Frequency Analysis of the Top Ten Significant Words Derived with the Suffix -keit in the DeReKo: A Corpus-Based Study in the Context of German as a Foreign Language

Selma Akol Göktaş[1]

Trakya University

Abstract

The ability to know the rules of word formation in a language and to use these rules in the process of understanding and producing more complex words is part of language skills. In German, it is possible to derive words of different types and with various meanings with morphemes with suffixes such as -keit, -ung, -ig, -lich. In this study, mixed research method was used in which both quantitative and qualitative data collection methods were used together. The study aimed to answer the following questions: What are the top ten most frequent words derived with the morpheme -keit in the DeReKo corpus? What are the significant conclusions from the data obtained from the frequency analysis of the DeReKo corpus? What is the importance of frequency analysis in the context of German as a Foreign Language? What is the importance of corpus-based studies in the context of German as a Foreign Language? In this study, first of all, the frequency of words derived with the statistically significant morpheme -keit was investigated through frequency analysis in the DeReKo corpus. According to the results of the frequency analysis in DeReKo, the top ten words with the highest frequency are "Möglichkeit, Möglichkeiten, Öffentlichkeit, Tätigkeit, Schwierigkeiten, Aufmerksamkeit, Wirklichkeit, Unabhängigkeit, Fähigkeiten and Notwendigkeit". Finally, the importance of corpus-based studies and frequency analysis for GFL is emphasized.

Key Words

Corpus linguistics • DeReKo • Frequency analysis • -keit suffix

[1] **Correspondance to:** Trakya University, Faculty of Education, Edirne, Türkiye. E-mail: selmaakol@trakya.edu.tr **ORCID:** 0000-0002-1755-1322

Introduction

Our ever-changing world requires different names and concepts for new events and new situations. For example, with the Covid-19 pandemic affecting the whole world in 2019, all languages have gained new concepts. Vocabulary must be constantly supplemented to adapt to changing situations and meet the requirements of the language. For this reason, new vocabularies are created by combining the morphemes, the smallest meaning-bearing structural unit of a language, according to certain patterns (Horstmann, Settinieri & Freitag, 2020; Meer & Pick, 2019; Meibauer, Demske, Geilfuß-Wolfgang, Ramers, Rothweiler and Steinbach, 2007; Pelz, 1992; Schäfer, 2015; Spillmann, 2000). In the use of such words, the receiver hears part of a word with which he is already familiar and can usually interpret it correctly on first hearing. For German, the words "möglich (possible), öffentlich (public)" are examples of this situation. By adding the morpheme -keit to these adjectives, new words and new word types "Möglichkeit (possibility)" and "Öffentlichkeit (public)" can be formed In this way, language learners develop the ability to correctly understand unheard sentences and create their own new sentences characterizing the creative nature of the language. Morphs, on the other hand, are achieved by breaking words into their parts, through morphs new words are formed in a language, new words can be derived by adding to existing words (Meibauer et. al., 2007; Meer & Pick, 2019; Dipper, Klabunde & Mihatsch, 2018; Horstmann et. al., 2020). In the literature, word derivation and word formation are discussed with different concepts in German and in subfields of linguistics such as morphology, semantics, syntax and grammar: Nominalization (Nominalisierung or Substantivierung) (Bussmann, 2002), derivation (Derivation) (Fleischer & Barz, 2012; Pelz, 1990), word formation (Wortbildung) (Hentschel & Weydt, 2003; Hentschel, 2021; Karatas, 2005). At the same time, different morphs can function in different ways, and these morphs are then called allomorphs (Harnisch & Trost, 2009). The suffixes -gil and -ler are examples of this in Turkish: "annemgil/annemler geldi (my parents came)". What is meant in this sentence is that my mum and her entourage have arrived. The Turkish suffix -gil and -ler adds plural meaning to the word to which it is added in Turkish. Depending on the context of the conversation, it can be my mum and my dad or my mum and the people with her (my brother, my aunt, my uncle, my mum's friends etc.). In German, on the other hand, the suffixes -st and -est, which are added as superlative form (Superlativform) suffixes that rank adjectives, function as allomorphs: dunkel ⇒ die dunkl-e Schokolade ⇒ die dunkel-st-e Schokolade (Harnisch & Trost, 2009). In German, there are morphemes like Kopf (head) that have a single meaning and can stand alone, and these are also called lexemes (Spillmann, 2000; Meibauer et. al., 2007; Dipper et. al., 2018; Horstmann et. al., 2020). Those that cannot be used independently and have more grammatical functions are also called grammatical morphemes. In German, Genetiv -s, -es are examples of these morphemes.

