

Korpora als Datenressource und Analysegegenstand in der modernen pädagogischen Lexikographie

Korpusi kot podatkovni vir in predmet analize v sodobni pedagoški leksikografiji

Corpora as a data source and a research subject in modern pedagogical lexicography

MILKA ENČEVA

Univerza v Mariboru, Filozofska fakulteta, Oddelek za germanistiko,
Koroška cesta 160, SI 2000 Maribor, milka.enceva@um.si

DOI <https://doi.org/10.18690/978-961-286-019-6.9>

ISBN 978-961-286-810-9

Zahvaljujoč hitremu razvoju informacijske tehnologije so digitalni korpusi postali standard v korpusnem jezikoslovju. Pridobljena spoznanja se uporabljajo v prevodoslovju, leksikografiji in poučevanju jezikov, še posebej poučevanju tujih jezikov. Prispevek se posveča vprašanju, v kolikšni meri je leksikografija v zadnjih dveh desetletjih sledila razvoju informacijske tehnologije in s tem tudi korpusnemu jezikoslovju. Pri tem je raziskovalni interes usmerjen v pedagoško leksikografijo. Z gradnjo obsežnih digitaliziranih besedilnih korpusov, kot so DeReKo Inštituta za nemški jezik v Mannheimu, jedrni korpus Digitalnega slovarja nemščine (DWDS) in drugi, je nastala dobra podlaga za sestavo enojezičnih in dvojezičnih spletnih slovarjev za potrebe učenja nemškega jezika. Kljub obsežnim besedilnim korpusom in visoko razviti tehnologiji pa še vedno ni na voljo ustreznega nabora slovarskih iztočnic, ki bi lahko služil kot podlaga za sestavo spletnih učnih slovarjev. Težavo predstavljajo predvsem kriteriji za izbor, ki so tudi tema prispevka.

Ključne besede: leksikografija, učni slovarji, jezikovni korpusi, nemščina kot tuji jezik, osnovno besedišče

Owing to the rapid development of information technology, digital text corpora have become a standard in modern corpus linguistics in the last 20 years. The insights gained have been used in translation studies, in lexicography and in language teaching, in particular in foreign language teaching. The paper deals with the following question: To what extent has lexicography kept pace with the development of information technology and in turn corpus linguistics in the last 20 years? Our research will focus on pedagogical lexicography. Large digital text corpora such as DeReKo of the German

Language Institute in Mannheim, the German core corpus of German digital dictionary (DWDS) and many others were used to create a sound basis for the compilation of both monolingual and bilingual German online learner's dictionaries. However, in spite of comprehensive text corpora and highly developed technology, still no appropriate lemma lists are available that could be used for compiling online learner's dictionaries. The paper will discuss the criteria for selecting lemmas, one of the main challenges in the field.

Key words: lexicography, learner's dictionaries, text corpora, German as a foreign language, basic vocabulary

1 Einführung

Der vorliegende Beitrag grenzt das breite Feld der pädagogischen Lexikographie ein und konzentriert sich auf die Fragestellung, wie und inwieweit korpusbasierte Sprachdaten bei der Erstellung von Online-Lernerwörterbüchern (OLWB) eingesetzt werden. Textkorpora spielen als Primärquelle im lexikographischen Prozess eine zentrale Rolle. Moderne Wörterbücher, die nicht korpusgesteuert, korpusgestützt oder mindestens korpusvalidiert sind, wirken antiquiert (Engelberg/Lemnitzer 2009: 238).¹ Seit den 90er Jahren bekommt die Lexikographie wichtige Impulse von den sich mit rasantem Tempo entwickelnden Informations- und Kommunikationstechnologien. Heutzutage ist die automatische Beschaffung von Sprachdaten ziemlich einfach, weil wissenschaftlich aufbereitete und strukturierte Textkorpora zusammengestellt werden und das Web selbst als eine riesengroße Textgrundlage genutzt werden kann. Dank der modernen Korpuslinguistik können umfangreiche digitale Korpora recherchiert und als Datenquellen verwendet werden. In der Lexikographie gewinnen korpusbasierte² Recherche- und Analysemethoden immer mehr an Bedeutung. Da drängt sich aber die Frage auf, ob die Lexikographen von den zur Verfügung stehenden elektronischen Ressourcen bei der Erstellung von Online-Wörterbüchern optimal Gebrauch machen?

