

## A Cross-Linguistic Corpus-Assisted Discourse Study of COVID-19 Vaccine News

Ahmet Ergün

Boğaziçi University

This paper combines the quantitative methods of corpus linguistics and the qualitative methods of critical discourse analysis to examine the news on coronavirus vaccines in English and Turkish. I have compiled two corpora for my research. The first is called “COVax-EN”, which consists of 787 English news on COVID vaccines taken from *BBC News*. The second is called “COVax-TR”, constituted by 505 Turkish news on COVID vaccines taken from *BBC News Türkçe*.

The first part of the data analysis consists of corpus techniques, mainly keyword analysis. Keyword analysis demonstrates what terms and concepts stand out in each corpus in relation to their respective reference corpora. Table 1 and Table 2 respectively display the first 5 keywords of COVax-EN and COVax-TR, their concordance results, percentage, and keyness within the corpus, and then introduce the corresponding terms in the other corpus, their concordance results, percentage, and keyness within that corpus. The reference corpora for COVax-EN and COVax-TR are “English Web 2021 (enTenTen21)” and “Turkish Web 2020 (trTenTen20)”, respectively.

**Table 1.***Keywords of COVax-EN*

KWIC in COVax-EN	Concordance results	Percent of the whole corpus	Keyness	KWIC in COVax-TR	Concordance results	Percent of the whole corpus	Keyness
Omicron	1248	0.24%	1	Omicron	335	0.11%	2
Jab	1358	0.26%	2	Aşı	3491	1.2%	10
Booster	1589	0.31%	3	Hatırlatıcı	30	0.0099%	110
Covid	3110	0.6%	4	Covid	919	0.3%	11
JCVI	120	0.023%	5	JCVI	28	0.0093%	71

**Table 2.***Keywords for COVax-TR*

KWIC in COVax-TR	Concordance results	Percent of the whole corpus	Keyness	KWIC in COVax-EN	Concordance results	Percent of the whole corpus	Keyness
BioNTech	404	0.13%	1	BioNTech	51	0.0099%	30
Omicron	335	0.11%	2	Omicron	1248	0.24%	1
Sinovac	286	0.095%	3	Sinovac	5	0.00097%	458
AstraZeneca	342	0.11%	4	AstraZeneca	114	0.022%	20
Moderna	285	0.094%	5	Moderna	125	0.024%	17

The comparison of keywords showcases that the term Omicron stands out almost equally in both corpora in that it is the first keyword in COVax-EN and the second in COVax-TR. The rest of the research consists of the qualitative examination of concordance lines of the term Omicron. By analyzing with which lexical patterns the Omicron is used, I compare the variant's representation in English and Turkish news. I adopt a critical discourse analysis approach to evaluate the significance of these lexical patterns. The findings from the concordance analyses are analyzed and the corresponding features across corpora are presented in the paper. The emphasis on the origin point of Omicron, the inadequacy of two doses of vaccine against Omicron, transmissibility and severity of the variant, the wave and tsunami metaphors, and the specification that Omicron is a Covid variant are the common themes that are contrastively analyzed across the two corpora. This cross-linguistic research showcases how the topic of Omicron is handled in different language services of the same source channel, namely *BBC News*.

Keywords: Corpus Linguistics, Discourse Analysis, English, Turkish, COVID-19 Vaccines