

Korpusbasierte Analyse zur Wortschatzvielfalt in regionalen und universellen DaF-Lehrwerken¹

Hakan Dolan , Eskişehir – Şerife Çelikkaya , Eskişehir

 <https://doi.org/10.37583/diyalog.1499761>



Abstract (Deutsch)

Die Verbesserung des Wortschatzes der Schüler im Sprachunterricht ist eines der Hauptziele des Sprachunterrichts. Um dieses Ziel zu erreichen, haben sich verschiedene Trends zur Verbesserung des Wortschatzes herausgebildet. Ein Trend ist die Bestimmung der Worthäufigkeit. Die vorliegende Studie wurde durchgeführt, um das häufig Vorkommen und die Austeilung der Wörter in Hinsicht auf Wortarten in den regional und universell herausgegebenen DaF-Lehrwerken festzustellen und ein Korpus für DaF-Lehrwerke auszuarbeiten. Die Worthäufigkeit in DaF-Lehrwerken wird mithilfe der internetbasierten Software Sketch Engine ermittelt. Im Korpus werden die Tests Log-Likelihood (LL) und Odd Ratio (OR) zur Definition von Wörtern verwendet, um die Besonderheit von Wortarten Token/Typ Ratio festzusetzen. Das aus diesen Daten und Analysen erstellte Korpus wird als HD-Korpus bezeichnet, und es wird angenommen, dass dies einen erheblichen Einfluss auf die Herstellung von Lehrwerken haben wird. Bei der Häufigkeitsanalyse nach Wortarten reihen sich die Wörter als Substantive, Verben, Adjektive, Adverbien, Präpositionen und Konjunktionen. Das Token/Type-Verhältnis im Korpus hat allerdings zu dem Ergebnis gebracht, dass im Gegensatz zu den anderen Wortarten Adverbien in der Mehrzahl sind, obwohl die Substantive mit den gehobenen Frequenzen die wenigsten Unterschiedlichkeit zeigen. Diese Ergebnisse lassen darauf schließen, dass die Vielfalt der Wortarten durch die häufige Verwendung von Wörtern nicht beeinflusst wird. Darüber hinaus zeigen die LL- und OR-Rankings eine hohe Ähnlichkeitsrate zwischen Lehrbüchern und erstelltem Korpus und weisen auf die Frequenz und Besonderheit der Wörter in den Lehrwerken in HD-Korpus hin.

Schlüsselwörter: Wortart, Wortschatzvielfalt, Worthäufigkeit, DaF-Lehrwerke, Korpus.

Abstract (English)

Corpus-based analysis of vocabulary diversity in regional and universal GFL textbooks

Enhancing vocabulary acquisition in language classrooms is a fundamental goal of language education. In pursuit of this objective, diverse approaches have been developed, one of which includes the analysis of word frequency. The present study was conducted to determine the frequency and distribution of words in terms of parts of speech in regional and universal GFL textbooks and to develop a corpus for GFL textbooks. The frequency of words in GFL textbooks is determined using the internet-based software

Einsenddatum: 30.03.2024

Freigabe zur Veröffentlichung: 15.06.2024

¹ Die in diesem Beitrag behandelten Daten basieren auf einem Teil der Magisterarbeit von Hakan Dolan (Die Lehrbücher im Deutsch als Fremdsprache: eine korpusbasierte Analyse der lexikalischen Vielfalt - Anadolu Universität/Türkei, 2022), deren Erstellung unter Betreuung von Asst.Prof.Dr. Şerife Çelikkaya erfolgte.

Sketch Engine. In the corpus, the tests Log-Likelihood (LL) and Odd Ratio (OR) are used to define words in order to determine the specificity of word types of Token/Type Ratio. The corpus, formulated from these data and known as the HD corpus, is projected to have a considerable impact on the development of textbooks. In the frequency analysis by word type, the words rank as nouns, verbs, adjectives, adverbs, prepositions, and conjunctions. However, the token/type ratio in the corpus has led to the conclusion that, in contrast to the other word types, adverbs are in the majority, although the nouns with the higher frequencies show the least diversity. These results suggest that the diversity of word types is not affected by the frequent use of words. In addition, the LL and OR rankings show a high similarity rate between textbooks and created corpus, indicating the frequency and specificity of words in the textbooks in HD corpus.

Keywords: *Word type, vocabulary variety, word frequency, GFL textbooks, corpus*

EXTENDED ABSTRACT

It has been evident that vocabulary functions as a key component in foreign language learning. It is possible to create the condition for mastery of a foreign language only if the four language skills are engaged effectively. Likewise, it is unlikely to neglect the fact of vocabulary used in all language skills. Learner proficiency level is also linked to vocabulary size regarding the Common European Framework of Reference for Languages (CEFR).

Despite recent advancements in technology, the influence of foreign language textbooks on foreign language teaching remains indisputable. Hence, examining and developing textbooks are noteworthy in terms of not only four basic language skills but also in vocabulary teaching. The quantitative research of the vocabulary in textbooks, or anticipated to be included in them, is greatly advanced by corpus linguistics. Corpus linguistics conducts frequency studies using data from specific sources. Thus, corpus linguistics should be evaluated with the great importance of such frequency studies.

There should be more focus on frequency studies by displaying an intensive interest in vocabulary and teaching. Within the framework of this study, it is evident that there is a lack of a detailed corpus tailored for German as a foreign language. Accordingly, this study attempts to contribute to teaching vocabulary in the FSU and to connect gaps in the literature. Suggestions for determining the most frequently used words during textbook design can be provided based on the corpus developed from this study.

This corpus-based study aims to contrast the frequently used parts of speech and their frequencies in DaF textbooks published in Turkey and Germany. This case study has been designed to examine the frequency of words in DaF textbooks, complemented by an analysis of the corpus distribution by parts of speech and the specificity of vocabulary employed.

Within this context, the study primarily centers on examining the frequency of words in DaF textbooks, analyzing the distribution of their types, and assessing the status of the corpus created. Regarding word frequency, the Turkish DaF textbook corpus (TDaF corpus) has been created for the DaF textbooks published in Turkey and the German DaF textbook corpus (DDaF corpus) is generated for the DaF textbooks published in Germany. By combining these two sub-corpora, HD corpus has been created and aimed to become a corpus for DaF textbooks. Therefore, a framework has been established to outline the types and frequencies of words in DaF textbooks, which are crucially taught in the target language.

Based on the findings of this study, HD corpus has been discovered to consist of 435059 tokens and a total of 58017 different parts of speech with more than one repetition based on the identified parts of speech. As it has been revealed with a result of the analysis carried out, the most commonly used parts of speech in textbooks have been expressed as nouns. This sequence has been followed by verbs, adjectives, and adverbs. Furthermore, prepositions and conjunctions have been considered the least common part of speech in foreign language textbooks.

Another important finding of the research are the results obtained after the Token/Type R (TTR) analyzed to reveal the word size and the ratio of the lexical frequency profile of the examined German textbooks. The volumetric space a word occupies does not mean that it is diverse and rich in terms of genre. It aims to answer the question whether it displays the repetition of words in textbooks or the richness of vocabulary or not. Based on the research data, the TTR results reveal that the adverbs in the German textbooks constructing the HD corpus show more diversity than the frequency profiles. On the other side, the nouns have demonstrated little variety in the German textbooks although they have a high volumetric frequency, which means similar words are repeated more than once.

Owing to the created HD corpus, it is possible to offer foreign language teachers and material designers both in Germany and in Turkey a wide range of help. Nevertheless, by means of its objectives, it is likely to consider the scope of this research as limited. In particular, an extensive applied study can be conducted to determine the presence and frequency of learned words based on the written and oral performance of learners mastering German as a foreign language.

1 Einleitung

Die Kenntnis der Wörter und Redewendungen einer Sprache ist ein wichtiger Bestandteil beim Erlernen einer Sprache und bei der Beherrschung von Sprachkenntnissen. Das Wort ist ein wichtiges Element des Wortschatzes, das das Geflecht der Sprache bildet. Auch mit Muttersprachlern fehlerfrei richtig zu kommunizieren wird ein Grundwortschatz benötigt (Runte 2015: 11). Ein Wort wird als ein Laut oder eine Gruppe von Lauten definiert, die eine Bedeutung tragen und eine aktive Rolle beim Aufbau des Satzes spielen (Kantemir 1997: 178). Nation (2001) betonte, dass es zum Erkennen des Wortes notwendig ist, seine Bedeutung(en), gesprochene und geschriebene Formen, Morphologie, grammatikalische Kategorie, Kollokationen, Reihenfolge, Konnotationen und Häufigkeit zu kennen. Diamond / Gutlohn (2006) stellten fest, dass Menschen mit einem großen Wortschatz effektiver Bedeutungen nutzen, denken und Beziehungen zwischen Wörtern herstellen können. So ist es zu sagen, dass es wichtig ist, sich der formalen, semantischen und funktionalen Merkmale des Wortes bewusst zu sein, um das Wort zu kennen und über einen reichen Wortschatz zu verfügen. Der Wortschatz bildet die Grundlage jeder sprachlichen Kommunikation und ist daher für den Fremdsprachenunterricht von entscheidender Bedeutung (Schmitt 2000). Zimmerman (1997: 5) betonte auch, dass das Wort die Grundlage der Sprache ist und für den Sprachenlerner einen äußerst wichtigen Platz einnimmt.