¹ *Korpusgesteuert* ist ein Wörterbuch, wenn alle eingesetzten Angaben anhand des Textkorpus gewonnen wurden. In *korpusgestützten* Wörterbüchern werden zusätzlich weitere Quellen benutzt, wenn im als Primärquelle gewählten Korpus gesuchte Sprachdaten nicht zu finden sind. *Korpusvalidiert* ist ein Wörterbuch, wenn ein Korpus nur zur Klärung von Zweifelsfällen herangezogen wird (Engelberg/Lemnitzer 2009: 238).

² Neben der Bezeichnung *korpusbasiert* kommt in neuester Zeit in der Fachliteratur wenn auch selten die Bezeichnung *korpusgeleitet* bzw. *datengeleitet* vor. In Lange/Okamura/Scharloth (2015) werden beide Bezeichnungen nicht synonym verwendet. *Datengeleitet* wird im Sinne von *korpusgesteuert* und *korpusbasiert* im Sinne von *korpusgestützt* gebraucht. Im Beitrag wird die Bezeichnung *datengeleitet* nicht gebraucht.

2004 stellt Carolin Müller-Spitzer fest, dass die bestehenden elektronischen Wörterbücher weit hinter den Möglichkeiten des elektronischen Mediums zurückbleiben (Müller-Spitzer 2004: 141). Im Beitrag wird untersucht, wie die Situation 13 Jahre nach dieser Aussage in Bezug auf (zweisprachige) OLWB aussieht. Der Ist-Zustand wird mit dem Soll-Zustand bzw. Kann-Zustand in Beziehung gesetzt, sodass Schlussfolgerungen für die Erstellung neuer moderner (Lerner-)Wörterbücher gezogen werden. Dabei gilt unsere Aufmerksamkeit in den folgenden Ausführungen nur den Wörterbüchern für Fremdsprachenlerner bzw. für Deutschlernende. Als wichtiges Qualitätskriterium eines Wörterbuchs steht in unseren Ausführungen die Lemmasektion und die eng damit verbundenen Ansätze zur Ermittlung des Grundwortschatzes einer Sprache als auch die dafür nötige Vorarbeit zur Zusammenstellung von geeigneten Korpora im Vordergrund.

2 Lemmasektion

Textkorpora als zentrale Primärquelle werden bei der Lemmasektion, bei der Differenzierung eines Lemmas in seine Lesarten, bei der Auswahl passender Gebrauchsbeispiele, bei der Untersuchung semantischer Beziehungen, bei der Identifizierung von Phraseologismen (Kollokationen, Funktionsverbgefügen, Sprichwörtern etc.) verwendet (Engelberg/Lemnitzer 2009: 239f.). Im vorliegenden Beitrag können wir aus Platzgründen nicht auf alle aufgezählten Bereiche eingehen, sondern wir setzen uns nur mit der Anwendung von Korpora bei der Lemmasektion elektronischer Lernerwörterbücher auseinander.

Bei der Lemmasektion werden zuerst die Wörter aus einem Textkorpus nach der Häufigkeit ihres Vorkommens absteigend sortiert und als Häufigkeits- bzw. Frequenzlisten dargestellt. Die Wörter mit den höchsten Frequenzrängen sollten den sogenannten Kernwortschatz (Grundwortschatz) einer Sprache bilden. Im Folgenden wird in Anlehnung an Lange/Okamura/Scharloth (2015) zwischen Kernwortschatz und Grundwortschatz allerdings unterschieden. Der Begriff Kernwortschatz wird bei einer zweckfreien Beschreibung verwendet, der Begriff Grundwortschatz hingegen in sprachdidaktisch ausgerichteten Texten. Da im Beitrag Lernerwörterbücher für Fremdsprachenlerner behandelt werden, kommen beide Bezeichnungen vor.

