

## **Methods of analysing opinionated texts**

This thematic volume is devoted to exploring various methods of analysing opinionated texts. It encompasses methodological proposals of how to examine opinions, whether expressed explicitly or implicitly, ingrained in qualitative and quantitative traditions as well as those embedded in computational approaches to discourse studies. The computational approaches to the textual analysis presented in this volume rest not only on now widely used corpus studies as they extend this trend in discourse analysis to more advanced, AI-based tools that involve sentiment analysis, topic modelling and embeddings. The topics represented by the data deployed in the articles revolve around evaluative language and the exploration of polarity, which comprise: select words or phrases explicitly expressing positive or negative values (Aneta Dłutek; Danuta Stanulewicz and Konrad Radomyski), implicitly encoding negative evaluation through irony (Anna Bączkowska, Katarzyna Kukowicz-Żarska, Aneta Dłutek and Iwona Benenowska) or idiomatic expressions (Joanna Redzimska), as well as social minorities, such as immigrants and/or refugees (Anna Bączkowska et al.; Stanisław Os and Tomasz Wąs) and misogynic discourse (Dagmara Solska and Jan Sawicki).

Four papers in this volume involve qualitative analysis. A paper by Monika Tosik (Kopytowska) focuses on framing television news coverage and triggering affective proximity in the audience. Aneta Dłutek's investigation, in turn, is based on a survey. A study examining various aspects of ironic language is proposed by Anna Bączkowska, Katarzyna Kukowicz-Żarska, Aneta Dłutek and Iwona Benenowska. Finally, Stanisław Os and

Tomasz Waś apply a typology stemming from Critical Discourse Analysis to verify opinions about immigrants. Quantitative methods of discourse analysis are present in five contributions. Corpus-related tools were applied in four studies proposed by Monika Tosik, Danuta Stanulewicz and Konrad Radomyski, Joanna Redzimska, and Anna Bączkowska, Katarzyna Kukowicz-Żarska, Aneta Dłutek and Iwona Benenowska. The same authors also resorted to machine learning approaches to sentiment analysis. Word embeddings, an AI-based analysis of how the meanings of words are related, are present in two papers by Anna Bączkowska et al., and Dagmara Solska and Jan Sawicki. Finally, topic modelling was explored by two papers (Anna Bączkowska, Katarzyna Kukowicz-Żarska, Aneta Dłutek and Iwona Benenowska; Dagmara Solska and Jan Sawicki).

The paper written by **Monika Tosik** (Kopytowska) opens the volume with a mixed approach (qualitative-quantitative) to opinionated texts that rests on the well-ingrained media studies theory of framing and the (cognitive) distance-reduction Media Proximization Approach, which she applies to television discourse, specifically to the news coverage on terrorism in East Africa. The Author is interested in how the news frames and constructs narratives to shape the public perception of Western audiences and to make the news cognitively and affectively closer to the audience. The analysis of frame-building in news coverage starts with the examination of keywords (with tools used in corpus linguistics) in texts presented by U.S. television on Garissa University College attack in Kenya in 2015, which is followed by the identification and description of several frames. The Author concludes that keywords and their modifiers act as proximation triggers that contribute to the creation of emotionally engaging narratives.

Another qualitative study, by **Aneta Dłutek**, investigates judgmental adjectives *glupi* (“stupid”) and *mądry* (“wise”), and it rests on a survey administered among 58 respondents. The Author verifies opinions expressed by secondary school and university students. She was interested in how the respondents

used evaluative meanings encoded by these adjectives in social interaction, towards whom or in what circumstances. The research shows that the respondents prefer the negative adjective in self-opinions, i.e., when talking to themselves about their mental or intellectual condition, and when expressing negative evaluation for momentary situations. Interestingly, neither of the adjectives mentioned above is employed with reference to strangers. Moreover, the Author notices that in Poland, it is a common practice to evaluate one's own knowledge or mental abilities in a pejorative way rather than boast about them.

The contribution by **Danuta Stanulewicz** and **Konrad Radomyski** utilises corpus linguistics tools and machine learning available in MS Azure Cloud Services in their exploration of sentiment analysis of opinions regarding the ELSA language learning application. They focus on the frequency of positively and negatively charged adjectives occurring in opinions retrieved from the Google Play Store. In their corpus of over 25 thousand comments, they notice that 87.62% of the authors of the comments express their positive opinions, giving ELSA five or four stars, and 86.86% of comments express a positive valuation. The mean sentiment score calculated for the whole corpus adds up to 0.72, confirming ELSA's overall positive evaluation. Interestingly, the Authors notice that the most frequent adjectives tend to be general evaluative words.