When we look at all these concepts, we come across the concepts of word, morpheme, lexical unit and morph. Since a word structure needs a lexical unit core structure and since it is necessary to place this word structure in a lexical unit category, every word must contain at least one root (Stamm) (Schäfer, 2015). In addition, the concept of affix is also used for morphs other than non-root word structures. German has a wide variety of affixes for word formation. Affixes are categorized in three structures. These are prefixes (Präfix), suffixes (Suffix) and separable affixes (Zirkumfix). One of these suffixes is the suffix -keit, which helps to derive nouns from adjectives. The suffix -keit is a variant of the suffix -heit. It comes with the adjectives -ig, -bar, -sam, -lich and is a suffix that helps to

derive nouns from adjectives (Fleischer & Barz, 2012). The suffix -keit is a suffix used with -heit to nounize adjectives in German. Suffixes such as -keit (Fähigkeit (ability)), Abhängigkeit (dependence)), -heit (Schönheit (beauty)), -ung (Begrüßung (greeting), Forschung (research)), prefixes such as ver- (verlaufen (run, proceed)), auf- (aufbereiten (edit, prepare)), her- (herstellen (produce, create)) and prefixes such as ge-...-e (Gerede (talk, gossip)), ge-... -ig (gefügig (compliant)), be-...-ig (begradigen (straighten)), and Zirkumfixes are important within the scope of German as a Foreign Language (Meibauer et. al., 2007; Nacak, 2020; Pelz, 1992; Spillmann, 2000). As can be seen, there are many ways of word derivation. Word derivation has an important place in developing four basic language skills. These four basic language skills are divided into two as receptive and productive skills. In German as a Foreign Language classes, "receptive skills" and "productive skills" i.e. listening and reading comprehension as well as speaking and writing skills are acquired and need to be developed (Funk et al., 2014; Rösler, 2012). Vocabulary knowledge plays an important role in this process. Therefore, most frequent words derived with the suffix -keit were analyzed using corpus linguistic analysis.

Vocabulary knowledge and derivation, the use of different types of words, occurs both orally and in writing in a variety of text types, including everyday texts, e.g. in newspaper or news texts, Wikipedia articles, literary texts, but also in spoken expressions and musical texts (Kabatnik, 2020; Storrer, 2013). In the DeReKo corpus, the largest corpus of the German language, the frequency of words derived with the statistically significant -keit morpheme was investigated through frequency analysis. The top ten words with the highest frequency were identified by the frequency analysis method in DeReKo, respectively. Hence, the importance of corpus-based studies and frequency analysis for GFL[1] lessons is emphasized. In addition, it is thought that this study will set an example for corpus linguistic research on suffixes such as -keit, -ung, -ig, -lich, which are used to derive words of different types and meanings in German.

Purpose and Importance of the Study

The ability to know the rules of word formation in a language and to use these rules in the process of understanding and producing more complex words is part of language skills. In German, it is possible to derive words of different types and with various meanings with morphemes with suffixes such as -keit, -ung, -ig, -lich. The use of these words has an important position for students learning German as a Foreign Language in terms of using them in the text and understanding such words in the text. Word derivation is very important especially in language teaching in terms of analyzing and understanding structures, and in terms of text reception and production. For students who want to improve their vocabulary knowledge, knowing the structures of word derivation from different types of words makes it easier for them in language learning. In other words, knowledge of deriving new words from different types of words is necessary for students in order to improve their vocabulary knowledge. The aim of this study is to identify the top ten most frequent words derived with the -keit morpheme through corpus linguistic analysis and to underscore the significance of corpus-based studies and frequency analysis for the teaching of German as a Foreign Language using these sample words. The questions that were sought and investigated in the research are as follows:

---

[1] German as a Foreign Language

- What are the top ten most frequent words derived with the morpheme -keit in the DeReKo corpus?

- What are the significant results according to the frequency analysis of the DeReKo corpus?

- What is the importance of frequency analysis in the context of GFL?

- What is the importance of corpus-based studies in the context of GFL?