Wenn versucht wird, den Begriff Kernwortschatz einzelsprachlich genau zu definieren, stößt man auf Schwierigkeiten. Rosengren (1976: 330f.) schlägt vor, diejenigen Wörter zum Kernbereich zu zählen, die in allen Situationsklassen vorkommen können und nicht nur auf eine bestimmte Situationsklasse

beschränkt sind. Der Begriff Situationsklasse selbst wird jedoch nicht definiert und die Formulierung »die in allen Situationsklassen vorkommen können« ist auch sehr allgemein. Da stellt sich auch die Frage, ob nur die am häufigsten vorkommenden Wörter den Grundwortschatz einer Sprache ausmachen.

2.1 Ansätze zur Bestimmung des Grundwortschatzes

Da bei der Bestimmung des Grundwortschatzes bisher verschiedene Kriterien berücksichtigt wurden, können drei unterschiedliche Ansätze unterschieden werden: der kommunikativ-pragmatische, der lexikographische und der frequenzorientierte Ansatz.

2.1.1 Der kommunikativ-pragmatische Ansatz

Dieser Ansatz beruht auf Kategorien aus dem Fachbereich Sozialwissenschaften und betrifft dazu die pragmalinguistische Sprechakttheorie. Es wird von typischen kommunikativen Situationen (Alltagssituationen) und Sprechintentionen ausgegangen. Ihnen werden die entsprechenden sprachlichen Mittel einschließlich der dabei gebrauchten Wörter zugeordnet. Die von Peter Kühn formulierte leitende Frage weist auf die bei der Wörterselektion zu berücksichtigenden Aspekte hin:

Welches lexikalische Material [...] benötigt ein Sprecher/Schreiber [...], um in der Situation [...] über das Thema [...] in der Rolle [...] die kommunikative Intention [...] mithilfe des Kommunikationsmodus [...] erfolgreich durchzuführen? (Kühn 1989: 232)

Leider wird der Begriff *Situation* weder theoretisch noch empirisch fundiert. Die Bezeichnung *Alltagssituation* wird nicht genau festgelegt, es wurden auch keine empirischen Untersuchungen gemacht, mit deren Hilfe bestimmte Alltagssituationen hätten ermittelt werden können.

Der kommunikativ-pragmatische Ansatz spielt im Bereich Deutsch als Fremdsprache nach wie vor eine wichtige Rolle. Das zeigt sich u. a. darin, dass internationale Deutschprüfungen des Goethe-Instituts auf dieser Basis beruhen, was zur Folge hat, dass auch die einschlägigen Lehrwerke für Deutsch als Fremdsprache diesem Ansatz verpflichtet sind.

Wegweisend für den kommunikativ-pragmatischen Ansatz im Bereich DaF sind *Kontaktschwelle Deutsch als Fremdsprache* (Baldegger/Müller/Schneider 1980) und *Profile Deutsch* (Glaboniat et al. 2005).

2.1.2 Der lexikographische Ansatz

Bei diesem Ansatz geht es darum, auf der Grundlage von Sekundärquellen (Wortschatzsammlungen und bestehenden Wörterbüchern) einen Grundwortschatz, einen Kernbereich zu ermitteln. Der lexikographische Ansatz war nie dominant, hatte aber jederzeit seine Vertreter. Das Forschungsvorhaben von Schnörch (2002) ist es zum Beispiel, aus sieben bereits erstellten Grundwortschatzen einen neuen zu extrahieren. Problematisch erweist sich, dass als empirische Basis nur sekundäre Quellen eingesetzt werden, was die Beibehaltung deren Unzulänglichkeiten zur Folge hat und sich dann negativ auf die Ermittlung eines Kernbereiches bzw. eines Grundwortschatzes auswirkt.

