

Ethical Aspects of Analyzing Kazakh Political Discourse

Gulmira Bekmanova[1,2(✉)], Banu Yergesh[1,2], Assel Omarbekova[1,2], Yerkyn Ongarbayev[2], and Altanbek Zulkhazhav[1,2]

[1] Faculty of Information Technologies, L.N. Gumilyov Eurasian National University, 010000 Astana, Kazakhstan
gulmira-r@yandex.kz

[2] L.N. Gumilyov Eurasian National University, 010000 Astana, Kazakhstan

Abstract. This article delves into the analysis of hotbeds points in political discourse, examining the diverse forms they can take. The primary focus revolves around the officiality and openness of these discourse sources. When addressing open sources, particularly instances where candidates in various elections openly and officially express their viewpoints, the article asserts that there is no violation of legal or ethical norms in the scientific analysis of texts, videos, or audio information for research purposes. The article emphasizes transparency, ethical considerations, and adherence to legal regulations in the context of studying and analyzing political discourse, particularly within the dynamic landscape of social networks. The guiding principles outlined here aim to ensure an ethical and responsible approach to the analysis of Kazakhstani political discourse in social media platforms. (funding by grant № AP19679847).

Keywords: Political Discourse · Database · Kazakh Language · Ontology · Ethical aspects of analyzing Kazakh political discourse

1 Introduction

The premise for this research is grounded in works related to the identification of terrorist threats on social media [1], sentiment analysis of texts [2–6], and the development of morphological analyzers [7, 8]. Researchers from various countries actively engage in this topic, recognizing that citizens often obtain information about state policies from mass media. In the era of widespread internet usage, where users not only communicate but also express opinions on social platforms, the discourse of politicians through social media has become a primary means for individuals to articulate their civic and political positions.

Efforts have been made to analyze the content flow concerning political discourse. For instance, a study [9] investigates the pragmatic and social functions of specific forms of ironic quotes used by social media users to mock political voices. Methods are applied for quantitatively assessing media bias regarding preferred stances on economic policies and consistency with the ideological positions of different political parties, influencing public understanding of politics [10].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14819, pp. 138–145, 2024.
https://doi.org/10.1007/978-3-031-65282-0_9

Publicly available data has been analyzed to critically evaluate how platforms targeting voters through online advertising either amplify or moderate the spread of political advertisements [11]. Another study [12] examines how language used by members of the UK Parliament changes after social media posts and reactions on the internet.

Sentiment analysis methods at the aspect level (ALSA - Aspect-level sentiment analysis) are employed to determine subjective sentiments of entities within texts [13]. Additionally, deep semantic and contextual knowledge networks (DSCNet) are explored in research [14], introducing a novel model with deep context.

An analysis of fake content, based on a frequency model of abstract terms in a reverse context using a diffuse neural network (Surf Scale), is proposed in a study [15], focusing on social content from public forums.

In Kazakhstan, sentiment analysis of official and unofficial information sources based on mood analysis of texts has been studied by researchers from L.N. Gumilyov Eurasian National University [16, 17], Al-Farabi Kazakh National University [18], and the International University of Information Technologies [19].

The literature review and analysis of journal articles from the first and second quartiles over the last three years indicate a significant interest in extracting meaning from texts and understanding conversation contexts. While some works on artificial intelligence related to political discourse exist, available applications do not yet fulfill functions such as emotion analysis or conversation context extraction. In Kazakhstan, few studies focus on sentiment analysis and cyberbullying, but comprehensive research on detecting political discourse, extracting semantics, and sentiments in social networks for the Kazakh language has not been conducted.

The scientific results obtained are crucial for addressing the scientific and technological needs of Kazakhstan in the context of ubiquitous digitization. As active activities transition from real life to the digital realm, which is prominently represented in social networks, monitoring and analyzing activities in social networks have become essential. Leading global social networks, like Twitter and Facebook, and individual states, such as the USA, the UK, and Russia, actively develop such tools.

