

# Korpusbasierte Worthäufigkeitslisten und Wortschatz – eine quantitative und qualitative Analyse am Beispiel des Fremdsprachenlehrwerkes »Deutsch – ein Hit! 1«

*Panagiotis Lymperakakis und Andromachi Sapiridou*

## Zusammenfassung

In dieser Studie werden die Ergebnisse einer Untersuchung präsentiert, die am Beispiel des griechischen Lehrwerkes »*Deutsch – ein Hit! 1*« eine quantitativ-qualitative Analyse des Wortschatzes eines Fremdsprachenlehrwerks Deutsch auf Korpusbasiertheit vorgenommen hat. Sie soll einen ersten Beitrag im Hinblick auf Forschungen zu »korpusbasierten Frequenzangaben aus DaF-Materialien« leisten. Als Datengrundlage wurden die DeReWo Grundformenliste des Instituts für Deutsche Sprache Mannheim (IDS) und die Wortformenliste des Leipziger Wortschatzprojektes der Universität Leipzig herangezogen. Die Analyse wurde mit Hilfe des Softwareprogramms »RANGE« durchgeführt. Die Untersuchungsergebnisse zeigen, dass der im Lehrwerk enthaltene Wortschatz lediglich zu ca. 38 % den 2000 häufigsten Grundformen der deutschen Sprache entspricht. Sie liefern aber nicht nur rein korpusbasierte Frequenzangaben, sondern können auch helfen, den Wortschatz in künftigen Lehrwerken dem tatsächlichen Gebrauch und Bedarf anzupassen.

*»Alles, was man wissen muss, um Erkenntnisse über Sprache zu gewinnen und zu vermitteln, steckt in den Sprachen selbst.«  
(Perkuhn/Belica 2006: 7)*

## 1. Einleitung

Der Gedanke, Korpora und Worthäufigkeitslisten für Deutsch als Fremdsprache zu benutzen, ist nicht neu. Bereits auf der Grundlage des Kaeding-Häufigkeitswörterbuches von 1891 wurde von Ortman (1975) eine Häufigkeitsliste der meistgebrauchten Wortformen erstellt. Auf dieser wurde u. a. der bekannte Zertifikatswortschatz erstellt, welcher bis heute in vielen DaF-Lehrwerken Eingang findet (vgl. Tschirner 2005: 136). Der Hintergrund für dieses Vorgehen bestand darin, dass die am häufigsten gebrauchten Wörter einer Sprache vermittlungsrelevanter

sind als solche, die weniger häufig gebraucht werden. Laut Tschirner basieren jedoch die meisten Grund- und Aufbauwortschätze größtenteils auf der Grundlage dieser ersten sprachstatistischen Ergebnisse und sind demnach veraltet und für den heutigen Sprachgebrauch wenig repräsentativ.

Die Verwendung von Korpora bzw. Korpusdaten spielt in der Fremdsprachenvermittlung und der Fremdsprachenerwerbsforschung eine zunehmende Rolle (vgl. Lüdeling/Walter 2009: 1). Lüdeling und Walter konstatieren zu Recht, dass es bislang keine korpusbasierten Frequenzangaben aus DaF-Lehrmaterialien gibt (Lüdeling/Walter 2009: 12). Aus diesem Grund soll mit der folgenden Untersuchung ein erster Beitrag hierzu geleistet werden.

In Anlehnung an Nation (2001) haben empirische Untersuchungen ergeben, dass die häufigsten 2000 Lexeme ca. 90 % der Lexeme von Alltagsgesprächen, ca. 87 % von literarischen Texten und ca. 80 % von Zeitungstexten erfassen. Nation führt unter Berücksichtigung eines Kosten-Nutzen-Vergleichs an, dass beim Wortschatzerwerb im Fremdsprachenunterricht die Konzentration auf die Wörter zu lenken ist, die auch tatsächlich am häufigsten in einer Sprache gebraucht werden. Daraus folgt, dass die Wörter zu ermitteln sind, die am häufigsten in der deutschen Sprache gebraucht werden, die dann von den Lernern zu lernen sind.

Das übergeordnete Ziel eines jeden Fremdsprachenunterrichts ist es, die Lerner u. a. kommunikativ handlungsfähig zu machen. Folglich müsste der Wortschatz, der gelehrt und gelernt wird, einerseits Wörter enthalten, die in typischen Kommunikationssituationen vorkommen, und andererseits dem typischen Wortschatzgebrauch des Zielsprachenlandes entsprechen.

Da die wichtigste Quelle für den Erwerb des neuen Wortschatzes ohne Zweifel das Lehrwerk bildet, erschien es uns interessant, den Wortschatz eines DaF-Lehrwerks auf seine Korpusbasiertheit hin zu untersuchen. Untersuchungsgegenstand bildete dabei der erste Band des Fremdsprachenlehrwerkes »Deutsch – ein Hit! 1«. Es wurde mit Hilfe eines Softwareprogramms auf seine Orientierung auf korpusbasierte Worthäufigkeitslisten hin überprüft. Dabei wurde der behandelte Wortschatz im Lehrwerk mit den ersten 2000 häufigsten deutschen Wörtern verglichen.

Im ersten Schritt sollen zunächst das Fremdsprachenlehrwerk (2.), die Worthäufigkeitslisten (3.) und das zur Analyse benutzte Software-Programm (4.) vorgestellt werden. Um eine computer-

gestützte Analyse zu ermöglichen, mussten Änderungen vorgenommen werden, die einerseits auf funktionalen Parametern des Softwareprogramms und andererseits auf dem Fremdsprachenlehrwerk und den Worthäufigkeitslisten an sich beruhen (5.). Dies betrifft die methodische Vorgehensweise der Analyse. Im Anschluss (6.) erfolgen eine quantitative und qualitative Analyse des Lehrwerkes auf der lexikalischen Ebene. Einige grundsätzliche Probleme, die bei der Studie auftraten, werden in (7.) behandelt. In (8.) wird ein Ausblick auf Folgestudien eröffnet.

Curriculare Richtlinien des griechischen Bildungsministeriums und des Pädagogischen Instituts bezogen auf den Wortschatz bildeten keine Kriterien für die vorliegende Untersuchung.

