

JUPP MÖHRING/FRANZISKA WALLNER

Wortschatzlisten auf dem Prüfstand

Dieser Beitrag beschäftigt sich mit verschiedenen Ansätzen der Konstruktion von Grundwortschatzbeständen für den Bereich Deutsch als Fremdsprache. Am Beispiel des kommunikativ-pragmatisch erstellten Wortschatzes, der in Profile Deutsch erfasst ist, und zwei frequenzorientierten Listen, der DeReWo Wortliste und der Herder/BYU Häufigkeitsliste, werden anhand einer empirischen Gegenüberstellung Stärken und Schwächen der verschiedenen Ansätze herausgestellt. Resümierend werden in diesem Beitrag Desiderate sowie mögliche Lösungsansätze zur validen und empirisch abgesicherten Grundwortschatzbestimmung zur Diskussion gestellt.

1 Einleitung

Nachdem frequenzorientierte Verfahren der Sprachbeschreibung im Zuge der kommunikativ-pragmatischen Wende an Bedeutung verloren hatten, wächst durch die neuen Möglichkeiten digitaler Korpora und Recherchertools das Interesse an empirisch abgesicherten, sprachwissenschaftlichen Analysen. Hierzu zählt beispielsweise die Identifikation von sprachlichen Mitteln, Strukturen und Wortschatzbeständen, die für Fremdsprachenlernende Relevanz besitzen.

Auch die Bestimmung des Grundwortschatzes bleibt davon nicht unberührt. Es werden neue quantitative Verfahren entwickelt, die eine differenzierte Erfassung besonders häufiger und gebräuchlicher¹ Wortschatzelemente ermöglichen. Es stellt sich die Frage, ob und inwieweit diese Verfahren die derzeit dominierenden kommunikativ-pragmatischen Verfahren ergänzen oder gar ersetzen sollten. Dazu werden nach einer kurzen Einführung in die Bestimmung des Grundwortschatzes zwei frequenzorientierte Wortschatzlisten (DeReWo- und Herder/BYU-Liste) mit dem lexikalischen Inventar des dem kommunikativ-

¹ Die reine Häufigkeit wurde als Kriterium für die Relevanz von Wortschatzelementen kritisch hinterfragt und durch weitere Kriterien wie beispielsweise die Streubreite eines häufigen Lemmas über verschiedene Genre, Textsorten und Texte oder die chronologische Verteilung erweitert. Die Ausdehnung von Frequenzanalysen um diese Kriterien wird auch mit den Termini Gebräuchlichkeit, Gebrauchswert oder Usage-Koeffizient erfasst. Der Begriff *Häufigkeit* wird dabei auch oft als Oberbegriff für Verfahren genutzt, die verschiedene gebrauchorientierte Kriterien mit einschließen (vgl. u. a. KOESTERS GENSINI 2009).

pragmatischen Ansatz verpflichteten *Profile Deutsch* verglichen. Ziel ist es, Vor- und Nachteile der beiden Verfahren zur Ermittlung des Grundwortschatzes genauer zu erfassen und empirisch zu untermauern.

2 Ansätze zur Erfassung des Grundwortschatzes

Der Grundwortschatz ist eine „Sammlung von Wörtern [...], die auf einer Kompetenzskala passiv und teilweise aktiv beherrscht werden müssen“ (PERLMANN-BALME 2010: 111). Wie diese Sammlung zustande kommt und wie Wortschatzbestände einzelnen Niveaustufen auf Sprachkompetenzskalen wie dem Gemeinsamen europäischen Referenzrahmen zugeordnet werden können, ist allerdings noch nicht abschließend geklärt. So lässt sich die Debatte, was zum Grundwortschatz einer Sprache gezählt werden sollte, wie der Grundwortschatz einer Sprache erfasst werden könnte und welche Bedeutung einem solchen Grundwortschatz für das Fremdsprachenlehren und -lernen beizumessen ist, mittlerweile seit über 30 Jahren verfolgen. Einigkeit herrscht bislang nur darüber, dass reine Wortschatzlisten im Sprachunterricht eher ungeeignet sind (vgl. KOESTERS GENSINI 2009: 195), sondern in erster Linie eine Grundlage für die Unterrichtsvorbereitung und für die Entwicklung von Lehrwerken, Unterrichtsmaterialien und Curricula darstellen. Hier sind sie jedoch eine äußerst gewichtige Informationsbasis – da sie schließlich auf das lexikalische Inventar ganzer Lerngenerationen Einfluss nehmen können. Der Frage nach der Auswahl des in einen Grundwortschatz aufzunehmenden lexikalischen Inventars sollte daher mit großer Sorgfalt, Präzision und Verantwortung nachgegangen werden.

In der Vergangenheit standen bei der Bestimmung des Grundwortschatzes zunächst quantitative Verfahren im Vordergrund. Während frühe frequenzorientierte Arbeiten auf manuellen Auszählungen beruhen (KAEDING 1898), werden inzwischen korpuslinguistische Verfahren zur Ermittlung der häufigsten Wortschatzelemente genutzt (vgl. JONES/TSCHIRNER 2006, QUASTHOFF/FIEDLER/HALLSTEINDÓTTIR 2011). Dabei kommen mittlerweile auch Verfahren zum Einsatz, die neben der absoluten Vorkommenshäufigkeit weitere Faktoren, wie den Gebrauchswert (Usage-Koeffizienten) der Wortschatzelemente, berücksichtigen. Dieser gibt Auskunft darüber, wie regelmäßig ein Wortschatzelement innerhalb eines Korpus auftritt (vgl. KOESTERS GENSINI 2009: 199). So kann es durchaus vorkommen, dass ein Wort zwar eine hohe Frequenz aufweist, sein Gebrauch jedoch auf nur einen Text innerhalb eines Korpus beschränkt ist. In diesem Fall wäre der Gebrauchswert gering und die Relevanz dieses Wortschatzelements deutlich niedriger als die eines hochfrequenten Wortes

