

# **Empirische Methoden und Beschreibungsverfahren der korpusbasierten Phraseografie**

---

**Doris SAVA**

Doz. Dr. habil.; Lucian-Bloga-Universität Sibiu/Hermannstadt;  
E-Mail: dorissava71@yahoo.com

**Abstract:** The modern lexicography must meet certain requirements in order to ensure an adequate lexicographical coverage of the presented lexical inventory. The relevant published literature dedicated to the profile of lemmas within the current lexicographical practice reveals shortcomings that reflect the gap between the current linguistic reality and the usual lexicographical practice. From the user's perspective and his consulting needs modern metalexicography stresses the need for a fair lexicographical exposure, empirically grounded, capable of indicating peculiarities and restrictions of the usage within the language. Parting from theoretical considerations, regarding the relevance of the criteria that define the quality of a lexicographical work, the principle of addressability and the current phraseographic practice, the article presents directions and ways of researching for the phraseological inventory in terms of relevant peculiarities for the mono- or bilingual lexicographical practice. Exemplary there are discussed modern empirical methods meant to improve the current lexicographical practice insisting upon the evidence provided by the corpus as object and means of verifying the hypothesis of theoretical or applied research. The analysis of occurrences extracted from corpora is a valuable tool for the lexicographer because it allows the identification of the specific characteristics belonging to fixed structures within usual multiple contexts, patterns of use, concordances and co-selections. An authoritative body of linguistic evidence allows observations of mutations occurring in use or frequency of distributions, as well as identifying outlying

areas of phraseology which so far have been insufficiently investigated or ignored by the lexicographical practice.

**Key words:** phraseologism, lexicographical practice, empirical methods and processes of collection/interpretation, reference corpus

Die Metalexikografie<sup>1</sup> diskutiert aus der Sicht der Benutzerbedürfnisse und der Benutzerfreundlichkeit die Besonderheiten der Phraseologismen und ihre Darstellung in ein- oder zweisprachigen Wörterbüchern.<sup>2</sup> Zu den lexikografisch relevanten Aspekten gehören: (1) die Orientierung im Wörterbuch; (2) die lexikografische Erfassung und Darstellung phraseologischer Spezifik (Nennformgestaltung; Präsentation phraseologischer Variation; Berücksichtigung der externen Valenz; Erfassung spezifischer Gebrauchsmuster/Darbietung der formal-semantischen und pragmatischen Restriktionen; Markierungspraxis); (3) die Berücksichtigung der Benutzerbelange. Mit der Festlegung des Benutzerkreises/der Bedürfnisse sind die adressatenspezifische Ausrichtung der Makrostruktur und die Gestaltung der Informationen verbunden, die in der Mikrostruktur dargeboten werden.<sup>3</sup>

<sup>1</sup> Die Theorie und Methodologie der *Metalexikografie* hat sich in den letzten Jahrzehnten stark entwickelt. Gegenstand der *Wörterbuchforschung* (oder *Metalexikografie*) sind alle lexikografischen Tätigkeiten und die Endergebnisse lexikografischer Bemühungen.

<sup>2</sup> Vgl. u.a. Blanco, Carmen Mellado (Hg.): *Theorie und Praxis der idiomatischen Wörterbücher*. Tübingen 2009. Aspekte der lexikografischen Darstellung von Phraseologismen im Wörterbuch betreffen die Gesamtheit der Wörterbücher, unabhängig davon, ob diese allgemein oder phraseologisch bzw. ein- oder zweisprachig sind oder einen speziellen Wörterbuchtyp bzw. ausgewählte Bereiche von Wörterbuchtypen. In diesem Zusammenhang kann die Frage nach der Nennform angeführt werden, die alle Wörterbücher betrifft, während z.B. die Äquivalenzangaben für die bilinguale Lexikografie relevant sind.

<sup>3</sup> Für Neuauflagen vorhandener Wörterbücher oder für die Konzeption geplanter ist das Heranziehen zuverlässiger Lexika, Fachquellen und

Die moderne Phraseografie muss bestimmten Anforderungen genügen, sodass die lexikografische Praxis ein semantisches *und* pragmatisches Verständnis der aufgenommenen Phraseologismen gewährleisten sollte. Dabei ist auch der jeweiligen Zielsetzung Rechnung zu tragen, je nachdem, ob es sich um ein auf Produktion oder Rezeption ausgerichtetes Wörterbuch handelt oder um ein für Muttersprachler konzipiertes oder für den ausländischen Benutzer gedachtes. Ein modernes lexikografisches Werk sollte die Ergebnisse der theoretischen und praktischen Phraseologie berücksichtigen und an die Bedürfnisse der lexikografischen Praxis und der anvisierten Benutzergruppen anpassen.<sup>4</sup> Die lexikografische Erfassung der vielfältigen und

---

Textkorpora grundlegend – wie auch das Erarbeiten einer Beschreibungsweise, welche die syntaktische Formenvielfalt und das Funktionspotential dieser Ausdrucksmittel angemessen zu erfassen vermag. Ein modernes Phraseolexikon sollte Informationen zu den Gebrauchspräferenzen und -restriktionen der Phraseologismen explizit anbieten. Durch die Ausarbeitung eines einheitlichen lexikografischen Konzepts für die phraseologische Wörterbuchpraxis und damit eines empirisch und auch wissenschaftlich abgesicherten phraseologischen (Lern-)Wörterbuchs kann den praktischen Bedürfnissen der Deutsch Lernenden gedient werden.

- <sup>4</sup> Die Auseinandersetzung mit der lexikografische Darbietung von Phraseologismen macht auf ihre mangelhafte Erfassung und Beschreibung in mono- und bilingualen Wörterbüchern aufmerksam. Die kritischen Anmerkungen zur phraseografischen Wörterbuchpraxis, die in der germanistischen metalexikografischen Forschung anzutreffen sind, betreffen Fragen der Lemmatisierung, der Nennformgestaltung und der Gebrauchsmarkierung phraseologischer Einheiten. Bemängelt wurde u.a., dass z.B. die Erklärung eines Phraseologismus durch einen anderen erfolgt, dass bei Mehrfachlemmatisierung die Bedeutungsangaben nicht immer übereinstimmen, dass Phraseologismen mit voneinander abweichenden Markierungen an verschiedenen Stellen im Wörterbuch vorkommen oder dass Inkonsequenzen der Handhabung lexikalischer Varianten registriert werden können. Bei der kritischen Durchkämmung unterschiedlicher Wörterbucheinträge konnten auch Ungenauigkeiten ermittelt werden, z.B. wenn eine Äquivalenz zwischen Phraseologismus

für den produktiven Gebrauch phraseologischer Einheiten relevanten pragmatischen Angaben, d.h. der pragmatischen Spezifik, ist für den Lexikografen eine Herausforderung.<sup>5</sup>

Aus der Sicht des Benutzers des Deutschen als Fremdsprache wurde eine reflektiertere – auch *pragmatische* – lexikografische Erfassung gefordert, welche die Textproduktion und -rezeption erleichtern würde. Dies bedeutet, dass das Lexikografenteam das Restriktionsprofil und die Besonderheiten der Verwendung von Phraseologismen ermitteln und eine Markierung dieser Phraseologismen vornehmen muss. Dies erfordert, dass der Phraseologismus hinsichtlich seines potenziellen Restriktionsprofils überprüft werden muss.<sup>6</sup>

Um beurteilen zu können, ob ein phraseologisches Nachschlagewerk benutzerfreundlich ist, kann die kritische Durchsicht der in der Wörterbucheinleitung explizit formulierten Ansichten zur lexikografischen Bearbeitung der Phraseologismen und/oder die Analyse der Kodifizierungspraxis hilfreich sein.<sup>7</sup> Die Auseinandersetzung mit lexikografisch relevanten Aspekten weist einerseits auf die enge Verbindung zwischen

---

und nichtphraseologischer Paraphrase suggeriert wurde. Auch kann das Fehlen einer obligatorischen Komponente eines lexikografisch erfassten Phraseologismus registriert werden, was seinen aktiven Gebrauch verhindern würde.