The fundamental difference of the project idea from existing global analogs lies in the fact that for each natural language, such systems are created almost from scratch due to the grammatical, syntactic, and semantic peculiarities of each language. To conduct any semantic natural language processing, it is necessary to create software tools for its implementation, such as morphological analyzers, syntactic analyzers, sentiment analyzers, and various synonymizers within subject domains. For the effective operation of such tools, formalized morphological, syntactic, and sentiment models of natural language, as well as extensive knowledge bases within subject domains, such as political discourse, are required.

2 Related Works

In the article [20] investigates a network of social media accounts orchestrated by Russian operatives during the 2016 American presidential elections. These accounts aimed to manipulate political discourse by masquerading as legitimate social actors. The authors propose a novel method for tracking the “social footprint” left by these actors, utilizing

artificial intelligence to achieve an impressive 88% prediction accuracy. In work [21] authors delve into motivational speech acts within Joe Biden's political speeches. By analyzing language use and persuasion techniques, the study sheds light on communication strategies employed by the U.S. President. Article [22] explores the intersection of lobbying and environmental crimes, focusing on the Brazilian mining sector. The authors examine how corporate political activities impact environmental offenses. In the paper article [23] investigates the use of scientific discourses and markers in managing shark attack risks on Reunion Island. By analyzing language patterns, the study provides insights into risk communication strategies. Article [24] dissects how routine government releases and clarifications unexpectedly impact public opinion. Using a Chinese Ministry of Commerce release encouraging the storage of necessities as a case study, the authors explore shifts from panic to banter in public perception. Manuscript [25] explores feminist political economy and feminist critical discourse analysis as methodologies in critical nursing research. The author examines how these approaches can enhance understanding within the nursing field. This article [26] investigates how Nigerian female political leaders (NFPLs) use self-presentation strategies on social media to present themselves as powerful voices for the voiceless and active alternative leaders. The study analyzes discursive strategies such as self-promotion, ingratiation, exemplification, intimidation, and supplication1. "In Other News: China's International Media Strategy on Xinjiang—CGTN and New China TV on YouTube" [27] study delves into China's foreign-facing propaganda through the YouTube channels of CGTN and New China TV, focusing on Xinjiang2. "The multinational as a myth-prince of the global south: Writing back an emancipating imaginary to the global north" by Manuel Hensmans in Organization. In research [28] theorizes multinationals' role in writing back an emancipating imaginary from the Global South to the Global North, using the case of Huawei in China and the European Union3. The article [29] explores the negotiation of boundaries between state and society in early Republican Turkish cities.

3 Research Methods

Achieving the research goal is carried out through computer processing of the Kazakh language, which includes:

– construction of formal models describing grammar and sentiment, as well as the creation of knowledge bases (concepts and their relationships) on election topics within given subject areas.

Expected results:

1. A text corpus of political discourse of at least 300,000 words.
2. Ontological models for election topics: "Election advertising", "Speech of political candidates", "Election debates".
3. Knowledge bases on election topics: "Election advertising", "Speech of political candidates", "Election debates" with a volume of at least 2000 dictionary entries.
4. Logical rules of inference from knowledge bases.
5. Software application for a processor for processing official and unofficial information sources of political discourse based on a graphematic, morphological, syntactic

analyzer of texts in the Kazakh language with a synonymizer function on election topics: “Election advertising”, “Speech of political candidates”, “Election debates”.
6. Software application sentiment analyzer of official and unofficial information sources of political discourse based on sentiment analysis of texts in the Kazakh language.

The research hypothesis is Ontological models on election topics: “Election advertising”, “Speech of political candidates”, “Election debates”, knowledge bases on these topics, as well as rules for logical inference from knowledge bases.

The expected results are given in the order in which the studies were conducted.

The most important experiments include the verification and refinement of ontological models on election topics, the correctness and completeness of the logical rules of knowledge bases, and then the knowledge base built according to these models for the correctness and completeness of semantic features.

– collecting data to create a text corpus of political discourse;
– collection of primary information on election topics: “Pre-election advertising”, “Speech of political candidates”, “Pre-election debates”;
– Ontological modeling;
– computational logic methods for creating knowledge bases based on formal rules
– methods for formalizing the grammar of the Kazakh language and recording logical rules of inference from the knowledge base;
– graphematic analysis [32];
– morphological analysis [7, 8, 33];
– syntactic analysis [34];
– sentiment analysis [2, 3, 6].