Der Umfang des Wortschatzes ist ein Faktor, der das Sprachniveau des Lernenden bestimmt (GER 2001: 110). Harmer (1991) stellte fest, dass die Sprachstruktur die Funktion eines „Skeletts“ hat und der Wortschatz die Funktion von „lebenswichtigen Organen und Fleisch“, die dieses Skelett bilden. Aus diesem Grund kann man sagen, dass Wortschatzlernen einen besonderen Stellenwert beim Sprachenlernen einnimmt.

Die bedeutende Rolle der Lehrwerke im Fremdsprachenunterricht ist trotz heutiger fortschrittlicher Technik nicht zu zweifeln. Daher ist es auch von Bedeutung, die fremdsprachigen Lehrwerke sowohl auf Basis der grundsätzlichen Sprachkompetenzen (Dolan / Çelikkaya 2024), als auch in Bezug auf den Wortschatz zu analysieren. Für eine effektive Lehre im Fremdsprachenunterricht sind in diesem Zusammenhang Korpusstudien erforderlich (Lymperakakis / Andromachi 2012: 139). Das Verstehen der gesprochenen und geschriebenen Texte der erlernten Sprache besteht nach Ansicht von Tschirner (2008: 195) in einem signifikanten Zusammenhang mit der Häufigkeit der verwendeten Wörter. Bei den quantitativen Untersuchungen im Rahmen von vermittelten bzw. zu vermittelndem Wortschatz in Lehrwerken nimmt die Korpuslinguistik eine äußerst wichtige Rolle.

Die Korpuslinguistik wird verwendet, um Häufigkeitsstudien mit Daten aus bestimmten Quellen durchzuführen (Denham / Lobeck 2013: 317). Korpuslinguistik kann als ein Bereich definiert werden, der Sprachstudien auf der Grundlage von Beispielen des „realen“ Sprachgebrauchs umfasst (McEnery / Wilson 2001: 1). Wie bei allen wissenschaftlichen Untersuchungen fügt der Ausgangspunkt realer Daten in korpusbasierten Studien der Forschung eine empirische Dimension hinzu (Tognini-Bonelli 2001: 2). Es wurden verschiedene Korpora erstellt, indem die Daten eingesetzt

werden, die aus bestimmten Quellen mittels Computersoftware erzeugt wurden. Diese Quellen können sich je nach Untersuchungsart variieren. Sie können z.B. Tonaufnahmen von Personen, literarische Texte, Bücher, schriftliche Texte usw. sein. Über qualitative Ansätze hinaus liefert die Korpuslinguistik auch Häufigkeitsinformationen über einen bestimmten Sprachgebrauch durch den Einsatz quantitativer Methoden in elektronisch verarbeiteten Texten.

Da der Fremdsprachenunterricht korpusbasiert gestaltet werden kann, spielt die Korpuslinguistik eine wichtige und effektive Rolle im Fremdsprachenunterricht. Nach Gut (2007: 1) erhalten die Schüler im korpusbasierten Unterricht im Vergleich zum regelbasierten Unterricht die Ermutigung, sich durch die Verwendung von Sprachkorpora muttersprachlichen Sprachgebrauch sowie sprachliche Regeln und Zusammenhänge in den Bereichen Wortschatz, Grammatik und Aussprache eigenständig anzueignen. Besonders hervorgehoben wird darin, dass der auf Korpus beruhende Unterricht eine effektive Methode des induktiven Lernens ist (Todd 2001). Nach Gut (2007: 7) soll der korpusbasierte Unterricht auch die Sprachkompetenz fördern und diese kann dazu beitragen, die Motivation zu steigern und dadurch den Lernerfolg zu verbessern.

Ein weiteres Einsatzgebiet von Korpora im Fremdsprachenunterricht besteht darin, dass sie bei der Vorbereitung des Unterrichts und der Entwicklung von Materialien unterstützend wirken. Sprachkorpora sollen hier als Zusatz oder sogar als Ersatz für konventionelle Lehrwerke und Übungsmaterialien dienen. Nach Gut (2007: 7) ist der Grund dafür, dass die Menge der muttersprachlichen Daten in den Korpora die tatsächliche Häufigkeit von sprachlichen Strukturen und Regularitäten darstellt. Frequenzlisten können in der Didaktik zur empirischen Absicherung eines Lehrplans verwendet werden (Römer 2008: 114, Tschirner 2005, 2008). Lüdeling / Walter (2009: 12) gaben an, dass die Häufigkeit von sprachlichen Strukturen der entscheidende Bestandteil von Lehrmaterialien auf Korpusbasis ist.

Die vorhandenen deutschsprachigen Korpora werden aus verschiedenen Daten erzeugt. Eines der größten deutschen Korpora „Digitales Wörterbuch der deutschen Sprache“ (DWDS) umfasst drei Komponenten: Textkorpora, Wörterbücher und Wortstatistiken. Die Korpora in DWDS bestehen aus 127.196.441 Dokumenten, die unter verschiedenen Kategorisierungen von Zeiträumen und Überschriften stehen. Die Manuskripte in DWDS bestehen im Ganzen aus 2.956.196.774 Sätzen und 44.511.102.227 Token (DWDS 2021). Beispielsweise ist DWDS-Kernkorpus ein Referenzkorpus und besteht aus 79.190 Manuskripten, 5.821.108 Phrasen und 121.494.429 Token. Unterhaltungsliteratur, Wissenschaft, Verwendungsliteratur und Zeitungen zwischen 1900-1999 sind die Textsorten, aus denen die Daten des DWDS-Kernkorpus bestehen.

Aus politischen Reden wurde auch ein anderes Korpus erstellt, dessen Daten zu den Jahren 1982-2020 gehören. Dieses Korpus enthält Reden unterschiedlicher Abgeordneter aus unterschiedlichen Ländern und besteht aus 15.240 Belege, 1.424.295 und 27.012.455 Token. Ein weiteres großes Korpus, das verschiedene Arten deutscher Texte aus dem 16. bis 20. Jahrhundert enthält, ist das Deutsche Textarchiv (DTA)

(Deutsches Textarchiv 2021). Die Textsorten dieser Korpora sind Wissenschaft, Verwendungsliteratur, Zeitungen und Belletristik. Das Deutsche Referenzkorpus auch bekannt als DeReKo ist außerdem ein weiteres elektronisches Korpus des Leibniz-Instituts für Deutsche Sprache. Dieses Korpus besteht aus belletrischen, wissenschaftlichen und populärwissenschaftlichen Texten und umfasst insgesamt 50,6 Milliarden Wörter. Spezialkorpora sind neben diesen genannten Referenzkorpora auch vorhanden. Als Beispiel hierfür kann das Korpus Dolmetschen im Krankenhaus, DiK, angeführt werden. Es besteht aus 91 Interviews zwischen deutschsprachigen Ärzten und medizinischem Personal und den Patienten deren Muttersprache Türkisch und Portugiesisch sind (Clarin-D 2021).

Bei der Betrachtung der Korpusforschungen ist es festzustellen, dass es kein breit erstelltes Korpus für Deutsch als Fremdsprache gibt. Wenn die Literatur für das Korpus durchsucht wird, wird deutlich, dass der Großteil der Studien zur Feststellung des Wortschatzes von Fremdsprachenlernenden und Muttersprachlern auf Kompetenzniveaus beruhen.

Henzl (1979) führte eine Studie mit Lernern von Englisch als Fremdsprache und englischen Muttersprachlern durch, in der er Geschichten verwendete, die in Tschechisch, Englisch und Deutsch erzählt wurden. Als Forschungsergebnis wurde hervorgehoben, dass die Muttersprachler mehr dazu neigen, verschiedene Verben mehr als einmal zu verwenden (Thomas 2005: 2). Anfang der 2000er Jahre wurde das Institut für Korpuslinguistik an der Humboldt-Universität gegründet, um die ersten Anstrengungen zur Etablierung eines deutschen Korpus zu starten (Schmidt 2011: 555), das sich auf die Lehre von Deutsch als Fremdsprache konzentrierte (vgl. Lüdeling, Seanna, Hagen, Karin, Maik 2008; Walter / Grommes 2008; Fandrych / Tschirner 2007).

