

Identifying keywords and phrases in German COVID-19 Twitter discourse

Julia Schilling
Universität Hamburg

julia.schilling@uni-hamburg.de

Robert Fuchs
Universität Hamburg

robert.fuchs@uni-hamburg.de

The COVID-19 pandemic has upended life around the globe, leading to intense public debate and a flurry of lexical innovation across many languages. (Socio-)Linguists quickly started to document and analyze COVID-19 discourse (Baines et al. 2021; Saraff et al. 2021), but there is as yet no systematic analysis of the lexical items and discourse patterns that characterize German COVID-19 discourse. We address this research gap through a systematic comparative analysis of public discourse during the COVID-19 pandemic. We identify not just distinct keywords and phrases linked to the pandemic but also track their development over time.

The starting point of the analysis is a contrastive keyword analysis of the discourse of every month of 2019 with its equivalent in 2020 and 2021, comparing pandemic with pre-pandemic discourse, while filtering out seasonal effects (e.g. discussion of snow in January). Our data comprises over 10 million geotagged tweets from Germany from over 160,000 unique Twitter users, producing a corpus of approximately 250 million words. Rather than collecting tweets based on a pre-existing list of keywords, we use a data-driven approach to identify COVID-19 related n-grams ($1 \leq n \leq 4$). We then assign these keywords to semantic fields such as COVID-19 NAMES (e.g. *Coronalage*), PUBLIC HEALTH INSTRUCTIONS (e.g. *Ausgangsbeschränkungen*), and VACCINATION and examine their development over time using statistical measures such as median and skewness of the distribution.

This analysis yielded over 800 1-grams, 500 2-grams, 130 3-grams, and 80 4-grams related to the COVID-19 pandemic. Results indicate that the lexis of COVID-19 discourse in German tweets significantly varies not only over time, but also within semantic fields of discourse and across regions.

References: Saraff, S.; Singh, T.; Biswal, R. (2021): Coronavirus Disease 2019: Exploring Media Portrayals of Public Sentiment on Funerals Using Linguistic Dimensions. *Frontiers in Psychology* 12:626638. • Baines, A.; Ilttefaq, M.; Abwao, M. (2021): #Scamdemic, #Plandemic, or #Scaredemic: What Parler Social Media Platform Tells Us About COVID-19 Vaccine. *Vaccines* 9(421), 1–16.