mit gleichmäßiger Verteilung in allen Texten des Korpus. Die Erfassung des Gebrauchswertes wird zunehmend ausdifferenziert. Wurde zunächst die Distribution eines Wortschatzelements in unterschiedlichen Textsorten ermittelt (JONES/TSCHIRNER 2006), werden neuerdings auch Kriterien wie temporale und themenübergreifende Stabilität in die Auswahl der für den Grundwortschatz relevanten Elemente einbezogen (vgl. OKAMURA/LANGE/SCHARLOTH 2012).

In bewusster Abgrenzung zum frequenzorientierten Ansatz entwickelte sich der kommunikativ-pragmatische Ansatz. Dabei wurde argumentiert, dass eine hohe Frequenz nicht unbedingt mit kommunikativer Relevanz korreliert und damit ein frequenzorientierter Grundwortschatz nicht den Kommunikationsbedürfnissen von Fremdsprachenlernenden gerecht werden kann (vgl. OKAMURA/LANGE/SCHARLOTH 2012). Daher orientiert sich die Wortschatzselektion im Rahmen der kommunikativ-pragmatischen Verfahren an Sprachhandlungen und ist stark funktional ausgerichtet.

In der DaF-Praxis findet seit längerer Zeit die kommunikativ-pragmatische Herangehensweise breite Anerkennung und bildet die Grundlage für verschiedene DaF-Zertifikate (bspw. *Zertifikat Deutsch*) und Referenzwerke (*Kontaktschwelle Deutsch als Fremdsprache* 1980/85, *Profile Deutsch* 2005). Kritisiert wird jedoch, dass die Auswahl und Festlegung der thematischen Kategorien sowie die Zuordnung des Wortschatzes zu einzelnen Themen und Niveaustufen wenig transparent und nicht empirisch abgesichert ist (vgl. u. a. ALTMAYER 2004, FUNK 2005, QUETZ 2010, OKAMURA/LANGE/SCHARLOTH 2012).

3 Wortschatzliste ist nicht gleich Wortschatzliste

Für das Deutsche existiert eine ganze Reihe verschiedener Wortschatzlisten, die in unterschiedlichem Maße den oben skizzierten Ansätzen verpflichtet sind. Da die kommunikativ-pragmatischen Verfahren aufgrund der angesprochenen Intransparenz und der mangelnden empirischen Absicherung in der Kritik stehen, soll im Folgenden überprüft werden, inwieweit sich die Wortschatzauswahl in dem vordergründig kommunikativ-pragmatisch ausgerichteten Referenzwerk *Profile Deutsch* tatsächlich von frequenzorientierten Wortschatzlisten unterscheidet.

Die Entscheidung fiel auf *Profile Deutsch*, da es sich dabei um die Umsetzung des Gemeinsamen europäischen Referenzrahmens (GeR) für die deutsche Sprache handelt. Es bildet eine wichtige Grundlage zur curricularen Planung, für die Entwicklung von Lehrwerken und Unterrichtsmaterialien und zur Zusammenstellung von Materialien für Tests und Prüfungen auf einer bestimmten Stufe (vgl. FUNK 2005: 122).

Profile Deutsch bietet eine umfangreiche Zusammenstellung und Einstufung des Wortschatzes für die Stufen A1 bis B2. Das lexikalische Inventar in *Profile Deutsch* wird auf der beiliegenden CD-ROM in vier verschiedenen Teilkapiteln präsentiert: Sprachhandlungen, Allgemeine Begriffe, Thematischer Wortschatz und Wörterbuch. Die Autoren weisen darauf hin, dass es sich dabei um eine künstliche Trennung handelt und zwischen diesen einzelnen Teilkapiteln Überschneidungen existieren (GLABONIAT et al. 2005: 69). Diese sind – mit Ausnahme des Wörterbuchs – wiederum in einzelne Themenbereiche untergliedert, wobei die Gruppierungen im Wesentlichen auf die Einteilung von *Kontaktschwelle Deutsch als Fremdsprache* (BALDEGGER u. a. 1980) zurückgehen. Die Auswahl der sprachlichen Mittel, die in den einzelnen Teilkapiteln aufgelistet und eingestuft werden, folgt einerseits anderen Katalogen, Listen und Referenzwerken (GLABONIAT et al. 2005: 69). Andererseits beruht sie auch auf intersubjektiven Entscheidungen des Autorenteam (ebd.). Detaillierte Informationen zur Auswahl und Zuordnung der sprachlichen Mittel zu den einzelnen Niveaustufen können leider nicht nachgelesen werden.

Zum Vergleich der sprachlichen Mittel, die in *Profile Deutsch* enthalten sind, wurden die Herder/BYU Häufigkeitsliste² sowie eine vom Institut für Deutsche Sprache bereitgestellte frequenzorientierte Wortschatzliste (DeReWo) herangezogen.

