APPLICATIONS OF COMPUTATIONAL LINGUISTICS AND LANGUAGE CORPORA IN OPINION STUDIES: EDITORIAL TO RESEARCH IN LANGUAGE 2025 SPECIAL ISSUE SECTION

EDITED BY

BARBARA LEWADOWSKA-TOMASZCZYK
University of Applied Studies in Konin

MARCIN TROJSZCZAK
University of Applied Studies in Konin

Table of Contents of the special issue:

Introduction to Applications of Computational Linguistics and Language Corpora in Opinion Studies

This volume presents a collection of six original papers exploring diverse applications of computational linguistics (CL) and language corpora, including Large Language Models (LLM) tools, in the field of Opinion Studies. Opinion Studies, Opinion studies collect data on public attitudes, beliefs, and preferences and among their methodology they broadly encompass computational linguistics and language corpus methodology, focusing on sentiment detection and stance mining. They have become increasingly vital for understanding human communication and societal trends. The papers included herein, first delivered at the *Contacts & Contrasts (C&C) 2025* bi-annual conference on 24-26 March, 2025 organized by the Department of Language and Communication at the University of Applied Sciences in Konin, partly instantiate research conducted within the COST Action OPINIONS CA21129, and demonstrate the power of CL and corpus (i.a., LLM) methods and the rich insights derived from language corpora across various languages and domains, pushing the boundaries of current research in these fields.

Overview of Papers

The contributions in this volume span several critical areas, with methodologies heavily reliant on computational techniques and specialized corpora:

The paper *Hate Speech and Ideological Text Detection in Kazakh* by Milana Bolatbek, Shynar Mussiraliyeva, and Moldir Sagynay- from al-Farabi Kazakh National University, in Almaty, Kazakhstan, addresses the pressing social issue of hate speech within the Kazakh language. The core methodology leverages computational techniques, specifically Machine Learning to analyze Kazakh language data. This significant contribution supports research in text classification and social media safety in under-resourced languages.

The paper *Testing Chat GPT on terminology generation, defining, translation, and ontology creation in German, English and Polish* by Barbara Lewandowska-Tomaszczyk and Grzegorz Pawłowski focuses on the challenging task of ontology generation—creating a formal representation of knowledge in a domain of eletrotechnology —across three distinct languages: Polish, English, and German, via the intermediate steps. The research methodology involved testing Chat

GPT_pro and ChatGPT-4 (a Large Language Model tool) on key tasks including terminology generation, defining, translation, and eventually, the creation of the ontology itself. An Assistant called SLA (Special Language Assistant) was created and described in the paper for the purpose of analysing prompts, recognising context and intent, and processing data using language models. This is highly relevant not only for enhancing knowledge graphs and structuring semantic data in opinion-rich domains but can also serve as a handy tool in getting familiarity with restricted domains such as electrotechnology by novice lexicographers of ontoterminological interest as well as by translators and interpreters. The paper introduces a special assistant tool to perform these tasks to make them run smoothly.

The third paper in this Special Issue Papers *Classification of Ideological Texts in the Kazakh Language Using Machine Learning and Transformers*, written by the authors of the first paper: Shynar Mussiraliyeva, Milana Bolatbek, Kymbat Baisylbayeva of al-Farabi Kazakh National University in Almaty, Kazakhstan, demonstrates that hybrid neural network models as the most effective solution in ideological texts classification, with DistilBERT performing most effectively among transformer models. The finding can help developing automatic monitoring and filtering systems in language texts to efficiently identify ideologically charged content.

The contribution *Linguistically Packaging Opinion: A Quantitative Token-Based Analysis of Adjective Structure in English* by the Spanish author Antoni Brosa-Rodriguez of the Universitat Rovira i Virgili in Tarragona, delves into the nuanced structure and function of adjectives in English examining the distribution and characteristics of three linguistic structures across three different semantic types of adjectives in English: opinion-based (evaluative), mixed (dimensional), and non-opinion (objective) adjectives The methodology employed is a Quantitative Token-Based Analysis and the author also observes register variations, with web language showing distinctive distributional patterns for opinion adjectives compared to more varied texts. It should be emphasized that understanding how evaluative adjectives are used in language data is fundamental to sentiment analysis and automatically extracting subjective information from texts.

Katarzyna Fronczak of the University of Lodz (Poland), the author of the next contribution, also focuses on adjectival structures as the carrier of stance judgments and investigates this issue in a different types of discourse. In her paper *"Behind Every Successful Woman Is a Tribe of Other Successful Women" – A Preliminary Corpus-Assisted Study of Evaluative Adjectives in Women Entrepreneurs' Blogs.* the author investigates the deployment of evaluative adjectives within a specific, real-world domain: women entrepreneurs' blogs. The methodology is a preliminary Corpus-Assisted Study, providing valuable insights into how female communication shapes and expresses opinion.

The final paper in the present Special Issue *Over-the-phone Community Interpreting — necessary evil?* deals with telephone interpreting. The author Przemysław Boczarski of Lodz University of Technology (Poland), researches over-the-phone (OTP) community interpreting and, basing on large corpus data of

his own, collected over his job as an OTP interpreter, asks the questions concerning the utility of this community service by exploring the linguistic and technical aspects of telephone interpreting. The topic's relevance to Opinion Studies is tied to communication integrity, the automatic detection of tone and stance in interpreted speech, and, eventually, the creation of specialized corpora for cross-cultural communication analysis.

Significance

Collectively, the papers collected in the present Special Issues of the journal *Research in Language*, underscore the multidisciplinary nature of Opinion Studies. They highlight the necessity of robust language corpora and sophisticated computational models to tackle complex identification and applicational tasks, ranging from ethical concerns like hate speech to foundational linguistic structure and professional communication This volume can thus serve as a resource for researchers and practitioners interested in the intersection of natural language processing (NLP), corpus linguistics, and the automated understanding of human opinion and subjective language.

Barbara Lewandowska-Tomaszczyk and Marcin Trojszczak
Special Issue editors
November 2025