Başaran, B. (2022). Measuring Word Frequency in Language Teaching Textbooks using LexiTürk. *International Online Journal of Education and Teaching (IOJET)*, *9*(1). 571-583.

Received : 12.09.2021
Revised version received : 14.12.2021
Accepted : 17.12.2021

MEASURING WORD FREQUENCY IN LANGUAGE TEACHING TEXTBOOKS USING LEXITÜRK

Research Article

Bora Başaran (ORCID ID: **0000-0003-0251-5895**).

The *Ministry* of National *Education*, Turkey

bbasaran@meb.gov.tr

Biodata:

Bora Başaran has been with the Faculty of Education at Anadolu University since 1999 and tried to create and disseminate knowledge in the Department of Foreign Language Education. He has been working as the Attaché of Education of the Turkish Republic in Frankfurt since September 2017.

MEASURING WORD FREQUENCY IN LANGUAGE TEACHING TEXTBOOKS USING LEXITÜRK

Bora Başaran

bbasaran@meb.gov.tr

Abstract

Vocabulary is a fundamental component of language usage, and study into its interactions with other aspects of language competence is an essential topic of language teaching. There are strong relationships between vocabulary and language competency measurements. Learners with larger vocabularies are better at a variety of language abilities than those with lower vocabularies. As a result, it may be stated that vocabulary knowledge is inextricably tied to total language proficiency. This suggests that the quantity of words we know has an impact on how much text we can comprehend.

The more often a word is used, the more polysemy and irregular morphology it is likely to have. One of a word's quantifiable qualities is how extensively it is used. Based on this measurable attribute, the word's prevalence and frequency can be considered a guiding reference. Analyzing the frequency of recurrence of a specific word or phrase is the most basic sort of corpus analysis. Words frequently occur together, forming collocations, colligations, and other word combinations. Exploring such trends is another sort of corpus analysis that one may perform. This is also known as chunks, n-grams, or lexical bundles.

In particular, when selecting which word should be prioritized for language learners. If this model is adopted, foreign language students will be taught essentially the most often used words. The development and application of measurement like an analysis tool can assist developers or researchers in the dilemma of preferences arising from the inevitable use of a word corpus. For teaching after the stage of creating a corpus and shaping the language textbook's content with the marked and dense frequency words from a corpus are discussed in this study, and a tool which is created by the author is presented to and for the scientific community.

Keywords: Word Frequency, Language Teaching, Comparing Corpora, LexiTürk,

1. Introduction

Lexicography attempts to re-present words that were previously divided by the alphabet using various representation approaches that reorder them depending on actual language use and context. The process of word recontextualization may be explained by looking at it from two perspectives: macro and microstructure. It's essential to remember that corpus isn't only a 'container' for words; it may also allow language and words to expand beyond their presumed limits. Morphology is closely linked to other language levels such as syntax, text, discourse, and also the listing of words in any dictionary. Focusing on words that are challenging by the subject and type of text can also eliminate the various associations that unite individual words with other words in the relevant language. This is linked to the indefinable differences in quantities between the morpheme and the syntactic unit (Storjohann, 2015).

Teachers and learners may interpret the relation between the words delivered in a textbook, as in a unit, theme, and general linguistic context, differently than the writers of the textbooks intended. The vocabulary preferences of professional textbook writers are based on a linguistic

use that must be portrayed as realistically as possible and depending on the specific term and context (Üstün, B. & Tanrıkulu, L., 2021). These preferences can be defined as usual preferences, usually using extensive corpora of words and topics, and created from comprehensive texts. Those who study the essential foreign language from a textbook, on the other hand, frequently access and interpret descriptive terms in a prescriptive manner and view the current explanation as a prescription for the language acquisition process. Textbooks, in this perspective, are instruments that give direction for the proper and wrong usage of words in terms of establishing standards and determining the study's focus.

Over the years, a diverse typology of words has emerged in textbooks and even micro-frame dictionaries that goes well beyond the typical divide between bilingual and monolingual dictionaries.

Kuhn's dictionary categorization, based on dictionary use functions, is widely referenced in this field (Engelberg, Lemnitzer 2009).