In der Literatur wurden im Bereich DaF und DaZ verschiedene Korpusstudien durchgeführt und Schmidt (2011: 560) hat das Lernerkorpus mit Datentypen und Erstellern zusammengefasst. Heide Wegener erstellte das Augsburger DaZ-Korpus, dessen Daten auf 12 Kindern basieren, die Türkisch, Russisch, Polnisch als Muttersprache sprechen und Deutsch erworben haben (Schmidt 2011: 560). Hingegen umfasst das Dimroth-Korpus von Christian Dimroth (Schmidt 2011: 560) die L2 Daten von Erwachsenen, die Kroatisch, Türkisch und Russisch als Muttersprache sprechen. Ein anderes Korpus von Judy Belz ist Telecorp, der Lernertexte von Studenten zwischen 2000-2005 als Daten enthalten und insgesamt 1,5 Millionen Token haben (Schmidt 2011: 560). Inzwischen wurde das Korpus Learning Prosody in a Foreign Language (Leap) von Ulrike Gut entwickelt, das auf Sprechfähigkeit der Studenten basiert (Schmidt 2011: 560). Für dieses Korpus wurden insgesamt 359 Audioaufnahmen von 131 Studenten verwendet. Corpus of Learner German (CleG-Korpus) von Ursula Maden-Weinberge ist ein weiteres Korpus, das Aufsätze von Schülern enthält (Schmidt 2011: 560). Das Korpus stützt sich auf Daten aus 451 Artikeln von 126 Studenten im 6. Semester im Bereich DaF. An der Universität Barcelona wurde andererseits ein Korpus von Marta Fernandenz-Villaeuva und Oliver Strunk erstellt (Schmidt 2011: 560). Das Korpus Varcom basiert auf die Daten von 1520 verfassten Texten und Videoaufnahmen, in denen 66 Studenten Spanisch, Katalanisch und Deutsch gesprochen haben. Zwischen 2001-2004 wurde das

Korpus Fehlerannotierte Lernerkorpus (Falko) von Freie Universität, Georgetown University und Humboldt Universität erstellt, das auf schriftlichen Texten fortgeschrittener Lerner des Deutschen als Fremdsprache basiert und frei zugänglich ist. Betrachtet man all diese Erklärungen ist es zu sehen, dass keine Studie zum Verhältnis von Worthäufigkeit zu den Lehrwerken durchgeführt ist. Aus diesem Grund wurde es als notwendig erachtet, solch eine Studie durchzuführen, um die Lücke in der Literatur zu schließen. Der aus der Studie resultierende Korpus zeigt die Wörter, die am häufigsten bei der Erstellung von Lehrwerken verwendet werden. Mit dieser Arbeit werden sich Unterschiede und Ähnlichkeiten der Wortarten daneben die Verwendungsfrequenzen in DaF-Lehrwerken zum Ziel gesetzt, die in der Türkei und in Deutschland herausgebracht wurden.

Zu diesem allgemeinen Zweck lauten die primäre Forschungsfragen, auf der die aktuelle Studie (Dolan 2022) basiert, wie folgt:

- 1- Wie haben sich Wortschatztypen und Verwendungshäufigkeit in Lehrwerken für den Unterricht Deutsch als Fremdsprache (DaF) in der Türkei und in Deutschland (von 2009 bis 2020) verändert und welche Unterschiede und Gemeinsamkeiten gibt es zwischen ihnen?
- 2- Wie haben sich Wortschatztypen und Verwendungshäufigkeit in Lehrwerken für den Unterricht DaF in der Türkei (von 2009 bis 2020) verändert?
- 3- Wie haben sich Wortschatztypen und Verwendungshäufigkeit in Lehrwerken für den Unterricht DaF in Deutschland (von 2009 bis 2020) verändert?
- 4- Welche Unterschiede gibt es zwischen den Lehrwerken für den Unterricht DaF in der Türkei und Deutschland in Hinsicht auf Wortarten?
- 5- Welche Gemeinsamkeiten gibt es zwischen den Lehrwerken für den Unterricht DaF in der Türkei und Deutschland in Hinsicht auf Wortarten?

Im Rahmen der Studie werden die folgenden Teilfragen beantwortet:

1. Wie oft werden die Wörter im Korpus für DaF-Lehrbücher verwendet, die im Rahmen dieser Studie erstellt wurde?
2. In welchem Verhältnis stehen die Token/Typ-Relationen in den DaF-Lehrwerken zueinander?
3. Welches sind die am häufigst verwendeten Wortarten in den DaF-Lehrbüchern?

2 Methode

In diesem Teil der Studie werden Design und Gegenstand der angewandten Forschung, Techniken der Datenerhebung-analyse und statistische Auswertungen zusammengefasst.

2.1 Forschungsdesign

Diese Studie wurde als Fallstudie konzipiert, um die Worthäufigkeiten, -spezifität und -verteilung nach Wortarten des aus DaF-Lehrbücher gewonnenen Korpus zu ermitteln. Fallstudien werden von Chimilar (2010: 582) als methodologischer Ansatz vorgestellt, der eine eingehende Studie umfasst, die mehrere Datensammlungen verwendet, um systematisch Informationen darüber zu sammeln, wie ein beeinträchtigt System funktioniert.

Ebenso lassen sich Fallstudien als ein Verfahren bezeichnen, das eine begrenzte Systemstruktur sorgfältig bestimmt und analysiert (Merriam 2013: 193). Bei einer Fallstudie werden bestimmte Situationen ausführlich untersucht und Situationen im Zusammenhang festgesetzt, die eine qualitative Forschungsmethode darstellen (Creswell 2007: 73). Fallstudien werden auf der anderen Seite von Davey (1991: 1) als eine Forschungsmethode definiert, bei der Ereignisse und Situationen systematisch erforscht und aufgezeigt werden, weshalb sich das Ereignis dermaßen mit den erzielten Folgen hervorgegangen ist und wofür sich in der angehenden Studie konzentriert werden sollte. In Bezug auf Wortfrequenzen wurden zwei Korpora erstellt: das türkische DaF-Lehrwerk-korpus (TDaF-Korpus) und das deutsche DaF-Lehrwerk-korpus (DDaF-Korpus), die aus regional und universell veröffentlichten DaF-Lehrwerken bestehen. Darüber hinaus wird durch die Zusammensetzung der beiden Subkorpora (TDaF- und DDaF-Korpus) ein Korpus namens HD-Korpus erstellt und ein allgemeines Korpus für DaF-Lehrwerke dargestellt. Mit diesem Korpus wird versucht, die Ermittlung der Art und Frequenz von Wörtern in DaF Unterrichtsmaterialien zu bestimmen, welche in der gelernten Sprache vermittelt werden sollten.

2.2 Die Lehrwerke

Insgesamt 36 DaF-Lehrbücher (siehe Anhang) wurden in dieser Studie untersucht. 12 von ihnen sind türkische DaF-Lehrbücher für Sekundarstufe und übrige 24 deutsche DaF-Lehrbücher für Jugendliche von den vier bekanntesten Verlagen (Hueber, Cornelsen, Klett und Langenscheidt) in Deutschland. Es gibt zwei Hauptgründe für die unterschiedliche Anzahl von Lehrbüchern zwischen den Ländern. Der erste Grund dafür ist die geringe Anzahl von Lehrbüchern, die in der Türkei veröffentlicht wurden. Der zweite Grund dafür ist die Absicht, für die Erstellung von Lehrbuchkorpus mehr Daten zur Verfügung zu stellen.

Die Auswahl der DaF-Lehrbücher erfolgte auf dem Niveau A1 bis B1, da in der Türkei Deutsch als erste Fremdsprache an Schulen mit Vorbereitungsklasse von A1 bis B1; Deutsch als erste Fremdsprache an Schulen ohne Vorbereitungsklasse von A1 bis

A2; Deutsch als zweite Fremdsprache an Schulen mit Vorbereitungsklasse von A1.1 bis A2.1 und Deutsch als zweite Fremdsprache an Schulen ohne Vorbereitungsklasse von A1.1 bis A1.2 unterrichtet wird (MEB 2018). Die von A1.1 bis B1.2 eingeteilten Lehrbücher werden zusammengetragen und als A1, A2 und B1 bearbeitet.