2.1.3 Der frequenzorientierte Ansatz

Beim frequenzorientierten Ansatz wird versucht, den Grundwortschatz über Häufigkeitszählungen in Textkorpora zu bestimmen. Vom Nachteil ist dabei die Favorisierung des frequenzorientierten Kriteriums bei der Ermittlung des Grundwortschatzes und dessen Gleichsetzung mit der kommunikativen Wichtigkeit.

Für die deutsche Sprache gibt es mehrere Häufigkeitslisten, die den Grundwortschatz darstellen sollten. Das erste Häufigkeitswörterbuch der deutschen Sprache wurde von Kaeding (1897) erstellt, 1975 wurde es von Ortmann überarbeitet und als *Hochfrequente Wortformen* herausgegeben. Letztere dienen als Grundlage für die Wortschatzliste des Goethe-Instituts für das *Zertifikat Deutsch* (1992), die später in *Profile Deutsch* (2005) übernommen wurde. Es ist erstaunlich, dass die genannten Häufigkeitslisten, welche den Grundwortschatz der deutschen Sprache darstellen sollten, teilweise auf den mehr als 100 Jahre alten Ergebnissen von Kaeding beruhen. Mit dem Einsatz korpuslinguistischer Verfahren in der Fremdsprachendidaktik gewinnt der frequenzorientierte Ansatz wieder an Bedeutung. An dieser Stelle sei auf zwei Ressourcen hingewiesen: auf das Häufigkeitswörterbuch von Jones/Tschirner mit dem Titel *A Frequency Dictionary of German. Core vocabulary for learners* (2006) und auf die Sammlung *Der Grund- und Aufbauwortschatz Deutsch als Fremdsprache nach Themen* von Tschirner (2008).

Seit 2000 haben sich Frequenzlisten des deutschen Wortschatzes etabliert, die auf besonders großen Korpora beruhen. Die Grund-/Wortformenliste DeReWo des Instituts für deutsche Sprache Mannheim wurde auf der Grundlage des zurzeit (2016) mit über sechs Milliarden Wörtern repräsentativsten

Referenzkorpus der deutschen Sprache DeReKo erstellt. Bubenhofer (2012) untersucht, welche Anteile des häufigsten Vokabulars der Sprache sich in Lehrbüchern für Deutsch als Fremdsprache wiederfinden und zieht dafür die DeReWo-Wortformenliste heran. Er stellt fest, dass in der DeReWo-Wortformenliste teilweise Wörter aus dem Alltag der Deutschlernenden fehlen, was mit dem großen Anteil von Zeitungstexten im DeReKo zu erklären ist. Deshalb macht er darauf aufmerksam, dass beim Zusammenstellen von Korpora verschiedene Textsorten ohne übertriebene Fokussierung auf einen einzigen Kommunikationsbereich zu berücksichtigen sind (ebenda: 27). Da liegt die Vermutung nahe, dass eine nur auf die Frequenz beschränkte Ermittlung des Grundwortschatzes nicht die erwarteten Ergebnisse absichern kann, es sei denn, der Frequenzbegriff wird neu definiert.

2.1.4 Der mehrdimensionale Frequenzbegriff

Im Rahmen des Forschungsprojektes *Basic German Vocabulary for Foreign Language Learners: A data-driven Approach*³ wurden Kriterien für die Wortschatzselektion und Methoden zur Berechnung des frequenzbasierten Grundwortschatzes erarbeitet. Die Lexemfrequenzen werden als Grundlage für die Berechnung des Kernwortschatzes gebraucht. Der Frequenzbegriff wird hierbei jedoch anders aufgefasst als von den anderen Vertretern des frequenzorientierten Ansatzes. Es geht um einen mehrdimensionalen Frequenzbegriff und nicht nur um eine Lemmaliste, geordnet nach der Rangfolge der relativen Frequenzen, die als Kriterium für die Aufnahme in den Kernwortschatz dient. Die Frequenz bzw. frequenzorientiert wird allgemeiner als die Distribution von Lexemen bzw. lexikalischen Morphemen betreffend verstanden und durch die Dimensionen Häufigkeit, Stabilität und Produktivität bestimmt. Demzufolge zählen zum Kernwortschatz diejenigen Lexeme,