When one examines the vocabulary used in textbooks from a typological standpoint, one may see that;

- The textbooks contain a vocabulary that is transmitted based on their units and subjects, such as seasonal definitions, proper nouns, antonyms, verbs, fundamental vocabulary, and related terms,

- In terms of the variety re-presented by textbook units and topics, dialects, technical language, literary language, author's language, and colloquial language, for example, frequently necessitate the use of other levels of the language in question, such as morphology and syntax, hidden, written, or spoken word. This is difficult to perform in a textbook since spoken language documentation is significantly more complicated than written language documentation.

-Foreign language textbooks usually focus on a few key pieces of information in specific ways in this respect. For instance, consider spelling, pronunciation, meaning, etymology, and semantic contexts.

It is self-evident that the set of words in textbooks does not directly correlate to the vocabulary of the language's community of speakers or that textbooks cannot include a language's whole vocabulary. One of the key reasons for this is that around half of all words in writings, and potentially even in speech, are only used once (Hapax Legomena). Foreign language textbooks often need to include a more familiar, subject-oriented vocabulary with a particular frequency and distribution in this respect. New word formations are constantly added to languages, and no textbook can capture them all.

The living dimension and development of languages are happening even faster today, and textbooks can often lag behind the development of the vocabulary described as social vocabulary, mainly due to the design and production processes.

2. Literature Review

Corpus linguistics is described as "the compilation and analysis of corpora" (Cheng 2012), which are vast collections of "naturally occurring language texts chosen to characterize a state or a variety of language" (Sinclair, 1991). Even though corpus linguistics is a relatively new area, it has transformed language studies (Hunston, 2002), because it has provided new means

of analyzing and explaining language use. The author highlights that corpora are collections of texts preserved in an electronic format, allowing academics to conduct automatic searches and obtain insights into the structure and regularity of naturally occurring language using specific methods and software (concordancers).

Because much corpus linguistics research has been done on English texts, this study will focus on corpora that reflect Turkish. It's important to note, too, that other languages have corpora, and that some nations (such as Germany and Turkey) have a long history of corpus-based research.

Electronic possibilities, which have been enthusiastically accepted by linguistics, lexicographers, and dictionary users in recent years, have shown ways to combine the lexicographic description of the word with non-dictionary discourses. Text, and speech worlds, thus enabling the use of language that goes beyond individual words or little words. Textbook authors typically organize language use around predetermined text sets (corpora), including keywords, discourse-typical syntax, metaphorical expressions, talk and text evaluations, and various other characteristics of a discourse made possible by these electronic potentials.

Corporeal Linguistics' core argument is the construction of a text corpus or corpus of texts based on a specified design and uploading them to a computer for use in linguistics research and descriptions. The goal is to create a miniature model of the language from the total of the texts recorded on the computer. To do so, a linguistically explicit criterion or criteria for selecting and organizing language samples must be determined and used. After forming the corpus by integrating texts from several language segments according to a certain design, the corpus is computer-aided analyzed. As a consequence, data-driven decisions concerning various aspects of language are made. Due to the difficulty of data collecting and scope in corpus linguistics, studies in the range of "Word Frequency," also known as "Language thrift law" in Turkish, have primarily been conducted in the educational sphere. For Turkish, İlyas Göz's (2003) "Word Frequency Dictionary of Written Turkish."

"A Study on the Written Vocabulary of 5th, 8th and 11th Grades in Uşak Merkez Primary School" conducted by Mustafa Çıplak (2005), Altan Avkapan's (2006) study titled "A Research on the Vocabulary of 11th Grade Students in Secondary Education" and by Ufuk Aşık and the studies of Belkıs Kılınçarslan (2009), "A Study on Word Frequency in Orhan Veli's Poems."

Gökhan Ölker's "Word Frequency Dictionary of Written Turkish (Between 1945-1950)" are examples of other Turkish word frequency studies. Among the Turkish corpus studies, "A Study on Developing a Corpus in the Computer Environment," also known as the Middle East Technical University Turkish Corpus, was undertaken under the supervision of Bilge Say and solely contains written language data. It's a two-million-word corpus generated by taking 2000-word samples from various writings, transferring them to an electronic environment, and tagging them. The "Turkish National Corpus" is another corpus research.

It is also known as Mersin University corpus. It is a corpus financed by TUBITAK (The Scientific and Technological Research Council of Turkey) and developed by Mersin University Linguistics academics. This simultaneous and broad corpus, which was started in 2008 and the introductory edition published in 2012, has a capacity of 50 million words and includes 95 percent written and 5% spoken samples in many sectors and categories between 1990 and 2009. The "Oral Turkish Corpus (STD)," supported by TUBITAK between 2008 and 2010, is an online project aimed at studying current Turkish in a computer environment by examining a database with a volume of one million words comprising of face-to-face or telephone conversations.