- <sup>5</sup> Über das Verstehen einer festen Wortverbindung hinaus muss die lexikografische Beschreibung auch den korrekten Gebrauch der kodifizierten Phraseologismen bei der Textproduktion sicherstellen.
- <sup>6</sup> Hierfür sind u.a. die Textkorpora des Instituts für deutsche Sprache in Mannheim zu empfehlen.
- <sup>7</sup> Die Sichtung des Wörterbuchvorspanns und mehrerer phraseologischer Reihen mit jeweils einer gemeinsamen Basiskomponente verdeutlicht, welche phraseografische Entscheidungen getroffen wurden und ob die Lexikografen sowohl in der Wörterbucheinleitung als auch in den einzelnen Eintragungen dem Benutzer vorführen, worin die Besonderheiten dieser sprachlichen Einheiten und deren lexikografische Bearbeitung bestehen. Zu den intensiv diskutierten Problemen der Phraseografie zählen u.a. die Prinzipien ihrer Lemmatisierung und

Sprachgebrauch und Wörterbuchpraxis hin, andererseits wird deutlich, dass die Ergebnisse der theoretischen und angewandten Phraseologie für die Wörterbuchpraxis wichtig sind. Trotz des verstärkten Interesses an der Phraseografie, das in den 1990er-Jahren registriert werden konnte, wurde immer wieder betont, dass die praktische Lexikografie die Erkenntnisse der (theoretischen und angewandten) Phraseologie bei der Aus- und Überarbeitung von Wörterbüchern *nicht* in vollem Umfang berücksichtigt.<sup>8</sup>

Bei der Durchsicht phraseografischer Werke kann eine Diskrepanz zwischen der theoretischen Problematik und der phraseologischen Praxis unschwer festgestellt werden. Der Grund liegt in der Tatsache, dass die Phraseografie<sup>9</sup> die sprachtheoretischen Erkenntnisse viel zu wenig berücksichtigte, die Theorie der Phraseologie ihrerseits die konkrete Wörterbucharbeit viel zu selten in den Mittelpunkt ihres Interesses stellte. Obwohl

---

der Nennformfixierung, semantische und pragmatische Angaben, Restriktionsmarkierungen oder die Markierungspraxis.

<sup>8</sup> Die Behandlung und Darstellung der Phraseologismen im Wörterbuch wurde aus unterschiedlichen Perspektiven thematisiert, wobei die Benutzerperspektive verstärkt das Interesse der Wörterbuchforschung erweckte. Von den Bedürfnissen des Benutzers ausgehend wurden verschiedene Wörterbuchtypen definiert, bei der es vor allem um die Auswahl und Darstellung der für die jeweilige Benutzergruppe relevanten Informationen geht, da bezüglich der Darstellung von Phraseologismen jeweils andere Anforderungen an ein Wörterbuch für die Rezeption und eines für die Produktion gestellt werden, ebenfalls ob das Wörterbuch für fremdsprachliche Benutzer oder Muttersprachler konzipiert worden ist.

<sup>9</sup> *Phraseografie* als Parallelbildung zum Terminus *Lexikografie* ist ein Teilbereich der Phraseologie/Lexikografie, der die lexikografische Erfassung, Beschreibung und Darstellung von phraseologischen Einheiten im ein- und zweisprachigen Wörterbuch erfasst. Der Status der *Phraseografie* in der Forschungsliteratur ist nicht eindeutig geklärt. Manche Autoren begreifen sie als Teilbereich der angewandten bzw. praktischen Phraseologie, andere schreiben ihre Aufgabenstellung der allgemeinen lexikografischen Beschreibungstheorie zu.

die Erkenntnisse zum textsortenorientierten Vorkommen von Phraseologismen die Diskussionen zur phraseografischen Beschreibungspraxis angeregt hat, sind pragmatische Beschreibungen in Wörterbüchern (auch hierzulande) ungenügend berücksichtigt worden.<sup>10</sup>

Angesichts der Tatsache, dass in den herkömmlichen Wörterbüchern die Semantik der Phraseologismen oft – und fast ausschließlich – durch deren denotative Bedeutung erfasst wird, hat sich in der metalexikografischen Diskussion verstärkt die Forderung nach einer sprachhandlungstheoretisch orientierten Beschreibung durchgesetzt.<sup>11</sup> Mit Recht wurde gefordert, dass Angaben zu den Sprecher-/Schreibereinstellungen und den kommunikativen Funktionen von Phraseologismen in Wörterbüchern – unabhängig von ihrem Typ – unbedingt notwendig sind, um die hinsichtlich der lexikografischen Kodifizierung von Phraseologismen immer wieder beklagte Diskrepanz zwischen Sprachwirklichkeit und Wörterbuchpraxis zu mindern.<sup>12</sup> Ein Nachschlagewerk, das verlässliche Informationen zu den Gebrauchspräferenzen und -restriktionen der kodifizierten

---

<sup>10</sup> Der richtige Gebrauch eines Phraseologismus erfordert die Darstellung des sprachlichen und situativen Kontextes in Form pragmatischer Angaben/Kommentare (gebrauchssemantische Beschreibung). Für den Benutzer des Deutschen als Fremdsprache müssen die Informationen in einem Nachschlagewerk differenzierter gestaltet werden als für den Muttersprachler. Pragmatische Erkenntnisse haben zwar einen „breiten Eingang“ in die Theorie der Phraseografie gefunden, „es mangelt jedoch an der konsequenten praktischen Umsetzung der gewonnenen Postulate“ (Filatkina 2007, S. 152) in der allgemeinen und phraseologischen Wörterbuchpraxis. Vgl. Filatkina, Natalia: Pragmatische Beschreibungsansätze. In: Burger, Harald/Dmitrij Dobrovol'skij/Peter Kühn/Neal R. Norrick (Hgg.): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin 2007, S. 132-158.

<sup>11</sup> Vgl. Filatkina 2007, S. 148.

<sup>12</sup> Vgl. hierzu u.a. Blanco, Carmen Mellado/Patricia Bujan/Claudia Herrero (Hgg.): *La fraseografía del S. XXI. Nuevas propuestas para el español y el alemán*. Berlin 2010.

Phraseologismen explizit anbietet, würde eine erhöhte Qualität des Wörterbuchs und dessen Benutzerfreundlichkeit sichern.<sup>13</sup> Für die praktische bilinguale Phraseografie der Sprachen Deutsch und Rumänisch ist daher die Ausarbeitung eines modernen phraseologischen Wörterbuchs, das zuverlässiger und benutzerfreundlicher als bisherige Werke ist, *dringend* zu fördern. Eine verbesserte Erfassung der Gebrauchsweisen<sup>14</sup> der

<sup>13</sup> Gebrauchsrestriktionen sind in den Nachschlagewerken für einen Nichtmuttersprachler nicht immer klar genug formuliert. Durch die Vernachlässigung der Information zur Valenzfähigkeit des Phraseologismus und die Reduzierung seiner Nennform kann der Wörterbuchbenutzer nicht immer einschätzen, ob als Handlungsträger eine Person oder auch eine Institution, Einrichtung usw. genannt werden kann. Die Nennformgestaltung sollte Restriktionen kenntlich machen und die korrekte Nennform sollte so formuliert werden, dass Angaben zur externen (syntaktisch-semantischen) Valenz auch berücksichtigt werden. Durch die Nennformformulierung müssen dem Nichtmuttersprachler u.a. auch Hinweise auf geschlechtsspezifische Gebrauchsrestriktionen geboten werden.

<sup>14</sup> Die Sichtung rumänischer Wörterbücher verdeutlicht, dass sich pragmatische Informationen eher auf stilistische Markierungen beschränken. Viele Phraseologismen, die (eindeutig) Beschränkungen unterliegen, wurden nicht entsprechend markiert. In vielen Fällen fehlen lexikografische Angaben zur stilistischen Bewertung oder es werden konstruierte Anwendungsbeispiele geboten, denen der Wörterbuchbenutzer nicht entnehmen kann, ob der Phraseologismus z.B. eine geschlechtsspezifische Beschränkung aufweist. Kontextbeispiele sollten nicht nur die Semantik des Phraseologismus verdeutlichen, sondern auch typische Ausfüllungen der durch die Valenz bestimmten Leerstellen. Der Vorteil aktualisierter Beispiele gegenüber der von der lexikografischen Theorie und praktischen Phraseografie geforderten neutralen Nennform liegt darin, dass Valenzangaben wie *etw.* oder *jmd.* lexikalisch besetzt werden. Der Nachteil besteht jedoch darin, dass keine klaren Aussagen gemacht werden, was zum Phraseologismus gehört und was nicht, d.h. darüber zu entscheiden, welche Komponenten zur phraseologischen Einheit i.e.S. gehören und welche nur fakultativ sind und in der Wendung erscheinen, um den Satz zu vervollständigen oder um den Phraseologismus pragmatisch zu verdeutlichen. Vgl.