Method

Research Design

The research was conducted using corpus linguistic methodology, in which both quantitative and qualitative data were analysed (Hirschmann, 2019; Lemnitzer & Zinsmeister, 2015; Scherer, 2006). This method was preferred because it allows for an in-depth investigation of the topic to be researched (Lüdeling & Kytö, 2009; Schmidt, 2022). In this part of the study, information about the corpus linguistic analysis method, data collection tools and data analysis is given.
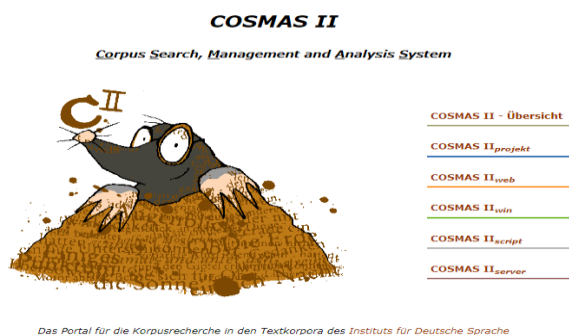
Data Collection Tools

In corpus linguistics, the interest is not in individual texts, but rather in the use of language in large groups of texts. This means that corpus linguistics is empirically based and works with quantitative methods (Bubenhofer, 2009; 2018). For this reason, corpus linguistic analysis method was preferred in the study. The DeReKo corpus was preferred as the data collection tool and its features are as follows:

The Corpus Query Processor COSMAS II and the DeReKo[2] Corpus: The DeReKo corpus for the German language, which has existed since 1964, is an electronic archive and corpus of the German written language, maintained and regularly developed by the Institut für Deutsche Sprache, Mannheim (Institute for German Language, Mannheim).

Figure 1

Corpus Query Processor COSMAS II

- DeReKo is the world's largest and comprehensive electronic corpus of German current language corpora for linguistic purposes (Kupietz, Belica, Keibel & Wittt, 2010; Kupietz, Lüngen, Kamocki & Witt, 2018).

---

[2] Deutsches Referenzkorpus

- The main corpus of German, compiled by the Leibniz Institute for the German Language in Mannheim, consists of written language,

- DeReKo is available to everyone and is the most comprehensive corpus of German,

- It includes many text types (fiction, scientific, popular science texts, newspapers, novels, etc.) and is constantly being developed,

- It is accessible free of charge via the COSMAS II[3] (Corpus Search, Management and Analysis System) platform.

Table 1

Methodology Followed in the Study

Determination of Research Questions and Corpus
DeReKo (W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuakquisitionen)
Quantitative and Qualitative Data Analysis
Frequency Analysis / Data Interpretation

Absolute Frequency	Relative Frequency	Frequency Class

The DeReKo reference corpus, which can be accessed through the COSMAS II portal, shows 5 different frequency degrees of words (absolute frequency, relative frequency, probability distribution, difference coefficient and frequency class) (Keibel, Kupietz & Belica, 2008; Keibel & Kupietz 2009). In this study, absolute, relative and frequency class analysis were used as frequency analysis.

Data Analysis

First of all, the archive "W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuakquisitionen)", which covers a total of 44,653,080 texts, was chosen for querying the DeReKo reference corpus, which consists of corpora created for many different purposes and from different text types and is collected in 18 different archives in COSMAS II. This archive was chosen because it is the largest of the 18 archives, it contains contemporary texts from the 18th century to the present day, and most importantly, it covers a broad range of text types. Thus, it provided the opportunity to see words formed with the -keit morpheme in different text types. The search query to find words formed with the -keit morpheme in corpora in this archive is "&-keit". In DeReKo, this sign, called the basic form operator or lemmatization operator, queries for new instances of word formation of a word-without-conjugation or word-forming morphemes entered in the "&" search query. With this query, words derived with the -keit morpheme in singular (Möglichkeit (possibility)) or plural (Möglichkeiten (possibilities)) form are found. After this query, the frequency analysis of the words derived with the -keit morpheme was performed and the absolute frequencies were obtained. In the frequency analysis, the words were ranked from the most frequently used word to the least frequently used word according to their absolute frequencies (query setting: nach Häufigkeit absteigend[4]). As a result of the frequency analysis, the 10 most frequently used words were found to be as follows: „Möglichkeit (possibility), Möglichkeiten (possibilities), Öffentlichkeit (the public), Tätigkeit (activity), Schwierigkeiten (difficulties),