Taner Sezer developed the first edition of the general TS Corpus, which had 491 million words, accessible online on March 1, 2012. The TS Corpus is a general-purpose corpus organized by word type, morpheme, and root word tags and provides a lot of ease to the user. Historical Corpus of Old Turkish and Karakhanid Turkish (ETKT-D) 400-450 thousand words and covering over 600 years (7-13th century), created by transferring the written texts of Orhon Turkish, Uyghur Turkish, and Karakhanid Turkish to electronic environment Another essential study, the Electronic Corpus of Pre-Islamic Turkish Texts (Vorislamische Alttürkische Texte: Elektronisches Corpus' VATEC'), was conducted between 1999 and 2003 under the direction of Marcel Erdal with the support of the German Research Foundation and included texts from the Uyghur Turkish period. And he was marking them lexically and syntactically. The density of German literature differs somewhat from that of Turkish. For example, research efforts based on audio cassette recordings from 1997 were transformed into a database in 2005, and contemporary methodological underpinnings in the relevant linguistics field were built in 2006. Organizing the collected data and researching today's corpus linguistics as lexical data was experienced in 2007, and experimental researches became the subject in 2008.

3. How does one decide if a word is suited to be used in a textbook?

Vocabulary helps to shape the internal structure of texts and performs a variety of roles connected to language use at the discourse level. The following paragraphs will show how analysis may emphasize the relevance of vocabulary in the formation and organization of discourse by giving instances of vocabulary elements that contribute to textual cohesiveness.

While computers are adept at evaluating enormous volumes of data, they are unable to explain why a particular trait is implemented in a particular manner. This means that the accuracy of the textbook corpus is heavily reliant on the textbook authors' analytical abilities. The accountability and role of the textbook creator is especially important in the case of selected corpora, because they are frequently assembled and evaluated by the same person.

Natural language processing technology allows us to gather statistical data about words, such as information on the use of word labels, which aids us in obtaining the information required to integrate it into a dictionary macrostructure. In terms of being measurable, how extensively a word is used or how common it is considered fundamental. When a word is frequently used, one common idea or strategy is to teach those words to the language learner first. This strategy is based on frequency and frequency data, which one can assess using the corpus methods of a word. This is valuable information for textbook authors throughout the theme and text editing process. From this perspective, the evaluation of word lists in language learning and teaching is based on pedagogical benefit expectations. Language teaching and learning were indeed complicated and dynamic processes, and there is no specific set of words that can help foreign language learners in this regard. Contexts, competence levels, learning aptitude, cultural backgrounds, and personal ambitions of Students' will most likely all have different needs.

As a result, any claim concerning pedagogical usefulness must be supported by actual evidence. For example, combining usefulness for learners with frequency information (Gardner and Davies 2014; Lei and Liu 2016) is a crucial strategy in this paradigm since "word frequency is a more trustworthy predictor of utility than pure intuition" (Gamier and Schmitt 2015). Frequency information is essential for today's modern language learning theories (Ellis 2014). Starting with the teaching plan stage, students are exposed to more frequent terms, and the importance of using them in oral and writing communication is emphasized (Gamier and Schmitt 2015). Studies on the idea that a set of words will be encountered repeatedly in

meaning-oriented interactions and the lexical scope of texts can be easily examined. In terms of educational applications of word definition, while defining vocabulary is definitely valuable for working with dictionaries, more concrete research in specific domains is needed to support the assertion that it is useful beyond this function. However, it is believed that more work is needed to link the word identification needs of learners in various contexts such as reading, writing, speaking, and so on with systematic, data-driven research. There is a need for a nonlinear n-gram construction or the concept of a vector space model. It is thought that the textbook designers should not expect it to be thoroughly examined. Computational linguistics is the study of language models. It is a conceptual framework for comparing all objects and then applying them to text processing tasks, rather than waiting for textbook authors to discuss concepts related to word-frequency (tf-idf) using computational linguistics and performing semantic analysis. It is hoped that tools like Lexitürk (www.lexiturk.com), which is designed as a lean corpus analysis interface that can support their own formal methods, will be helpful in allowing book designers to analyze language use in human and human behavior, formulate hypotheses, and evaluate linguistic data (corpora) at a later stage.