## 2. Das Fremdsprachenlehrwerk »Deutsch – ein Hit! 1«

Das griechische Fremdsprachenlehrwerk »Deutsch – ein Hit!« besteht aus jeweils drei Kurs-, drei Arbeits- und drei Lehrerhandbüchern für die jeweils drei gymnasialen Stufen des griechischen Gymnasiums. Untersuchungsgegenstand bildet im Rahmen dieser Untersuchung nur der erste Band, also »Deutsch – ein Hit! 1«. Aus den curricularen Richtlinien und der Notwendigkeit eines Fremdsprachenlehrwerkes für griechische Fremdsprachenlerner, entstand das Lehrwerk »Deutsch – ein Hit! 1«, das seit dem Schuljahr 2006 in der ersten gymnasialen Klassenstufe des griechischen Gymnasiums eingesetzt wird. »Deutsch – ein Hit! 1« richtet sich an jugendliche Anfänger ab 12 Jahren. Zielgruppe sind demnach griechische Fremdsprachenlerner, die in ihrem Heimatland Deutsch als Fremdsprache lernen. Das Lehrbuch führt zur GER-Niveaustufe A1. Es umfasst ca. 180 Seiten und besteht aus einem übersichtlichen

Inhaltsverzeichnis, aus dem die Inhalte und deren Relevanz für Kommunikation, Textsorten, Grammatik, Wortschatz, Strategien usw. leicht entnommen werden können. Das Lehrbuch besteht aus 10 Lektionen. Laut dem Lehrerhandbuch (LHB) ist die erste Lektion eine Einstiegslektion, die Lektionen 1 bis 3 bilden die erste Einheit des Lehrbuches, die Lektionen 5 bis 7 bilden die zweite Einheit. Die Lektionen 4 und 8 sind sogenannte Plateaulektionen und dienen der Wiederholung und freien Anwendung der bereits gelernten Sprachphänomene, wobei Projektarbeit im Vordergrund steht. Die Lektionen der beiden Einheiten bestehen aus jeweils einem A-Modul als Einstieg, mehreren B-Modulen (neuer Lehrstoff), C-Modulen (Grammatik) und D-Modulen (Landeskunde, Sprachmittlung und Strategien). Die Einheitslektionen enden mit einer Grammatik- und einer Wortschatzrubrik. Am Ende des Lehrbuches befindet sich ein gesonderter Anhang zu jeder einzelnen Lektion, die gemäß dem LHB als zusätzliches Übungsmaterial entweder zur weiteren Vertiefung oder zur Differenzierung dienen soll. Das Lehrbuch schließt mit einer Grammatikübersicht, einer alphabetischen Wortliste und einer kleinen Übersicht mit Verben, die Konjugationsunterschiede zu den regelmäßigen Verben aufweisen. Das Lehrbuch zeichnet sich durch eine große Bildervielfalt und Farbigkeit aus; an vielen Stellen wird für zusätzliche Übungen auf das Arbeitsbuch verwiesen.

Den einführenden Bemerkungen des LHB ist weiterhin zu entnehmen, dass »*Deutsch – ein Hit! 1*« dem kommunikativ-pragmatischen Ansatz verpflichtet ist und Aspekte des interkulturellen An-

satzes sowie der Multikulturalität mit einbezieht. Dass es sich hierbei um ein modernes Lehrbuch handelt, lässt sich anhand folgender Punkte demonstrieren: Nicht nur die vier Fertigkeiten Leseverstehen, Hörverstehen, Sprechen und Schreiben spielen eine wichtige Rolle, sondern auch Sprachmittlung findet in diesem Lehrbuch Eingang, was in vielen anderen modernen Lehrbüchern gar nicht berücksichtigt wird. Ferner wird das Ziel der Autonomie des Lerner durch spezifische Lernstrategien verfolgt, dabei stehen Motivation und Aktivität der Lerner im Vordergrund, selbstentdeckendes und spielerisches Lernen werden ebenfalls gefördert. Darüber hinaus haben Lernende Wahl- und Entscheidungsmöglichkeiten sowie die Möglichkeit, in kleinen Projekten zu arbeiten.

### 3. Die DeReWo korpusbasierte Grundformenliste des IDS Mannheim und das Wortschatzportal der Universität Leipzig

Für die Analyse des Wortschatzes auf Korpusbasiertheit wurden als Arbeitsgrundlage die DeReWo Grundformenliste und die Wortformenliste des Leipziger Wortschatzprojektes herangezogen. Im Folgenden soll auf beide Worthäufigkeitslisten eingegangen werden.

#### 3.1 Die DeReWo Grundformenliste

Als Arbeitsgrundlage diente zum einen die DeReWo Grundformenliste (derewo-v-30000g-2007-12-31-0.1)<sup>1</sup> des Instituts für Deutsche Sprache Mannheim (IDS), welche auf dem Deutschen Referenzkorpus (DeReKo) basiert. DeReKo stellt weltweit die größte Datensammlung deutscher Texte dar und soll als Grundlage für empirisch linguistische Untersuchungen der deutschen Sprache dienen

1 <http://www.ids-mannheim.de/kl/derewo/derewo-v-30000g-2007-12-31-0.1.zip>

(vgl. Kupietz/Keibel 2009). Die DeReWo Grundformenliste beinhaltet 30.000 Einträge in Grundformen und die Daten stammen aus den Jahren 1990–2007. Repräsentativität, Aktualität und Ausgewogenheit sind mit diesem Referenzkorpus für den Untersuchungsgegenstand weitestgehend soweit gesichert, als dass eine umfangreiche elektronische Stichprobe deutschsprachiger Texte aus Gegenwart und jüngerer Vergangenheit der BRD, der ehemaligen DDR, der Schweiz und Österreichs darin enthalten sind. Das Deutsche Referenzkorpus (DeReKo) enthält belletristische, wissenschaftliche und populärwissenschaftliche Texte, eine große Anzahl an Zeitungstexten sowie eine breite Palette weiterer Textarten (vgl. Perkuhn 2009). In Anlehnung an empirischen Untersuchungen sollen für die quantitative und qualitative Analyse nur die ersten 2000 Einträge dieser Liste herangezogen werden.

### 3.2 Die Wortformenliste des Leipziger Wortschatzprojektes

Unter dem Wortschatzportal der Universität Leipzig kann die Wortliste mit den 10000 meist gebrauchten Wörtern der deutschen Sprache heruntergeladen werden, die im Rahmen eines Wortschatzprojektes<sup>1</sup> der Abteilung »Automatische Sprachverarbeitung des Instituts für Informatik« der Universität Leipzig<sup>2</sup> entstanden ist (Quasthoff/Wolff 1999). Sie wurde auf der Grundlage der gesammelten Daten (Texte großer Online-Nachrichtenportale u. ä.) des Wortschatzprojektes mit automatisierten Mitteln (Zerlegen der Texte in Sätze und Wörter, Zählen der Anzahl des Vorkommens aller

Wörter usw.) erstellt. Da die Häufigkeitsliste aus Wort- und nicht aus Grundformen besteht, mussten die Einträge zuerst lemmatisiert werden. Dabei mussten so viele Wortformen lemmatisiert werden, bis die ersten 2000 Grundformen aus der Leipziger Wortformenliste erreicht wurden.

Die Entscheidung, diese beiden Listen für die Analyse des Wortschatzes heranzuziehen, basierte darauf, dass sowohl die Leipziger Liste als auch die DeReWo-Liste online frei zur Verfügung stehen und dass beide Frequenzlisten versuchen, die deutsche Standardsprache widerzuspiegeln. Ein Korpus der Jugendsprache steht nicht online frei zur Verfügung. Eine Worthäufigkeitsliste, die auf einem Korpus der Jugendsprache basiert, wäre für die Analyse des Lehrwerkes auch nicht geeignet. Sie würde zahlreiche umgangssprachliche Einträge enthalten, die einerseits nicht der Standardsprache und andererseits nicht dem lexikalischen Bestand eines Lehrwerkes der GER-Niveaustufe A1 entsprechen.