Phraseologismen würde auch den Rückstand der rumänischen bilingualen Wörterbuchpraxis mit Deutsch aufholen.

Phraseologie *im* Wörterbuch [...] klingt weniger komplex als es de facto ist. Denn in Wirklichkeit und bei näherer Betrachtung ist die Angelegenheit vielschichtiger, und zwar nicht nur, wenn man den Begriff der Phraseologie im Sinne einer wissenschaftlichen Disziplin, also im Sinne von *Phraseologieforschung*, sondern auch und gerade dann, wenn man ihn als auf den Objektbereich ‚Phraseologie‘ bezogen versteht. Denn es ist ja keineswegs so, dass die Phraseologie quasi von sich aus in einem Wörterbuch bzw. Korpus enthalten ist und damit gewissermaßen ‚einfach so‘ für weitere Zwecke [...] zur Verfügung steht. Vielmehr unterscheiden sich Wörterbuch und Korpus (genauer gesagt: verschiedene Arten von Wörterbüchern und Korpora) in dieser und verschiedener anderer Hinsicht voneinander und stehen zudem möglicherweise in komplexen Wechselbeziehungen miteinander.<sup>15</sup>

In den letzten Jahrzehnten haben elektronische Möglichkeiten bei der Erstellung und Auswertung von Korpora oder bei der Ausarbeitung von Wörterbüchern an Bedeutung gewonnen.<sup>16</sup> Anfang der 1990er-Jahre hat die Forschung auf den Einsatz von Computern und Datenbanken bei der Ausarbeitung

---

Martin, Luis: *Phraseologie im zweisprachigen Wörterbuch Deutsch-Spanisch*. Frankfurt/Main 2001, S. 144.

<sup>15</sup> Jesenšek, Vida/Grzybek, Peter (Hgg.): *Phraseologie im Wörterbuch und Korpus/Phraseology in Dictionaries and Corpora*. Maribor 2014. Die Beiträge dieses Bandes gehen auf die traditionelle zweijährige Tagung der Europäischen Gesellschaft für Phraseologie (EUROPHRAS) zurück, die vom 27. 8. bis zum 31. 8. 2012 an der Universität Maribor stattfand. Hervorh. im Original.

<sup>16</sup> Die Nützlichkeit von Sprachkorpora in der linguistischen Forschung ist unbestreitbar. Sie sind auch wichtige Hilfsmittel bei der Erstellung oder Verbesserung von Wörterbüchern. Vgl. hierzu auch Blanco 2009. Dieser Sammelband bietet aus verschiedenen Perspektiven einen repräsentativen Einblick in den Forschungsstand und in die aktuellen Tendenzen der ein- und zweisprachigen Phraseografie (Korpusbasiertheit und verstärkte Benutzerorientiertheit).



lexikografischer Arbeiten hingewiesen und die Vorteile der Hinzunahme von elektronischen Maschinen auch für die phraseografische Praxis erläutert. Dabei wurden wiederholt die Mängel traditioneller Wörterbücher hervorgehoben. Bei der kritischen Bilanz betont die Forschung ausdrücklich, dass viele Desiderate der klassischen Lexikografie mit den modernen Möglichkeiten der Computerlexikografie erfüllt werden können.<sup>17</sup> Bereits 1980 postulierte Helmut Henne, dass die elektronische Datenverarbeitung im Dienste der lexikografischen Dokumentation der Theoriediskussion und einer verbesserten lexikografischen Praxis Vorschub leisten würde.<sup>18</sup>

Der Sammelband von Blanco/Bujan/Herrero (2010) dokumentiert den aktuellen Forschungsstand und gibt wichtige Impulse für die Weiterentwicklung der ein- und zweisprachigen Phraseografie des 21. Jahrhunderts.<sup>19</sup> Die lexikografische Darbietung wurde auch daraufhin untersucht, welche Antworten sie

<sup>17</sup> Die ältere Fachliteratur hat Überlegungen und konstruktive Vorschläge zur Erstellung einer Datenbasis formuliert, wie auch theoretisch begründete Verfahren, wobei auch auf das Fehlen einer Patentlösung für die endgültige Form der Datenbank hingewiesen wurde. Vgl. hier u.a. die Arbeiten von Dobrovol'skij in 1990er-Jahren zur Erstellung eines Thesaurus deutscher Idiome und zu den **Perspektiven der computer-gestützten Phraseografie**.

<sup>18</sup> Dazu vgl. Steyer, Kathrin: *Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten*. In: *Deutsche Sprache* 28/2000, S. 101-125.

<sup>19</sup> Darunter die Beschreibung des umfangreichen Projektes FRASESPAL, eines onomasiologisch gegliederten phraseologischen Wörterbuches des Deutschen und des Spanischen; Betrachtungen zur Erstellung eines Kollokationswörterbuchs für ausländische Spanischlernende und zu einem mehrsprachigen elektronischen Wörterbuch zu Phrasemen mit nicht austauschbaren Verben der Sprachen Spanisch, Französisch, Katalanisch, Englisch, Deutsch und Arabisch; Fragen der lexikografischen Erfassung von Kollokationen in plurizentrischen Sprachen und neue empirische Methoden zur korpusbasierten Phraseografie. Vgl. Blanco, Carmen Mellado/Patricia Bujan/Claudia Herrero 2010.

auf die Suchfragen in verschiedenen Wörterbuchbenutzungssituationen bereithält. Viele Nachschlagewerke werden als passive Wörterbücher genutzt, da sie eher auf die Rezeption von Phraseologismen ausgerichtet sind. Die Erstellung von aktiven phraseologischen Wörterbüchern, die sich von den passiven nicht nur darin unterscheiden, dass mehr Informationen über eine beschränkte Auswahl von Einheiten geboten werden, sondern auch darin, dass neue theoretische Erkenntnisse und Möglichkeiten (Metasprache und Kodifizierungsverfahren) in die phraseografische Praxis eingebunden werden, ist eine zentrale Aufgabe der modernen Phraseografie.<sup>20</sup>

Dräger (2010), der 42 phraseologische Nachschlagewerke – teilweise in mehreren Auflagen – untersucht hat, begründet, weshalb die historische und auch aktuellere deutschsprachige Phraseografie seit über 30 Jahren kritisiert wird. Dies liegt einerseits an den lexikografischen Standards und andererseits an den Anforderungen der Phraseologie. Die stichprobenartige Untersuchung belegt, „dass das Kopieren älterer Sammlungen und Lexika nicht nur Wissen konserviert, sondern auch zu einer Tradierung und Vermischung von Fehlerklärungen und Fehlmotivierungen geführt hat.“<sup>21</sup> Seine Wörterbuchanalyse führt zum Schluss, dass Hinweise auf einen phraseologischen Wandelprozess spärlich erfasst wurden und dass es „[d]as Nachschlagewerk, welches in jedem Fall eine verlässliche Auskunft gibt“ deshalb nicht gibt, da „keines die diaphasische und diatopische Dimension der Phraseologismen ausreichend aufdeckt.“ (S. 414).

---

<sup>20</sup> Vgl. Dobrovol'skij, Dmitrij: *Zur deutschen Phraseografie*. In: *Cahiers d'Etudes Germaniques* 23/1992, S. 161-172; hier S. 168.

<sup>21</sup> Dräger, Marcel: *Phraseologische Nachschlagewerke im Fokus*. In: Korhonen, Jarmo/ Wolfgang Mieder/Elisabeth Piirainen/Rosa Piñel (Hgg.): *EUROPHRAS 2008. Beiträge zur internationalen Phraseologiekonferenz vom 13.-16.8.2008 in Helsinki*. Helsinki 2010, S. 411-421; hier S. 411ff.