2.3 Datenerhebung und Analyse

Bei den in dieser Studie analysierten DaF-Lehrbüchern handelt es sich um Lehrwerke (A1-B1.1), die vom türkischen Bildungsministerium (MEB) für die Sekundarstufe genehmigt und von den vier großen Verlagen Hueber (A1.1-B1.2), Langenscheidt (A1-B1), Klett (A1-B1) und Cornelsen (A1-B1) für den Fremdsprachenunterricht in Deutschland veröffentlicht wurden. Die Wörter von den bestimmten DaF-Lehrwerken sind mit der webbasierten Software Sketch Engine, die eine Online-Software für Korpusmanager und Textanalyse ist, analysiert. Seit 2003 wird sie von Lexical Computing Limited mit den Programmiersprachen C++, JavaScript und Python erstellt. Die Sketch Engine bietet zurzeit sehr große Korpora in mindestens 90 Sprachen und ca. 50 Milliarden Frequenzen zur Verfügung und bietet in ihrer Datenbank mehr als 5.346.41.196 deutsche Wortfrequenzen.

Die ausgewählten regionalen und universellen DaF-Lehrwerke wurden in PDF-Dateiformat umgewandelt und auf den Computer übertragen, um die Analyse ohne Ablenkung durchzuführen. Die Inhalte außer den Themenseiten wie Wortlisten, Übungsseiten, Buchdeckel, Vorworten, Landkarten, Quellenverzeichnis, Inhalte ausgelassen. Die Wörter in PDF-Dateien wurden mit Hilfe von der OCR-Technologie durchsuchbar gemacht. In dem nächsten Schritt wurden die Bücher auf webbasierten Korpusmanager Sketch Engine hochgeladen und in 6 Wortklassen (Substantiv, Adjektiv, Konjunktion, Adverb, Präposition, Verb) analysiert. Als Referenzquelle für die Auswahl der Wortarten wurde das „das Häufigkeitwörterbuch der gesprochenen Sprache (HWB)“ von Arno Ruoff aus dem Jahr 2013 verwendet. Dieses Wörterbuch basiert auf 25 Jahren Vorarbeit und Tonbandaufnahmen am Tübinger Sprachenzentrum in Südwestdeutschland (Ruoff 2013: 3).

Häufigkeitswörterbuch gesprochener Sprache	Belege		Lemmata	
	abs.	%	abs.	%
Substantiv	54052	10,81	9398	59,95
Verb	105939	21,19	4414	28,16
Adjektiv	13995	2,80	1462	9,33
Adverb	50395	10,08	209	1,33
Konjunktion	62427	12,49	47	0,30
Präposition	27865	5,57	45	0,29
Partikel	30628	6,13	58	0,37
Artikel, Fragewort Pronomen	112474	22,49	43	0,27
Zahlwort	28056	5,61	–	–
Name	14169	2,83	–	–
Gesamt	500000	100,00	15676	100,00

Tab. 1: Wortartenverteilung und Anzahl der Lemmata von HWB (Ruoff 2013: 20)

Die 6 Wortarten [Substantiv (10,81%), Adjektiv (2,80%), Konjunktion (12,49%), Adverb (10,08%), Präposition (5,57%), Verb (21,19%)], die für diese Arbeit ausgewählt wurden, machen 62,94 % des Wörterbuches aus. Partikel gehört nicht zu diesen Wortarten. Die Ursache dafür liegt darin begründet, dass die Partikel den Lernenden des Deutschen als Fremdsprache und der Didaktik zahlreiche Schwierigkeiten bereiten. Eines dieser Probleme ist das Homonymenproblem, das zeigt, dass Partikel zu verschiedenen Klassen gehören kann (Busse 1992: 12). Im Wörterbuch sind die übrigen Wortarten [Artikel, Fragewort und Pronomen (22,49%), Zahlwort (5,61%) und Name (2,83%)] nur selten anzutreffen und haben nur eine begrenzte Anzahl von Lemmata (30,93%). Aus diesem Grund wurden sie nicht in die Analyse aufgenommen.

Für jede Wortart wurden nach der Analyse Frequenzlisten erstellt und in Excel-Formaten heruntergeladen, damit sie sich zur Analyse bereitstellen lassen. Die Wörter sind zu lemmatisieren, um die Wortarten auf die Korpora zu vergleichen. Eggers et al. (2015: 2) definiert Lemmatisierung als ein Verfahren, in dem die Begriffe bei sprachwissenschaftlichen Datenerhebungen in ihre lexikalische Hauptform verwandelt werden. Um eine korrekte Untersuchung durchführen zu können, klassifiziert Sketch Engine die konjugierten Verben, die Verben, die trennbar sind, mit ihren Präfixen und Hilfsverben als Grundform und zählt die Häufigkeit auf diese Weise. Die Adjektive (Nominativ, Akkusativ, Dativ und Genitiv), die in allen Kasusformen dekliniert werden, werden ebenfalls lemmatisiert.

Sketch Engine verfügt über einen Teil unter dem Namen Concordance für hochgeladenen Texten oder Textabschnitten, in dem die Wörter alphabetisch in den Sätzen untersucht und kontrolliert werden können. Die analysierten Präpositionen werden auch in Concordance eingehend überprüft. Partikeln werden in Sketch Engine automatisch unter Adverbien gezählt, jedoch werden die Partikeln in dieser Studie nicht hinzugefügt. Außerdem wird auch festgestellt, dass diese Engine nicht in der Lage ist, die zweiteiligen Bindewörter wie „zwar aber“ oder „sowohl als auch“ automatisch

analysieren zu können. Aus diesem Grund werden Bindewörter, die aus zwei Teilen bestehen, mit der Hand ausgeführt und auf die Tabelle eingetragen.

2.4 Statistische Auswertung

In dieser Studie wurden mit den 50 am häufigsten ausgewählten Wörtern aus dem TDaF-Korpus und dem DDaF-Korpus zwei Wortlisten erstellt, die das HD-Korpus bilden. Für die Ermittlung der Differenzen hinsichtlich der Spezifität der Wörter in diesen beiden Listen wurden die Wahrscheinlichkeitslog-Likelihood-Statistik (LL) und das Odds Ratio (OR)-Statistik für Effektstärken verwendet. Während die möglichen Nutzungsraten von Wörtern durch OR verglichen werden, wird LL gebraucht, um die im Korpus befindende Wörter zu beschreiben. Für die beiden Corpora werden die Worthäufigkeitslisten mithilfe Sketch Engine erstellt.

Zur Vergleichung der Ergebnisse wurde für OR und ebenso für LL dieselbe Häufigkeit verwendet. Es ist erwähnenswert, dass OR-Werte in Vergleichskorpora durch sehr geringe Frequenzen hart beeinflusst werden, d.h. die höchstrangigen Wörter im Vergleichskorpus kommen gar nicht oder mit einer Häufigkeit vor (Pojanapunya / Todd 2016: 16; Tomokiyo / Hurst 2003: 38).

Zur Berechnung in OR-Effektwahrscheinlichkeit von Wörtern mit niedriger Häufigkeit wurde die geringe Häufigkeitsschwelle, die das Wort in befindlichen Korpuskeln enthält, vor der OR-Berechnung als 5 akzeptiert. Diesbezüglich werden in der LL und der OR-Liste die ersten 50 Wörter genommen, denn es ist gewohnt, das Erste n für LL zu gebrauchen (Pojanapunya / Todd 2016: 16). Der Grund dafür ist, dass diese Werte von Dimensionen bewirkt und bloß die Erklärung des Forschers in bedenklich orientierten Studien bestätigt werden. In dieser Sachhaltung wurden die ersten 50 Wörter der Studie als geeignet für eine kritische und genreorientierte Recherche erachtet.

Um die LL zu berechnen, wird die folgende Kontingenztabelle eingesetzt.

	Korpus 1	Korpus 2	Insgesamt
Häufigkeit des Wortes	a	b	a+b
Häufigkeit anderer Wörter	c-a	d-b	c+d-a-b
Insgesamt	c	d	c+d

Tab. 2: Kontingenztabelle bei der Berechnung von LL (Rayson, Damon / Francis 2004: 3)

Bemerkungen:

Berechnung von OR

a = Frequenz eines Wortes in Korpus 1;
b = Frequenz eines Wortes in Korpus 2;
c = Gesamtanzahl der Wörter in Korpus 1;
d = Gesamtanzahl der Wörter im Korpus 2.

$$\text{OR} = (a / b) / (b / d) = ad / bc$$

a = Frequenz eines Wortes in Korpus 1;
b = Frequenz eines Wortes in Korpus 2;
c = Frequenz anderer Wörter in Korpus 1 (c-a);
d = Frequenz anderer Wörter im Korpus 2 (d-b).