4. LexiTürk

LexiTürk (www.lexiturk.com) is a tool for interactive calculation and visualization of the frequency for any text in the context of Zipf Law (Zipf G.K.) and the frequencies of word types in texts or by loaded corpora. LexiTürk visualizes frequency information for a form of interaction for two texts or corpora, interactively showing the degree to which, a lexical item occurs more often than expected (Scott, 2006) Thus, the tool provides measures of lexical diversity and is theoretically helpful for textbook authors. The construction of wordlists is based on information about the frequency of words gathered from corpora. These are lists of words or phrases that are rated by the frequency or number of times they appear in a corpus. It can be used to explore the relationship between possible lexical overlap and species diversity. LexiTürk is free and open to everyone's use.

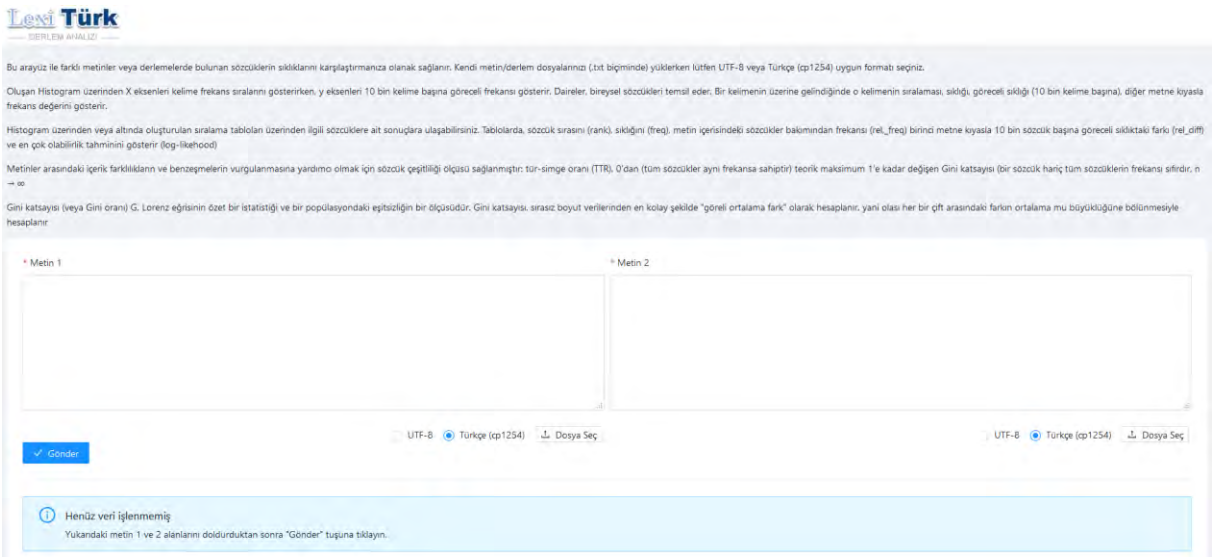

Figure 1. Default view of the interface

In the LexiTürk interface, which has a plain design, Turkish (cp1254) and UTF-8 character encodings are used as the industry standard. The Turkish character encoding formats are mainly used for Turkish in the fields expressed as Text 1 and Text 2.

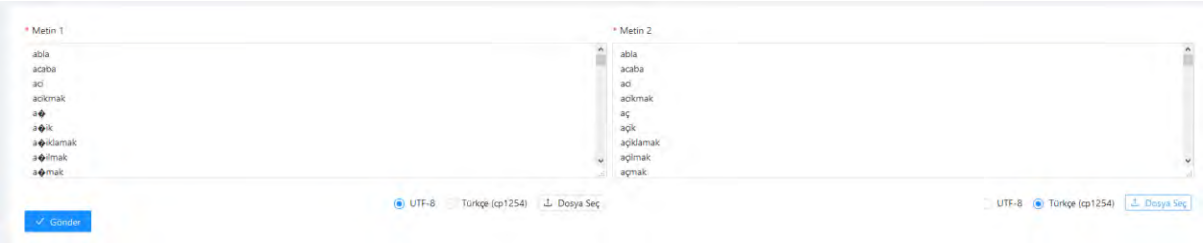

Figure 2. Character Encoding Preference view of the interface

Users can choose the encoding option that shows the letters correctly after uploading their text and collections in .txt format with the Choose File option and the "Choose File" option. After the files are uploaded to both Text fields, the analysis can be performed with the Submit button.