- die nicht nur häufig vorkommen, sondern über einen längeren Zeitraum gleichmäßig häufig auftreten,
- die nicht nur auf einen einzigen Themenbereich beschränkt sind, sondern in Texten aus unterschiedlichen Themenbereichen gleichmäßig distribuiert sind,
- die nicht nur als lexikalische Morpheme in vielen Ableitungen und Zusammensetzungen auftreten, sondern auch als Lexeme selbst häufig vorkommen

³ Das Forschungsprojekt *Basic German Vocabulary for Foreign Language Learners: A data-driven Approach* wurde durch die Japan Society for the Promotion of Science (JSPS) von 2011 bis 2015 finanziert.

und als lexikalische Morpheme häufiger als Zweitglied in Komposita verwendet werden (Lange/Okamura/Scharloth 2015: 6f.).

Am Beispiel der Lexeme *Frucht* und *Futter* wird veranschaulicht, wie die Lexemfrequenzen berechnet werden. Nach ihrer Vorkommenshäufigkeit gehören beide Lexeme zur gleichen Häufigkeitsklasse. Deshalb werden sie weiter nach ihrer Stabilität in Bezug auf Themenbereiche und nach ihrer Stabilität in Bezug auf die Zeit überprüft. Auf der Basis der Stabilitätswerte sollte das Lexem *Frucht* vor das Lexem *Futter* platziert werden. Es stellt sich heraus, dass Komposita mit *Frucht* häufiger verwendet werden als Komposita mit *Futter*. Wegen Unterschiede hinsichtlich der Stabilität und Produktivität wird trotz der Zugehörigkeit zur gleichen Häufigkeitsklasse *Frucht* Rang 3673 und *Futter* Rang 4729 zugewiesen (Lange/Okamura/Scharloth 2015: 9 ff.).

Besonders aufschlussreich ist das auf empirischen Daten basierende Ergebnis der Studie, dass es kein frequenzgesteuertes Kriterium für die Abgrenzung eines zentralen Wortschatzes von einem Fachwortschatz gibt. Aus diesem Grund vertreten Lange/Okamura/Scharloth (2015) die Ansicht, dass sich der Umfang von frequenzbasierten Grundwortschätzen für Deutsch als Fremdsprache ausschließlich nach didaktischen Kriterien richten kann. Auf den lexikographischen Prozess übertragen würde das heißen, dass bei der Bestimmung des im Lernerwörterbuch aufzunehmenden Wortschatzes kein frequenzgesteuertes Kriterium sondern didaktische Erwägungen im Vordergrund stehen.

Dieses Ergebnis wird im Beitrag als eine Bestätigung von Kempckes Schlussfolgerung angesehen, »daß es *den* Grundwortschatz nicht gibt« (Kempcke 1992: 174). Bei der Konzipierung von De Gruyter Wörterbuch *Deutsch als Fremdsprache* (Kempcke: 2000) setzt er sich mit Problemen bei der Bestimmung eines Grundwortschatzes auseinander, verzichtet auf die Bezeichnung Grundwortschatz und schlägt acht operationalisierbare Kriterien für die Lemmaselektion in einem Lernwörterbuch vor. Frequenz spielt dabei keine vorrangige Rolle.