Abb. 1: Berechnung von LL (Rayson, Damon / Francis 2004: 3)

Es ist offensichtlich, dass der 'c'-Wert der Wörterzahl in Korpus 1 und der 'd'-Wert der Wörterzahl (N-Werte) in Korpus 2 entspricht. Als Beobachtungswerte (O) werden 'a'- und 'b'-Werte bezeichnet. Erwartungswert (E) wird jedoch mit der nachstehenden Formel berechnet.

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

Abb. 2: Formel von Erwartungswert (Rayson, Damon / Francis 2004: 2)

Bei der neuen Forschung sind $N_1 = c$ und $N_2 = d$. Das bedeutet: $E_1 = c \cdot (a+b)/(c+d)$ und $E_2 = d \cdot (a+b)/(c+d)$. Die Größe der beiden Korpora wird bei der Berechnung der Erwartungswerte einberechnet. Nominalisierung der Daten vor der Anwendung der Formel ist also nicht notwendig. Nach der folgenden Formel ist der Log-Likelihood-Wert zu berechnen:

$$-2 \ln \lambda = 2 \sum_i O_i \ln \left(\frac{O_i}{E_i} \right)$$

Abb. 3: Formel von LL (Rayson 2021)

Die LL-Werte ergeben sich aus den obigen Berechnungen. Die Berechnung berücksichtigt die folgende Richtlinie von Paul Rayson:

95. Perzentil; 5%-Niveau; $p < 0,05$; kritischer Wert = 3,84
99. Perzentil; 1%-Niveau; $p < 0,01$; kritischer Wert = 6,63
99,9. Perzentil; 0,1 %-Niveau; $p < 0,001$; kritischer Wert = 10,83
99,99. Perzentil; 0,01% Niveau; $p < 0,0001$; kritischer Wert = 15,13

Abb. 4: LL-Werte (Rayson 2021)

3 Ergebnisse

Die Ergebnisse dieser Studie werden in zwei Teilen erläutert.

3.1 Frequenzen der Wortarten in HD-Korpus

Allgemeine Verteilung und die Frequenzen der Wortarten, aus denen das HD-Korpus entsteht, bildet das Ergebnis der quantitativen Analyse. Das HD-Korpus umfasst Lehrwerke für den Sekundarbereich in der Türkei sowie spezifische Lehrwerke der vier meistgenutzten Verlage in Deutschland. Die folgende Tabelle zeigt die Verteilung von Token und Typen der Bücher.

Türkische DAF-Bücher			Deutsche DAF-Bücher			HD-Korpus	
Token	Typen	%	Token	Typen	%	Token	Typen
90034	13884	23.94	345025	44133	76.06	435059	58017

Tab. 3: Die Aufteilung der Wörter Türkische-Deutsche DaF Bücher und in HD-Korpus

Gemäß der Tabelle ist ersichtlich, dass der HD-Korpus aus 435059 Token besteht und Wörter unterschiedlicher Art enthält, die mehr als einmal wiederholt werden. Der Anteil der weltweit veröffentlichten Lehrwerke am HD-Korpus beträgt 76,06%, während die regional veröffentlichten Lehrwerke mit 23,94 % eine geringere Anzahl ausmachen. Die Verteilung der Wörter im gesamten Korpus nach Wortarten ist in der Abbildung 5 dargeboten.

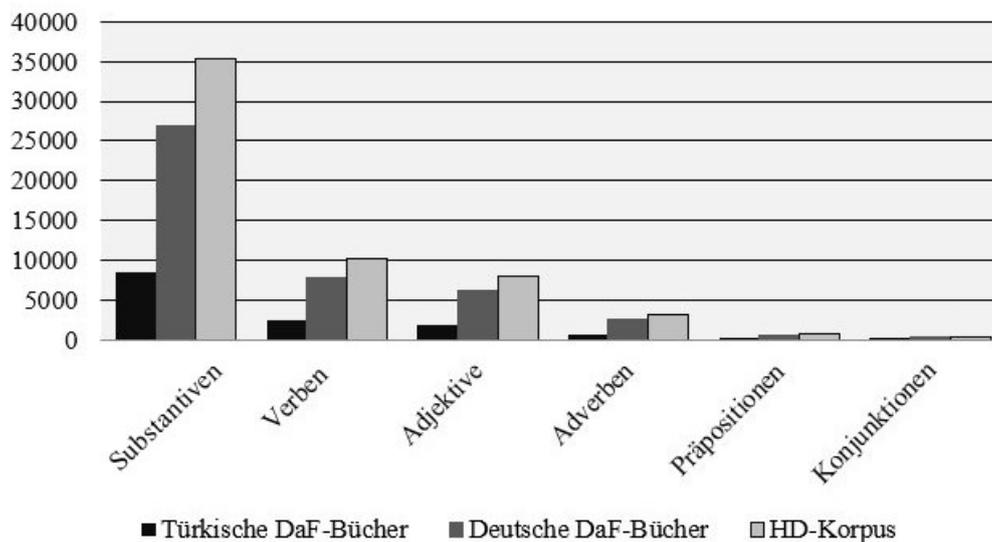


Abb. 5: Wortarteneinteilung im HD-Korpus

In Abbildung 5 ist zu sehen, dass Substantive nach ihren Arten den größten Anteil (n=35397; %=61,01) der Wörterverteilung im HD-Korpus in der ersten Reihe ausmachen. Die Verben (n=10193; %=17,56) sind an der zweiten Stelle und danach folgen die Adjektive (n=8006; %=13,80) und Adverbien (n=3259; %=5,62). Eine geringe lexikalische Auswirkung verfügen die Präpositionen (n=727; %=1.26) und Konjunktionen (n=435; %=0.75). Wenn die Lehrwerke, die das HD-Korpus bilden, verglichen werden, ist es möglich ähnliche Ergebnisse zu sehen. Die Wortgröße und

Verhältnisse des lexikalischen Häufigkeitsprofils in den Lehrbüchern sind in der nächsten Tabelle dargestellt.

In Tabelle 4 sind Token/Typ-Verhältnisse der Wortarten in den Lehrwerken aufgeführt.

Wortarten	Token	Type	TTR
Substantive	150307	35397	8.13
Verben	131129	10193	2.34
Adjektive	37585	8006	1.84
Adverbien	35396	3259	0.74
Präpositionen	52111	727	0.16
Konjunktionen	28531	435	0.09

Tab 4: Token/ Typenverhältnisse nach Wortarten

Wenn das TTR-Verhältnis an 1 liegt, wird das Vokabular von Büchern als groß angesehen, aber wenn die resultierende TTR größer ist, dann ist zu verstehen, dass die wiederholten Wortverwendungen weniger sind. Wenn eine Wortart aus 20 Wörtern besteht und diese alle unterschiedlich sind, dann ist es zu sagen, dass es zu einer „ideale“ TTR erreicht ist. Wenn sich dasselbe Wort zum Beispiel 20-mal wiederholt, ergibt sich ein Wert von $1/20 = 0,05$. Es wurde festgestellt, dass im HD-Korpus die gleichen Wörter (Substantive, Verben, Adjektive) mehrmals wiederholt wurden und große Lücken hatten. Gemäß den Ergebnissen aus der Tabelle, ist zu sehen, dass die Substantive (TTR = 8,13) in der ersten Reihe mit oft Verwendung den ersten Platz einnehmen; dann kommen Verben (TTR=2.34); Adjektive (TTR=1.84); Präpositionen (TTR=0,16) und Adverbien (TTR=0,74). Mit der geringen Wiederholung sind die Konjunktionen (TTR=0,09) in der letzten Reihe, aber sie haben in Zusammenhang auf den Wortreichtum die größte Varianz.

3.2 Häufig verwendete Wörter in DaF- Lehrwerken

In dieser Studie wurde aus den regional und universell veröffentlichten DaF- Lehrwerken mit den ersten 50 am häufigsten vorkommenden Wörtern eine Situationstabelle erarbeitet und für jede Art von Wörtern im Zielkorpus wurde LL berechnet. In Tabelle 5 lassen sich die ersten 50 Wörter nach LL und OR deutlich erkennen.

Ein genauerer Blick auf die beiden Listen zeigt, dass die 50 wichtigsten Wörter in den universellen und regionalen Lehrwerken mit LL und OR in der Rangliste übereinstimmen. Es wurde jedoch herausgefunden, dass die Wörter (arbeiten, als, Frage, Frau, Satz, dann, so, immer, mehr, nur, spielen) in den regionalen DaF- Lehrwerken nicht auf der LL-Liste stehen.