Trotz der Tatsache, dass es heutzutage viele Wörterbücher gibt, sind Fragen nach neuen Möglichkeiten in der (deutschen) Phraseografie offen. Die primäre Aufgabe phraseologischer Wörterbücher besteht u.a. auch in der „Lehr- und Lernbar-machung der idiomatischen Komponente der sprachlichen Kompetenz“ (Dobrovol’skij 1992, S. 161), sodass z.B. ein interessierter Deutschlerner durch die in den einschlägigen Wörterbüchern enthaltenen Informationen deutsche Phraseologismen situativ und grammatisch angemessen gebrauchen kann.<sup>22</sup>

Filipenko (2002) bringt Ausführungen zur Konzeption eines aktiven zweisprachigen phraseologischen Wörterbuchs unter Nutzung der Beispiele aus der Online-Datenbank der Textkorpora des Instituts für deutsche Sprache in Mannheim. Dabei weist Filipenko darauf hin, dass die Analyse von Textkorpora interessante Informationen bietet, die noch keinen Eingang in Lexika gefunden haben.<sup>23</sup> Belege aus Korpora haben den Vorzug, dass

<sup>22</sup> Für den aktiven Gebrauch von Phraseologismen i.e.S (Idiomen) sind nach Dobrovol’skij (S.161f.) folgende Fragen ausschlaggebend: (1) Welche phraseologische Einheiten (Idiome) werden heute von den Muttersprachlern tatsächlich verwendet? (2) In welchen Ko- und Kontexten werden diese gebraucht bzw. welche Informationen müssen im Wörterbuch fixiert werden, damit ein adäquater Gebrauch gewährleistet wird? Die Frage nach dem Usus fester Wortverbindungen würde nur dann angemessen beantwortet werden können, wenn man phraseologische Spezialwörterbücher für unterschiedliche Funktionsbereiche (Presse, wissenschaftlicher Text, Belletristik) ausarbeiten würde.

<sup>23</sup> Filipenko, Tat’jana: Beschreibung der Idiome in einem zweisprachigen Idiomatik-Wörterbuch (Deutsch-Russisch). In: Vollstedt, Marina (Hg.): *Das Wort. Germanistisches Jahrbuch GUS*, 2002, S. 43-62. Zu den Vorteilen der Analyse von Textkorpora vgl. u.a. Steyer, Kathrin: Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse. In: Haß-Zumkehr, Ulrike/Kallmeyer, Werner/Zifonun, Gisela (Hgg.): *Ansichten der deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag*. Tübingen 2002, S. 215-236.

sie die verschiedenen Verwendungen eines Wortes aufgrund realer Textbeispiele widerspiegeln.

Als innovativer Ansatz für die Erforschung fester Wortverbindungen gilt die Methode der statistischen Kookkurrenzanalyse, die in den 1980er-Jahren am Institut für deutsche Sprache (IDS) in Mannheim entwickelt wurde. Die statistische Kookkurrenzanalyse stellt ein Konzept dar, mit dem ein empirischer Zugang zum Sprachwissen ermöglicht werden soll. Neue Möglichkeiten dieses Konzepts stellt Steyer (z.B. 2000) an ausgewählten Beispielen dar.<sup>24</sup> Durch die korpusanalytische Auswertung großer Datenmengen ergeben sich für die elektronische Phraseographie neue und vielfältige Möglichkeiten. Die COSMAS-Kookkurrenzanalyse<sup>25</sup> erlaubt eine Verifizierung von Wortverbindungen z.B. hinsichtlich ihrer Usualität oder auch in Bezug auf noch nicht kodifizierte Erscheinungen. Sie erbringt darüber hinaus Erkenntnisse über den Status der Komponenten, der Kookkurrenzpartner und Domänengebundenheit, hinsichtlich der Sprechereinstellungen und -bewertungen oder typischer Kontextrealisierungen. Die Methode der statistischen Kookkurrenzanalyse ermöglichtes, feste Wortverbindungen mit ihren typischen Kontexten aus einem Korpus herauszufiltern und zu analysieren. Kookkurrenzen des Deutschen, wie sie in den IDS-Korpora auftauchen, werden mit mathematisch-statistischen Methoden herausgefiltert, geordnet und lexikografisch in Form von Kookkurrenzangaben aufgearbeitet. Diese Systematisierung erlaubt eine Kontextspezifizierung und das Erkennen useller Bedeutungen/von Musterkontexten. Auch

---

<sup>24</sup> Zur kookkurrenzbezogenen Herangehensweise als fundierte Basis für linguistische Untersuchungen, wie auch zur Korpusmethodik, zum linguistisches Modell und den lexikografische Perspektiven vgl. Steyer 2000.

<sup>25</sup> Mannheimer Computersystem zur Speicherung von Textkorpora *COSMAS (Corpus Storage Maintenance and Acces System)*. Die Basis bilden die Textkorpora des IDS, die weltweit größte Sammlung elektronischer Korpora von geschriebenen deutschsprachigen Texten.

unikale Komponenten und damit die Gebundenheit eines Elements an die jeweilige Wortverbindung und ihr Nicht-Mehrvorkommen außerhalb dieser können durch die Korpuserhebung überprüft werden. In Steyer (2002) wird ein integrativer Ansatz vorgestellt, der ein hochkomplexes automatisches Tool, die am Institut für deutsche Sprache in Mannheim entwickelte COSMAS-Kollokationsanalyse mit dem linguistischen Modell der usuellen Kookkurrenz zu verbinden versucht. Steyer systematisiert am Kookkurrenzfeld von *Hund* die typischen Verwendungsmuster. Durch die Analyse des Kookkurrenzfeldes (S. 231-234) und der damit verbundenen Kontextmuster lässt sich eine viel differenziertere Beschreibung der Bedeutung und der Gebrauchsspezifik der Fügungen vornehmen, als dies aufgrund der eigenen Kompetenz bzw. auf der Basis einzelner Texte möglich ist.

Steyer präsentiert in mehreren Arbeiten das linguistische Modell, die Korpusmethodik und die lexikografischen Perspektiven dieses Ansatzes. Dabei verweist die Autorin darauf, dass Kookkurrenzen (und somit feste Wortverbindungen i.w.S.) einerseits als Gegenstand, andererseits als heuristisches Verfahren für die linguistische Analyse und Beschreibung der Bedeutung und Verwendung sprachlicher Einheiten dienen können. Korpusinformationen bzw. automatische Analysewerkzeuge ermöglichen dem Linguisten, die für ihn relevanten sprachlichen Informationen zu extrahieren. Die Kookkurrenzanalyse dient nicht nur der Überprüfung von Phraseologismen im aktuellen Usus, sondern darüber hinaus auch der Verifizierung von Lesarten und der externen Valenz. Durch dieses Verfahren wird es möglich, auch synonyme oder antonyme Wortverbindungen zu eruieren. Für die festen Wortverbindungen wird folgendes Informationsbündel angeboten: Informationen zur wendungsinternen Struktur (syntaktische Transformationsmöglichkeiten und -anomalien, Angaben zur Negation und Komparation, zur Anzahl und zum Charakter der Basiselemente, Unikalität), Angaben zur Bedeutung und Verwendung von Phraseologismen,

Angaben zur Klassenzugehörigkeit (Phraseologismus, phraseologisch gebundene Formative, Funktionsverbgefüge, Zwillingsformeln usw.), Angaben zu phraseologischen Synonymen und Antonymen, Informationen zur sprachlichen Umgebung eines Phraseologismus oder zu korpusbasierten Variations- und Modifikationsmöglichkeiten. Besonders für den Bereich der korpusbasierten Pragmatik scheinen diese Methoden einen neuen Weg zur empirischen Verifizierung zu eröffnen. Die Methoden der Korpuserschließung – insbesondere die statistische Kookkurrenzanalyse – werden v.a. im *lexiko-Modul „Usuelle Wortverbindungen“* angewandt. Den Terminus *usuelle Wortverbindungen* (UWV) hat Kathrin Steyer 2000 für ihre Forschungsarbeiten entwickelt. Er umschreibt feste Wortverbindungen verschiedener Art, die in rekurrenten Strukturen auffällig sind. Diese UWV umfassen feste Wortverbindungen wie z.B. Kollokationen, Phraseologismen oder Sprichwörter. Solche Wortverbindungen sind für Lexikologie, Lexikografie, Übersetzungswissenschaften und Fremdsprachendidaktik von Interesse, da sie einzelsprachspezifisch geprägt sind.<sup>26</sup> Das von Steyer für lexikografische Zwecke angepasste bzw. vorgestellte linguistische Modell für Wortverbindungen berücksichtigt verschiedene Ergebnisse der Idiomatikforschung. Die Repräsentation von phraseologischen Wortverbindungen im Informationssystem bietet vielfältige Informationen (u.a. über Schreibung bzw. Aussprache, über