Die DeReWo-Liste beruht auf dem Deutschen Referenzkorpus (DeReKo), welches mit 5,4 Milliarden Textwörtern als die weltweit größte linguistisch motivierte Textdatenbank gilt. Diese umfasst ausschließlich geschriebene deutschsprachige Texte aus Deutschland, Österreich und der Schweiz. Einen Großteil davon bilden Zeitungstexte, es sind aber auch belletristische, wissenschaftliche und andere Texte enthalten.³ Für die vorliegende Untersuchung wurde die online verfügbare Grundformenliste von 2012 mit insgesamt 326.946 Einträgen ausgewählt.⁴ Sie wurde auf Grundlage einer DeReKo-Version von 2012 erstellt und umfasst einen Datenbestand, der sich über die vergangenen 30 Jahre erstreckt. Die enthaltenen Grundformen werden nach Häufigkeitsklassen angeordnet. Dabei wird die Frequenz eines Wortes in Bezug zum häufigsten Wort inner-

² Bereits TSCHIRNER (2006) hat die Wortschatzbestände der Herder/BYU-Liste mit den Wortschatzangaben in *Profile Deutsch* verglichen. Er kam u. a. zu dem Ergebnis, dass lediglich 60% der in *Profile Deutsch* für das Niveau A1-B1 (produktiv) enthaltenen Wörtern auch zu den häufigsten 4000 Wörtern des Deutschen gehören (TSCHIRNER 2006: 1283).

³ Ausführliche Informationen zum DeReKo können unter <http://www.ids-mannheim.de/kt/projekte/korpora/> nachgelesen werden.

⁴ Vgl. <http://www.ids-mannheim.de/kl/projekte/methoden/derewo.html>.

halb des Korpus gesetzt. Die Zuordnung eines Wortes x zur Häufigkeitsklasse N sagt aus, dass das häufigste Wort des Korpus 2^n mal häufiger vorkommt als das Wort x (vgl. BUBENHOFER 2012: 16). Innerhalb der Häufigkeitsklassen sind die Wörter in dieser Wortliste wiederum nach ihrer absoluten Häufigkeit geordnet.

Da die DeReWo-Liste auf ausschließlich schriftsprachlichen Texten beruht, wurde zusätzlich auf eine Wortschatzliste zurückgegriffen, die auf dem Herder/BYU-Korpus (TSCHIRNER/JONES 2005) basiert. Das Herder/BYU-Korpus ist mit einem Umfang von 4,2 Millionen Tokens deutlich kleiner als das DeReKo. Es enthält Zeitungstexte, Sach- und Fachtexte, literarische Texte und gesprochene Sprache zu jeweils gleichen Teilen (1 Millionen Tokens). Hinzu kommen Gebrauchstexte mit 200.000 Tokens. Die Texte des Herder/BYU-Korpus stammen aus der Zeit von 1989 bis 2002. Das beschriebene Herder/BYU-Korpus wurde bereits auf verschiedene Art und Weise lexikografisch verwertet (vgl. JONES/TSCHIRNER 2006; TSCHIRNER 2008a; TSCHIRNER 2008b; TSCHIRNER/MACKUS/MÖHRING/PFEIFER 2012). Die ursprüngliche Häufigkeitsliste, auf deren Grundlage verschiedene Wörter- und Übungsbücher entstanden, umfasste die häufigsten 4000 Wörter des Deutschen. Diese Liste wurde 2008 auf die häufigsten 5000 Wörter erweitert (vgl. TSCHIRNER 2008c) und umfasst damit das professionelle Wortschatzminimum (vgl. TSCHIRNER 2008c). Die Liste der häufigsten 5000 Wörter des Deutschen (insgesamt 5050 Einträge) wurde uns in digitaler Form zur Verfügung gestellt.⁵ Sie enthält im Original Informationen zum Rang und der relativen Häufigkeit/1 Million Tokens für jeden Eintrag.

Die DeReWo-Liste liegt im Original als Textdokument⁶ vor. Diese wurde in ein Microsoft Excel Datenblatt übertragen, wobei die einzelnen Informationen (Wortgrundform, Häufigkeitsklasse) in separaten Spalten aufgeführt sind. Da die Einträge (s. o.) innerhalb einer Häufigkeitsklasse wiederum nach der absoluten Häufigkeit geordnet sind, spiegelt die so entstandene Liste gleichfalls eine nicht nur nach Häufigkeitsklassen, sondern auch nach absoluter Häufigkeit geordnete Liste wieder. Um diese Reihenfolge für spätere Bearbeitungen der Liste nachvollziehbar zu halten, wurde in eine separate Spalte der Rang eingefügt. Der Umfang der DeReWo-Liste wurde analog zur Herder/BYU-Häufigkeitsliste dem professionellen Wortschatzminimum von 5050 Einträgen (vgl. TSCHIRNER 2008c) angepasst. Die Anzahl der ausgewählten

⁵ Wir bedanken uns an dieser Stelle bei Prof. Dr. Erwin Tschirner für die Bereitstellung der Herder/BYU-Häufigkeitsliste 1–5000 in digitaler Form.

⁶ derewo-v-ww-bll-320000g-2012-12-31-1.0.txt, zuletzt abgerufen am 10.03.2013.

Einträge entspricht damit auch in etwa der Menge an Einträgen, welche im Rahmen des korpusbasierten English Profile Projekts für niveauspezifische Wortschatzlisten A1-B2 aufgenommen wurden (vgl. CAPEL 2010: 5). Da für die weiteren Analysen reine Wortlisten genutzt werden sollten, war es erforderlich die DeReWo-Liste noch zu bereinigen, denn es waren noch Einträge mit gesonderter Anzeige des Flexionsparadigmas wie „welche(e, er, es)“ oder „vierte(r, s)“ enthalten. Diese wurden an die Aufführungsform der Herder/BYU-Liste angepasst.