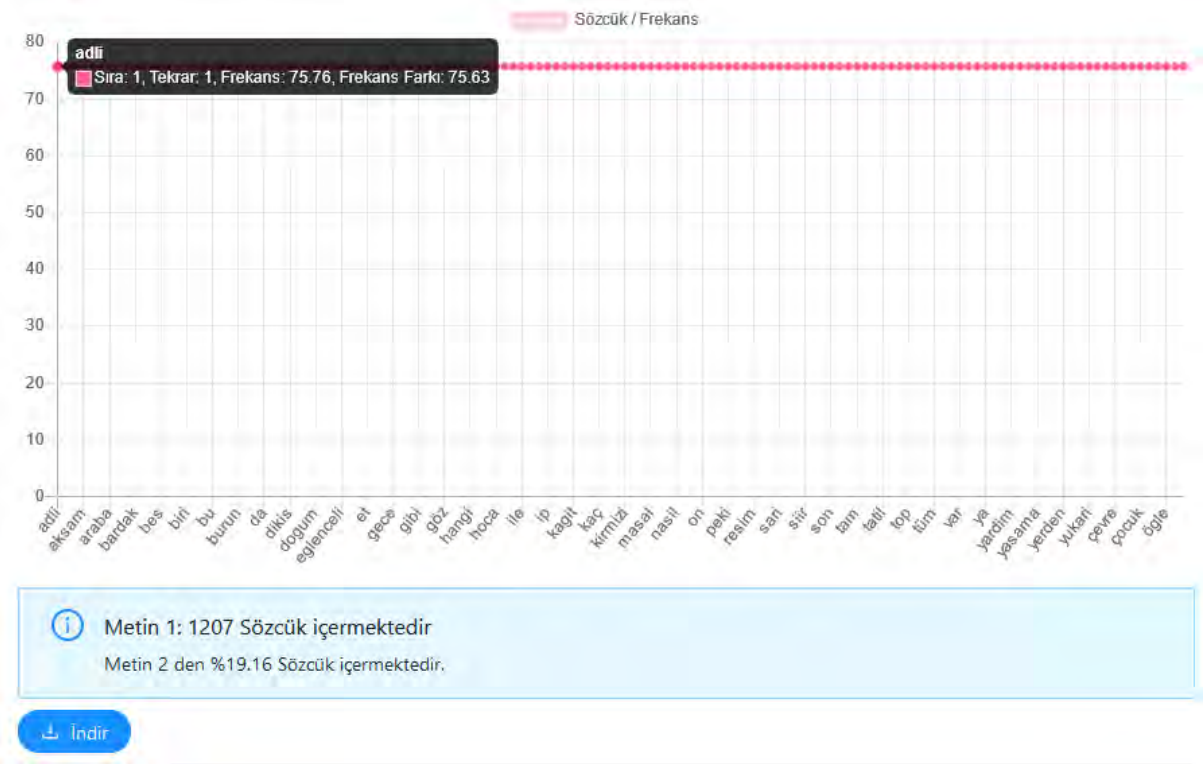

Figure 3. Frequency Graphic Distribution Display of the Interface

It shows the word order on the x-axis and the relative frequency (per 10,000 symbols) on the y-axis. Dots on graphs re-present individual word units and hovering over a word measures the word itself, its order, frequency, and relative frequency, as well as the log-likelihood measure [1,2,3] Sequence Number, Number of occurrences in Text and Corpus, Within Text

and Corpus. It is shown as four values expressed as frequency and Frequency Difference. The number of words from both texts (where the comparison is made) is located under the graphic. The distribution here can be downloaded as an excel file, or a list is created on the interface.