Aus dem Gesagten lässt sich schließen, dass eine hundertprozentige Übereinstimmung nicht in Frage kommt. Die geordneten LL-Listen weisen darauf hin, dass zumindest eine der Wortarten (Substantiv, Verb, Adjektiv, Adverb, Präposition und Konjunktion) in den Lehrwerken enthalten ist. Die Vokabeln, aus denen sich die LL-Liste

zusammensetzt, beinhalten deutlich Wörter aus dem HD-Korpus, der als ein Teil der Arbeit als „Häufigkeitswörterbuch“ entworfen wurde. Das sind die Wörter, die die Bedeutung des DDaF-Korpus aufweisen, der eines der Subkorpusse des HD-Korpus ist. Bei der Betrachtung der Tabelle 5 ist die Reihenfolge der Wortliste LL und OR-Lehrwerks auch bemerkenswert, denn die Wörter „sein, sprechen, müssen und Jahr“ stehen in derselben Reihenfolge.

rangiert von LL						rangiert von OR							
Rang	Wort	f(TDaF)	f(DDaF)	f(HD-Korpus)	LL	OR	Rang	Wort	f(TDaF)	f(DDaF)	f(HD-Korpus)	OR	LL
1	sein	4144	14529	18395	25.09	1.10	1	sein	4144	0	17083	0.30	5751.54
2	in	3530	9880	15506	243.78	1.38	2	und	4032	0	15226	0.33	4596.79
3	haben	2008	7923	17020	1.37	0.97	3	in	3530	0	5472	0.81	92.70
4	mit	1048	4009	8691	0.00	1.00	4	haben	2008	0	5204	0.48	822.42
5	zu	852	3576	7414	5.80	0.91	5	mit	1048	0	5178	0.25	2113.43
6	und	4032	3522	7302	3898.82	4.55	6	an	953	0	4427	0.27	1699.72
7	können	942	3424	6062	2.04	1.05	7	können	942	0	4145	0.28	1502.71
8	an	953	3294	5124	7.72	1.11	8	machen	897	0	4016	0.28	1483.48
9	machen	852	3208	4144	0.21	1.02	9	zu	852	0	3912	0.27	1484.03
10	für	521	2912	3759	68.61	0.68	10	lesen	829	0	3890	0.27	1508.99
11	gut	657	2515	3560	0.00	1.00	11	von	720	0	3685	0.24	1554.71
12	auf	632	2455	3277	1426.38	2.92	12	gehen	673	0	2521	1.29	32.75
13	von	720	2325	2907	15.63	1.19	13	hören	665	0	2500	1.29	31.50
14	hören	665	2136	2800	15.31	1.19	14	gut	657	0	2372	1.34	41.31
15	gehen	673	2063	2792	24.33	1.25	15	auf	632	0	2214	1.38	47.83
16	werden	829	1922	2607	135.90	1.66	16	schreiben	623	0	2203	1.37	44.60
17	lesen	829	1865	2547	150.87	1.71	17	mögen	589	0	2096	1.36	40.57
18	sprechen	555	1840	2420	1092.91	2.95	18	sprechen	555	0	2073	1.30	27.60
19	oder	378	1741	2323	10.92	0.83	19	für	521	0	1986	1.27	22.15
20	dann	226	1735	2069	115.09	0.50	20	nach	496	0	1900	1.26	20.26
21	finden	344	1688	2057	18.43	0.78	21	kommen	479	0	1897	1.22	14.54
22	kommen	479	1663	1967	3.56	1.10	22	Text	433	0	1838	1.14	5.73
23	so	151	1632	1886	198.14	0.35	23	Modul	413	0	1808	1.10	3.21
24	mögen	589	1609	1829	46.75	1.41	24	noch	385	0	1773	1.05	0.72
25	bei	247	1567	1772	61.14	0.60	25	auch	382	0	1656	1.12	3.58
26	aber	255	1520	1710	47.25	0.64	26	oder	378	0	1594	1.15	5.51
27	geben	343	1512	1654	5.64	0.87	27	sehr	375	0	1589	1.14	5.09
28	schreiben	623	1506	1647	87.29	1.59	28	essen	366	0	1581	1.12	3.65
29	sehen	50	1481	1629	403.80	0.13	29	Uhr	362	0	1553	1.13	4.05
30	müssen	357	1432	1615	0.60	0.96	30	müssen	357	0	1553	1.11	3.14
31	aus	307	1423	1523	9.52	0.83	31	Klasse	352	0	1525	1.12	3.33
32	nach	496	1285	1508	51.57	1.48	32	finden	344	0	1502	1.11	2.82
33	Satz	209	1241	1465	37.99	0.64	33	geben	343	0	1419	1.17	6.45
34	Frau	113	1238	1462	153.09	0.35	34	wollen	342	0	1416	1.17	6.37
35	Jahr	323	1206	1456	0.17	1.03	35	Jahr	323	0	1343	1.16	5.70
36	gern	263	1190	1441	6.16	0.85	36	über	319	0	1265	1.22	9.56
37	arbeiten	0	1178	1438	546.27	0.00	37	um	314	0	1249	1.22	9.13
38	als	164	1169	1386	64.59	0.54	38	Schule	313	0	1216	1.24	11.34
39	über	319	1119	1379	1.91	1.09	39	werden	308	0	1203	1.24	10.64
40	wollen	342	1098	1378	7.91	1.19	40	aus	307	0	1130	1.31	16.96
41	sehr	375	1091	1342	20.22	1.32	41	Dialog	302	0	1119	1.31	15.92
42	fahren	96	1085	1325	139.75	0.34	42	Bild	299	0	1096	1.32	16.99
43	sagen	258	1071	1318	1.35	0.92	43	Brage	286	0	1017	1.36	19.79
44	Frage	249	1033	1291	1.28	0.92	44	einmal	270	0	9764	0.13	2140.39
45	Uhr	339	1009	1246	15.52	1.29	45	gern	263	0	9755	0.13	2163.59
46	einmal	270	986	1225	0.49	1.05	46	sagen	258	0	9710	0.13	2167.87
47	immer	168	922	1169	19.88	0.70	47	lernen	258	0	9242	0.13	2016.96
48	schon	84	916	1169	112.56	0.35	48	aber	255	0	9156	0.13	2000.40
49	nur	119	911	1166	60.04	0.50	49	bei	247	0	8778	0.13	1908.52
50	spielen	214	903	1131	1.64	0.91	50	sollen	211	0	8684	0.12	2015.14

Tab. 5: Vergleich der ersten 50 Wörter der Subkorpora (TDaF und DDaF) von HD-Korpus

Nach diesen Feststellungen ist es zu sagen, dass in den DaF-Lehrwerken primär wichtige Wörter vermittelt werden sollten und diese Wörter einen beachtenswerten Einfluss haben. Die in der Tabelle 5 dargebotenen Wörter repräsentieren bloß einen kleinen Prozentsatz aller Wortarten und Varietäten im HD-Korpus. Betrachtet man die Häufigkeit der in den DaF- Lehrwerken verwendeten Wörter, so kann man sagen, dass die Differenz zwischen den beiden in der Tabelle 5 dargestellten Listen gering ist. So ist es auch zu sehen, dass jedes Wort OR-Werte nahe 1 und weit 0 LL-Werte hat. Ein nahe 0 liegender LL-Wert bedeutet, dass die verwendeten Wörter nicht einen ausreichenden Beleg ausgeben, die im

Unterricht beigebracht werden sollten. Deswegen waren in dieser Studie besondere Sonderfälle nötig, um die entstellten OR-Werte zu verhindern.

4 Schlussfolgerung und Diskussion

Die Studie lieferte eine Antwort auf die Frage, wie oft die Wörter im Korpus für DaF-Lehrbücher verwendet werden. Als Ergebnis dieser Forschung kann gesagt werden, dass in der ersten Reihe die Substantive, dann die Verben, Adjektiven, Adverbien danach Präpositionen im Korpus für DaF-Lehrbücher stehen. Die am wenigsten vorkommenden Wortarten in den Fremdsprachenlehrwerken sind die Konjunktionen.

Darüber hinaus beantwortete die Studie auch die Frage, welches die am häufigst verwendeten Wortarten in den DaF-Lehrbüchern sind. Der Grund, warum Substantive unter den deutschen Wörtern am häufigsten verwendet werden, kann durch die Struktur der deutschen Sprache aufgeklärt werden. Die Ergebnisse dieser Forschung stimmen mit den von Klosa (2013: 14) durchgeführten Forschungsergebnissen überein, indem er diesen Sachverhalt damit aufklärt und sagt, dass das Deutsche eine Wortstruktur hat, die hauptsächlich aus zusammengesetzten Wörtern besteht, was für die Wortableitung geeignet ist. Auch deuteten Fleischer / Barz (2012: 2) darauf hin, dass Substantive bezüglich auf Wortarten den größten Platz einnehmen und sie zeigen an, dass die Adjektive in der zweiten Reihe stehen. Das Faktum, dass Adjektive den größten Platz in Bezug auf Wortarten nach Substantiven behalten, spricht den Ergebnissen dieser Studie entgegen.