---

<sup>26</sup> Das Modul „Usuelle Wortverbindungen des Deutschen“ ist ein zentrales Konzept des Projekts „Wissen über Wörter“ (Institut für deutsche Sprache Mannheim), ein hypertextbasiertes Informationszentrum, das auf den umfassenden IDS-Korpora basiert und das den deutschen Wortschatz umfassend dokumentieren und linguistisch erläutern wird. Dieses Informationszentrum unterscheidet sich im Aufbau und Nutzungsmöglichkeiten grundlegend von den herkömmlichen Printwörterbüchern aber auch von anderen elektronischen Wörterbüchern (vgl. Steyer 2000, S. 101f.). Steyer plädiert für eine strengere terminologische Schäfe bei der Unterscheidung zwischen statistisch erhobenen Kookkurrenzen und linguistisch interpretierten Kollokationen.

Bedeutung, Verwendung und Grammatik, Geschichtliches und Sachliches, Sprachkritisches und Normatives). Das System verfügt über eine komplexe Verlinkungsstruktur, die u.a. Verweise von einem Lemma zum anderen, zum Glossar, zu Publikationen und Korpusbelegen erlaubt. Für die festen Wortverbindungen sind Angaben zur Grammatik, Informationen zu den einzelnen Elementen, aber auch zur wendungsinternen Struktur (syntaktische Transformationsmöglichkeiten und -anomalien, Angaben zur Negation und Komparation, zur Anzahl und zum Charakter der Basiselemente, zur Unikalität), Hinweise zur Bedeutung und Verwendung phraseologischer Einheiten, zur Klassenzugehörigkeit (Funktionsverbgefüge, Zwillingsformeln usw.), zu phraseologischen Synonymen und Antonymen und zur sprachlichen Umgebung einer Fügung, zu korpusbasierten Variations- und Modifikationsmöglichkeiten vorgesehen. Zu Recht fordert Steyer (2000, S. 107), dass nur eine Korpuserhebung Aufschluss darüberbringen kann, wie die Grundform eines Phraseologismus beschaffen ist und welche Variationen bzw. Modifikationen möglich sind.<sup>27</sup> Steyer spricht ironisch in diesem Zusammenhang

<sup>27</sup> Die Verständlichkeit eines Phraseologismus und sein richtiger Gebrauch werden über die Bedeutungsparaphrase bzw. durch pragmatische Kommentare hinaus auch durch Informationen gewährleistet, die der Nennform explizit oder implizit zu entnehmen sind. Sichtet man die im Wörterbuch fixierten Nennformen, so ist auffallend, dass diese in vielen Fällen für einen Nichtmuttersprachler nicht klar genug formuliert sind. Die Frage nach der Nennform von Phraseologismen variiert von Wörterbuch zu Wörterbuch, aber auch innerhalb ein und desselben Nachschlagewerks. Die Phraseografie hat mehrmals darauf hingewiesen, dass der Phraseologismus im Wörterbuch in einer Neutralform angeführt werden muss. Darunter versteht man für verbale Einheiten meistens die Infinitiv-Präsens-Form, für substantivische Phraseologismen die Form im Nominativ Singular, für adjektivische Phraseologismen die Nominativ-Maskulin-Form. Jede andere nicht neutrale, d.h. aktualisierte Form ist ein Hinweis für den Benutzer, dass für die Wendung bestimmte morphosyntaktische Restriktionen vorliegen, d.h. der Phraseologismus kann nicht in einem beliebigen Kasus, Tempus oder



auch von einem „vererbten kollektiven Beispielgedächtnis“, das in der traditionellen Printlexikografie immer wieder anzutreffen ist und das wenig oder gar nicht dem aktuellen Gebrauch in der Sprache entspricht. Dabei verweist Steyer mit Nachdruck darauf, dass viele Phraseologismen einer empirischen Überprüfung im aktuellen Sprachgebrauch bedürfen.<sup>28</sup> Korpusbasierte

---

einer beliebigen Person verwendet werden. Wenn keine Restriktionen vorliegen, sollte die aktualisierte Form gemieden werden, damit der Benutzer keine falschen Schlüsse über den Gebrauch der Wendung ziehen soll. Die Nennform muss den Wörterbuchbenutzer über mögliche Usus-Restriktionen informieren. Daher sollte eine benutzergerechte Nennformgestaltung syntaktisch-semantische Angaben zu den Kontextpartnern explizit und vollständig darbieten. Oft findet der Wörterbuchbenutzer unterschiedliche Informationen darüber, welche Komponenten einen Phraseologismus ausmachen. Für die Klärung der Frage nach den fakultativen und obligatorischen Komponenten des Phraseologismus, die in die Nennform aufgenommen werden sollen, und für eine vollständige und korrekte Kodifizierung sind Recherchen der modernen elektronischen Textkorpora wie auch das Sammeln von Belegen aus modernen Quellen (z.B. aus der aktuellen Presselandschaft) hilfreich.

<sup>28</sup> Damit die Nennform eine sichere Textproduktion gewährleisten kann, müssen durch die Wahl der Nennform klare Aussagen über mögliche Restriktionen in der Verwendung mancher Phraseologismen gemacht werden und dem Wörterbuchbenutzer die wichtigsten Informationen zur Einbettung des Phraseologismus im Satz geliefert werden. Mit den Komponenten *jmdm.*, *jmdn.*, *jmds.*, *jmd.*, *etw.* werden sowohl semantische Kriterien (u.a. Personen- und Sachbezug) als auch syntaktische Kriterien (u.a. Aktantenrealisierung, Satzgliedfunktion) bei der Einbettung fester Wortverbindungen in den Ko(n)text berücksichtigt. Eine wichtige Ergänzung zur Valenzbeschreibung von verbalen Phraseologismen besteht darin, dass die entsprechenden Nennformen auch Informationen über die Art des Subjekts enthalten. Bei einer Person als Subjekt muss u.a. differenziert werden nach: Geschlecht (weiblich vs. männlich) und Alter (Kinder, Erwachsene). Bei geschlechtsbedingten Restriktionen wie z.B. in *unter dem Pantoffel stehen* kann die Subjektstelle nur von einem Maskulinum eingenommen werden, während in [*jmd.*] *trägt jmdn. auf Händen* die Subjektstelle von einem



Kookkurrenzanalysen stellen ein wichtiges lexikografisches Arbeitsinstrument für die Gewinnung von Lesarten, semantischen Merkmalen und Eigenschaften der Lemmata dar.

Die vielfältigen Möglichkeiten, die ein solches offenes System bieten, werden insbesondere im Bereich der pragmatischen Angaben evident, die für diese Erscheinungen relevant und die nicht mit den von Wörterbuchgeneration zu Wörterbuchgeneration weitergegebenen stilistischen Markierungen zu vergleichen sind. Dazu Steyer (2000, S. 122f.):

Mit Hilfe korpusbasierter Kookkurrenzextraktionen und -analysen wird es möglich sein, zum einen durch Empirie und nicht mehr nur durch Introspektion zu einer Systematisierung sprachlichen Wissens zu gelangen, zum anderen aber auch, pragmatische Bedeutungen zu rekonstruieren, die das Wörterbuch in keinem Fall aufbereiten kann [...].