Während mit der DeReWo-Liste und der Herder/BYU-Liste bereits maschinenlesbare Wortschatzsammlungen/Daten zur Verfügung standen, musste aus dem lexikalischen Angebot von *Profile Deutsch* (Menüpunkt „Sprachliche Mittel“) zunächst eine Wortschatzliste generiert werden. Dabei haben wir uns auf die Bereiche „Allgemeine Begriffe“ (AB), „Thematischer Wortschatz“ (TW) und „Wörterbuch“ (WB) beschränkt.⁷ Die Wortschatzbestände aus diesen Bereichen wurden zunächst extrahiert und in maschinenlesbare Daten umgewandelt. Berücksichtigung fanden ausschließlich Einträge mit einer Niveaueingabe. Während der Bereich WB eine allgemeine Niveaueingabe zu jedem Eintrag enthielt, wurde in den Bereichen AB und TW in rezeptive und produktive Fertigkeiten unterschieden. Für die weitere Analyse war hierbei die Niveaueingabe zu den rezeptiven Fertigkeiten maßgeblich, u. a. weil es nicht zu jeder rezeptiven Niveaueingabe eine produktive Niveaueingabe gab und ein möglichst großer Wortschatzbereich erfasst werden sollte. Die Wortschatzlisten enthielten in dieser Form jedoch noch zahlreiche (uneinheitliche) Formatierungen, Affixe, Mehrworteinheiten und Metainformationen, welche ein quantitatives Vergleichen untereinander sowie mit den oben beschriebenen frequenzorientierten Listen unmöglich machte. Sie mussten deshalb überarbeitet und bereinigt werden, so dass reine Wortlisten für die einzelnen Bereiche/Niveaus vorlagen. Dabei wurden metasprachliche Grammatikangaben (z. B. „Nullartikel“, „Nominalgruppe + selbst“ oder „wo- + Präposition“) exkludiert, die aufgeführten, i. d. R. nicht idiomatischen Mehrwortverbindungen (z. B. *weg sein, nicht fahren, wie groß, von...bis, es geht* etc.) wurden in die einzelnen Grundformen aufgelöst.

Für die Untersuchung standen uns somit neben den beiden frequenzorientierten Listen (DeReWo und Herder/BYU) drei Listen (AB, TW, WB) zum lexikalischen Inventar von *Profile Deutsch* zur Verfügung.

7 Der Menüpunkt „Sprachliche Mittel“ umfasst außerdem den Bereich „Sprachhandlungen A1–B2“. Dieser wurde für die Untersuchung jedoch ausgeklammert, da er komplexe Strukturen sowie viele metasprachliche Angaben enthält und somit keine geeignete Grundlage für eine Wortschatzliste darstellt.

Ziel der Untersuchung war es, zu überprüfen, inwieweit die frequenzorientiert erhobenen Wortschatzsammlungen mit dem kommunikativ-pragmatisch entwickelten Wortschatzangebot von *Profile Deutsch* Schnittmengen aufweisen. Dabei haben wir uns im Rahmen dieses Beitrags auf den in *Profile Deutsch* für das Niveau A1 dargestellten Wortschatz beschränkt. Diese Beschränkung beruht auf der Annahme, dass der für das Niveau A1 notwendige Grundwortschatz von besonderer Wichtigkeit ist und den weiteren Lernerfolg maßgeblich bestimmt (vgl. TSCHIRNER 2006: 1278f.). Demnach sollten – folgt man frequenzorientierten Ansätzen – die häufigsten Wörter vertreten sein (ebd.).

In einem ersten Schritt wurde hierfür die Schnittmenge aus der Herder/BYU- und der DeReWo-Liste gebildet.⁸ Bei dieser Schnittmenge, im folgenden Kernwortschatz genannt, handelt es sich um einen doppelt empirisch abgesicherten Wortschatzbestand – basierend auf unterschiedlich großen und nach verschiedenen Kriterien aufgebauten Korpora. Etwa 75% der beiden Wortschatzlisten stimmen überein. Das zeigt, dass trotz der unterschiedlichen Verfahren und Kriterien beim Aufbau der Korpora und den daraus resultierenden Wortlisten ein großes Maß an Übereinstimmung herrscht.

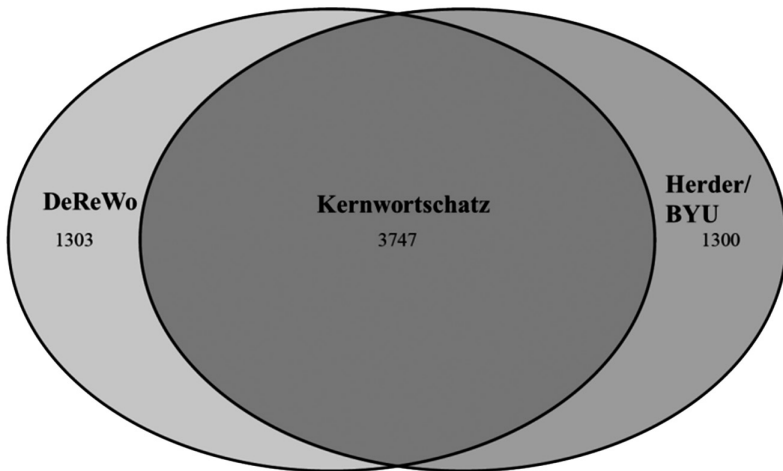


Abbildung 1: Schnittmenge DeReWo mit Herder/BYU

⁸ Die empirische Auswertung der verschiedenen Wortschatzbestände wurde mit Hilfe der Software R vorgenommen. Für die Unterstützung bei der technischen Umsetzung bedanken wir uns herzlich bei David Petroff.