### 4. Das Software-Programm »RANGE«

Für die quantitative und qualitative Analyse wurde das Software-Programm »RANGE« (Heatley/Nation/Coxhead 2002) verwendet<sup>3</sup>. »RANGE« kann Wörter (Vokabeln) aus bis zu 32 verschiedenen Texten miteinander vergleichen und verfügt über folgende Funktionalitäten: Verteilungshäufigkeit, Frequenzhäufigkeit, Bestimmung der Wortfamilien, Überprüfung der Textdeckung anhand von Wortlisten, Überprüfung der Wortschatzdeckung verschiedener Texte usw. Das Programm listet alle Wörter in

1 <http://wortschatz.uni-leipzig.de/html/wliste.html>

2 <http://www.asv.informatik.uni-leipzig.de>

3 Für den Hinweis, das Software-Programm »RANGE« zu verwenden, danken wir Thomaï Alexiou und Nikolaos Konstantakis, die eine Analyse für das Englische als Fremdsprache durchgeführt haben (vgl. Alexiou/Konstantakis 2007: 59–66).

einem Ranking nach »in der Liste x enthalten« und »in keiner Liste enthalten« auf. Die Einspeisung der zu überprüfenden Texte muss in txt-Format erfolgen. Der Band »*Deutsch – ein Hit! 1*« steht online auf der Webseite des Pädagogischen Instituts zur Verfügung und konnte mit der copypaste Funktion in das Programm als txt-Datei hineinkopiert werden.<sup>1</sup>

Aufgrund des Funktionsumfangs des Softwareprogramms<sup>2</sup> mussten für die Analyse folgende Punkte berücksichtigt werden: Das Programm ist nicht in der Lage, das Sonderzeichen »ß« zu erkennen. Es differenziert nicht zwischen unterschiedlichen syntaktischen Funktionen, die ein Wort innerhalb eines Satzes übernehmen kann (z. B. Präposition »zu« als Infinitivkonjunktion, als Vergleichspartikel, als Verbalpräfix oder als Adverb). Darüber hinaus erfolgt keine Differenzierung zwischen Groß- und Kleinschreibung. Folglich unterscheidet das Programm nicht zwischen Verben und nominalisierten Verben. Weiterhin erfolgt keine Differenzierung zwischen mehrdeutigen Wörtern. Bei all diesen Fällen zählt »RANGE« alle unterschiedlich vorkommenden Wortformen unabhängig von ihrer syntaktischen und semantischen Funktion zu ein und demselben Eintrag, weshalb es bei solchen Einträgen zu einer erhöhten Häufigkeit führen kann. Darüber hinaus werden Zahlen und Doppelnennungen als Error ausgegeben, welches die Entfernung dieser Einträge erforderte. Aus all dem ergab sich eine Vorarbeit, die darin be-

stand, Einträge entweder zu entfernen oder entsprechend zu bearbeiten. Dies betraf nicht nur die Worthäufigkeitslisten, sondern auch das Fremdsprachenlehrwerk.

## 5. Aufbereitung des Fremdsprachenlehrwerkes und der korpusbasierten Worthäufigkeitslisten

Für die softwaregestützte Analyse mussten sowohl das Fremdsprachenlehrwerk »*Deutsch – ein Hit! 1*« als auch die Worthäufigkeitslisten aufbereitet werden.

### 5.1 Das Fremdsprachenlehrwerk

Für die Analyse des ersten Bandes wurden alle Lektionen und der gesonderte Anhang zu jeder einzelnen Lektion berücksichtigt. Die Grammatikübersicht und die alphabetische Wortliste wurden nicht miteinbezogen. In Anlehnung an Tschirner (vgl. 2005: 136) sind Grundformen<sup>3</sup> für den Fremdsprachenunterricht interessanter als Wortformen, da der Lernaufwand für ein neues Lexem deutlich größer ist als beispielsweise für eine weitere grammatisch realisierte Form eines bereits bekannten Wortes. Wortformenlisten wären demnach repräsentativer, wenn sie auf Grundformen reduziert würden. Deswegen mussten alle im Lehrwerk enthaltenen Wortformen auf ihre Grundformen zurückgeführt (lemmatisiert) werden. Das Lehrbuch enthält zahlreiche griechische Erläuterungen, die entfernt werden mussten. Entfernt wurden auch Anglizismen oder Wörter aus anderen Fremdsprachen, die nicht Bestandteil der deutschen Sprache sind. Da das Pro-

1 Online unter: <http://www.pi-schools.gr/lessons/german>

2 »RANGE« wurde für die Bewertung von englischen Texten entwickelt, kann aber auch für andere Sprachen verwendet werden.

3 Grundformen sind alle unflektierten Simplizia verschiedener Wortarten, unflektierte Ableitungen und Komposita. Das heißt, Substantive sind grundsätzlich auf ihren Nominativ Singular, Verben auf ihren Infinitiv, Präverbfbügungen auf den Infinitiv der jeweiligen Präverbfbügung und Adjektive auf ihren Positiv zurückzuführen.

gramm keine Zahlen erkennt, mussten alle Zahlensymbole entfernt werden. Für die Analyse war ebenfalls notwendig, das Sonderzeichen »ß« durch »ss« zu ersetzen.

Weitere Änderungen waren:

- Entfernung von Personennamen, Firmenbezeichnungen und Straßennamen, weil die hohe Anzahl dieser Eigennamen die Ergebnisse negativ beeinflussen würde, zumal u. E. Eigennamen zum Erlernen einer Sprache nicht notwendig erscheinen. Städte- und Ländernamen wurden hingegen beibehalten.
- Expansion von Abkürzungen in ihre Vollformen
- Entfernung von Internetadressen
- Ersatz diskontinuierlicher durch kontinuierliche Konstruktionen (... *biege ... ab* durch ... *abbiegen*)
- Ersatz von Getrennschreibung durch Zusammenschreibung
- Entfernung zahlreicher Grammatikbegriffe (aus dem gleichen Grund wie bei den Eigennamen) und
- Korrektur vorhandener Druckfehler

## 5.2 Aufbereitung der korpusbasierten Worthäufigkeitslisten

Aufgrund des softwarespezifischen Funktionsumfangs mussten beide Worthäufigkeitslisten wie folgt angepasst werden, um eine Analyse mit dem Software-Programm »RANGE« zu ermöglichen:

- Rückführung der Wortformen in Grundformen (Leipziger Liste)
- Entfernung von Groß- und Kleinschreibung (Ein, ein)
- Entfernungen von Doppelnennungen (ein, ein)
- Ersatz des Sonderzeichens »ß« durch »ss«
- Entfernung von Eigennamen bis auf Städte- und Ländernamen
- Expansion von Abkürzungen in Vollformen

- Entfernung nicht-üblicher Abkürzungen
- Entfernung von Anglizismen (soweit vorhanden).

## 6. Quantitative und qualitative Analyse des Wortschatzes

Nachdem nun das Softwareprogramm »RANGE«, die Worthäufigkeitslisten und die Aufbereitung der Listen und des Fremdsprachenlehrwerkes vorgestellt wurden, werden im Folgenden zunächst die Analyseergebnisse präsentiert. Das Fremdsprachenlehrwerk bildete in seiner lemmatisierten Form den zu überprüfenden Basistext, während die aufbereitete Grundformenliste »DeReWo« und die lemmatisierte Wortliste des Leipziger Wortschatzprojektes die Basislisten darstellten. Es wurde überprüft, inwieweit sich der im Fremdsprachenlehrwerk enthaltene Wortschatz mit den 2000 häufigsten Einträgen beider Listen deckt. Nach der quantitativen folgt die qualitative Analyse.

