpora eingesetzt werden (können), um konkrete Fragestellungen in der Lexikographie zu lösen, versuchen die Autoren von zehn Beiträgen in dem vorliegenden Sammelband zu beantworten. Es handelt sich um Referate, die während des „10. Kolloquiums zur Lexikographie und Wörterbuchforschung“ gehalten wurden sind. Die Tagung mit dem Thema „Korpora in der Lexikographie – Stand und Perspektiven“ fand an der Adam-Mickiewicz-Universität in Poznań (Polen) am 19. und 20. Oktober 2018 mit Beitragenden aus Deutschland, Italien, Polen, der Slowakei, Slowenien, Spanien und Südafrika statt.

In seinem englischsprachigen Beitrag verweist **Rufus Gouws** (Stellenbosch) auf die konstant bedeutende Rolle der Korpora in der Planung sowie Zusammenstellung von Wörterbüchern. Dabei hebt er hervor, dass die Online-Umgebung sowohl neue Herausforderungen als auch Gelegenheiten bietet, die Verwendung von Korpora zu erweitern, sodass sowohl der Lexikograph als auch der Benutzer direkt darauf zugreifen und sie verwenden können. Gouws betont dabei, dass die Implementierung dieser Prozeduren auch einen neuen Blick auf den lexikographischen Prozess und verschiedene Wörterbuchstrukturen erfordert. Einen Schwerpunkt in der metalexikographischen Forschung sollten dem Autor zufolge Erforschungen von Strukturen für neue lexikographische Prozesse, Verwendungen von Korpora und neue Methoden zum Abrufen von Informationen bilden.

**Vida Jesenšek** (Maribor) diskutiert in ihrem Beitrag drei lexikographisch relevante, jedoch wenig berücksichtigte Aspekte der phraseologischen Bedeutung: Motiviertheit, Vagheit sowie Sprachhandlungspotential inklusive Sprechereinstellungen. Den Beitrag eröffnet die Skizzierung der prinzipiellen Komplexität der phraseologischen Bedeutung (Abschnitt 2), um im weiteren Teil, anhand der Beobachtung der aktuellen lexikographischen Behandlung ausgewählter Phraseme in allgemeinen Wörterbüchern mit Deutsch (Duden online, DWDS) und im speziellen deutschen phraseologischen Wörterbuch (Duden 11, Redewendungen), zu zeigen, wie unzulänglich die bisherige lexikographische Praxis ist (Abschnitt 3) und wie sie an Qualität gewinnen kann (Abschnitt 4).

---

<sup>1</sup> Krishnamurthy, Ramesh. 2006. Corpus Lexicography. In: Brown, Keith (ed.): *Encyclopedia of language and linguistics*. Bd. 3. 2. Aufl., Amsterdam, Elsevier: 250–254.

Im Rahmen ihres Beitrags „Lexikographische Behandlung von ausgewählten nicht lemmatisierten deutschen Idiomen“ beweist **Elżbieta Dziurewicz** (Poznań), dass Corpora innerhalb der Phraseographie sowohl bei den Fragen der Makro- als auch Mikrostruktur gewinnbringend Einsatz finden und damit zu einer verbesserten lexikographischen Beschreibung beitragen können. Die Autorin unterzieht einleitend konkrete phraseologische Kandidaten einer Frequenzanalyse anhand von DeReKo. Die frequentesten Phraseologismen werden in einer unter Deutschsprachigen aus Deutschland, Österreich und der Schweiz durchgeführten Onlinebefragung berücksichtigt. Auf der Basis der Frequenzanalyse und der Befragung stellt sich heraus, dass nur einige Einheiten von der ursprünglichen Liste Kriterien der Frequenz und Geläufigkeit erfüllen und deswegen Eingang in ein phraseologisches Wörterbuch finden sollten. Mithilfe des Korpus werden nachfolgend typische Umgebung, Varianten, Modifikationspotential und der thematische Bereich von sieben üblichen Phraseologismen untersucht.

Der Beitrag von **Carolina Flinz** (Mailand) zeigt, dass Corpora die lexikographische Primärquelle von *Tourlex*<sup>2</sup> sind, d. i. eine zweisprachige wiki-basierte Online-Ressource zur Tourismusfachsprache. In der lexikographischen Quellenstruktur von *Tourlex* sind, wie die Autorin schreibt, sowohl kleine als auch große zweckgebundene Corpora für die Erarbeitung spezifischer lexikographischer Schritte eingesetzt worden. *Tourlex* kann als dynamisches Wörterbuch funktionieren, indem weitere (kleinere und größere) Corpora hinzugefügt werden. Im Laufe des lexikographischen Prozesses haben sich auch die Benutzung des Internets als zusätzlicher Recherchemöglichkeit und die Kombination eines korpusbasierten (*corpus-based*) und eines korpusgesteuerten Ansatzes (*corpus-driven*) als fruchtbar erwiesen.