Der 3747 Einträge umfassende Kernwortschatz wurde mit einer Wortschatzliste abgeglichen, die alle Einträge (891) aus den drei Bereichen AB, TW und WB in *Profile Deutsch* auf dem Niveau A1 umfasst. Dabei hat sich gezeigt, dass 61% der in *Profile Deutsch* für A1 angegebenen Wörter auch Bestandteil des Kernwortschatzes sind (Abb. 2).

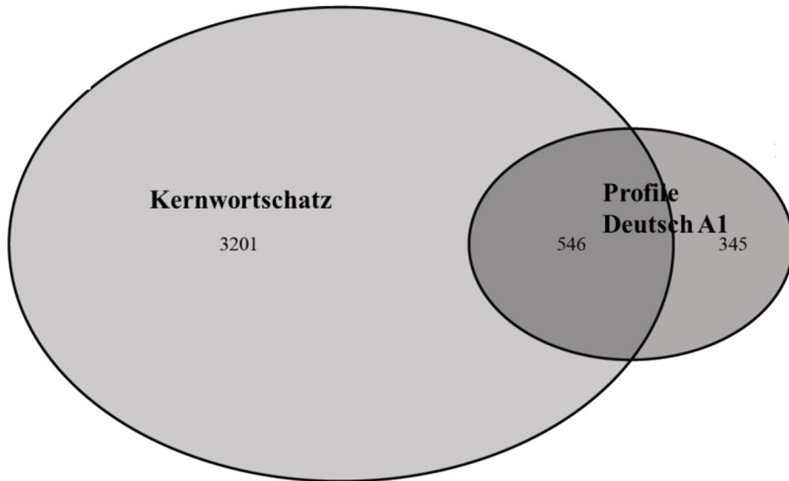


Abbildung 2: Schnittmenge zwischen Kernwortschatz und *Profile Deutsch* A1

Damit gehören ca. 39 % des A1-Wortschatzes in *Profile Deutsch* nicht zu den häufigsten Wörtern des Deutschen. Beispiele für Wörter, die in *Profile Deutsch* A1 gelistet sind, ohne im Kernwortschatz zu erscheinen, sind: *blöd, licht, Kalb, Marmelade, Parkverbot, Raucher* und *zäh* (vgl. auch Tabelle 1).

Da in der häufigkeitsorientierten Fremdsprachendidaktik dafür plädiert wird, die häufigsten Wörter auch als Erstes zu lernen (vgl. NATION 2001, TSCHIRNER 2006: 1283), wurde das Wortschatzinventar von *Profile Deutsch* A1 im Folgenden separat mit den 1000 häufigsten Einträgen der DeReWo- und Herder/BYU-Liste abgeglichen. Hier hat sich gezeigt, dass 38% der häufigsten 1000 Einträge in der DeReWo-Liste und 41% der häufigsten 1000 Einträge in der Herder/BYU-Liste in *Profile Deutsch* A1 enthalten sind. Um zu prüfen, inwieweit die Schnittmengen übereinstimmen, wurden zusätzlich DeReWo 1000 und Herder 1000 miteinander verglichen. 73% der beiden Listen stimmen überein. 54% des in *Profile Deutsch* A1 erfassten Wortschatzes befinden sich weder unter den 1000 häufigsten DeReWo-Einträgen noch sind sie Bestandteil der 1000 häufigsten Einträge der Herder/BYU-Liste (vgl. Abb. 3).

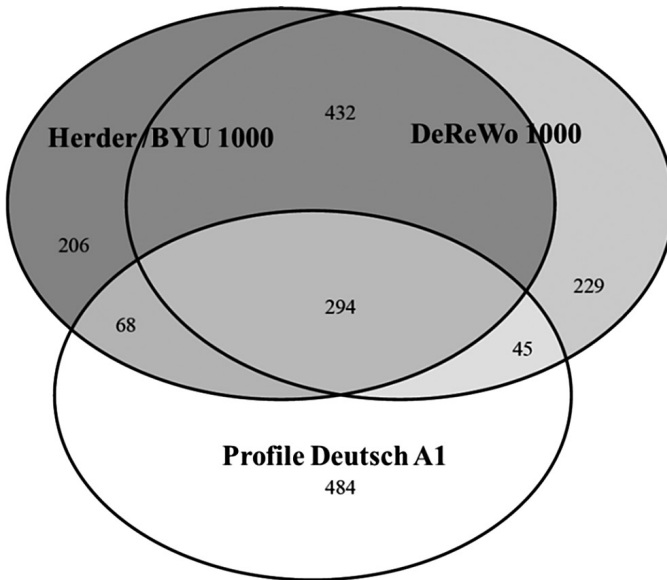


Abbildung 3: Schnittmenge von *Profile Deutsch A1*, DeReWo 1000 und Herder/BYU 1000

Abschließend wurde der *Profile Deutsch*-Wortschatz der einzelnen Bereiche AB, TW und WB (Niveau A1) separat der Schnittmenge der 1000 häufigsten Einträge aus der DeReWo- und Herder/BYU-Liste (insgesamt 726 Wörter) gegenübergestellt. Die größte Deckung zeigte sich mit 55% für den Bereich AB, gefolgt von WB mit 40% und TW mit lediglich 28%. Der „Thematische Wortschatz“ in *Profile Deutsch* weist damit die größten Diskrepanzen mit häufigkeitsorientierten Wortlisten auf.