### 6.1 Quantitative Analyse

Die folgende Tabelle enthält die Analysewerte, die mit Hilfe des Softwareprogramms »RANGE« berechnet wurden.

Tabelle 1: Deckung des Wortschatzes mit den 2000 häufigsten Wörtern

Lehrwerk »Deutsch – ein Hit! 1«	DeReWo (Grundformenliste)	Leipziger Wortliste (lemmatisiert)
a. 1–1000	454 (27,8 %)	432 (26,5 %)
b. 1001–2000	175 (10,7 %)	193 (11,8 %)
enthalten (a+b)	629 (38,5 %)	625 (38,3 %)
c. nicht enthalten	1003 (61,5 %)	1007 (61,7 %)
insgesamt	1632 (100 %)	1632 (100 %)

Das Fremdsprachenlehrwerk besteht aus insgesamt 1632 Grundformen. 454 (27,8 %) Grundformen tauchen in den ersten 1000 Einträgen der DeReWo Grundformenliste und 432 (26,5 %) der

Abb. 1: Schnittmenge der Worthäufigkeitslisten

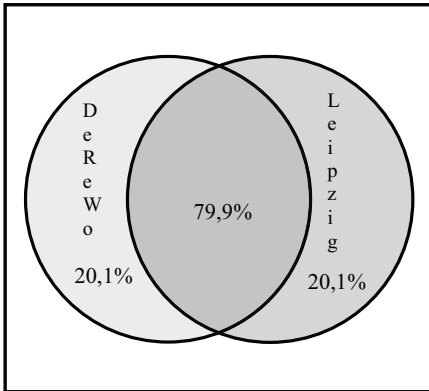
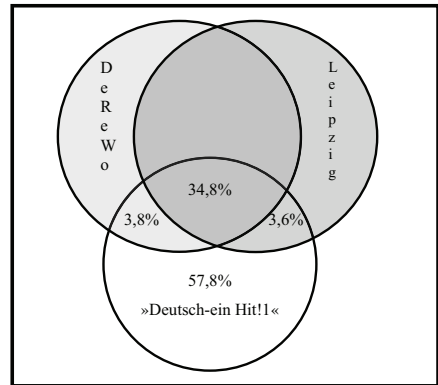


Abb. 2: Schnittmenge des Lehrwerkes und der Worthäufigkeitslisten



lemmatisierten Wortliste des Leipziger Wortschatzprojektes auf. Für die nächsten 1000 Einträge hingegen ist die Deckung niedriger. Nur 175 (10,7%) Einträge des Lehrwerkes entsprechen den Einträgen der DeReWo-Liste und 193 (11,8%) der Leipziger Liste. Insgesamt tauchen somit 629 (38,5%) Einträge des Lehrwerkes in der DeReWo-Liste und 625 (38,3%) Einträge in der Leipziger Liste auf. Daraus folgt, dass mehr als die Hälfte der Einträge des Fremdsprachenlehrwerkes nicht zu den 2000 häufigsten Wörtern zählen: 1003 (61,5%) in Bezug auf die DeReWo-Liste und 1007 (61,7%) auf die Leipziger Liste. Interessant ist hier, dass ein Vergleich der Treffer der DeReWo-Liste und der Treffer der Leipziger Liste einen Unterschied von nur 0,02% ergab, also eine geringe Differenz von nur 4 Einträgen. Eine Analyse der beiden Wortlisten hat ergeben, dass von den 2000 Einträgen beider Listen 1598 Einträge gleich sind (siehe Abb. 2). Das macht einen prozentualen Wert von fast 80% (79,9%) aus. Beide Listen enthalten jeweils 402 Einträge, die nicht in der jeweils anderen Liste enthalten sind. Sie unterscheiden sich demnach bezogen auf die ersten 2000 Einträge um 20,1%.

Aus dem folgenden Venn-Diagramm wird die Deckung des Wortschatzes aus dem Fremdsprachenlehrwerk *»Deutsch – ein Hit! 1«* mit den beiden Worthäufigkeitslisten graphisch verdeutlicht.

Wir bemerken, dass von den DeReWo-Treffern (628) und den Leipziger-Treffern (625) 568 Grundformeneinträge des Lehrwerkes in beiden Listen auftauchen. Das macht eine gemeinsame Schnittmenge von über einem Drittel aus. Der Anteil der Grundformen des Lehrwerkes, der nur in der DeReWo-Liste aber nicht in der Leipziger Liste vertreten ist, beläuft sich auf 3,8% (62 Einträge) und vice versa 3,6% (58 Einträge).

Zusammenfassend kann folglich gesagt werden, dass der im ersten Band des Lehrwerkes enthaltene Wortschatz in seiner Grundform zu ca. 38% den 2000 häufigsten Wörtern der deutschen Sprache entspricht. Um aber ein aussagekräftiges Fazit ziehen zu können, muss die quantitative Analyse durch eine qualitative Analyse ergänzt werden.

## 6.2 Qualitative Analyse

Die quantitative Analyse hat – wie bereits oben erwähnt – eine Deckung von

ca. 38,5% bzw. 38,3% ergeben. Aber was genau bedeutet diese Zahl für die Bewertung eines Lehrwerkes? Um eine qualitative Aussage über die Worthäufigkeits-Orientierung des Lehrbuches treffen zu können, müssen die Treffer und Nicht-Treffer im Einzelnen analysiert werden. Hier sei erwähnt, dass eine Auflistung der gesamten Treffer und Nicht-Treffer an dieser Stelle aus Platzgründen nicht möglich ist. Deswegen soll im Rahmen der qualitativen Analyse eine Stichprobe gegeben werden. Dabei stehen bestimmte Fragestellungen im Vordergrund. Des Weiteren wurden für die qualitative Analyse die Frequenzhäufigkeiten der Listen in der unveränderten Fassung nicht berücksichtigt. Dieses Vorgehen ist darin begründet, dass durch die Aufbereitung und Entfernung der Einträge die ursprüngliche Reihenfolge dieser Einträge mit den Originallisten nicht mehr übereinstimmen, aber auch, dass zur Neuberechnung die absoluten Frequenzhäufigkeiten zu den einzelnen Einträgen nicht zur Verfügung stehen. Die DeReWo Grundformenliste arbeitet mit Häufigkeitsklassen<sup>1</sup> und die Leipziger Wortformenliste listet die Einträge in einem Ranking auf, ohne die absoluten Werte anzugeben. Deswegen soll in der gesamten Analyse der Schwerpunkt auf die jeweils ersten 2000 Einträge gelegt werden, die nach der Aufbereitung beider Listen ermittelt wurden.