Sıralama	Sözcük	Tekrar	Frekans	Frekans Farkı
1	adli	1	75.76	75.63
2	agiz	1	75.76	75.63
3	aile	1	75.76	75.51
4	aksam	1	75.76	75.63
5	alt	1	75.76	75.63
6	altinci	1	75.76	75.63
7	araba	1	75.76	75.38
8	ayni	1	75.76	75.13
9	az	1	75.76	75.63
10	bardak	1	75.76	75.63

< 1 2 3 4 5 ••• 14 > 10 / sayfa

Figure 4. Frequency List Distribution Display of the Interface

Frequency analyses also enable comparisons between corpora and texts and between different words in a narrower context – for example, detecting a word's tendency to appear more frequently inside a text and prioritizing it as a highlighted or 'prototype' phrase in this regard. Frequency analyses can also be performed on textbook grammar, such as determining which tenses are more common or comparing the intensities of use of various word kinds. It will be possible to generate keyword lists by compiling word lists based on the frequency numbers of each word as a whole. It would also be able to do concordance-based analyses to explain why particular words are more frequent than others in the context of the researchers' "frequencies do not explain themselves" approach. Researchers will be able to do a fundamental analysis of texts and corpora containing Turkish words using the LexiTürk interface in the future, without needing to learn other software environments or programming languages for statistical computation and graphics, such as R.

5. Discussion and Conclusion

In this study corpus linguistics are established as an approach that uses corpora, or large, systematic, and computer-readable collections of language data. In addition, the corpus-based analysis examined as advantage. Finally, an online tool is created and presented that can be used by textbook authors or researchers for free via web-based interfaces.

Contemporary lexicography can be defined as a constant quest, through methodological improvements, to break the isolation and stereotype of words while studying vocabulary and analysis. Within this context, artificial lexical isolation of words measured in terms of text and speech is likewise desirable and required. Words are, in essence, the essential tools of language production, reflection, and policy in a speaking community. LexiTürk (www.lexiturk.com)

offers all scholars the ability to readily observe and compare the logical frequency formation between two different texts or corpora without requiring additional application or statistical calculation skills. Designers and scholars can use the book to interactively investigate word frequencies, revealing changes in content and dialogue. The word calculated by the interface based on the criteria can be understood in terms of text length and shared vocabulary and the variety of information content and use of function words. It is possible to determine how the words in a corpus built with the interface are included in a textbook using the interface. With LexiTürk, for example, it will be possible to identify the density of the most frequently used Turkish words in a coursebook prepared for Turkish as a Foreign Language and where they appear, which will be helpful for both researchers and coursebook designers. In the context of the unit/topic, there may be a discrepancy between prescriptive word interpretation and explanatory vocabulary in textbooks. In light of the analyses, the accurate word or language can be determined and corrected in a new edition of the design and the appropriate textbook. LexiTürk's ability to examine and compare word sequences and frequencies in various texts and collections can supplement lectures and studies explaining Zipf distributions and mathematical features. The method can also be used to provide evidence for discourse parallels or differences in literary, historical, or cultural studies while examining textual material.

References

Aksan, Yesim and Mustafa Aksan. (2009). Building a national corpus of Turkish: Design and implementation. Working Papers in corpus-based linguistics and language education No:3. 299-310 Tokyo: Tokyo University of Foreign Studies.

Abdurrahman Güzel, Özay Karadağ, Mehmet Kurudayıoğlu. (2005). İlköğretim 5. Sınıf Türkçe Ders Kitaplarının Kelime Hazinesi Bakımından Niceliği ve Ortak Kelime Hazinesi Kazandırması Bakımından Yeterlilikleri. XIV. Eğitim Bilimleri Kongresi-Kongre Kitabı. (Ed.: Hüseyin Kıran). C.2. Pamukkale Üniversitesi Eğitim Fakültesi. Anı Yayıncılık, Ankara, 28-30 Eylül 2005.

Ali İ. Tekcan , İlyas Göz. (2006). Türkçe Kelime Normları, Boğaziçi Üniversitesi Yayınevi,

Altan, Avkapan. (2006). Orta Öğretim 11. Sınıf Öğrencilerinin Kelime Hazinesi Üzerine Bir Araştırma, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü.(Unpublished Master Thesis).

Başaran, Bora (2008). Der Wortschatz, die Qual der Wahl für den türkischen Lehrbuchautor, Estudios Filologicos Alemanes, Volumen 15, Fenix Editora, ISSN 1578 -9438, Sevilla, 189-198.

Başaran, Bora. (2008). Die Problematik der Zuordnung der DaF Lehrwerke nach den Kompetenzstufen, 7 Congresso Brasilerio de Professores de Alemao.

Başaran, Bora. (2008). Almanca Ders Kitabı Tasarımında Yazılım Desteği: Sözcük Seçimi Boyutu, VIII. International Educational Technology Conference I.E.T.C. 2008, Eskişehir.