The collection of primary (source) information is carried out on the Internet, as well as through the analysis of printed media publications. Ontological models will also be created on election topics, including concepts and relationships.

Previously, a group of authors was developing methods for analyzing political discourse in social networks in the Kazakh language. Ontological models were chosen for modeling, since this model allows taking into account different aspects of the subject area and taking into account semantics [30, 31].

The reliability and reproducibility of information is ensured by the use of only official sources to create ontological models and knowledge bases, as well as cross-verification of the created knowledge bases by members of the research group and discussion of the results at researchers’ seminars. The knowledge base also contains a database of words and dictionary entries of the subject area with semantic features.

4 Ethical Aspects of Analyzing Kazakh Political Discourse

When analyzing hotbeds of political discourse, which can be of different types, the officiality and openness of these sources of discourse are first taken into account. When we are talking about open sources, where, for example, candidates in various elections openly and officially express their point of view, we are not talking about violation of any legal or ethical standards in the scientific analysis of texts, video, audio information in scientific purposes.

Political discourse analysis in social media necessitates a careful examination of ethical considerations. This paper outlines key principles to be observed during the analysis, focusing on privacy, transparency, fairness, informed consent, cultural sensitivity, and compliance with legal regulations.

In the analysis of open sources of political discourse, where candidates express their views openly and officially, no legal or ethical norms are violated. However, when processing comments, consent is obtained, and the confidentiality of participants is strictly maintained.

Achieving transparency and openness involves providing clear information about the analysis methods, openly stating the research objectives, and ensuring clarity regarding the study's context.

Fairness and impartiality are paramount, excluding bias during the analysis and interpretation of data. Multiple perspectives and political beliefs are considered.

Informed consent is sought from participants where possible, ensuring a comprehensive understanding of the research goals by all stakeholders.

Minimizing potential negative consequences for participants in discussions is a priority, protecting against psychological or social impacts.

Cultural sensitivity is crucial, avoiding stereotypes or biased assessments in the analysis.

Transparency in the use of Artificial Intelligence (AI) technologies is maintained by openly discussing the analysis conducted through AI and addressing possible algorithm limitations and errors.

Discrimination based on political beliefs, social status, or other factors is strictly prohibited.

Ethical review of research projects is conducted before the analysis begins.

Adherence to laws and regulations governing data usage and research conduct is mandatory.

These aspects should be considered as guiding principles to ensure the ethical and responsible nature of the analysis of Kazakh political discourse in social media.

5 Conclusion

In conclusion, the ethical considerations outlined in this article serve as a foundational framework for the analysis of political discourse, specifically within the context of social networks. Emphasizing the principles of officiality, openness, and respect for privacy, the article underscores the importance of maintaining ethical standards while conducting scientific analyses of various media types. The principles of transparency, fairness, informed consent, and cultural sensitivity are pivotal in navigating the complexities of political discourse analysis in the Kazakhstani context. As technology and artificial intelligence play integral roles in these analyses, an ongoing commitment to transparency and adherence to ethical guidelines becomes imperative for responsible and conscientious research in the ever-evolving landscape of political discourse in social media.

Funding. This research is funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant no. AP19679847).