---

[3] Leibniz-IDS (2023), https://cosmas2.ids-mannheim.de/cosmas2-web/
[4] Decreasing according to frequency

Aufmerksamkeit (attention), Wirklichkeit (reality), Unabhängigkeit (independence), Fähigkeiten (skills) and Notwendigkeit (necessity)"

Results

Statistically Most Frequent Words Derived with the Suffix -keit in German

A frequency analysis was performed on the DeReKo corpus to query the two statistically most frequent words derived with the suffix -keit in German. The results of the analysis are shown in Figure 2:

Figure 2

Results of Word Query with the Morpheme -keit in the DeReKo (W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuakquisitionen)) Corpus

Liste zu: &-keit : 138.339 Wortformen, nach Häufig. absteigend	
Möglichkeit	1.306.073
Möglichkeiten	859.657
Öffentlichkeit	820.457
Tätigkeit	394.918
Schwierigkeiten	380.830
Aufmerksamkeit	376.902
Wirklichkeit	357.907
Unabhängigkeit	226.058
Fähigkeiten	224.761
Notwendigkeit	203.083
Persönlichkeit	160.668
Wahrscheinlichkeit	154.656
Persönlichkeiten	154.609
Fähigkeit	141.026
Räumlichkeiten	128.515
Glaubwürdigkeit	119.903
Nachhaltigkeit	117.649
Tätigkeiten	116.272
Abhängigkeit	107.718
Sehenswürdigkeiten	103.968
und 138319 weitere	9.398.325

As can be seen in Figure 2, the analysis of the DeReKo corpus with the search query "&-keit" yielded a total of 138,339 word forms. The frequency analysis was ordered from the most frequent to the least frequent word in the corpus. According to this result, the most frequent word is "Möglichkeit" with an absolute frequency of 1,306,073 and the second most frequent word is "Möglichkeiten" with an absolute frequency of 859,657. The third most frequently used word in this query is 'Öffentlichkeit' with an absolute frequency of 820,457. These words are followed by other words „Tätigkeit (a.f.[5]=394.918), Schwierigkeiten (a.f.=380.830), Aufmerksamkeit (a.f.=376.

---

[5] Absolute frequency

902), Wirklichkeit (a.f.=357.907), Unabhängigkeit (a.f.=226.058), Fähigkeiten (a.f.=224.761), Notwendigkeit (a.f.=203.083)".

The Significant Results According to the Frequency Analysis of the DeReKo Corpus

Table 2 shows the absolute frequency, relative frequency and frequency class results of these two words according to the frequency analysis.

Table 2

Frequency Analysis Results in the DeReKo Corpus

	Möglichkeit (possibility)	Möglichkeiten (possibilities)	Öffentlichkeit (public)	Öffentlichkeiten* (publics)
Absolute Frequency	1.306.073	859.657	820.457	1.015
Relative Frequency (%)	0.020120%		0,007631%	
Frequency Class[6]	8	9	9	18

The most frequently used word is "Möglichkeit" with an absolute frequency of 1,306,073. According to these results, it is remarkable that the second most frequently used word is the plural of the first word. Another striking point is that the word "Öffentlichkeiten" is found 1015 times in this corpus. When the word "Öffentlichkeit" is analyzed, it is observed that "Öffentlichkeiten", which is the plural expression of this word in German language and which is actually incorrectly used in this way, is also present in the corpus. Since the meaning of the word "Öffentlichkeit" is already a plural expression (public, public opinion), the plural of this word is not used in German and can be interpreted as a spelling mistake or an expression error. It is noteworthy here that the word, which is considered incorrect according to German spelling rules, is found in the DeReKo corpus with an absolute frequency value of 1.015. Since the relative frequency value shows the ratio of the investigated word to the words found in the whole corpus, these two words were also analyzed in terms of relative frequency. According to the analysis of the relative frequencies of these two words, it is observed that "Möglichkeit/-en" is a word that is used significantly more than "Öffentlichkeit/-en" (0.007631%) with a ratio of 0.020120%. According to the frequency class analysis, "Möglichkeit" is in 8 frequency classes, "Möglichkeiten and Öffetlichkeit" are in 9 frequency classes, while "Öffentlichkeiten" is in 18 frequency classes. For "Möglichkeit", it can be interpreted that it is used quite frequently according to the frequency class of 8.