Kempckes Kriterien stellen eher eine Zusammenfassung von Leitprinzipien dar, denen bei der Erstellung eines jeden kleineren allgemeinen Wörterbuchs einschließlich der Lernerwörterbücher Rechnung getragen werden sollte. Er legt Prioritäten fest, indem er bestimmt, dass allgemeinsprachlicher Wortschatz vor fachsprachlichem Wortschatz, stilistisch Neutrales vor stilistisch Markiertem, Überregionales vor Regionalem, Gebräuchliches vor Veraltetem, Seltenem rangiert, grammatische, semantische, stilistische Schwierigkeiten vor Regulärem, wortbildungsträchtige und polyseme Lemmata vor isolierten und

monosemen rangieren, dass Basiswörter und ihre wichtigsten Ableitungen vor Komposita dominieren und selektive oder exhaustive Behandlung von Wortarten nötig ist (ebenda: 175). Solche Prinzipien sind jedoch zu allgemein, damit sie als verlässliche Grundlage für Entscheidungen über die Aufnahme von Lexemen in das Wörterbuch dienen können. Sie machen die Lemmaselektion intransparent und die Ergebnisse anfechtbar.

Bei einem Vergleich von sieben Grundwortschatzlisten bzw. Lernwortschatzlisten, die im Zeitraum von 1969 bis 2009 erschienen sind,⁴ wurde festgestellt, dass mehr als die Hälfte der Lexeme nur in einer der Listen vorkommen und nur 164 Lexeme in allen Listen zu finden sind. Dieses Ergebnis wurde auch dadurch bestätigt, dass beim Vergleich von je zwei der Grundwortschatzlisten die Anteile der in beiden Listen vorkommenden Lexeme in einem sehr weiten Bereich von 13 % bis 78 % liegen (Lange/Okamura/Scharloth 2015: 3f.). Daraus lässt sich schließen, dass die einzelnen Grundwortschatzlisten nach sehr unterschiedlichen Kriterien erstellt wurden.

Wenn man die oben erwähnte Aussage von Carolin Müller-Spitzer (2004) auf die Bestimmung von Grundwortschatzen bezieht, so wird klar, dass in diesem Bereich großer Nachholbedarf besteht. Vor allem ist wichtig, dass für die Erstellung von Grundwortschatzlisten neben geeigneten Selektionskriterien das eingesetzte Textkorpus eine entscheidende Rolle spielt.

2.2 Zusammenstellung des Textkorpus

Aus Platzgründen wird im Folgenden lediglich auf zwei Korpora eingegangen: auf das *Herder/BYU-Korpus* (Tschirner/Jones 2005) und auf das Korpus, entstanden im Rahmen des Projekts *Basic German Vocabulary for Foreign Language Learners: A data-driven Approach* (2015).

Das Herder/BYU-Korpus ist in Kooperation des Herder-Instituts mit der Birg-ham Young University entstanden. Es umfasst ca. 4,2 Mio. Wortformen und

⁴ Es geht um folgende Grundwortschatzlisten: Hiratsuka, S./Hatori, H.: *4000 Wörter Deutsch zum praktischen Gebrauch*, Tokyo 1969; Baldegger, M./Müller, M./Schneider, G.: *Kontaktschwelle Deutsch als Fremdsprache*, Berlin 1980; James, C./James, Ch.: *Basic German Vocabulary*. Berlin 1991; Reimann, M./Dinsel, S.: *Großer Lernwortschatz Deutsch als Fremdsprache. Deutsch-Englisch*, Ismaning 2006; Lübke, D.: *Lernwortschatz Deutsch. Deutsch-Englisch*, Ismaning 2008; Tschirner, E.: *Deutsch als Fremdsprache. Grund- und Aufbauwortschatz nach Themen*, Berlin 2008; Feuerle, L. M./Schmidt, C. J./Weiss, E.: *Schaum's Outline of German Vocabulary*. New York 2009.

das daraus resultierende Häufigkeitswörterbuch verzeichnet 4000 Lexeme. Das Korpus besteht aus Zeitungs-, Sach-, Fach-, Gebrauchs- und literarischen Texten sowie aus einem Subkorpus der gesprochenen Sprache, wobei ein Verhältnis von 70 : 20 : 10 zwischen deutschen, österreichischen und Schweizer Texten angestrebt wurde. Das Ziel war, ein ausgewogenes und repräsentatives Korpus der deutschen Gegenwartssprache zu schaffen (Tschirner 2005). Leider gibt es dazu aber keine empirischen Untersuchungen, welche kommunikativen Gattungen, Register und Stile als repräsentativ für die moderne deutsche Sprache angesehen werden sollen. Ausgewogenheit und Repräsentativität erweisen sich folglich als vage Begriffe, die unterschiedlich interpretiert werden können.