**Tamás Kispál** (Göttingen) diskutiert in seinem Beitrag „Zugriff auf Korpusbelege in deutschen einsprachigen Onlinewörterbüchern aus der Perspektive des Deutschen als Fremdsprache“ aus einer lernerlexikographischen Perspektive erstens die Frage, welche deutschen Onlinewörterbücher einen integrierten Zugriff auf Korpusbelege bieten. Zweitens wird die Frage beantwortet, wie Wörterbuchportale diese Belege in der virtuellen Mikrostruktur positionieren und wie sie das Auffinden des entsprechenden Moduls ermöglichen. Dabei ist hier zudem der Aufbau des Moduls der Korpusbelege in den Onlinewörterbüchern von Interesse. Schließlich stellt sich der Autor die Frage, welchen Mehrwert diese Korpusbelege für die Benutzer haben, die Deutsch als Fremdsprache lernen. Dabei werden die gestellten Fragen in zwei einsprachigen deutschen Onlinewörterbüchern, in DWDS und ELEXIKO, exemplarisch untersucht.

Im Beitrag von **Barbara Komenda-Earle** (Szczecin) werden lexikographische Einträge zu den ältesten Phraseologismen des Deutschen am Beispiel der Somatismen untersucht. Die Formen der Somatismen werden in Anlehnung an Wörterbücher des Mittelhochdeutschen, das Frühneuhochdeutsche Wörterbuch und das

---

<sup>2</sup> <<https://wiki.uni-mannheim.de/tourlex/>>

„Deutsche Wörterbuch“ von Jacob und Wilhelm Grimm (1854–1961) verfolgt. Ziel des Beitrags ist es zu ermitteln, wie Phraseologismen in den Wörterbüchern berücksichtigt werden sowie inwieweit ältere historische Wörterbücher des Deutschen über Somatismen, insbesondere Kinegramme und Pseudokinegramme, Aufschluss geben können.

**Ryszard Lipczuk** (Szczecin) untersucht in seinem Beitrag geflügelte Worte, die von Johann Wolfgang von Goethe, Friedrich Schiller, Heinrich Heine, Wilhelm Busch, Bertolt Brecht und anderen Autoren stammen. Es wurde überprüft, inwieweit sie in den heutigen Wörterbüchern registriert wurden. Als Grundlage der Untersuchung dienen: das 10-bändige Dudenwörterbuch (1999) sowie die deutsch-polnischen Großwörterbücher Pons (2007) und PWN (2010), darüber hinaus (für das Deutsche) die Büchmann-Lexika und (für das Polnische) die Lexika von Markiewicz/Romanowski sowie Markiewicz.

Die Beiträge von **Meike Meliss** (Mannheim/Santiago de Compostela) und **Christine Möhrs** (Mannheim) präsentieren Hintergründe des LeGeDe-Projektes<sup>3</sup>. Dabei versteht sich die LeGeDe-Ressource als Prototyp für die lexikographische Darstellung der Besonderheiten des standardnahen, gemeinsprachlich gesprochenen Deutsch in der Interaktion, wie es in natürlichen Gesprächen im privaten und institutionellen Kontext genutzt wird. Dadurch wird eine lexikographische Sprachdokumentation korpusbasierter lexikalischer Besonderheiten des gesprochenen standardnahen Deutsch ermöglicht, was primär für Gesprächsforschende, Lexikologen, Sprachwissenschaftler und Sprachlehrende von großem Interesse sein kann. Ziel des Beitrages von Meike Meliss ist es dabei, die einzelnen Angaben als lexikographische Artikelsegmente detailliert zu beschreiben und damit die lexikographische Mikrostruktur der elektronischen LeGeDe-Ressource vorzustellen. Christine Möhrs zeigt Grundannahmen auf, vor deren Hintergrund im LeGeDe-Projekt eine korpusbasierte Methode zur Ermittlung von geeigneten Stichwortkandidaten für den in der Projektlaufzeit entwickelten LeGeDe-Prototyp erarbeitet wurde.

**Petra Oboňová** (Trnava) analysiert in ihrem Beitrag die Kollokationen der deutschen Fußballsprache, die in den Live-Kommentaren der deutschen Bundesliga in der Saison 2018–2019 benutzt wurden. In der empirischen Analyse geht die Autorin von der theoretischen Annahme aus, dass die aus einer Basis und einem Kollokator bestehenden Kollokationen als „typische, spezifische und charakteristische Kombinationen von Wörtern“ (Hausmann 1985: 118) zu betrachten sind. Das Analyseergebnis ist ein Glossar mit Kollokationen der deutschen Fußballsprache, in dem Basen samt ihren Kollokatoren zu finden sind.

---

<sup>3</sup> <<https://www.owid.de/legede/>>

Der Band wird mit einem Beitrag zur Geschichte des *Kolloquiums zur Lexikographie und Wörterbuchforschung* abgerundet, das auf 10 Editionen in fünf Ländern Ost- und Südeuropas zurückblicken kann. Dessen Werdegang wird von **Milka Enčeva** und **Vida Jesenšek** (Maribor) abschließend präsentiert.

Poznań, im Februar 2021

Michał Piosik  
Janusz Taborek  
Marta Woźnicka