References

1. Bekmanova, G., Yelibayeva, G., Aubakirova, S., Dyussupova, N., Sharipbay, A., Nyazova, R.: Methods for analyzing polarity of the Kazakh texts related to the terrorist threats. In: International Conference on Computational Science and Its Applications. ICCSA 2019, pp. 717–730 (2019). https://doi.org/10.1007/978-3-030-24289-3_53
2. Yergesh, B., Bekmanova, G., Sharipbay, A.: Sentiment analysis of Kazakh text and their polarity. Web Intell. **17**(1), 9–15 (2019). https://doi.org/10.3233/WEB-190396
3. Bekmanova, G., Yergesh, B., Sharipbay, A.: Sentiment analysis model based on the word structural representation. In: Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q., Zhong, N. (eds.) Brain Informatics. LNCS (LNAI), vol. 12960, pp. 170–178. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86993-9_16
4. Bekmanova, G., Yergesh, B., Sharipbay, A., Mukanova, A.: Emotional speech recognition method based on word transcription. Sensors **22**(5) (2022). https://doi.org/10.3390/s22051937
5. Yergesh, B., Bekmanova, G., Sharipbay, A.: Sentiment analysis on the hotel reviews in the Kazakh language. In: Proceedings of the 2nd International Conference on Computer Science and Engineering, UBMK 2017, pp. 790–794 (2017). https://doi.org/10.1109/UBMK.2017.8093531 (2017)
6. Yergesh, B., Bekmanova, G., Sharipbay, A., Yergesh, M.: Ontology-based sentiment analysis of Kazakh sentences. In: Gervasi, O., et al. (eds.) Computational Science and Its Applications – ICCSA 2017. LNCS, vol. 10406, pp. 669–677. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6_47
7. Zhetkenbay, L., Sharipbay, A., Bekmanova, G., Kamanur, U.: Ontological modeling of morphological rules for the adjectives in Kazakh and Turkish languages. J. Theor. Appl. Inf. Technol. **91**(2), 257–263 (2016)
8. Bekmanova, G., et al.: A uniform morphological analyzer for the Kazakh and Turkish languages. In: CEUR Workshop Proceedings, pp. 20–30 (2017)
9. Zappavigna, M.: Social media quotation practices and ambient affiliation: weaponising ironic quotation for humorous ridicule in political discourse. J. Pragmatics **191**, 98–112 (2022). https://doi.org/10.1016/j.pragma.2021.12.003. ISSN 03782166
10. Sen, A., et al.: Analysis of media bias in policy discourse in India. In: ACM International Conference Proceeding Series. 4th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies, COMPASS 2022, vol. Par F180472, pp. 57–77 (2022). https://doi.org/10.1145/3530190.3534798. ISBN 978-145039347-8
11. Papakyriakopoulos, O., Tessono, C., Narayanan, A., Kshirsagar, M.: How algorithms shape the distribution of political advertising: case studies of Facebook, Google, and TikTok. In: AIES 2022 - Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. 5th AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society, AIES 2022, pp. 532–546 (2022). https://doi.org/10.1145/3514094.3534166. ISBN 978-145039247-1
12. Bollenbacher, J., Loynes, N., Bryden, J.: Does United Kingdom parliamentary attention follow social media posts? EPJ Data Sci. **11**(1) (2022). Art. no. 51. https://doi.org/10.1140/epjds/s13688-022-00364-4. ISSN 21931127
13. Phan, H.T., Nguyen, N.T., Hwang, D.: Aspect-level sentiment analysis: a survey of graph convolutional network methods. Inf. Fusion **91**, 149–172 (2023). https://doi.org/10.1016/j.inffus.2022.10.004. ISSN 15662535
14. Al-Absi, A.A., Kang, D., Al-Absi, M.A.: Sentiment analysis and classification using deep semantic information and contextual knowledge. Comput. Mater. Continua **74**(1), 671–691 (2023). https://doi.org/10.32604/cmc.2023.030262. ISSN 15462218

