

USE OF FREQUENCY DATA FOR LANGUAGE LEARNERS

Dr. Emrah ÖZCAN, Yıldız Technical University

eoacan@yildiz.edu.tr ORCID: 0000-0001-9340-5485

Recently data is becoming a key component in everything, and language studies are no exception. In order to gather language data, many linguists rely on corpus studies, which provides important language examples from real life usage. This data is also important for lexicographical studies, in other terms dictionary studies, for revealing the patterns of vocabulary in that language. Among these patterns is the frequency value of a word. Frequency is defined as “the value that shows how frequently a word is cited in the corpus” (Y. Aksan, Mersinli, & Yaldir, 2011, p. 397). For such purposes frequency dictionaries are compiled, e.g., "Frequency Dictionary of Turkish: Core Vocabulary for Learners" (Y. Aksan, Aksan, Mersinli, & Demirhan, 2017) and “A frequency dictionary of contemporary American English: word sketches, collocates, and thematic lists” (Davies & Gardner 2010). The former frequency dictionary is one of the outputs of Turkish National Corpus (TUD) (Y. Aksan et al., 2012) and the latter is based on COCA (Corpus of Contemporary English) (Davies, 2008). These dictionaries help language learners to find out which words are the most frequently used in that specific language based on corpus data. In addition to specific frequency dictionaries, it would be beneficial for language learners if general dictionaries should include frequency data. Considering the studies conducted in languages with a large number of speakers, such as English, it is seen that frequency values are commonly used in both general dictionaries and pedagogical dictionaries, albeit in different ways. General corpus data (Faaß, 2017), which has been used in dictionaries since the 1960s, manifests itself in the form of sharing frequency data with dictionary users, such as Longman and Collins learner dictionaries, beginning from the 1990s (Kilgarriff, 1997). Using of frequency values as an auxiliary element in the general dictionaries will provide positive benefits for language learners, especially when there is no learners’ dictionary in that specific language.

In this study, how the frequency values obtained from the corpus data are used will be discussed by showing examples from both Turkish and English. This data is invaluable for language learners from the fact that according to Zipf’s Law and Pareto Distribution, most frequently used words play an important role in understanding the language with a limited number of words. Therefore, language learners should focus on studying most frequently used words in the target language in order to gain momentum.

REFERENCES

- Aksan, D. (2004). Linguistics and Turkish Writings. multilingual
- Aksan, Y., Aksan, M., Seatless, A., Sezer, T., Mersinli, Ü., Demirhan, U. U., ... Yıldız, İ. (2012). Construction of the Turkish National Corpus (TNC). In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012). Eight International Conference on Language Resources and Evaluation (LREC 2012), presented paper, Istanbul, Turkey. Retrieved from <http://www.lrec-conf.org/proceedings/lrec2012/papers.html>.
- Aksan, Y., Aksan, M., Mersinli, Ü. and Demirhan, U. U. (2017). A frequency dictionary of Turkish: Core vocabulary for learners. Routledge frequency dictionaries. Milton Park, Abingdon, Oxon; New York: Routledge.
- Aksan, Y., Mersinli, U. and Yaldir, Y. (2011). Vocabulary Frequencies in Primary Education Turkish Textbooks Corpus and Turkish National Language Corpus Sample. In V. D. Günay, F.

- Özden, B. Çetin, and F. Yıldız (Ed.), *Studies on Teaching Turkish* (pp. 397-408). Izmir: DEDAM Publications.
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Davies, M. and Gardner, D. (2010). *A Frequency Dictionary of Contemporary American English: Word Sketches, Collocates, and Thematic Lists*. 1st ed. London ; New York: Routledge.
- Faaß, G. (2017). Lexicography and corpus linguistics. In *The Routledge Handbook of Lexicography*. Routledge. doi:10.4324/9781315104942.ch8
- Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2), 135-155.