---

[6] In the Frequency Distribution Class (Häufigkeitsklasse) all words are classified according to their frequency, which is approximately close to each other. There are 30 frequency classes and they are numbered 0,1,2,3,4.... The important thing here is that the smaller the number, the higher the frequency. That is, words in frequency class 0 are the most frequent words, and in German this class includes certain articles (der, die, das, des, dem, den) including their inflections (Keibel, Kupietz & Belica, 2008; Keibel & Kupietz 2009).

Figure 3

Results of the word query[7] Möglichkeit/-en in the DeReKo corpus

Since the word "Möglichkeiten" is the plural form of the word "Möglichkeit", the absolute frequency value of the most frequent word derived with the -keit morpheme is 2,165,730. Figure 3 also shows that this word was found in 996 corpora in 1,737,420 texts between 1997 and 2022.

The Importance of Frequency Analysis and Corpus-Based Studies in the Context of GFL

The words used in texts assembled with the aid of a computer can be compared with the case of the word under investigation or with a specific list of words (e.g. the vocabulary found at the back of textbooks). These words can then be categorised according to the learning level of the students. It is also important that the corpus is large in scope. This facilitates the conceptualisation of linguistic terms. It is important to count the frequency of certain linguistic units and to evaluate this by the frequency of their occurrence in the text (Perkuhn, 2021). Students should try to find unknown words themselves from the context in the text rather than looking them up directly from the dictionary. Here they can benefit from corpora. Corpus linguistics methodology can also support by showing inter-word context, or even inter-word context. In this context, in addition to counting word frequencies, collocations are also important and have an important place (Clarke & Nation, 1980; Mukherjee, 2002). According to the didactic methodological principles developed by Baumert (2016), it makes sense to first start with vocabulary teaching

---

[7] Search Query: Möglichkeit*, 2 aktivierte von 796 Wortformen, nach Häufigkeit absteigend [Possibility*, 2 activated from 796 word forms, decreasing according to frequency]

according to frequency, then determine word groups and their frequencies, and systematically expand word combinations. Finally, it is important to follow the order from easy to difficult, from simple to complex, from known to unknown in vocabulary teaching.

It is important to know frequently used words when learning a foreign language, but in some cases it is also necessary to know and use specific words and expressions. Here, it is not enough to just know the word frequencies and create course materials according to this frequency group; instead, it is necessary to identify collocations, show word groups, create a list of word groups and give examples of their use in the text. Of course, for this, it is first necessary to know and determine the frequency of word usage. In this context, it can be said that word frequency ratings and analyzes have an important place in German as a foreign language.

Discussion, Conclusion & Suggestions

To answer the questions, words derived with the suffix "-keit" in the DeReKo corpus were analyzed. According to the results obtained from the frequency analysis in this corpus, frequent words derived with the suffix -keit were identified. According to the results of the frequency analysis in DeReKo, the first most frequent word is "Möglichkeit" and the second most frequent word is plural form of the same word "Möglichkeiten" (Figure 2 and Table 2). According to this analysis, the other words in the top ten are „Öffentlichkeit (The public), Tätigkeit (Activity), Schwierigkeiten (Difficulties), Aufmerksamkeit (Attention), Wirklichkeit (Reality), Unabhängigkeit (Independence), Fähigkeiten (Skills) and Notwendigkeit (Necessity)" respectively. Based on these results, a number of suggestions are made, emphasizing the importance of corpus-driven research, corpora and corpus linguistic analysis (e.g. frequency analysis) for German as a Foreign Language.

- Top ten most frequent words derived with the suffix -keit have been selected for further investigation of their use in context due to their frequency in German, as an example for further research on derivation.

- These results can be investigated in different corpora and a comparison can be made as to whether the results match or not. For example, in the Falko corpus, which is an error analysis annotated corpus, the most frequent word derived with the suffix -keit by learners of German can be investigated. And what kind of errors they make as a result of word derivation can be determined.

- In general, through corpus linguistic analysis, the frequency of constructions derived with the suffix -keit in German shows its importance for German learners. By automatically identifying statistically significant structures, word derivation forms can be adapted to the concept of German textbooks.

- Word derivation, which plays an important role in both receptive and productive skills, should not be limited to linguistics, morphology, semantics and syntax, but should also be addressed in detail in textbooks.

- After data collection and analysis of the selected sample words, corpus data can be used for preliminary evaluations and analysis with texts for the lesson.