**Anna Bączkowska, Katarzyna Kukowicz-Żarska, Aneta Dłutek,** and **Iwona Benenowska** apply various methods, both qualitative and quantitative, in order to uncover implicitly encoded opinions through ironic comments retrieved from the Polish subReddit. Drawing on existing neo-Gricean typologies of irony, they propose their modified categories adapted for social media's syntactically- and lexically-simplified discourse. In their data, they identify predominating topics of ironic comments, types and functions of irony as well as types of reactions to ironic comments. A deeper scrutiny of topics, particularly the hidden ones (the so-called "third space"), was conducted using the topic modelling Mallet tool, while the expressed emotions

were verified with machine-based sentiment analysis. Some corpus-based tools were employed to enrich the study (relative frequency ALDF, pronoun-noun ratio, and collocations). They concluded that lexical irony was predominant in their data, malice and jocular malice were the most frequent functions of ironic comments that usually addressed the third parties. They also noticed that the overall sentiment was negative, and some hidden topics were revealed (related to Ukrainians and the stereotypical role of women as child-rearing) with a machine-learning topic modelling tool.

**Joanna Redzimska** concentrates on colours (particularly *green*) as opinion markers. She based her investigation on a selection of metaphorical phrases retrieved from three language corpora representing British and American English. The Author assumes that opinionated texts, at least partially, convey opinions through evaluative language; this is also the claim expressed by Bączkowska et al. in this volume. Given this assumption, the analysis is ingrained in theoretical concepts of cognitive axiology, which she elaborates on in the first part of her paper. The Author conducts machine learning-based sentiment analysis (Free Sentiment Analyser) to check the axiological charge of the phrases under scrutiny. She concludes that colour terminology encoding metaphorical senses can convey opinions.

**Dagmara Solska** and **Jan Sawicki** deploy a Twitter-roBERTa-base Sentiment Analysis model to verify the polarity of gender-biased discourse retrieved from the MensRights subreddit. The dataset was extracted with the help of Python Reddit API Wrapper. Using the spaCy library for Natural Language Processing enabled the Authors to extract the top 300 n-grams to uncover the frequency distribution of two (bigrams), three (trigrams) and four consecutive words (quadrigrams). The recurring phrases of various lengths reveal a deep-seated animosity between genders and even gender conflicts discernible in anti-feminist, discrediting discourse.

The next article by **Bączkowska et al.** reports on research in progress that aims to analyse opinions seen in evaluative lan-

guage. It is a theoretical-analytical paper proposing a typology which categorises and describes words and phrases that carry values. Evaluative language is characterised by the Authors relying on philosophical and linguistic perspectives. The Authors offer a tripartite typology based on three research avenues stemming from media studies, linguistics and Critical Discourse Analysis, which they modify and combine to achieve a unified typology. They attest their proposal on a small dataset retrieved from an existing corpus of Facebook comments (the FRENK corpus) to conclude that the typology offered by this pilot study, while on the whole capable of identifying and labelling detailed information regarding polarity, needs further refinement, which the Authors plan to pursue in their future research.

In the last contribution, **Stanisław Os** and **Tomasz Waś** resort to Critical Discourse Analysis in their study devoted to opinion-mining based on newspaper articles. They explore 2066 user comments related to immigration policy in Germany. The focus of the study is on social actors the Authors identify in line with van Leeuwen's typology. The comments unravel how immigrants are conceived of by the host country's citizens. The Authors conclude that migrants are often described in the passive voice and in terms of objects, while the members of the German right-wing party AfD are presented as victims of the migration open policy and particularly of the "democratic opposition".

Overall, the volume reports on studies offering a combination of methods applicable to the analysis of large datasets of opinionated texts, resorting to mixed approaches relying on both corpus linguistics methods and cutting-edge computational linguistics tools as well as those typically deployed for smaller datasets that often provide deep insights. As this thematic volume shows, opinionated texts can be explored in a number of ways by resorting to a wide array of methods involving traditional and qualitative approaches along with quantitative and less known computational, algorithm-based techniques or a mixture of these. It is hoped that all the methodologies offered by the papers presented here can inspire and guide

further research in the field of discourse analysis and suggest future avenues in computational discourse studies.

*Anna Bączkowska*