Baş, Bayram. (2005). Yılında Yayımlanmış Bazı Çocuk Dergilerindeki Söz Varlığı Üzerine Bir Değerlendirme. XIV. Eğitim Bilimleri Kongresi-Kongre Kitabı, Anı Yayıncılık, Ankara, C.2. 28-30.

Bilge, Say, Umut, Özge, Kemal, Oflazer, Bilgisayar Ortamında Bir Derlem Geliştirme Çalışması. Akademik Bilişim 2002, Selçuk Üniversitesi, 6-8 Şubat 2002.

Cheng, W. (2012). Exploring Corpus Linguistics: Language in action. London and New York: Routledge.

De Schryver, G.-M., Joffe, D., Joffe, P. & Hillewaert, S. (2010). Do dictionary users really look up frequent words?—on the overestimation of the value of corpus-based lexicography. Lexikos, 16.

Desagulier, Guillaume. (2017). Corpus Linguistics and Statistics with R; Introduction to Quantitative Methods in Linguistics, Springer

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. Computational Linguistics 19,: 61–74

Ekin, Şen. (2014). Türkçenin Ekleri ve Bu Eklerin Sözlü Dildeki Kullanım Sıklığı. E-Dil Dergisi, Sayı 2,: 86-90.

Ellis, N. C. (2014). 'Frequency-based accounts of second language acquisition' in S. Gass and A. Mackey (eds): The Routledge Handbook of Second Language Acquisition. Routledge, 193-210.

Gardner, D. and M. Davies. (2014). A new academic vocabulary list,' Applied Linguistics 35, 305-27.

Garnier, M. and N. Schmitt. (2015). 'The PHaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses,' Language Teaching Research 19: 645–66.

Hans Peter Vietze, Ludwing Zenker, Ingrid Warnke. (1975). Rückläufiges Wörterbuch der türkischen Sprache. Leipzig.

Hayriye Memoğlu-Süleymanoğlu. (2006). Türkçenin ters sıklık sözlüğü, kurmay yayınları, 5.

Hilal, Kutlu. (2006). MEB İlköğretim 6, 7 ve 8. Sınıf Türkçe Ders Kitaplarında Yer Alan Metinlerin Söz Varlığı Açısından Değerlendirilmesi, Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, (Unpublished Master Thesis).

Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge: Cambridge University Press.

İpek Eğilmez. (2010). İlköğretim Türkçe Ders Kitaplarındaki Söz Varlığının İlköğretim Dördüncü Sınıf Öğrencilerinin Yazılı Anlatımlarına Aktarımı, Uludağ Üniversitesi, Sosyal Bilimler Enstitüsü. (Unpublished Master Thesis).

Wagener, P. and Bausch, K.-H. (1997). Tonaufnahmen des gesprochenen Deutsch. Dokumentation der Bestände von sprachwissenschaftlichen Forschungsprojekten und Archiven. Tübingen: Niemeyer. (= Phonai Band, 40).

Wolfer, S., Koplenig, A., Meyer, P. & Müller-Spitzer, C. (2014). Dictionary Users do Look up Frequent and Socially Relevant Words. Two Log File Analyses. In Proceedings of the XVI Euralex International Congress. Bolzano/Bozen, Italy, 281–290.

Fiehler, Reinhard and Wagener, Peter (2005). Die Datenbank Gesprochenes Deutsch (DGD). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion 6, 136-147.

Köhler, Reinhard. (2016). Korpuslinguistik. Zu wissenschaftstheoretischen Grundlagen und methodologischen Perspektiven. In: LDV-Forum 20/2, Berlin, New York: de Gruyter, 1-16.

Kallmeyer, Werner and Zifonun, Gisela. (2007). Sprachkorpora – Datenmengen und Erkenntnisfortschritt. 2007.

Kemal Oflazer. (2006). Bilişim Tabanlı Dil Bilimi Üzerine Ufuk Turu. Bilgisayar Destekli Dil Bilimi Çalıstayı Bildirileri. (14 Mayıs 2005), Ankara: TDK Yayınları.

Kemal Oflazer, B. Say, D. Z. Hakkani-Tür, G. Tür. (2000). Building a Turkish Treebank. Abeille A (haz.) Building and Exploiting Syntactically Annotated Corpora.

Kemal Oflazer. (1997). Natural Language Processing Research in Turkey. Proceedings of 3rd Telri European Seminar,Montecatini, Italy, Oct 16-18, 1997.