Die Ergebnisse zeigen, dass frequenzorientierte und kommunikativ-pragmatisch entwickelte Wortschatzsammlungen durchaus Schnittmengen aufweisen. Für das Niveau A1 konnte jedoch festgestellt werden, dass der Anteil des Wortschatzes, der für dieses Niveau in *Profile Deutsch* angeboten wird und gleichzeitig dem oben beschriebenen Kernwortschatz angehört, mit 61% recht gering ist. Die folgende Tabelle zeigt einige Beispiele von Wörtern, die nicht zum Kernwortschatz gehören, aber in *Profile Deutsch* gelistet sind. Ergänzt wurden, soweit verfügbar, Informationen zum Auftreten in den beiden frequenzorientierten Wortlisten.

Wort <i>Profile Deutsch A1</i>	Rang Herder/BYU-Liste	Rang DeReWo-Liste	Häufigkeitsklasse ⁹ DeReWo-Liste
<i>Bleistift</i>	5750	14883	15
<i>Diskette</i>	10213	21566	16
<i>Durst</i>	7429	11209	15
<i>Gabel</i>	5135	15999	15
<i>kopieren</i>	4385	6735	14
<i>licht</i>	nicht enthalten	10985	15
<i>mieten</i>	5022	6630	13
<i>Schulkollege</i>	nicht enthalten	59202	18
<i>zäh</i>	4358	6498	13

Tabelle 1: Wörter aus *Profile Deutsch A1*, die sich außerhalb des Kernwortschatzes befinden, mit Frequenzangabe

Einige dieser Einträge aus *Profile Deutsch* wurden unter kommunikativ-pragmatischen Gesichtspunkten sicherlich zu Recht auf das Niveau A1 eingestuft (bspw. *Durst*, *Gabel*, *mieten*). Obwohl sie nicht zum hochfrequenten Kernwortschatz gehören, kann angenommen werden, dass sie für elementare Kommunikationsbedürfnisse erforderlich sind. Dass sie in den frequenzorientierten Listen einem höheren Rang bzw. einer höheren Häufigkeitsklasse angehören, kann auch auf die Zusammensetzung der den Listen zugrunde liegenden Korpora zurückgeführt werden. So beruht die DeReWo-Liste auf einem überwiegend presssprachlich geprägten Korpus. Auch die im Herder/BYU enthaltenen Textsorten beinhalten offensichtlich einige Themenbereiche nicht, die Grundbedürfnisse und/oder dem Unterrichtskontext angemessene Begriffe enthalten. Andererseits gibt es zahlreiche Einträge, die auf Grund ihrer Häufigkeit zum Grundwortschatz gezählt werden sollten, aber nicht in *Profile Deutsch A1* aufgeführt werden. So unter anderem:

⁹ Für die Herder/BYU-Liste lagen keine Häufigkeitsklassen vor. Die Präsentation der Häufigkeitsklassen für die DeReWo-Liste wurde als nützlich erachtet, da die Angabe beispielsweise die Vergleichbarkeit zu anderen, mit Häufigkeitsklassen versehenen Wortlisten (z. B. QUASTHOFF et al. 2011) ermöglicht. Zudem wäre ein Blick auf andere Einträge der in Tabelle 1 aufgeführten Häufigkeitsklassen (13 bis 18) aufschlussreich hinsichtlich der Beurteilung der Relevanz der entsprechenden Einträge in *Profile Deutsch*.

Wort NICHT enthalten in <i>Profile Deutsch A1</i>	Niveau in <i>Pro- file Deutsch</i>	Rang Herder/ BYU-Liste	Rang DeReWo-Liste	Häufig- keitsklasse DeReWo-Liste
<i>Abschied</i>	B1	2955	2244	11
<i>Bedeutung</i>	B2	403	773	10
<i>Beispiel</i>	k.A. ¹⁰	89	298	8
<i>egal</i>	k.A.	1544	1998	11
<i>Kaffee</i>	B1.	1414	1977	11
<i>kriegen</i>	B1	498	2816	12
<i>laufen</i>	A2	248	327	9
<i>neben</i>	B1	244	251	8

Tabelle 2: Wörter, die zum Kernwortschatz zählen und nicht in *Profile Deutsch A1* enthalten sind¹⁰

Vor allem diejenigen Einträge, welche in beiden Häufigkeitslisten zu den häufigsten 1000 Einträgen gehören z. B.: *Bedeutung*, *Beispiel*, *laufen*, *neben* (vgl. Tabelle 2) sollten hinsichtlich ihrer Einstufung genauer überprüft werden.

4 Fazit

Es hat sich gezeigt, dass die Grundwortschatzlexikographie in Hinblick auf die Auswahl der zu erfassenden Lemmata noch Desiderate aufweist. Der kommunikativ-pragmatisch ermittelte Wortschatz von *Profile Deutsch* umfasst viele wichtige Einträge und entspricht in vielerlei Hinsicht den Ansprüchen von Fremdsprachenlernenden. Allerdings fehlen einerseits zahlreiche wichtige, hochfrequente Ausdrücke und andererseits werden auf dem Niveau A1 einige Ausdrücke geführt, die womöglich nicht von unmittelbarer Relevanz für Lernende auf diesem Niveau sind und selbst dem im Rahmen dieser Untersuchung ermittelten Kernwortschatz (etwa A1-B2) zum Teil nicht angehören. Damit können Lehrkräfte, Deutschlernende, Lehrmaterialentwickler und Curriculumplaner bislang nicht auf ein zuverlässig ermitteltes, valides und

¹⁰ Die Abkürzung k. A. bedeutet, dass das Wort in *Profile Deutsch* zwar enthalten ist, aber in den Bereichen AB, TW und WB keiner Niveaustufe zugeordnet ist.

mit den Niveaustufen des GeR kalibriertes Wortschatzinventar zurückgreifen. Die Vernachlässigung häufigkeitsorientierter Wortschatzlisten führt dazu, dass Lernenden programmatisch Wörter vermittelt werden, die nicht zu einer optimalen Textdeckung¹¹ und damit verbundenen Verstehensleistung beitragen. Das Verständnis alltagsprachlicher Texte und Textsorten ist jedoch von hoher Relevanz für den weiteren Wortschatzerwerb der Lernenden, bspw. den beiläufigen Lexikerwerb beim Lesen (vgl. TSCHIRNER 2006: 1279).