**Fragestellung 1: Welche Einträge der Worthäufigkeitslisten kommen im Lehrwerk mit der höchsten Frequenz vor?**

In der folgenden Abbildung werden die Wortlisteneinträge aufgelistet, die im Lehrwerk »Deutsch – ein Hit! 1« die höchste Frequenz aufweisen:

Table 2: Frequenz und Ranking des »Deutsch ein Hit! 1«-Wortschatzes

Einträge	Rang	Frequenz
D-	1.	1507
SEIN	2.	383
IN	3.	314
UND	4.	300
ICH	5.	262
SIE	6.	210
DU	7.	187
IHR	8.	132
AN	9.	125
EIN	10.	123

Bei den ersten 10 Einträgen des ersten Bandes, die in beiden Worthäufigkeitslisten mit dem höchsten Vorkommen auftauchen, handelt es sich um sog. Funktionswörter (Synsemantika), die keine lexikalische Bedeutung tragen und vor allem syntaktisch-strukturelle Funktionen erfüllen. Dass die Funktionswörter die höchste Frequenz aufweisen, kann damit begründet werden, dass sie auch die größte Anzahl aller vorkommenden Einträge nicht nur im Lehrbuch ausmachen, sondern

1 In Anlehnung an Perkuhn sollten Häufigkeiten nicht absolut angegeben werden: »Der Betrachter verbindet [mit absoluten Zahlen] eine Genauigkeit und eine Zuverlässigkeit der Aussage, die nicht gegeben ist. Aufgrund der Zusammensetzung der Datengrundlage können sich Verzerrungen bei den Wortformfrequenzen ergeben, die [...] beschriebenen Problemfelder können zusätzliche Verschiebungen bei den Grundformfrequenzen bewirken.« (Perkuhn 2009: 13) Mit beschriebenen Problemfeldern sind Groß- und Kleinschreibung, diskontinuierliche Konstituenten, adjektivisch gebrauchte Partizipien, Fremdwörter, Anglizismen, Eigennamen, Kurzwörter u. a. gemeint. Deswegen werden die Einträge der DeReWo Liste in Häufigkeitsklassen angegeben (vgl. ebd.).



auch generell in der deutschen Sprache. Der Eintrag »d« weist durch die Zusammenfassung aller Artikel *der, die, das, dem, den, des* die höchste Frequenz auf. Mit Funktionswörtern allein ist ein Fremdsprachenlerner aber nicht in der Lage, eine schriftliche oder mündliche Äußerung zu tätigen, obwohl sie wichtig sind, um überhaupt sprachliches Handeln zu ermöglichen. Für sprachliche Äußerungen sind auch Inhaltswörter (sog. Autosemantika) wichtig. So tauchen in Anlehnung an die Ergebnisse der Analyse folgende Inhaltswörter des Lehrwerkes als erste auf:

*Tabelle 3: Frequenz und Ranking der Inhaltswörter, die als erste im Band »Deutsch – ein Hit! 1« auftauchen*

Einträge	Rang	Frequenz
HABEN	11.	123
GUT	15.	95
KOMMEN	16.	92
MACHEN	25.	71
UHR	26.	71
KÖNNEN	27.	67
HEISSEN	28.	65
FAHREN	30.	63
GERN	31.	61
BERLIN	32.	58

**Fragestellung 2: Welche Einträge der aufbereiteten DeReWo Grundformenliste und der lemmatisierten Leipziger Wortliste kommen nicht im Lehrbuch vor?**

Aus der quantitativen Analyse ging hervor, dass der Anteil der Wortlisteneinträge, die nicht im Lehrbuch vorkommen, größer ist als die Anzahl der Einträge, die enthalten sind.

Wirft man einen genauen Blick in die Einträge beider Worthäufigkeitslisten, die nicht im Lehrbuch »*Deutsch – ein Hit! 1*« auftauchen, dann ist auffällig, dass ein großer Teil dieser Einträge aus dem wirtschaftlichen, politischen, sportlichen und rechtlichen Bereich kommen.<sup>1</sup>

Aus dem wirtschaftlichen Bereich:

*Abteilung, Aktie, Aktiengesellschaft, Arbeitgeber, Arbeitnehmer, Arbeitslose, Arbeitslosigkeit, Auftrag, Beschäftigter, Betrieb, Bilanz, Budget, Börse, Branche, finanziell, Finanzierung, Firma, Fusion, führen, führend, Führung, Geschäftsführer, Gewerkschaft, Gewinn, global, gründen, Gründung, Händler, herstellen, Hersteller, Industrie, investieren, Investition, Investor, Job, Konzern, Kooperation, Kredit, Krise, Management, Manager, Mitarbeiter, Organisation, Produktion, produzieren, Sanierung, Standort, Umsatz, Unternehmer, Unternehmen, Vertrag, Vertreter, Vorsitzender, Vorstand, Wettbewerb, Wirtschaft, wirtschaftlich*

Aus dem politischen Bereich:

*Abgeordneter, Amt, Ausschuss, Außenminister, Bezirk, Botschaft, Bund, Bundeskanzler, Bundesland, Bundesrat, Bundesregierung, Bundestag, bundesweit, Bündnis, Bürgermeister, Debatte, Finanzminister, Fraktion, Förderung, Gemeinde, Gemeinderat, Generalsekretär, Gipfel, Innenminister, Kanzler, Koalition, Kommission, Kommune, Kongress, Landesregierung, Landtag, liberal, Macht, Militär, militärisch, Minister, Ministerium, Ministerpräsident, NATO, Opposition, Parlament, Partei, Politik, Politiker, politisch, Präsident, Reform, regieren, Regierung, Regierungschef, regional, Sitzung, Sozialdemokrat, Stadtrat, Staat, staatlich, Union, Verband, verhandeln, Verhandlung, Wahlkampf, wählen, Wähler*

<sup>1</sup> Bei der Interpretation der Wörter, die in Frequenzlisten vorkommen, können einige Einträge mehrere Lesarten haben, wie z.B. *Börse* auch als *Geldbörse*, *Prozess* als *Gerichtsprozess* oder in einer idiomatischen Redewendung verwendet (*Prozess in Gang*) setzen usw.

Aus dem rechtlichen Bereich:

*Angeklagter, Anhängen, Anschlag, Anwalt, Auflage, Aussage, Experte, Gesetz, gesetzlich, Haft, Motiv, Mord, Opfer, Prozess, Staatsanwaltschaft, Szene, Tod, tot, Täter, Urteil, Verdacht*

Aus dem sportlichen Bereich:

*Bundesliga, Cup, Liga, Meister, Meisterschaft, Olympia, olympisch, Sieg, Sieger, Stadion, Turnier*

Um eine schnelle und hohe Textdeckung zu erreichen, müssen nach Nation (2001) die 2000 häufigsten Wörter zuerst gelernt werden. Dies setzt aber voraus, dass in den ersten Fremdsprachenlehrbüchern diese 2000 häufigsten Einträge auch in die zu behandelnden Lektionen eingebunden werden (können). Die Wortlisten weisen

jedoch aufgrund der zeitungslastigen Korpora eine Vielzahl von Einträgen auf, die für ein Fremdsprachenlehrwerk der GER-Niveaustufe A1 für Jugendliche (ab 12 Jahren) nicht relevant sind.