Kemal Oflazer. (2014). Turkish and its challenges for language processing, Lang Resources & Evaluation 48, 639-653.

Kertész, András, Rákosi, Csilla. (2008). Daten und Evidenz in linguistischen Theorien: Ein Forschungsüberblick. In: Kertész, A. Rákosi, Cs. (eds.): New Approaches to Linguistic Evidence. Pilot Studies / Neue Ansätze zu linguistischer Evidenz. Pilotstudien. Frankfurt am Main u.a.: Lang, 21–60.

Lei, L. and D. Liu. (2016). 'A new medical aca- demic word list: A corpus-based study with enhanced methodology,' Journal of English for Academic Purposes 22: 42–53.

Mustafa Aksan, Ümit Mersinli. (2011). A Corpus Based Nooj Module for Turkish, Proceedings of the Nooj 2010 International Conference and Workshop. Komotini, 29-39.

Mustafa Çıplak. (2005). Uşak Merkez İlköğretim 5., 8. ve 11. Sınıfların Yazılı Kelime Hazinesinin Belirlenmesi, Afyon Kocatepe Üniversitesi, Sosyal Bilimler Enstitüsü. (Unpublished Master Thesis).

Pilav, Salim. (2008). Üniversite Birinci Sınıf Öğrencilerinin Söz Varlığı Üzerine Bir Araştırma, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü. (Unpublished Master Thesis).

Rayson, P., Garside, R. (2000). Comparing corpora using frequency profiling. In: WCC '00: Proceedings of the workshop on Comparing corpora - Volume 9, 1–6 doi.org/10.3115/1117729.1117730.

Scott, M., Tribble, C.: Textual patterns. John Benjamins, Amsterdam (2006).

Stefanowitsch, Anatol. (2020). Corpus linguistics: A guide to the methodology (Textbooks in Language Sciences 7). Berlin: Language Science Press.

Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford: Oxford University Press.

Storjohann, Petra. (2015). Sinnrelationale Wortschatzstrukturen: Synonymie und Antonymie im Sprachgebrauch. *Handbuch Wort und Wortschatz*. De Gruyter,: 248–273, doi.org/10.1515/9783110296013-011.

Stubbs, M. (2010). Three concepts of keywords. In: Bondi, M., Scott, M. (eds.), Keyness in texts, 21–42. John Benjamins, Amsterdam.

Oğuz, Cesur. (2005). (Kastamonu İlinde Bir İnceleme) Pansiyonlu İlköğretim Öğrencileri Üzerinde Kelime Serveti Araştırması, Abant İzzet Baysal Üniversitesi, Sosyal Bilimler Enstitüsü. (Unpublished Master Thesis).

Özay Karadağ, Mehmet Kurudayıoğlu. (2006). 2005 Türkçe Programına Göre Hazırlanmış

İlköğretim Birinci Kademe Türkçe Ders Kitaplarının Kelime Hazinesi. XV. Ulusal Eğitim Bilimleri Kongresi, Muğla Üniversitesi, Eğitim Fakültesi, 13-15 Eylül 2006.

Ufuk, Aşık. (2007). Yabancılar İçin Temel Türkçe Söz Varlığının Oluşturulması, Dokuz Eylül Üniversitesi, Eğitim Bilimleri Enstitüsü. (Yayımlanmamış Yüksek Lisans Tezi).

Üstün, B. & Tanrıkulu, L. (2021). Eine statistische Analyse des Lehrwerks „Netzwerk a1 Deutsch als Fremdsprache" in Bezug auf die vier grundlegenden Sprachfertigkeiten, Vokabeln und Grammatikaktivitäten . Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi , 11 (1) , 147-162 . DOI: 10.30783/nevsosbilen.869432.

İlyas Göz. (2003). Yazılı Türkçenin Kelime Sıklığı Sözlüğü, Türk Dil Kurumu Yayınları.

Yeşim Aksan, Mustafa Aksan. (2009). Building a national corpus of Turkish: Design and implementation, Working Papers in corpus-based linguistics and language education, Cilt3, 299-310

Yesim Aksan, Mustafa Aksan. Ahmet Koltuksuz, Taner Sezer, Ümit Mersinli, Umut Ufuk Demirhan, Hakan Yilmazer, Gülsüm Atasoy, Seda Öz, Ipek Yildiz, Özlem Kurtoglu (2012). Construction of the Turkish National Corpus (TNC), LREC Konferansı, 3223-3227