Zudem spielt die präzise und empirisch abgesicherte Beschreibung von (Grund-) Wortschatzbeständen auch für die Konstruktion und Entwicklung von Sprachtests eine maßgebliche Rolle. Wie GROTHJAHN (2011) feststellt, wurden bspw. C-Tests „nach einem sprachlichen Abgleich mit *Profile Deutsch* dem GeR zugeordnet“ (ebd. 135). In Ermangelung alternativer Arbeitsinstrumente ist anzunehmen, dass auch zahlreiche andere Sprachtests für Deutsch als Fremdsprache auf *Profile Deutsch* und hierbei nicht zuletzt auf die dort verzeichneten sprachlichen Mittel zurückgreifen (vgl. GLABIONAT 2010: 133). Die in dieser Studie ermittelte Diskrepanz des in *Profile Deutsch* für das Niveau A1 angegebenen Wortschatzes mit häufigkeitsorientierten Wortlisten wirft hierbei Zweifel auf, ob ein eng an *Profile Deutsch* orientierter Sprachtest hinsichtlich des Wortschatzes als valide angesehen werden kann, oder ob bei der Testentwicklung und auch der Kalibrierung von Tests mit dem GeR empirische Ansätze zur Wortschatzerfassung unterstützend eingesetzt werden sollten.

Zusammenfassend lässt sich sagen, dass empirisch abgesicherte Verfahren für die Bestimmung niveauspezifischer Wortschatzbestände unerlässlich sind. Gleichzeitig sollten aber auch kommunikativ-pragmatische Verfahren Berücksichtigung finden, denn aufgrund der Zusammensetzung bislang verfügbarer Korpora gehören einige Wörter, welche für die Lernenden sicherlich den elementaren Kommunikationsbedürfnissen zuzuordnen sind, nicht unbedingt zu den häufigsten Wörtern der untersuchten Korpora. Zu klären, wie genau sich die verschiedenen Ansätze der lernerorientierten Wortschatzbestimmung ergänzen können, gehört zu den ausstehenden Aufgaben von Lexikographie, Korpuslinguistik und Wortschatzdidaktik. An dieser Stelle sind auch weitere

11 Unter Textdeckung versteht man den Anteil der Wörter eines konkreten Textes, der von einer konkreten Wortliste abgedeckt wird. Verschiedene Studien gehen davon aus, dass mindestens 90 bis 98% der Wörter eines fremdsprachlichen Textes bekannt sein müssen, um impliziten Wortschatzerwerb zu ermöglichen. Bei durchschnittlichen alltagsprachlichen Texten kann davon ausgegangen werden, dass mit den auf Grundlage eines großen Korpus ermittelten häufigsten Wörtern eine hohe Textdeckung erreicht wird (vgl. TSCHIRNER 2005: 144).

Überlegungen zu einer zielgruppenadäquaten Zusammensetzung von Korpora als Grundlage für die frequenzorientierte Bestimmung des Grundwortschatzes erforderlich. Interessante Ansätze bieten hierbei beispielsweise OKAMURA/LANGE/SCHARLOTH (2012), die im Rahmen des Projekts „Korpusgeleiteter Grund- und Aufbauwortschatz Deutsch“ u. a. „prototypische Kommunikationsdimensionen“ (ebd.: 38) einbeziehen und in mehrere Frequenzdimensionen (wie temporale und themenübergreifende Stabilität) differenzieren. Beachtenswert ist auch das English Profile Projekt, in welchem u. a. lernersprachliche Korpora für die niveauspezifische Einstufung von Wortschatz für das Englische als Fremdsprache Berücksichtigung finden (vgl. CAPEL 2010). Ein weiteres Desiderat stellt die quantitativ und qualitativ in der bisherigen Forschung oft vernachlässigte Beachtung verschiedener Lesarten dar, zu der bspw. semantische Annotationsverfahren bei der Aufbereitung der in Sprachkorpora gesammelten Daten beitragen können. Darüber hinaus sollten auch die in dieser Studie weitgehend unberücksichtigten bzw. in Einzelwörter aufgelösten Mehrworteinheiten stärker in den Blickpunkt der frequenzorientierten Grundwortschatzforschung gerückt werden.

Literaturverzeichnis:

- ALTMAYER, Claus (2004): Sprachkultur und Mehrsprachigkeit. Neuerscheinungen zur europäischen Sprachenpolitik, Teil 3 [Sammelrezension]. In: Zeitschrift für interkulturellen Fremdsprachenunterricht 9 (2). Online verfügbar unter <http://zif.spz.tu-darmstadt.de/> [16.03.2013].
- BALDEGGER, Markus/MÜLLER, Martin/SCHNEIDER, Günther (1980): Kontaktschwelle Deutsch als Fremdsprache. Berlin: Langenscheidt.
- BUBENHOFER, Noah (2012): Lehrwerke und Referenzwortschätze. Der Nutzen frequenzbasierter Grundwortschätze. In: Grundwortschatz Deutsch: Lexikografische und fremdsprachendidaktische Perspektiven. Hrsg. v. Saburo Okamura, Willi Lange u. Joachim Scharloth. Tokyo: JGG (Studienreihe der Japanischen Gesellschaft fuer Germanistik, 088), S. 13–27.
- CAPEL, Annette (2010): A1–B2 vocabulary: insights and issues arising from the English Profile Wordlists project. In: English Profile Journal 1 (1), S. 1–11.
- FUNK, Hermann (2005): Wortschatz im Europäischen Referenzrahmen und in „Profile Deutsch“ – nützliche Planungshilfe oder lexikalische Planwirtschaft? In: Deutsch als Fremdsprache. Konturen und Perspektiven eines Faches. Hrsg. v. Antje Heine, Mathilde Hennig u. Erwin Tschirner. München: IUDICIUM, S. 119–132.
- GLABONIAT, Manuela et al. (2005): Profile Deutsch. Gemeinsamer europäischer Referenzrahmen; Lernzielbestimmungen; Kannbeschreibungen; kommunikative Mittel;