**Fragestellung 3: Welche von diesen Einträgen könnten in Abhängigkeit zu den behandelten Lektionsthemen im Lehrwerk eingeflochten werden, unter der Voraussetzung, das Ziel sei, eine möglichst hohe Korpusbasiertheit zu erreichen?**

Es gibt dennoch in beiden korpusbasierten Worthäufigkeitslisten Einträge, die für die A1 Niveaustufe und in Abhängigkeit zu den behandelten Lektionsthemen geeignet wären, aber nicht im Lehrwerk auftauchen.

Tabelle 4: Einträge aus beiden Worthäufigkeitslisten, die im Lehrwerk eingeflochten werden könnten

»Deutsch – ein Hit! 1« Lektionen	DeReWo (IDS) (Grundformenliste)	Leipziger Wortliste (lemmatisiert)
Lektion 1: Erste Bekanntschaft	Junge	Junge
Lektion 2: Die Familie	Feuerwehr, Koch	Architekt
Lektion 3: Hobbys	Photo, Turnier, sportlich, Training	Trainer, Training
Lektion 5: Alles fertig für die Schule?	anfangen, aufgeben ( <i>Hausaufgaben</i> ), Farbe, grün, blau, weiß, leicht, schwer	Farbe, weiß, Geschäft, Kasse, leicht, schwer
Lektion 6: Schule und Alltag	Arbeit, Blatt, krank, Bett, Test	Arbeit, Blatt, Prüfung
Lektion 7: Guten Appetit!	wenig, viel	Rechnung
Lektion 9: Endlich Ferien!	baden, Hotel, See, Tourist	Dorf, Hotel, Meer

Wenn also das Lernziel wäre, die ersten 2000 häufigsten Wörter zu vermitteln, dann könnten diese Einträge in den einzelnen Lektionen eingeflochten werden.

**Fragestellung 4: Welche Lehrwerkseinträge tauchen weder in der DeReWo-Liste noch in der Leipziger Liste auf?**

Das Lehrbuch weist aber auch viele Einträge auf, die zwar nicht zu den 2000 häufigsten Wörtern gehören, aber sich

aus den behandelten Lektionsthemen ergeben und nicht in einem Lehrbuch fehlen dürfen:

*mögen, ergänzen, wann, Basketball, Arbeitsgemeinschaft, Bibliothek, Sekretariat, Heft, Hobby, Gitarre, Unterricht, bitte, Ferien, Sportplatz, Schultasche, Pinsel, Griechenland, griechisch, Oma, Athen, Großeltern, kochen, backen, Mathe, langweilig, Schulhof, Adresse, Buntstift, Klassenraum, schwimmen, Malkasten, Frühstück, Adressbuch,*

*schlafen, Supermarkt, Geschwister, Lieblingsfach, Schülerzeitung, Einladung, Großvater, Taschenrechner, Wörterbuch, Biologie, Mittagessen, Cousine, Mineralwasser, Nichte, Nefte, Reisebüro, Sonnenbrille, Stundenplan, Zeugnis, Badehose, Bikini, CD-Player, Geschenk, Sonnencreme, Telefonnummer, Tischtennis, E-Mail, Schreibwarenladen, Sehenswürdigkeit, Skifahren, Spaghetti usw.*

Darüber hinaus gibt es eine geringe Anzahl von Lehrwerkeinträgen, die ebenfalls weder in der einen noch in der anderen Liste vorkommen, aber für eine natürliche Kommunikationskompetenz sehr wichtig sind und auf keinen Fall fehlen dürfen. Angesprochen sind hier die Interjektionen, die auch von Muttersprachlern in der gesprochenen Sprache häufig gebraucht werden, wie z. B. *Äh!, Aha!, Au!, Hey!, Hm!, Na!, Och! und Puh!*. Diese sind für eine muttersprachähnliche Klangweise durchaus sehr wichtig, fehlen aber in den ersten 2000 Einträgen beider Worthäufigkeitslisten, da diese ausschließlich schriftliche Texte berücksichtigen.

Des Weiteren weist das Lehrwerk eine Anzahl an Einträgen auf, die in keiner der beiden Worthäufigkeitslisten vertreten sind und unter Umständen auch entfernt werden könnten. Es handelt sich hierbei um Einträge, die aus den vielen kleinen authentischen Texten des Lehrwerkes stammen und die von den Lehrwerksautoren weder als aktiver noch als passiver Wortschatz in der alphabetischen Wortliste gekennzeichnet sind. Auffällig ist, dass es sich bei den meisten Einträgen um Komposita handelt:

*Arztnotrufzentrale, Berufsfachschule, Berufsoberschule, Berufsschule, Bezirksamt, dirigieren, Erziehungsberechtigter, Fachoberschule, Fachschule, Gartenarbeitsschule, Impressionist, Landesverband, Militärareal, Schulgesetz, Solidarität, Stahlrohrmöbel, Surrealist, Kunstgewerbemuseum usw.*

An dieser Stelle könnte man sich fragen, ob Fremdsprachenlerner tatsächlich mit Wörtern konfrontiert werden sollten, die sie entweder gar nicht oder kaum in realen Kommunikationssituationen brauchen werden.

#### **Fragestellung 5: Wie frequent ist der aktive Wortschatz des Lehrwerkes?**

Für die quantitative und qualitative Analyse wurde bislang der gesamte Wortschatz des Lehrwerkes herangezogen. Aus didaktischer Sicht macht es durchaus Sinn, nur den aktiven Wortschatz (Lernerwortschatz) zu berücksichtigen, denn diesen müssen Fremdsprachenlerner lernen und aktiv beherrschen. Deswegen soll in der letzten Fragestellung der Lernerwortschatz gesondert betrachtet werden. Die folgende Abbildung zeigt die berechneten Werte des Lernerwortschatzes auf seine Korpusbasiertheit:

*Tabelle 5: Deckung des aktiven Wortschatzes mit den Worthäufigkeitslisten*

<b>Lernerwortschatz »Deutsch – ein Hit! 1«</b>	<b>DeReWo (IDS) (Grundformenliste)</b>	<b>Leipziger Wortliste (lemmatisiert)</b>
a. 1–1000	233 (36,6 %)	230 (36,1 %)
b. 1001–2000	74 (11,7 %)	66 (10,4 %)
Enthalten (a+b)	307 (48,3 %)	296 (46,5 %)
c. nicht enthalten	329 (51,7 %)	340 (53,5 %)
insgesamt (a+b+c)	636 (100 %)	636 (100 %)

Insgesamt beläuft sich der aktiv zu beherrschende Wortschatz auf 636 Grundformen, von denen 307 Einträge in der DeReWo Häufigkeitsliste und 296 in der lemmatisierten Leipziger Liste vorkommen. Die Einschränkung der Untersuchung auf den Lernerwortschatz führt zu einer höheren Deckung, als wenn der gesamte Wortschatz des Lehrwerkes in Betracht gezogen wird. Die Deckung beläuft sich in Bezug auf die DeReWo-Liste auf 48,3% und auf die lemmati-

sierte Leipziger Liste auf 46,5%. Etwas mehr als die Hälfte des Wortschatzes bewegt sich außerhalb beider Listen. Es handelt sich bei diesen Einträgen um

einen Wortschatz, der sich aus den Lektionsthemen des Lehrwerkes ergibt und folglich in beiden Listen unterrepräsentiert ist:

Tabelle 6: Einträge aus dem Wortschatz des Lehrwerkes, die in beiden Worthäufigkeitslisten unterrepräsentiert sind

Deutsch – ein Hit! 1 Lektionen	DeReWo (IDS) & Leipziger Wortliste (Grundformenliste) (lemmatisiert)
Lektion 1: Erste Bekanntschaft	Familiename, Griechenland
Lektion 2: Die Familie	Cousin, Großvater, Großmutter, Informatiker, Ingenieur, Nefte, Nichte, Oma, Onkel, Opa, Tante, Verkäuferin
Lektion 3: Hobbys	Basketball, Fußballmannschaft, Hobby, Jugendclub, Radfahren, Skateboard, Ski
Lektion 5: Alles fertig für die Schule?	Bleistift, Buntstift, Federtasche, Filzstift, Füller, Kassettenrekorder, Lieblingsfach, Malkasten, Pinsel, Schultasche, Stundenplan, Textmarker, Wörterbuch
Lektion 6: Schule und Alltag	Bibliothek, Biologie, Chemie, Chemielehrer, Computerraum, Erdkunde, Gymnasium, Klassenraum, Klassenzimmer, Kugelschreiber, Latein, Mathematik, Physik, Schulfach, Schulhof, Sekretariat, Sekretärin, Turnhalle, Turnschuh
Lektion 7: Guten Appetit!	Abendbrot, Apfelsaft, Appetit, braten, Bratkartoffeln, Bratwurst, Durst, Fisch, Fleisch, Getränke, Kartoffel, Ober, Rezept, Saft, Salat, Sauerkraut, Würstchen
Lektion 9: Endlich Ferien!	Bauernhof, Badehose, Bikini, Flugticket, Schnorchel, Sonnenbrille, Sonnencreme, Sonnenhut, Strandhandtuch, Schwimfflosse, Windjacke

Die höhere Deckung liegt einerseits der reduzierten Anzahl der Grundformen von 1632 auf 636 Einträge zugrunde und andererseits besteht bei den als aktiv gekennzeichneten Einträgen eine höhere Wahrscheinlichkeit, in beiden Worthäufigkeitslisten repräsentiert zu sein, als wenn der gesamte Wortschatz betrachtet wird.

## 7. Abschließende Bemerkungen zur Analyse des Lehrbuches<sup>1</sup>

Die 2000 (am) häufigsten (gebrauchten) deutschen Wörter sind im Grunde ge-

nommen nicht festzulegen, ohne sich auf eine konkrete Zeit, auf eine konkrete Sprachsituation, auf die ausgewerteten Quellen, Textsorten, Fachgebiete u. ä. zu beziehen. Beide Listen versuchen mehr oder weniger sich einer solchen Vorstellung zu nähern. Dass die Wortlisten sich nicht ähneln, liegt letztendlich auch an der unterschiedlichen Datengrundlage, auf der sie beruhen. Auch der Zeitraum beider Corpora ist unterschiedlich. Dementsprechend können sich zwei verschiedene Wortlisten auch nicht vollkommen ähneln. Allein daraus ergibt sich, dass

1 Für einige wichtige inhaltliche Hinweise danken wir Herrn Rainer Perkuhn (wissenschaftlicher Mitarbeiter im Bereich Korpuslinguistik des Instituts für Deutsche Sprache Mannheim).

sich die 2000 »häufigsten« Einträge von Korpus zu Korpus nie ganz decken können. Die Frage, welches Korpus bzw. welche korpusbasierte Frequenzliste als Grundlage für eine Analyse herangezogen werden sollte, ist also vordringlich zu beantworten.

Die Grundformenlemmatisierung sowohl des Fremdsprachenlehrwerkes als auch der Leipziger Wortformenliste war mit folgenden Entscheidungsproblemen verbunden: Die Aufbereitung der Listen bewirkt durch die Entfernung der Doppelnennungen, der Eigennamen – bis auf Städte- und Ländernamen –, der Anglizismen sowie der nicht üblichen Abkürzungen eine Aufrückung vieler Einträge. Die Lemmatisierung der Leipziger Liste hatte eine starke Reduzierung der Wortformen in Höhe von ca. 38 % zur Folge, was zu einer höheren Deckung führte. Weiterhin gestaltete sich die Zuordnung einiger Einträge zu ihrer Grundform schwierig, wie beispielsweise bei Mehrdeutigkeiten von Einträgen: Gehört der Eintrag »bestimmt« zu dem *Verbparadigma* »bestimmen« oder zum *Adjektiv* »bestimmt«? Angesprochen sind hier die Partizipien, die je nach Gebrauch entweder dem Verbparadigma untergeordnet oder als eigenes Lemma als Adjektiv angesetzt werden können. Ohne Kontext ist es schwierig, eine eindeutige Entscheidung zu treffen, da die Listeneinträge kontextlos aufgelistet werden. Die *kontinuierlichen* und *diskontinuierlichen Präverbfügungen* stellten kein Problem dar, da die Anzahl solcher Vorkommen im Lehrwerk sehr gering war, so dass sie ihrem Infinitiv der jeweiligen Präverbfügung manuell zugeordnet werden konnten. Die Aufarbeitung der Worthäufigkeitslisten hatte zur Folge, dass die ursprüngliche Reihenfolge der Liste verändert wurde. Mit der Lemmatisie-

rung erhält man zwar der äußeren Form nach eine Grundformenliste, aber die ursprüngliche Frequenz bezieht sich nur auf die Originalwortformen- bzw. Grundformenlisten, nicht jedoch auf die für die Analyse bearbeiteten Listen. Die Reihenfolge der Einträge spielte bei der Analyse jedoch keine wesentliche Rolle. Auch bei der Entfernung der zahlreichen Eigennamen bestanden folgende Entscheidungsschwierigkeiten: Handelt es sich bei dem Eintrag »Heuer« um einen Familiennamen »Herr/Frau Heuer« oder um die schweizerische Bezeichnung für »Heumacher« oder für den »Lohn eines Seemanns«? Ist mit dem Eintrag »Ernst« der Vorname oder die Grundform eines am Satzanfang stehenden Adjektivs gemeint? Bei diesen Zweifelsfällen wurde entschieden, die Einträge beizubehalten und sie nicht als Eigennamen zu kennzeichnen.

Die *Komposita* stellten bei den Vorüberlegungen zur Lemmatisierung ebenfalls ein Entscheidungsproblem dar. Hier galt es zwei Perspektiven zu beachten, auf der einen Seite die didaktische und auf der anderen die linguistische. Aus didaktischer Sicht wird ein Fremdsprachenlerner, der die Wörter »das Buch« und »die Schule« kennt, kein Problem haben, das Kompositum »das Schulbuch« zu verstehen. Folglich wäre aus diesem Gesichtspunkt eine Dekomposition notwendig. Linguistisch hingegen unterscheidet man bei den Komposita zwischen *lexikalisierten* Komposita und Komposita, die *weniger transparent* sind (produktive Komposita, die jederzeit aus den Wortbildungsregularitäten gebildet werden können). Erstere wären mit einer eigenen Grundform anzusetzen, letztere müssten einer Dekomposition unterzogen werden. Im Rahmen dieser Untersuchung wurde der Aspekt der Dekomposition vernachlässigt.