Die Frage „In welchem Verhältnis stehen die Token/Typ-Relationen in den DaF-Lehrwerken zueinander?“ ist eine der wichtigen Fragen dieser Studie. Eines der wichtigsten Ergebnisse dieser Studie sind die Summen der Token/Type R (TTR)-Analyse, die darauf abzielt, den Zusammenhang zwischen der Wortgröße der untersuchten deutschen Lehrwerke und dem lexikalischen Häufigkeitsprofil aufzuzeigen. Volumetrischer Raum, die von einem Wort eingenommen wird, heißt nicht, dass es in Hinsicht auf 55 Genre reich und vielfältig ist. Das Ziel ist, auf diese Frage eine Antwort zu geben: „Zeigt es die Vielfalt des Wortschatzes oder die Wiederholung von Wörtern in deutschsprachigen Lehrbüchern?“ Nach den Daten weisen die TTR-Ergebnisse auf, dass in den deutschsprachigen Lehrwerken stehende Adverbien, die das HD-Korpus erstellen, eine breitere Mehrheit besitzen als die Frequenzprofile. Obwohl Substantive eine hohe volumetrische Häufigkeit haben, haben sie im Gegensatz zu den Adverbien wenig Vielfalt in den deutschen Lehrwerken. Dies ist ein Beweis, dass gleichartige Wörter mehrmals wiederholt werden. Nach allem sind für die im Rahmen der aktuellen Studie erstellten DaF-Lehrwerke wichtige Erkenntnisse zur Validität der Wörter im Korpus identifiziert. Zu diesem Zweck wurden auf zwei unterschiedliche Listen LL- und OR-Tests angewendet. Ergebnisse LL schaffen die Möglichkeit die Wörter zu erkennen, die Wörter zu identifizieren, die im Vordergrund stehen und die während des Unterrichts gelehrt werden müssen. Hingegen OR trägt zur Identifizierung spezifischerer Wörter bei und ermöglicht die sprachlichen Merkmale der Zielsprache zu beschreiben. Es lässt sich aus den gewonnenen Resultaten schließen, dass die Ausdrücke „sein“, „sprechen“,

„Zustand“ und „Jahr“ bestimmt gelehrt werden müssen. In diesem Kontext kann festgehalten werden, dass der HD-Korpus nützlich sein kann, um bei der Erstellung von DaF-Lehrwerken Schlüsselwörter auszumachen.

Im Ganzen lässt sich sagen, dass das neu erstellte HD-Korpus für DaF-Lehrwerke einen bedeutenden Beitrag zum Lehren und Lernen von Deutsch im Ausland leisten wird. Es ist auch zu sagen, dass solche Lehrwerke mit tiefer Lexik, Worthäufigkeit, Verbreitung und Wortvielfalt viele Möglichkeiten bieten.

In den ausgewerteten DaF-Lehrwerken wurde festgestellt, dass die Worthäufigkeit aus Adjektiven, Verben, Substantive, Präpositionen, Konjunktionen und Adverbien besteht. Lernende, die über diese Art von Wissen verfügen, werden keine Schwierigkeiten haben, Methoden und Lösungen zu finden, die für sie in Bezug auf die Verwendung von Wörtern und deren Verwendung beim Sprachenlernen effektiv sein können.

Die Forschung hat versucht aufzuzeigen, wie wichtig es ist, die Wörter in den Lehrwerken zu untersuchen, die ein wichtiges Mittel für erfolgreiches Sprachenlernen und -lehren sind. Sie hat auch versucht zu belegen, dass die volumetrischen Größen der Vokabeln in den DaF- Lehrwerken ein indirektes Verhältnis zur Wortvielfalt haben. Diese Studie ergab, dass die in Lehrwerken häufig verwendeten Wortarten zwar Substantive sind, aber im Vergleich zu Adverbien in der Minderheit stehen, wenn es um Vielfalt geht. Diese Situation lässt erkennen, dass das gleiche Wort ein paarmal in unterschiedlichen Texten in den regional und universell veröffentlichten Lehrwerken wiederholt wird. Es zeigt, dass es eine Tendenz zur Steigerung der Namensvielfalt in Lehrwerken geben sollte. Das HD-Korpus liefert auf der Grundlage der Ergebnisse des LL- (Log-Likelihood) und des OR-Tests (Odds-Ratio) präzise Hinweise darauf, welche Wortarten in Lehrwerken am häufigsten vorkommen. Diese Situation gibt einen Hinweis für den Entwurf der zukünftigen DaF-Lehrwerke und macht die wichtigen Wörter deutlich.

Das betreffende HD-Korpus kann den Lernern, den Lehrern und auch den Schriftstellern im In- und Ausland verschiedene Hilfestellungen anbieten. Es können auch praktische Untersuchungen zu den Wörtern in den Lehrbüchern, die oft verwendet werden und den Schülern Lernschwierigkeit zeigen, durchgeführt werden und es können Erklärungen dazu gegeben werden, welche Strategien sie beim Lernen der Wörter anwenden sollten (Çelikkaya 2012). Um die Ergebnisse dieser Studie zu untermauern und das Vorhandensein und die Häufigkeit von Wörtern zu bestimmen, kann eine angewandte Forschung im Kontext alltäglicher Themen zur schriftlichen und mündlichen Leistung von Deutschlernenden durchgeführt werden. Es können auch praktische Untersuchungen zu den Wörtern in den Büchern, die oft verwendet werden und den Schülern Lernschwierigkeit zeigen, durchgeführt werden. DaF-Lernende können ihren Wortschatz unterstützen, erweitern und vertiefen, indem sie ihr eigenes Hochfrequenz-Wörterbuch erstellen und auf diese Weise kann die Wirkung der Worthäufigkeit auf den Lernvorgang recherchiert werden. Die möglichen Auswirkungen auf den Lernvorgang können ein dienliches und belehrendes Thema für weitere Studien sein.

Literaturverzeichnis

- Chmiliar, Linda** (2010): Multiple-case designs. A. J. Mills, G. Eupras / E. Wiebe (Eds.): *Encyclopedia of case study research* (pp 582-583). USA: SAGE Publications.
- Clarín-D Center:** Corpora, <https://www.clarin-d.net/en/corpora#BAS> (Letzter Zugriff: 12.11.2021).
- Creswell, John Ward** (2007): Qualitative inquiry & research design: *Choosing among five approaches* (2). USA: SAGE Publications.
- Çelikkaya, Şerife** (2012): Almanca Öğretmeni Adaylarının Sözcük Öğreniminde Kullandıkları Öğrenme Stratejileri, *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 28, 71-76.
- Davey, Lynn** (1991): The application of case study evaluations. *Practical Assessment, Research, and Evaluation*, 2, 1-3.
- Denham, Cristin / Lobeck, Anne** (2013): *Linguistics for Everyone: An Introduction*. USA.
- Deutsches Textarchiv:** Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Herausgegeben von der Berlin-Brandenburgischen Akademie der Wissenschaften. <https://www.deutschestextarchiv.de/> (Letzter Zugriff: 10.12.2021).
- Diamond, Linda / Gutlohn, Linda** (2006): *Vocabulary handbook*. California: Core Publications.
- Dietrich, Busse** (1992): Partikeln im Unterricht Deutsch als Fremdsprache Semantische und didaktische Probleme der Synsemantika. *Muttersprache* 102, 37–59.
- Digitales Wörterbuch der deutschen Sprache (DWDS):** Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart, hrsg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften, <https://www.dwds.de/> (Letzter Zugriff: 10.12.2021).
- Dolan, Hakan** (2022): *Die Lehrbücher im Deutsch als Fremdsprache: Eine korpusbasierte Analyse der lexikalischen Vielfalt* (Unveröffentlichte Magisterarbeit). Anadolu Universität Institut für Erziehungswissenschaften, Eskişehir.
- Dolan, Hakan / Çelikkaya Şerife** (2024): Wortfrequenzbasierte Analyse von Hörtexten im Deutschkursbuch „Menchen A1.1“. *International Journal of Language Academy*, 11 (5), 272-288.
- Eggers, Hans** u.a. (2015): *SALEM: Ein Verfahren zur automatischen Lemmatisierung deutscher Texte*. Tübingen: Max Niemeyer Verlag.
- Fandrych, Christian, Tschirner Erwin** (2007): Korpuslinguistik und Deutsch als Fremdsprache. Ein Perspektivenwechsel. *Deutsch als Fremdsprache* 44 (2007), S. 195-204.
- Fleischer, Wolfgang / Barz, Irmhild** (2012): *Wortbildung der Deutschen Gegenwartssprache*. (4. Auflage). Berlin und Boston: Walter de Gruyter GmbH & Co.
- Gemeinsamer europäischer Referenzrahmen für Sprachen (GER)** (2001): *lernen, lehren, beurteilen*. Übersetzt von Quetz, Jürgen in Zusammenarbeit mit Schieß, Raimund & Sköries, Ulrike. Berlin: Langenscheidt.
- Gut, Ulrike** (2007): Sprachkorpora im Phonetikunterricht. *Zeitschrift für interkulturellen Fremdsprachenunterricht*, 12(2).
- Harmer, Jeremy** (1991): *The practice of English language teaching*. New York: Longman.
- Henzl, Vera** (1979): Foreigner talk in the classroom. *International Review of Applied Linguistics* 17, 159-67.