15. Saranya, S., Usha, G.: A machine learning-based technique with intelligent WordNet lemmatize for Twitter sentiment analysis. Intell. Autom. Soft Comput. **36**(1), 339–352 (2023). https://doi.org/10.32604/iasc.2023.031987. ISSN 10798587
16. Yergesh, B., Kenzhina, L.: Analysis of the users' emotional state in social networks. In: ACM International Conference Proceeding Series, art. no. 3492654. 7th International Conference on Engineering and MIS, ICEMIS 2021 (2021). https://doi.org/10.1145/3492547.3492654. ISBN 978-145039044-6
17. Boranbayev, A., Shuitenov, G., Boranbayev, S.: The method of analysis of data from social networks using rapidminer. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) Intelligent Computing. AISC, vol. 1229, pp. 667–673. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52246-9_49. ISSN 21945357. ISBN 978-303052245-2
18. Sultan, D., Suliman, A., Toktarova, A., Omarov, B., Mamikov, S., Beissenova, G.: Cyberbullying detection and prevention: data mining in social media. In: 11th International Conference on Cloud Computing, Data Science and Engineering, Confluence, pp. 338–342 (2021). Art. no. 9377077. https://doi.org/10.1109/Confluence51648.2021.9377077. ISBN 978-073813160-3
19. Gulnara, B., Ilyas, Z., Gulnara, Z.: The development of a web application for the automatic analysis of the tonality of texts based on machine learning methods. In: International Conference on Control, Automation and Systems, vol. 2018, art. no. 8571950. 18th International Conference on Control, Automation and Systems, ICCAS 2018 (2018). ISSN 15987833. ISBN 978-899321515-1
20. Swed, O., Dassanayaka, S., Volchenkov, D.: Keeping it authentic: the social footprint of the trolls' network. Soc. Netw. Anal. Mining **14** (1) (2024). Art. no. 38
21. Al-Shboul, O.K., Al-Khawaldeh, N.N., Hamdan, H.J., Alqbailat, N.: Macro and micro analysis of motivational speech acts in Biden's political speech. Int. J. Soc. Cult. Lang. **12**(1), 362–373 (2024)
22. Santos, A.N., de Souza, G.M., Abdalla, M.M., Ferreira, A., Nogueira, N.J.: Lobbying and environmental crimes: an analysis based on the Brazilian mining sector. Extr. Ind. Soc. **17** (2024). Art. no. 101419
23. Losen, B.: Uses of scientific discourses and markers in shark attack risk management on Reunion Island. Marine Policy **161**, art. no. 106011 (2024)
24. Sun, Q., Chen, J., Gao, S.: From panic to banter: how do routine government releases and clarifications cause unexpected public opinion crisis—an analysis of public opinion toward a release by Chinese Ministry of Commerce encouraging the storage of necessities. J. Contingencies Crisis Manag. **32** (1) (2024). Art. no. e12530
25. Lebold, M.: Exploring feminist political economy and feminist critical discourse analysis as methodologies in critical nursing research. J. Adv. Nurs. **80**(3), 958–970 (2024)
26. Igwebuike, E.E., Chimuanya, L.: Self-promotion, ideology and power in the social media posts of Nigerian female political leaders. J. Lang. Polit. **23**(1), 67–90 (2024)
27. Alpermann, B., Malzer, M.: "In Other news": China's international media strategy on Xinjiang—CGTN and New China TV on YouTube. Mod. China **50**(2), 135–178 (2024)
28. Hensmans, M.: The multinational as a myth-prince of the global south: writing back an emancipating imaginary to the global north. Organization **31**(2), 269–294 (2024)
29. Hand, I.: If the municipality cannot do It!: Negotiating the Boundary between State and Society in Early republican Turkish cities. J. Urban Hist, **50**(2), 289 –306 e 7.1, 20220141 (2024)
30. Bekmanova, G., Omarbekova, A., Mukanova, A., Zulkhazhav, A., Zakirova, A., Ongarbayev, Ye.: Development of an ontological model of words in public political discourse. In: Proceedings of the 7th International Conference on Education and Multimedia Technology (ICEMT 2023), pp. 362–367. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3625704.3625720

31. Sairanbekova, A., Bekmanova, G., Omarbekova, A., Mukanova, A., Zulkhazhav, A.: The use of Python, Owlready, Sparql in processing the words ontological model of public political discourse. Comput. Syst. Commun. Technol. **49**, 45–53 (2024). https://doi.org/10.3233/ATDE240007
32. Sharipbay, A., Niyazova, R., Turebayeva, R.T., Razakhova, B., Zulkhazhav, A., Yelibayeva, G.: Graphematic analysis of Kazakh language text. J. Math. Mech. Comput. Sci. **3**(103), 90–102 (2019)
33. Yelibayeva, G., Mukanova, A., Sharipbay, A., Zulkhazhav, A., Yergesh, B., Bekmanova, G.: Metalanguage and knowledgebase for Kazakh morphology. In: Misra, S., et al. (eds.) Computational Science and Its Applications – ICCSA 2019. LNCS, vol. 11619, pp. 693–706. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24289-3_51
34. Sharipbay, A., Razakhova, A., Mukanova, A., Yergesh, B., Yelibayeva, G.: Syntax parsing model of Kazakh simple sentences. In: DATA 2019: Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems, pp. 1–5 (2019). https://doi.org/10.1145/3368691.3368745