- It is thought that it would be useful to add lexicology and corpus linguistics lessons to the current curriculum as a main lesson or elective lesson.

Ethic

I declare that the research was conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Author Contributions

All of the authors contributed equally in the article.

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

No scholarships or payments have been received from any institution for this article.

Note

This article is the extended version of the study presented as an oral presentation at the XVI. Türkischer Internationaler Germanistik Kongress (08-10 May 2023, Edirne/Turkey).

References

Adamzik, K. (2004). *Textlinguistik: Eine einführende Darstellung.* Tübingen, Germany: Max Niemeyer Verlag.

Baumert, A. (2016). *Leichte Sprache – Einfache Sprache. Literaturrecherche Interpretation Entwicklung.* Open Access. Hannover. https://serwiss.bib.hshannover.de/frontdoor/deliver/index/docId/697/file/ES.pdf. Accessed on 13.07.2023.

Bubenhofer, N. (2009). *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse.* Berlin, Germany: Walter de Gruyter Verlag.

Bubenhofer, N. (2018). Visualisierungen in der Korpuslinguistik. Diagrammatische Operationen zur Gegenstandkonstitution, -analyse und Ergebnispräsentation. In Kupietz, M. & Schmidt, T. (Hrsg.), *Korpuslinguistik.* (pp. 27-60). Berlin, Germany: Walter de Gruyter Verlag.

Bussmann, H. (2002). *Lexikon der Sprachwissenschaft*. Stuttgart, Germany: Kröner.

Clarke, D. F. & Nation, I. S. P. (1980). Guessing the meanings of words from context: strategy and techniques. *System* 8: 3, 211-220. https://doi.org/10.1016/0346-251X(80)90003-2

Das Deutsche Referenzkorpus – DeReKo. Ausbau und Pflege der Korpora geschriebener Gegenwartssprache. In: *Digitale Sprachwissenschaft.* Institut für Deutsche Sprache, Accessed on 13.07.2023.

Dipper, S., Klabunde, R. & Mihatsch, W. (Eds.). (2018). *Linguistik: Eine Einführung (nicht nur) für Germanisten, Romanisten und Anglisten.* Berlin, Germany: Springer-Verlag.

Fleischer, W. & Barz, I. (2012). *Wortbildung der Deutschen Gegenwartssprache*. 4. Aufl., völlig neu bearbeitet von Irmhild Barz unter Mitarbeit von Marianne Schröder. Berlin, Germany: Walter de Gruyter Verlag.

Funk, H., Kuhn, C., Skiba, D., Spaniel-Weise, D. & Wicke, R. E. (2014). *Aufgaben, Übungen, Interaktion.* Stuttgart, Germany: Klett Sprachen.

Harnisch, R. & Trost, I. (2009). Affix. In E. Hentschel & P.M. Vogel (Eds.), *Deutsche Morphologie.* (pp. 16-37). Berlin, Germany: Walter de Gruyter Verlag.

Hentschel, E. (2021). *Basiswissen deutsche Wortbildung.* Tübingen, Germany: Narr Francke Attempto Verlag.

Hentschel, E. & Weydt, H. (2003). *Handbuch Der Deutschen Grammatik*. Berlin, Germany: Walter De Gruyter Verlag.

Hirschmann, H. (2019). *Korpuslinguistik. Eine Einführung.* Stuttgart, Germany: Metzler.

Horstmann, S., Settinieri, J. & Freitag, D. (2020). *Einführung in die Linguistik DaF/DaZ.* Stuttgart, Germany: Verlag Ferdinand Schoningh.

Kabatnik, S. (2020). *Leistungen von Funktionsverbgefügen im Text. Eine korpusbasierte quantitativ-qualitative Untersuchung am Beispiel des Deutschen und Polnischen.* Tübingen, Germany: Narr.

Karatas, R. (2005). Morphologie: Die Lehre von den Bausteinen der Wörter. In J. Volmert (Eds.), *Grundkurs Sprachwissenschaft: Eine Einführung in die Sprachwissenschaft für Lehramtsstudiengänge.* (pp. 87-98). München, Germany: Wilhelm Fink Verlag.

Keibel, H., Kupietz, M. & Belica, C. (2008). Approaching grammar: Inferring operational constituents of language use from large corpora. In Š. František, F., Mirjam (Eds.). *Grammar & Corpora 2007: Selected contributions from the conference Grammar and Corpora*. (pp. 235-242). Prague: Academia.

