

WHAT ARE THE MAIN TOPICS RELATED TO #MATHEMATICS? COMPARATIVE
INSIGHT FROM TWITTER DATA OF THREE EX-YU COUNTRIES

Isidora Gatarić*, Ilija Milovanović & Ivana Isailović

Department of Psychology, Faculty of Philosophy, University of Novi Sad

Previous studies that dealt with the analysis of tweets from the Twitter platform suggested that sentence context, in which the key words can be obtained, provide significant information about the sentiment, opinion and/or attitudes that people have on specific topic. The opinion about mathematics, as one of the key subjects that accompanies us throughout schooling, is no exception. The aim of this research was to examine the context in which tweets about mathematics appeared in Serbian, Croatian and Slovenian Twitter data. These three languages, as part of the South Slavic linguistic group, share historical, cultural, and linguistic connections but exhibit significant differences in vocabulary, grammar, and usage, making them ideal for comparative analysis in natural language processing. The data were taken from three corpora of Tweets (Tweet-sr, Tweet-hr and Tweet-sl), of the three mentioned languages from the CLARIN platform (<https://www.clarin.si/ske/#open>). Total set of tweets where the word #mathematics appears were selected from three corpus, which consisted of 21 500 Serbian tweets 935 Croatian tweets and 256 Slovenian tweets. Before applying the topic modeling technique Latent Dirichlet allocation (LDA) in the Python programming language, the raw tweets were preprocessed and cleaned of punctuation marks, specific symbols, conjunctions, adverbs, prepositions, profanity etc. After that, we started to detect the number of different topics (contexts) in which the word #mathematics appears in tweets and compare the results between the three languages. The results suggest that in all three languages it is most optimal to extract three different dimensions (thematic units) within which #mathematics is mentioned. In the Serbian language, #mathematics is most often mentioned: i) as a school obligation, ii) in context of sentiment (e.g. love, hate etc.) and iii) as parenting topics related to comments about professors, books, school and children. Almost identical results were obtained in the Croatian and Slovenian languages, with the difference that in these two languages all three topics were focused more on sentiment and school obligations, while parental attitudes toward #mathematics do not appear as often as in the Serbian language.

Keywords: mathematics,; Twitter; ex-YU languages; natural language processing; topic modeling

* corresponding author e-mail: isidora.gataric@uns.ac.rs