Deshalb wurde im Projekt *Basic German Vocabulary for Foreign Language Learners: A data-driven Approach* bei der Zusammenstellung des Textkorpus von zwei prototypischen kommunikativen Grundkonstellationen ausgegangen – von mehrfach adressierenden und konzeptionell schriftlichen Texten sowie von persönlich adressierenden und konzeptionell mündlichen Texten. Das Korpus umfasst 845 000 Wörter; somit ist es wesentlich größer als das Herder/BYU-Korpus und es deckt mehrere Jahre (von 1990 bis 2012), damit eine valide Messung von Stabilität und Produktivität gewährleistet werden kann. Als Grundlage für die erste kommunikative Konstellation wurden Texte aus Print- und Online-Zeitungen und Zeitschriften herangezogen, für die zweite kommunikative Konstellation Texte aus Onlineforen. Im Vergleich zum Herder/BYU-Korpus, das mit Texten aus sechs verschiedenen Subkorpora zusammengestellt ist, werden mit diesem Korpus erheblich weniger Textsortentypen abgedeckt. Letzteres ist erst 2015 veröffentlicht worden und es gibt noch keine Untersuchungen, ob der auf seiner Grundlage ermittelte Grundwortschatz tatsächlich den lexikalischen Kernbestand der deutschen Sprache darstellen könnte.

3 Fazit

Im Beitrag wurden Ansätze zur Ermittlung des Kernwortschatzes der deutschen Sprache vorgestellt. Im Laufe der Ausführungen wurde klar, dass moderne Sprachtechnologien ermöglichen, große Sprachdatenmengen schnell und verlässlich zu bearbeiten, komplexe Probleme bei der Bestimmung des Kernwortschatzes sind aber nur dadurch nicht zu bewältigen. Mittlerweile hat es sich herausgestellt, dass der frequenzbasierte Ansatz nicht a priori abzulehnen ist, sondern er soll erneut überdacht werden. Der im Beitrag behandelte mehrdimensionale Frequenzbegriff bietet weitere Möglichkeiten zur Ermittlung *eines* Kernwortschatzes, ohne dass der Anspruch erhoben wird,

dadurch einen einheitlichen allgemeingültigen Kernwortschatz ermitteln zu können. Die Ergebnisse des japanischen Projekts *Basic German Vocabulary for Foreign Language Learners: A data-driven Approach* (2015) zeigen, dass es durchaus akzeptabel ist, von mehreren Grundwortschätzen auszugehen, je nachdem, welche Zwecke man verfolgt und welche Adressatengruppen anvisiert werden. Bei der Bestimmung des Grundwortschatzes sollen didaktische Erwägungen die entscheidende Rolle spielen. Für die pädagogische Lexikographie heißt das, dass sich der Nachholbedarf nicht nur oder nicht vorrangig auf die bessere Anwendung der digitalen Textkorpora beschränken darf, sondern dass neue Orientierungen in der Didaktik und elektronischen Lernlexikographie nötig sind.

Literatur

Markus BALDEGGER, Martin MÜLLER, Günther SCHNEIDER, 1980: *Kontaktschwelle Deutsch als Fremdsprache*. Berlin etc.: Langenscheidt.

Noah BUBENHOFER, 2012: Lehrwerke und Referenzwortschätze. Der Nutzen frequenzbasierter Grundwortschätze. *Grundwortschatz Deutsch: lexikografische und fremdsprachendidaktische Perspektiven*. Hg. Saburo Okamura, Willi Lange, Joachim Scharloth. Tokyo. 13–27.

Das Zertifikat Deutsch als Fremdsprache, 1992. Bonn etc.: Goethe-Institut.