Keibel, H. & Kupietz, M. (2009). Approaching grammar: Towards an empirical linguistic research programme. In: Minegishi, Makoto / Kawaguchi, Yuji (Eds.). *Working Papers in Corpus-based Linguistics and Language Education*. No. 3. (pp. 61-76). Tokyo: Tokyo University of Foreign Studies (TUFS).

Kupietz, M., Belica, C., Keibel, H. and Witt, A. (2010). The German Reference Corpus DeReKo: A Primordial Sample for Linguistic Research. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10).* (pp. 1848-1854).Valletta, Malta: European Language Resources Association (ELRA).

Kupietz, M., Lüngen, H., Kamocki, P. & Witt, A. (2018). The German Reference Corpus DeReKo: New Developments – New Opportunities. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, T. Tokunaga, Takenobu (Eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).* (pp. 4353-4360). Miyazaki: European Language Resources Association (ELRA).

Leibniz-Institut für Deutsche Sprache (2023). *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2023-I* (Release vom 31.03.2023). Mannheim: Leibniz-Institut für Deutsche Sprache. Retrieved from www.ids-mannheim.de/DeReKo

Lemnitzer, L. & Zinsmeister, H. (2015). *Korpuslinguistik. Eine Einführung* (3., überarbeitete und erweiterte Auflage). Tübingen, Germany: Narr.

Lüdeling, A. & Kytö, M. (Eds.). (2009). *Corpus Linguistics. An International Handbook.* Vol 1 und 2. (Reihe Handbücher zur Sprach- und Kommunikationswissenschaft) Berlin, Germnay: Mouton de Gruyter Verlag.

Meer, D. & Pick, I. (2019). *Einführung in die Angewandte Linguistik: Gespräche, Texte, Medienformate analysieren.* Stuttgart, Germany: J.B. Metzler Verlag.

Meibauer, J., Demske, U., Geilfuß-Wolfgang, J., Ramers, K.H., Rothweiler, M. & Steinbach, M. (2007). *Einführung in die germanistische Linguisitk*. 2. Aufl. Stuttgart, Germany: J.B. Metzler Verlag.

Mukherjee, J. (2002). *Korpuslinguistik und Englischunterricht. Eine Einführung.* Berlin u.a., Germany: Peter Lang.

Nacak, N. (2020). Almancadaki ikili Ad+Ad Birleşimlerinde önekimsiler. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi,* (Ö7), 468-476. https://doi.org/10.29000/rumelide.808743

Pelz, H. (1992). *Linguistik für Anfänger.* 10. Aufl. Hamburg, Germany: Hoffmann und Campe Verlag.

Perkuhn, R. (2021). Korpusfrequenzen und andere Metriken zur Strukturierung von DaF-Lehrmaterial. *Korpora Deutsch als Fremdsprache* 1(2), 116–136. https://doi.org/10.48694/tujournals-86

Rösler, D. (2012). *Deutsch als Fremdsprache: Eine Einführung*. Tübingen, Germany: J.B. Metzler.

Schäfer, R. (2015). *Einführung in die grammatische Beschreibung des Deutschen* (Textbooks in Language Sciences 2). Berlin, Germany: Language Science Press.

Scherer, C. (2006). *Korpuslinguistik*. (2., aktual. Aufl. erschienen 2014.). Heidelberg, Germany: Universitätsverlag Winter.

Schmidt, C. (2022). Korpuslinguistik und Mündlichkeit. Methodische und technische Herausforderungen bei der Erstellung eines fachspezifischen Korpus zur Verständigung über Literatur im Deutschunterrichtsdiskurs auf der Grundlage archivierter Transkripte Korpora Deutsch als Fremdsprache. *Korpora Deutsch als Fremdsprache*, 2(1), 81-94. https://doi.org/10.48694/kordaf-55

Spillmann, H.O. (2000). *Einführung in die germanistische Linguistik: Germanistische Fernstudieneinheit 5.* Berlin, Germany: Druckhaus Langenscheidt.

Storrer, A. (2013). Sprachstil und Sprachvariationin sozialen Netzwerken. In B. Frank-Job, A. Mehler & T. Sutter (Eds.), *Die Dynamiksozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen anBeispielen des WWW.* (pp. 331-366). Wiesbaden, Germany: VS Verlag.