- Kantemir, Enise** (1997): *Yazılı ve Sözlü Anlatım*. Ankara: Engin Yayınevi.
- Klosa, Annette** (2013): *Wortbildung im elektronischen Wörterbuch*. Tübingen: Narr Francke Attempto Verlag.
- Lüdeling, Anke / Walter, Maik** (2009): Korpuslinguistik für Deutsch als Fremdsprache. *Sprachvermittlung und Spracherwerbsforschung*, 1-37.
- Lüdeling, Anke**, u.a. (2008): Das Lernerkorpus Falko. *Deutsch als Fremdsprache* 45 , S. 67-73.
- Lymperakakis, Panagiotis / Andromachi, Sapiridou** (2012): Wortschatz und Worthäufigkeit – wie frequent ist eigentlich der Lernwortschatz aus DaF-Lehrwerken für Jugendliche? *Deutsch als Fremdsprache* 03, 139-149.
- McEnery, Tony / Wilson, Andrew** (2001): *Corpus Linguistics: An Introduction*, Edinburgh, Edinburgh University Press.
- Merriam, Sharan** (2013): *Qualitative research: A guide to design and implementation*. Jossey-Bass.
- Milli Eğitim Bakanlığı (MEB)**, (2018): Almanca Dersi Öğretim Programı. <https://mufredat.meb.gov.tr/> (Letzter Zugriff: 19.04.2024)
- Nation, Ian / Stephan, Paul** (2001): *Learning vocabulary in another language*. Cambridge University Press.
- Pojanapunya, Pumjaporn / Todd, Richard Watson** (2016): Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory* 14, S.1-35.
- Rayson, Paul** u.a. (2004): Extending the Cochran rule for the comparison of word frequencies between corpora. *JADT 2004: 7es Journées internationales d'Analyse statistique des Données Textuelles*.
- Rayson, Paul**: How to calculate log likelihood. <https://ucrel.lancs.ac.uk/llwizard.html> (Letzter Zugriff: 12.11.2021).
- Römer, Ute** (2008): *Corpora and language teaching*. Lüdeling, Anke und Merja Kytö (Hg.), 112-131.
- Runte, Maren** (2015): *Lernerlexikographie und Wortschatzerwerb*. Leck: CPI books GmbH.
- Ruoff, Arno** (2013): *Häufigkeitwörterbuch gesprochener Sprache: Gesondert nach Wortarten, alphabetisch, rückläufig alphabetisch und nach Häufigkeit geordnet*. Berlin: Walter de Gruyter.
- Schmidt, Karin** (2011): Lernerkorpora: Ressourcen für die Deutsch-als-FremdspracheForschung. Eğit, Yadigar (Ed.): *Globalisierte Germanistik: Sprache, Literatur, Kultur: Tagungsbeiträge*; XI. Türkischer Internationaler Germanistik-Kongress (2009), 555-573.
- Schmitt, Norbert** (2000): *Vocabulary in language teaching*. Cambridge University Press.
- Thomas, Dax**: Type-token Ratios in One Teacher's Classroom Talk: An Investigation of Lexical Complexity. <https://www.birmingham.ac.uk/Documents/college-artslaw/cels/essays/language-teaching/DaxThomas2005a.pdf> (Letzter Zugriff: 21.11.2021)
- Todd, Richard** (2001): Induction from self-selected concordances and self-correction, *System* 29: 1, 91-102.
- Tognini Bonelli, Elena** (2001): *Corpus Linguistics at Work*, Amsterdam ve Philadelphia, John Benjamins.
- Tomokiyo, Takashi / Hurst, Matthew** (2003): A language model approach to keyphrase extraction. *Proceedings of the ACL [Association for Computational Linguistics] 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment [MWE'03]*. Sapporo: Association for Computational Linguistics, 33-40.
- Tschirner, Erwin** (2008): Das professionelle Wortschatzminimum im Deutschen als Fremdsprache. *Deutsch als Fremdsprache* 45, 195–208.

Walter, Maik / Grommes, Patrick (2008): *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer.

Zimmerman, Cherly Boyd (1997): *Historical trends in second language vocabulary instruction*. Cambridge University Press.

ANHANG-1

<i>Nr.</i>	<i>Lehrwerk</i>	<i>Verlag</i>	<i>Erscheinungsjahr</i>
1	Deutsch für Gymnasien A1.1 Schülerbuch	ATA	2020
2	Wie Bitte A1.1 Lehrbuch	MEB	2020
3	Wie Bitte A1.2 Lehrbuch	MEB	2016
4	Plus Deutsch A1.1 Schülerbuch	KOZA	2019
5	Deutsch macht Spaß A1.2 Schülerbuch	MEB	2018
6	Deutsch macht Spaß A2.1 Schülerbuch	MEB	2018
7	Deutsch macht Spaß A2.2 Schülerbuch	MEB	2018
8	Mein Schlüssel zu Deutsch B1.1 Schülerbuch	MEB	2018
9	Schritt für Schritt Deutsch A1.1	ATA	2016
10	Deutschstube A1.1 Schülerbuch	MEB	2014
11	Deutsch ist Spitze A1.1	MEB	2010
12	Deutsch ist Spitze A1.2	MEB	2011

Tab. 6: Auswahl der türkischen DaF-Lehrbücher

ANHANG-2

<i>Nr.</i>	<i>Lehrwerk</i>	<i>Verlag</i>	<i>Erscheinungsjahr</i>
1	Pluspunkt Deutsch Leben in Deutschland Kursbuch	Cornelsen	2017
2	Pluspunkt Deutsch Leben in Deutschland Kursbuch	Cornelsen	2017
3	Pluspunkt Deutsch Leben in Deutschland Kursbuch	Cornelsen	2017
4	Pluspunkt Deutsch Kursbuch A1	Cornelsen	2007
5	Pluspunkt Deutsch Kursbuch A2	Cornelsen	2007
6	Pluspunkt Deutsch Kursbuch B1	Cornelsen	2007
7	Schritte International Neu 1 A1.1 Kursbuch	Heuber	2016
8	Schritte International Neu 2 A1.2 Kursbuch	Heuber	2016
9	Schritte International Neu 3 A2.1 Kursbuch	Heuber	2016
10	Schritte International Neu 4 A2.2 Kursbuch	Heuber	2016
11	Schritte International Neu 5 B1.1 Kursbuch	Heuber	2016
12	Schritte International Neu 6 B1.2 Kursbuch	Heuber	2016
13	Menschen A1.1 Kursbuch	Heuber	2012
14	Menschen A1.2 Kursbuch	Heuber	2012
15	Menschen A2.1 Kursbuch	Heuber	2013
16	Menschen A2.2 Kursbuch	Heuber	2013
17	Menschen B1.1 Kursbuch	Heuber	2014
18	Menschen B1.2 Kursbuch	Heuber	2014
19	Genial Deutsch für Jugendliche Kursbuch A1	Klett & Langenscheidt	2013
20	Genial Deutsch für Jugendliche Kursbuch A2	Klett & Langenscheidt	2013
21	Genial Deutsch für Jugendliche Kursbuch B1	Klett & Langenscheidt	2013
22	Logisch Deutsch für Jugendliche A1	Klett & Langenscheidt	2009
23	Logisch Deutsch für Jugendliche A2	Klett & Langenscheidt	2009
24	Logisch Deutsch für Jugendliche B1	Klett & Langenscheidt	2011

Tab. 7: Auswahl der deutschen DaF-Lehrbücher