## 8. Ausblick

Diese Untersuchung sollte einen Beitrag in Richtung »korpusbasierte Frequenzangaben aus DaF-Materialien« leisten. Die Untersuchung hat ergeben, dass der Band »Deutsch – ein Hit! 1« eine Deckung von 38,3 % bzw. 38,5 % und für den Lernerwortschatz im Speziellen von 48,3 % bzw. 46,5 % aufweist. Die Untersuchungen von englischen Lehrwerken von Alexiou und Konstantakis haben hingegen höhere Deckungswerte ergeben. Sie liegen zwischen 79 % und 85 %. Dies ist in erster Linie darauf zurückzuführen, dass types<sup>1</sup> untersucht wurden und nicht Grundformen (vgl. Alexiou/Konstantakis 2007). Für zukünftige Untersuchungen ist vorgesehen, Lehrwerke für Deutsch als Fremdsprache für griechische Lerner kontrastiv zu vergleichen. Hierbei soll der aktive und passive Wortschatz auf den Prüfstand gestellt werden, indem beide auf ihre Korpusbasiertheit hin analysiert werden, um der interessanten Fragestellung nachzugehen, wie frequent der Wortschatz in Fremdsprachenlehrwerken des Deutschen als Fremdsprache ist. Untersuchungsergebnisse eines Vergleichs mehrerer Lehrwerke könnten eventuell dazu verhelfen, bessere Aussagen über die optimale Auswahl des didaktischen Wortschatzes in DaF-Lehrwerken zu machen.

## Literatur

Alle URLs wurden am 29.12.2009 überprüft

Alexiou, Thomai; Konstantakis, Nikolaos: »Lexis for Young Learners: Are we heading for frequency or just common sense?« In: Tsangalidis, Anastasios (Hrsg.): *Selection of papers for the 18th Symposium of Theoretical and Applied Linguistics*. Aristot-

le University of Thessaloniki, 2007, 59–66.

Heatley, Alex; Nation, I. S. Paul; Coxhead, Averil: *RANGE and FREQUENCY programs* (2000). Online unter: <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>.

Kupietz, Marc; Keibel, Holger: »The Mannheim German Reference Corpus (DeReKo) as a basis for empirical linguistic research«, *Working Papers in Corpus-based Linguistics ans Language Education* 3 (2009), 53–59. Tokyo: Tokyo University of Foreign Studies (TUFS) Online unter: [http://cblle.tufs.ac.jp/assets/files/publications/working\\_papers\\_03/section/053–059.pdf](http://cblle.tufs.ac.jp/assets/files/publications/working_papers_03/section/053–059.pdf).

Linke, Angelika; Nussbaumer, Markus; Portmann, Paul R.: *Studienbuch Linguistik*. 5., erweiterte Auflage. Tübingen: Niemeyer, 2004 (RGL, 121).

Lüdeling, Anke; Walter, Maik: *Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung*. (2009) Online unter: <http://www.linguistik.hu-berlin.de/institut/professuren/korpus-linguistik/mitarbeiterinnen-anke/pdf/LuedelingWalterDaF.pdf>.

Nation, Paul: *Learning vocabulary in another language*. Cambridge: Cambridge University Press, 2001.

Ortmann, Wolf Dieter: *Hochfrequente deutsche Wortformen*. München: Goethe Institut, 1975.

Perkuhn, Rainer; Belica, Cyril: »Korpuslinguistik – Das unbekannte Wesen oder Mythen über Korpora und Korpuslinguistik«, *Sprachreport* 1 (2006). Online unter: <http://www.ids-mannheim.de/pub/autoren/ids/perkuhn.html>.

Perkuhn, Rainer; Belica, Cyril; Kupietz, Marc; Keibel, Holger: *DEREWO: Korpusbasierte Grundformenliste*. Mannheim: IDS, 2007. Online unter: <http://www.ids-mannheim.de/kl/derewo/derewo-v-30000g-2007-12-31-0.1.zip>.

Perkuhn, Rainer: *Allgemeine Anmerkungen zur Reihe DeReWo – Korpusbasierte Wortlisten*. Mannheim: IDS, 2009. Online unter: <http://www.ids-mannheim.de/kl/derewo/derewo-general-remarks.pdf>.

1 Worthäufigkeiten können auch in tokens und types gezählt werden. Tokens sind einzelne sprachliche Äußerungen, die in einem Text vorkommen (jedes Vorkommen einer Wortform). Types hingegen sind alle sprachlichen Äußerungen, die in einem Text nur einmal realisiert werden (vgl. Linke/Nussbaumer/Portmann 2004: 62).

Quasthoff, Uwe; Wolff, Christian: »Korpuslinguistik und große einsprachige Wörterbücher«, *Linguistik online* 3/2 (1999). Online unter: [http://www.linguistik-online.de/2\\_99/quasthoff.html](http://www.linguistik-online.de/2_99/quasthoff.html).

Quasthoff, Uwe; Heyer, Gerhard (Projektverantwortliche): *Wortformenliste des Leipziger Wortschatzprojektes, Universität Leipzig*. Stand: 2001. Online unter: <http://wortschatz.uni-leipzig.de/html/wliste.html>.

Stai, Daniela; Kapoti, Chariklia; Spyropoulou, Panagiota; Pasisi, Ekaterini: »*Deutsch – ein Hit! 1*«. KURSBUCH für die 1. griechische Gymnasialstufe. 1. Auflage. Athen: O. E. D. B., 2006.

Stai, Daniela; Kapoti, Chariklia; Spyropoulou, Panagiota; Pasisi, Ekaterini: »*Deutsch – ein Hit! 1*«. LEHRERHANDBUCH für die 1. griechische Gymnasialstufe. 1. Auflage. Athen: O. E. D. B., 2006.

Tschirner, Erwin: »Korpora, Häufigkeiten, Wortschatzerwerb«. In: Heine, Antje; Hennig, Mathilde; Tschirner, Erwin (Hrsg.): *Deutsch als Fremdsprache – Konturen und Perspektiven eines Fachs*. München: Iudicium, 2005, 133–149.

*Panagiotis Lymparakakis*

Geb. 1975; 2000–2005 Studium der Betriebswirtschaftslehre (Dipl.-Betriebswirt); seit 2007 Student an der Abteilung »Deutsche Sprache und Philologie« mit Schwerpunkt »Linguistik und Didaktik« an der Aristoteles Universität Thessaloniki und Stipendiat für hervorragende Studienleistungen; dieser Fachbeitrag entstand auf dem Hintergrund der Diplomarbeit mit dem Titel »Moderne Wortschatzarbeit und korpusbasierter Wortschatz unter Begutachtung des Fremdsprachenlehrwerkes ›Deutsch – ein Hit! 1‹«.

*Andromachi Sapiridou*

Assistenzprofessorin an der Abteilung für deutsche Sprache und Philologie der Aristoteles Universität Thessaloniki/Griechenland, mit dem Schwerpunkt DaF-Didaktik. Besondere Forschungsinteressen: Lehrwerkanalyse, Interkulturelle Didaktik, Sprachpolitik.