Die Funktionen vorgesehener Angaben zur Kommunikationssituation bestehen in einer differenzierteren Einordnung präferierter kommunikativer Situationen für die Verwendung eines Lemmas; dies wären u.a. Angaben zur Diskursspezifik, zur Textsortenspezifik, zur Fachsprachenzugehörigkeit, zur Gruppen- und Situationsspezifik, zu Einstellungen und Bewertungen des Sprechers, zum Illokutionspotential. Was so eine korpusbasierte Analyse zur Beschreibung leistet, kann ein phraseologisches Wörterbuch nicht. In Zukunft sollen alle usuellen Wortverbindungen umfassend aufbereitet und mit linguistischen Angaben versehen werden.

Das Deutsche Referenzkorpus (DeReKo) des IDS gehört gegenwärtig weltweit zu den größten und wichtigsten

---

Maskulinum besetzt ist und die Akkusativstelle von einem weiblichen Nomen. Ist das Subjekt eine Sachbezeichnung, erscheint als Markierung *etw.*, wenn sowohl eine Sach- als auch eine Personenbezeichnung möglich ist, so erscheint als Markierung *jmd.* bzw. *etw.*, wobei die Form von *jmd.* vorangestellt ist. Wird *etw.* vorangestellt, so signalisiert dies, dass die Belegung ‚Mensch‘ selten ist.

sprachwissenschaftlichen Sammlungen des geschriebenen Deutsch. Es wird fortlaufend weiterentwickelt. Das Analysemodul ist seit 1995 in das IDS-Korpusrecherchesystem COSMAS integriert und online aufrufbar. Ein erstes Anwendungsgebiet ist die korpuslinguistisch abgesicherte Lexikografie. Die Kookkurrenzanalyse bietet der Lexikografie Möglichkeiten zur Inventarisierung, Klassifikation und Beschreibung usueller Wortverbindungen (z.B. Phraseologismen, Sprichwörter, kommunikative Formeln, Funktionsverbgefüge). Ackermann (2015) konnte am Beispiel von Kookkurrenzanalysen belegen, dass – durch außersprachliche Faktoren bedingt – Frauen nicht mehr das „schwache Geschlecht“ repräsentieren.<sup>29</sup> Dabei ist auf die unterschiedlichen Bedeutungen von *stark* hinzuweisen (physische/geistige Stärke), wie auch auf die Relevanz des äußeren Erscheinungsbildes bzw. von Schönheitsidealen bei Frauen und zunehmend auch bei Männern (vgl. hierzu das auffällige Vorkommen von *muskulös*). Die Kookkurrenzanalysen verdeutlichen auch, dass die Frau immer noch als abhängig vom Mann wahrgenommen wird (präferiertes systematisches Vorkommen von *geschieden*, *verheiratet* oder *alleinstehend* mit *Frau*). Die

<sup>29</sup> Zur korpuslinguistischen Untersuchung der Geschlechterrollen von Mann und Frau im Zeitraum 1996-2013 und zur Rekonstruktion des Wandels von typischen Rollenbildern vgl. Ackermann, Fabian: *Gehören nun die Männer an den Herd? Anmerkungen zum Wandel der Rollenbilder von Mann und Frau*. In: *IDS Sprachreport* 4/2015, S. 12-15. Die Genderlinguistik als junge Forschungsrichtung ist u.a. bemüht, den Einfluss kultureller, sozialer und sprachlicher Faktoren auf stereotype Rollenvorstellungen aufzuzeigen. Dabei werden auch lexikografische Quellen unter dem Aspekt der Transponierung von Gender-Auffassung durch Sprache untersucht. Vgl. hierzu u.a. Nübling, Damaris: *Zur lexikografischen Inszenierung von Geschlecht. Ein Streifzug durch die Einträge von Frau und Mann in neueren Wörterbüchern*. In: *Zeitschrift für Germanistische Linguistik* 37/2009, S. 593-633. Nübling erfasst die (internationale) Forschungslage zur lexikografischen Geschlechterinszenierung, wobei neuere einsprachige Wörterbücher des Deutschen analysiert werden.

korpuslinguistische Untersuchung der Rollenbilder von Mann und Frau belegt, dass „klare Tendenzen vorhanden sind, die auf eine graduelle Verschmelzung der Rollenbilder von Mann und Frau hindeuten. Dies könnte ein Anzeichen dafür sein, dass sich die Gesellschaft zunehmend in Richtung der Gleichstellung von Mann und Frau bewegt.“ (Ackermann 2015, S. 15). Bei der kritischen Durchsicht diverser Wörterbücher können Diskrepanzen zwischen der lexikografischen Fixierung von Gender-Restriktionen und der Verifizierung dieser Markiertheit im aktuellen Sprachgebrauch registriert werden.<sup>30</sup>

Deutliche Anzeichen für einen Wandel im Gebrauch ausgewählter Phraseologismen, die sich primär auf den Menschen beziehen, lassen die Frage nach der Abgrenzung zwischen Restriktion und Gebrauchspräferenz aufkommen. Wie die Restriktionen auf Subjekt- und Objektpositionen [+ mask.] vs. [+ fem.] ausgestaltet sind und ob Gebrauchspräferenzen aufgehoben werden, muss an ausreichenden Textkorpora überprüft werden. Auch kann in vielen Fällen von einer allgemein gültigen oder exklusiven Restriktion nicht ausgegangen werden. Gebrauchspräferenzen müssen durch Einzelanalysen überprüft werden, da ein und dasselbe wendungsinterne (geschlechtstypisch markierte) Element nur bedingt eine Geschlechtsrestriktion bei anderen mit dieser Komponente gebildeten Phraseologismen hervorruft.<sup>31</sup>

<sup>30</sup> Die Wörterbuchanalysen lassen den Schluss zu, dass viele Wörterbücher Informationsdefizite aufweisen. In vielerlei Hinsichten wurden die Mängel traditioneller Wörterbücher hervorgehoben, wobei Beispielklassiker der Phraseologie herangezogen wurden. Exemplarisch ist untersucht worden, inwieweit eine aus der Sicht der potenziellen Benutzer und der Spezifik phraseologischer Erscheinungen angemessene lexikografische Darstellung ausgewählter Phraseologismen gewährleistet wird.

<sup>31</sup> Vgl. hierzu auch Šichová, Kateřina: „Kann er ihr Hörner aufsetzen?“ Zu Geschlechtsspezifität und Restriktionen von deutschen und tschechischen somatischen Phrasemen. In: *Aussiger Beiträge – Germanistische Schriftenreihe aus Forschung und Lehre: Lexikologie und*

Ziel des an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) beheimateten Projekts *Digitales Wörterbuch der deutschen Sprache* (DWDS) ist die Schaffung eines umfassenden, frei zugänglichen Informationssystems, das Auskunft über den deutschen Wortschatz in Vergangenheit und Gegenwart gibt. Die Beschreibung der Kollokationen im DWDS-Wörterbuch (DWDS = Digitales Wörterbuch der deutschen Sprache) durch statistische Methoden erfolgt vollständig korpusbasiert. Das Herzstück der Kollokationsbeschreibung im DWDS-Wörterbuch ist das Wortprofil, das Ergebnis einer automatischen syntaktischen und statistischen Analyse großer Korpora. Es liefert einen Überblick über die statistisch relevanten syntagmatischen Beziehungen eines Wortes.<sup>32</sup> Geyken (2011) erläutert das DWDS-Wortprofil aus Nutzersicht und bietet einen Ausblick auf die Potenziale des Wortprofils.<sup>33</sup> Die Berechnung des DWDS-Wortprofils erfolgt in drei Etappen: Festlegung der zu extrahierenden syntaktischen Relationstypen, Extraktion der Relationen mittels einer automatischen syntaktischen Analyse und Bewertung der statistischen Signifikanz der extrahierten Relationen. Die Voraussetzung für die Aufnahme einer Relation in das Wortprofil ist, dass dafür wenigstens vier Belege im Korpus vorkommen.

Die Korpora, die als Datengrundlage des DWDS-Wortprofils dienen, sind frei wählbar. Die Zusammensetzung und Größe der Korpora sind jedoch für das Wortprofil relevant, da die extrahierten syntaktischen Relationen die im Korpus vorkommenden syntaktischen Nachbarn des Wortes widerspiegeln. Somit erhöht ein breites, nach Textsorten ausgewogenes Korpus (ein

---

*Lexikografie. Aktuelle Entwicklungen und Herausforderungen* 7/2013, S. 211-236.