Stefan ENGELBERG, Lothar LEMNITZER, 2009: *Lexikographie und Wörterbuchbenutzung*. 4., überarbeitete und erweiterte Auflage. Tübingen: Stauffenburg.

Lois M. FEUERLE, Conrad J. SCHMIDT/ Edda WEISS, 2009: *Schaum's Outline of German Vocabulary*. New York: McGraw Hill.

Manuela GLABONIAT et al., 2005: *Profile deutsch. Lernzielbestimmungen, Kannbeschreibungen und kommunikative Mittel für die Niveaustufen A1, A2, B1, B2, C1 und C2 des Gemeinsamen europäischen Referenzrahmens für Sprachen*. Berlin: Langenscheidt.

Shigeo HIRATSUKA, Hisahiro HATORI, 1969: *4000 Wörter Deutsch zum praktischen Gebrauch*. Tokyo: Hakusuisha.

Carol JAMES, Charles JAMES, 1991: *Basic German Vocabulary*. Berlin: Langenscheidt.

Randall L. JONES, Erwin TSCHIRNER, 2006: *A Frequency Dictionary of German. Core vocabulary for learners*. London, New York: Routledge.

Friedrich W. KAEDING, Hg., 1897: *Häufigkeitwörterbuch der deutschen Sprache*. Steglitz bei Berlin: Selbstverlag.

Günter KEMPCKE, 1992: Organisationsprinzipien und Informationsangebote in einem Lernerwörterbuch. *Lexikontheorie und Wörterbuch*. Hg. Ursula Brauße, Dieter Viehweger. Tübingen: Niemeyer. 161–243.

Günter KEMPCKE, 2000: *De Gruyter Wörterbuch: Deutsch als Fremdsprache*. Berlin, New York: de Gruyter Verlag.

Peter KÜHN, 1989: Pragmatische Aspekte der Grundwortschatzbestimmung. *Neuphilologische Mitteilungen* 81. Helsinki: Modern Language Society. 230–239.

Willi LANGE, Saburo OKAMURA, Joachim SCHARLOTH, 2015: Grundwortschatz Deutsch als Fremdsprache. Ein datengeleiteter Ansatz. *Deutscher Wortschatz – beschreiben, lernen, lehren. Beiträge zur Wortschatzarbeit in Wissenschaft, Sprachunterricht, Gesellschaft*. Hg. Jörg Kilian, Jan Eckhoff. Frankfurt am Main: Peter Lang. 203–222.

Diethard LÜBKE, 2008: *Lernwortschatz Deutsch. Deutsch-Englisch*. Ismaning: Hueber.

Carolyn MÜLLER-SPITZER, 2004: Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikografischen Prozessen. *Lexikographica* 19. Tübingen: Niemeyer. 140–168.

Wolf D. ORTMANN, 1975: *Hochfrequente deutsche Wortformen*. München: Goethe-Institut.

Monika REIMANN, Sabine DINSEL 2006: *Großer Lernwortschatz Deutsch als Fremdsprache. Deutsch-Englisch*. Ismaning: Hueber.

Inger ROSENGREN, 1976: *Ein Frequenzwörterbuch der deutschen Zeitungssprache. Die Welt. Süddeutsche Zeitung*. 2 Bde. LiberLäromedel/Gleerup: Lund.

Ulrich SCHNÖRCH, 2002: *Der zentrale Wortschatz des Deutschen. Strategien zu seiner Ermittlung, Analyse und lexikografischen Aufarbeitung*. Tübingen: Narr.

Erwin TSCHIRNER, 2005: Korpora, Häufigkeitslisten, Wortschatzerwerb. *Deutsch als Fremdsprache – Konturen und Perspektiven eines Faches*. Hg. Antje Heine, Mathilde Hennig, Erwin Tschirner. München: Iudicium. 133–149.

Erwin TSCHIRNER, 2008: *Grund- und Aufbauwortschatz Deutsch als Fremdsprache nach Themen*. Berlin: Cornelsen.