- Niveau A1-A2, B1-B2; C1-C2; [CD-ROM Version 2.0 mit Begleitbuch]. Berlin [u. a.]: Langenscheidt.
- GLABONIAT, Manuela (2010): Internationale DaF-Zertifikate. Ein Überblick. In: *Deutsch als Fremdsprache* 47 (3), S. 131–142.
- GROTJAHN, Rüdiger (2011): C-Tests – Aspekte der Validität. In: *Deutsch als Fremdsprache* 48 (3), S. 131–137.
- Institut für Deutsche Sprache, Mannheim (Hrsg.) (2013): Korpusbasierte Wortgrundformenliste DeReWo, v-ww-bll-320000g-2012–12–31–1.0, mit Benutzerdokumentation. Mannheim: IDS Mannheim.
- JONES, Randall L./TSCHIRNER, Erwin (2005): *The Herder-BYU electronic corpus of contemporary German*. Leipzig: Herder-Institut.
- JONES, Randall L./TSCHIRNER, Erwin (2006): *A frequency dictionary of German. Core vocabulary for learners*. London: Routledge (Routledge frequency dictionaries).
- KAEDING, Friedrich Wilhelm (Hrsg.) (1898): *Häufigkeitwörterbuch der deutschen Sprache*. Steglitz bei Berlin: Selbstverl. des Hrsg.
- KOESTERS GENSINI, Sabine E. (2009): Der deutsche Grundwortschatz zwischen Lexikologie und Sprachdidaktik. In: *Deutsch als Fremdsprache* 46 (4), S. 195–202.
- NATION, Paul (2001): *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- OKAMURA, Saburo/LANGE, Willi/SCHARLOTH, Joachim (2012): Methoden der Bestimmung des Kernwortschatzes Deutsch. In: *Grundwortschatz Deutsch: Lexikografische und fremdsprachendidaktische Perspektiven*. Hrsg. v. Saburo Okamura, Willi Lange u. Joachim Scharloth. Tokyo: JGG (Studienreihe der Japanischen Gesellschaft fuer Germanistik, 088), S. 29–44.
- PERLMANN-BALME, Michaela (2010): Grundwortschatz. In: *Fachlexikon Deutsch als Fremd- und Zweitsprache*. Hrsg. v. Hans Barkowski u. Hans-Jürgen Krumm. Tübingen: Francke (UTB Sprachwissenschaft, 8422), S. 111.
- QUASTHOFF, Uwe/FIEDLER, Sabine/HALLSTEINSDÓTTIR, Erla (2011): *Frequency Dictionary German/Häufigkeitwörterbuch Deutsch*. Leipzig: Leipziger Universitätsverlag.
- QUETZ, Jürgen (2010): Der Gemeinsame europäische Referenzrahmen als Grundlage für Sprachprüfungen. Eine kritische Beschreibung des Status quo. In: *Deutsch als Fremdsprache* 47 (4), S. 195–201.
- TRIM, John/QUETZ, Jürgen (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen. Lernen, lehren, beurteilen*. Berlin [u. a.]: Langenscheidt.
- TSCHIRNER, Erwin (2005): Korpora, Häufigkeitslisten, Wortschatzerwerb. In: *Deutsch als Fremdsprache. Konturen und Perspektiven eines Faches*. Hrsg. v. Antje Heine, Mathilde Hennig u. Erwin Tschirner. München: IUDICIUM, S. 133–152.
- TSCHIRNER, Erwin (2006): Häufigkeitsverteilungen im Deutschen und ihr Einfluss auf den Erwerb des Deutschen als Fremdsprache. In: *Atti del XII Congresso Internazionale di Lessicografia. Proceedings – XII Euralex International Congress*. Torino, Italia,

- September 6th-9th 2006. Hrsg. v. Elisa Corino, Carla Marelllo u. Christina Onesti. Alessandria: Edizioni dell'Orso, S. 1277–1288.
- TSCHIRNER, Erwin (2008a): Grund- und Aufbauwortschatz Deutsch als Fremdsprache nach Themen – Lernwörterbuch. Berlin: Cornelsen (lextra).
- TSCHIRNER, Erwin (2008b): Grundwortschatz Deutsch als Fremdsprache nach Themen – Übungsbuch. Berlin: Cornelsen (lextra).
- TSCHIRNER, Erwin (2008c): Das professionelle Wortschatzminimum im Deutschen als Fremdsprache. In: Deutsch als Fremdsprache 45 (4), S. 195–207.
- TSCHIRNER, Erwin/MACKUS, Nicole/MÖHRING, Jupp/PFEIFER, Fleur (2012): Aufbauwortschatz Deutsch als Fremdsprache nach Themen – Übungsbuch. Berlin: Cornelsen (lextra).