<sup>32</sup> Das DWDS-Wortprofil kann über die Webseite des lexikalischen Informationssystems DWDS ([www.dwds.de](http://www.dwds.de)) genutzt werden.

<sup>33</sup> Geyken, Alexander: Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora. In: Abel, Andrea/Renata Zanin (Hgg.): *Korpora in Lehre und Forschung*. Bozen/Bolzano 2011, S. 115-137.

allgemeinsprachliches Referenzkorpus), die Qualität des Wortprofils. Spezialkorpora oder spezielle Zeitungskorpora liefern andere Wortprofile als Referenzkorpora. Mit der Größe der Korpora „steigt nicht nur der Variantenreichtum der Gebrauchskontexte, sondern auch die Anzahl der Belege für ein Suchwort bzw. einen Suchausdruck“ (Geyken 2011, S. 116). Wortprofile sind in der Regel nur aussagekräftig, wenn das Lemma wenigstens 500, besser jedoch 1000 Mal im Korpus auftaucht. Unter dieser Zahl ist die Aussagekraft eines Wortprofils nur begrenzt, da viele syntaktische Relationen eher nur ein oder zwei Mal vorkommen und somit kaum nachweisbar ist, dass es sich bei den extrahierten syntaktischen Relationen um typische Beispiele und nicht um Zufallsfunde handelt. Mit wachsender Korpusgröße erhöht sich auch die Anzahl verschiedener Wörter, die hochfrequent im Korpus vorkommen. Eine Korpusgröße von 100 Millionen laufenden Textwörtern ist zu klein, um eine ausreichende Anzahl von Wortprofilen zu extrahieren. So gibt es beispielsweise im 100 Millionen Textwörter umfassenden DWDS-Kernkorpus nur etwa 5.000 Lemmata, die mehr als 1.000 Mal vorkommen, sodass größere Korpora nötig sind, wenn man qualitativ ausreichende Wortprofile für eine Stichwortanzahl erstellen möchte. Knapp ein Fünftel aller Wortprofile basiert auf Lemmata, die mindestens 1000 Mal im Korpus vorkommen, also ausreichend für ein umfassendes Wortprofil sind. Als Schwellwert für die Mindestanzahl des Vorkommens einer syntaktischen Relation wurde die Zahl 4 gewählt. Damit soll verhindert werden, dass okkasionelle Verbindungen fälschlicherweise in das Wortprofil aufgenommen werden. Bei der Erstellung eines hinreichend großen Wortprofils ist man auf Spezialkorpora oder auf eine Mischung aus Referenz- und Zeitungskorpora angewiesen. Der derzeitige Prototyp des DWDS-Wortprofils beruht auf einer Mischung eines Referenz- und eines Zeitungskorpus mit einer Gesamtgröße von 500 Millionen Tokens: dem DWDS-Kernkorpus, einem nach Textsorten ausgewogenen und zeitlich

gleichmäßig gestreuten Referenzkorpus der deutschen Sprache des 20./21. Jahrhunderts (ca. 110 Millionen Tokens) und dem ZEIT-Archiv (1946-2009, ca. 400 Millionen Tokens). Daraus wurden für etwa 90.000 Lemmata Wortprofile extrahiert mit etwa 2.000.000 syntaktischen Relationen.

Nach der digitalen Aufbereitung des historischen Wörterbuchs von Jacob und Wilhelm Grimm, des Etymologischen Wörterbuchs des Deutschen und des Wörterbuchs der deutschen Gegenwartssprache (WDG; 1961-1977) werden die in den Projekten DWDS und Deutsches Textarchiv erstellten Korpora in das DWDS-Informationssystem integriert und gemeinsam mit den Wörterbüchern abfragbar gemacht. Es handelt sich hierbei um das zeitlich und nach Textsorten ausgewogene DWDS-Kernkorpus (100 Millionen Textwörter), das historische Referenzkorpus des Deutschen Textarchivs (ca. 50 Millionen Textwörter) und die vorwiegend aus elektronischen Zeitungsquellen stammenden DWDS-Ergänzungskorpora (*Bild, FAZ, SZ, Welt, Zeit* mit über 2,5 Milliarden Textwörtern). Ziel des DWDS-Wörterbuchs ist die Erstellung eines synchronen Wörterbuchs mit historischen Anteilen (2013-2024), wobei der Grundbestand des DWDS-Wörterbuchs überarbeitet werden soll.

Im Unterschied zu einer reinen Kookkurrenzanalyse werden im DWDS-Wortprofil die syntaktischen Kontexte berücksichtigt. Ein wesentlicher Mehrwert des Wortprofils besteht daher darin, dass alle extrahierten Relationen stets mit den dazugehörigen Satzkontexten im Korpus verknüpft erscheinen und einen Überblick über den Verwendungszeitraum und die semantisch-pragmatischen Kontexte ermöglichen, in denen die syntaktische Relation vorkommt.

Um im Wortprofil statistisch signifikant zu sein, muss ein Wort in einer gewissen Nähe des Suchwortes vorkommen und in einer relevanten syntaktischen Relation (z.B. Adjektiv-Nomen, Verb-Objekt, Genitivattribute von Nomen oder Verb-Präpositionalphrase-Verbindungen) mit dem Suchwort stehen. Das

Wortprofil basiert auf der von Adam Kilgarriff eingeführten Sketch Engine (2004). Das Wortprofil ist auf der Plattform [www.dwds.de](http://www.dwds.de) frei verfügbar. Die Darstellung der Relationen erfolgt als Schlagwortwolke oder in Tabellenform. Diese Wortverbindungen werden zusammen mit den Fundstellen im Korpus, in denen sie auftreten, angezeigt.

Trotz der späteren Etablierung der Korpuslinguistik<sup>34</sup> im deutschsprachigen Raum verdanken ihr die Phraseologie und die Phraseografie wichtige Anregungen. Als Forschungsinstrument dient die Korpuslinguistik der empirischen Untersuchung linguistischer Fragestellungen und der Überprüfung von Hypothesen durch authentische Texte (Korpora). Ein Korpus sollte für den fraglichen Sprachausschnitt repräsentativ sein und den aktuellen Sprachgebrauch reflektieren.

Die Vorteile der Nutzung von Korpora sind offensichtlich. Bestimmte Annahmen in der traditionellen Forschung können durch den von den modernen Methoden eröffneten Zugang zu empirischen Daten belegt/widerlegt oder relativiert werden. Die Auswertung von Datenmengen verdeutlicht zudem wichtige Aspekte des Gebrauchs von Phraseologismen, wobei Verwendungskontexte, die eingehendere semantische Analysen von Phraseologismen erlauben, gleichfalls erhellt werden können. Durch Korpusanalysen können auch Fragen der (ausgeprägten) Polysemie und (Pseudo)Synonymie empirisch gestützt beantwortet werden. Andererseits werden Grenzbereiche der Phraseologie in den Blick genommen werden, die bis dato nicht eingehend erforscht wurden (z.B. Funktionsverbgefüge).

Digitalisierte Korpora stellen große Datenmengen bereit, was eine Verlagerung des Interesses von polylexikalischen Einheiten mit stark ausgeprägter Idiomatizität und Festigkeit zu

<sup>34</sup> Die Korpuslinguistik bezweckt die Erfassung, Beschreibung und Erklärung sprachlicher Erscheinungen natürlicher Sprachen durch die Analyse authentischer Texte, die in Korpora zusammengefasst sind. Vgl. u.a. Lemnitzer, Lothar/Zinsmeister, Heike: *Korpuslinguistik. Eine Einführung*. Tübingen 2006, S. 9.



schwach oder nicht idiomatischen Wortverbindungen bewirkte.<sup>35</sup> Die Ermittlung der Grundform bei Mehrworteinheiten, von Varianten und fakultativen Elementen wird durch eine Korpusanalyse entscheidend erleichtert. Gleichfalls die (genauere) Beschreibung der denotativen Bedeutung, wie auch die Erfassung von Markierungsangaben, die durch relevante Korpusbeispiele belegt werden könnten. Elektronische Datenbanken vermögen u.a. bei phraseologischen Wörterbüchern das Problem der sekundären Alphabetisierung zu lösen. Dadurch entfällt auch die Forderung nach eigenständigen Wörterbüchern für jeden Phraseologismus-Typ, ebenfalls die Frage, welcher Typ in welchem Wörterbuch unter welchem Stichwort nachzuschlagen ist, da diverse Wörterbuchtypen in eine Datenbank integriert/verlinkt werden können.<sup>36</sup> Die Datenbanken können außerdem als Ausgangsbasis zur Erstellung von benutzerspezifischen (gedruckten) Wörterbüchern dienen. Somit würden empirisch fundierte Erkenntnisse Eingang in die Erforschung der Spezifik des phraseologischen Inventars und auch in eine verbesserte lexikografische Bearbeitung in Nachschlagewerken finden.<sup>37</sup>

---

<sup>35</sup> Z.B. sind Kollokationen i. e. S. verstärkt in das Forschungsinteresse gerückt.

<sup>36</sup> Bei der Äquivalentdarbietung sind viele Informationen zur semantischen und pragmatischen Differenzierung sowie zu grammatischen und distributionellen Eigenschaften notwendig. In einem Printwörterbuch ist das in einem wünschenswerten Umfang nicht möglich. Die Lösung stellt deshalb ein virtuelles, hypertextbasiertes Wörterbuch dar, das die Revision und Erweiterung bestehender Printwörterbücher anzuregen vermag.

<sup>37</sup> Zu den Möglichkeiten und Grenzen der Korpusanalyse für die Lexikografie vgl. den Beitrag von Antje Heine (2009, S. 233-250) im Sammelband von Blanco, Carmen Mellado/Patricia Bujan/Claudia Herrero (Hgg.) 2010. Heine distanziert sich von der oft geäußerten Hoffnung, die Korpuslinguistik würde zwangsläufig eine „Revolutionierung der Wörterbücher und damit der Wörterbuchlandschaft“ (S. 234) einleiten und plädiert für eine Weiterentwicklung und Erweiterung der Korpora in Zusammenarbeit mit anderen Wissenschaftszweigen.



Obwohl die Nutzung von Korpusportalen für lexikografische Zwecke begrüßt und als dringend erforderlich eingestuft wurde, warnt Piirainen (2009) davor, bei der Beschreibung von Phraseologismen ausschließlich auf Textkorpora zurückzugreifen. Sie fordert daher auch den zusätzlichen Einsatz von Informantenbefragungen zur Absicherung der Ergebnisse und plädiert für die verstärkte Einbeziehung von Dialekten und Umgangssprachen, um das Übergewicht von Schriftsprache und Textkorpora auszugleichen.<sup>38</sup>

Verbesserte Korpora können z.B. durch Informantenbefragungen und die Zusammenarbeit von Phraseologen und Lexikografen erreicht werden und implizite auch zu besseren (Print)Wörterbüchern führen.

Sollen die Wörterbücher von Grund auf verbessert, ja sogar revolutioniert werden, muss man zu ihrer elektronischen Form übergehen, um die Kluft zwischen dem phraseologisch-phraseografisch Erwartbarem und dem in Printwörterbüchern Machbarem und von den Verlagen auch Zugestandenem überwinden zu können. Da elektronischen Datenbanken nicht so enge Grenzen wie gedruckten Wörterbüchern gesetzt sind, könnte in ihnen die eingangs erwähnte Diskrepanz zwischen Wunsch und Wirklichkeit gemindert werden.<sup>39</sup>

## Literatur

Ackermann, Fabian: *Gehören nun die Männer an den Herd? Anmerkungen zum Wandel der Rollenbilder von Mann und Frau.* In: *IDS Sprachreport 4/2015*, S. 12-15.

---

Eine Erweiterung und ein Ausbau der Korpora sind notwendig, um die Dominanz der Pressesprache auszugleichen.

<sup>38</sup> Vgl. Piirainen, Elisabeth: Dialektale Phraseografie – Randerscheinung, Ergänzung oder Herausforderung einer modernen Phraseografie? In: Blanco, Carmen Mellado (Hg.): *Theorie und Praxis der idiomatischen Wörterbücher.* Tübingen 2009, S. 83-100.

<sup>39</sup> Blanco, Carmen Mellado (Hg.): *Theorie und Praxis der idiomatischen Wörterbücher.* Tübingen 2009, S. 2.

- Blanco, Carmen Mellado (Hg.): *Theorie und Praxis der idiomatischen Wörterbücher*. Tübingen 2009.
- Blanco, Carmen Mellado/Patricia Bujan/Claudia Herrero (Hgg.): *La fraseografía del S. XXI. Nuevas propuestas para el español y el alemán*. Berlin 2010.
- Dobrovol'skij, Dmitrij: *Zur deutschen Phraseografie*. In: *Cahiers d'Etudes Germaniques* 23/1992, S. 161-172.
- Dräger, Marcel: *Phraseologische Nachschlagewerke im Fokus*. In: Korhonen, Jarmo/ Wolfgang Mieder/Elisabeth Piirainen/Rosa Piñel (Hgg.): *EUROPHRAS 2008. Beiträge zur internationalen Phraseologiekonferenz vom 13.-16.8.2008 in Helsinki*. Helsinki 2010, S. 411-421.
- Filatkina, Natalia: *Pragmatische Beschreibungsansätze*. In: Burger, Harald/Dmitrij Dobrovol'skij/Peter Kühn/Neal R. Norrick (Hgg.): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin 2007, S. 132-158.
- Filipenko, Tat'jana: *Beschreibung der Idiome in einem zweisprachigen Idiomatik-Wörterbuch (Deutsch-Russisch)*. In: Vollstedt, Marina (Hg.): *Das Wort. Germanistisches Jahrbuch GUS*, 2002, S. 43-62.
- Geyken, Alexander: *Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora*. In: Abel, Andrea/Renata Zanin (Hgg.): *Korpora in Lehre und Forschung*. Bozen/Bolzano 2011, S. 115-137.
- Jesenšek, Vida/Grzybek, Peter (Hgg.): *Phraseologie im Wörterbuch und Korpus/Phraseology in Dictionaries and Corpora*. Maribor 2014.
- Lemnitzer, Lothar/Zinsmeister, Heike: *Korpuslinguistik. Eine Einführung*. Tübingen 2006.
- Martin, Luis: *Phraseologie im zweisprachigen Wörterbuch Deutsch-Spanisch*. Frankfurt/Main 2001.
- Nübling, Damaris: *Zur lexikografischen Inszenierung von Geschlecht. Ein Streifzug durch die Einträge von Frau und Mann in neueren Wörterbüchern*. In: *Zeitschrift für Germanistische Linguistik* 37/2009, S. 593-633.

- Piirainen, Elisabeth: Dialektale Phraseografie – Rander-scheinung, Ergänzung oder Herausforderung einer moder-nen Phraseografie? In: Blanco, Carmen Mellado (Hg.): *Theorie und Praxis der idiomatischen Wörterbücher*. Tübingen 2009, S. 83-100.
- Ptashnyk, Stefaniya/Erla Hallsteinsdottir/Noah Bubenhofer (Hgg.): *Korpora, Web und Datenbanken. Corpora, Web and Databases: Computergestützte Methoden in der modernen Phraseologie und Lexikographie. Computer-Based Methods in Modern Phraseology and Lexicography*. Baltmannsweiler 2010.
- Šichová, Kateřina: „Kann er ihr Hörner aufsetzen?“ Zu Geschlechtsspezifik und Restriktionen von deutschen und tschechischen somatischen Phrasemen. In: *Aussiger Bei-träge – Germanistische Schriftenreihe aus Forschung und Lehre: Lexikologie und Lexikografie. Aktuelle Entwick-lungen und Herausforderungen 7/2013*, S. 211-236.
- Steyer, Kathrin: *Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten*. In: *Deutsche Sprache* 28/2000, S. 101-125.
- Steyer, Kathrin: Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse. In: Haß-Zumkehr, Ulrike/Kallmeyer, Werner/Zifonun, Gisela (Hgg.): *Ansichten der deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag*. Tübingen 2002, S. 215-236.

## Internetquellen

Digitales Wörterbuch der deutschen Sprache: <http://www.dwds.de/>; 30.06.2016.