

Doctoral School of Information and Biomedical Technologies
Polish Academy of Sciences (TIB PAN)

SUBJECT:

Automatic balancing of gender forms for content in Polish

SUPERVISOR:

dr hab. inż. Grzegorz Borowik
grzegorz.borowik@nask.pl
NASK-PIB, Centrum Badań i Rozwoju
Kolska 12, 01-045 Warszawa

dr Inez Okulska
inez.okulska@nask.pl
NASK-PIB, Centrum Badań i Rozwoju
Kolska 12, 01-045 Warszawa

DESCRIPTION:

Contemporary cultural and social changes prompt the adaptation of the Polish language to the evolving reality. To achieve this goal, computer tools allowing for the automatic balancing of grammatical gender forms in Polish texts can be useful. The doctoral thesis will utilize research methods to analyze, develop, and implement algorithms for the automatic balancing of grammatical gender forms in Polish texts. Specifically, syntactic analysis of the Polish language, statistical analysis, machine learning techniques, and natural language processing will be applied.

Moreover, the development of algorithms for the automatic balancing of grammatical gender forms in Polish can positively impact content analysis in other Slavic languages, contributing to a better understanding of these languages and cultures. This innovation will expand research on Slavic languages and increase knowledge about them.

The analysis of grammatical gender forms in Polish will also improve the quality of machine translations, which is particularly important in today's globalized world. The conclusions resulting from the doctoral thesis will have practical applications in natural language processing, machine translations, and text editing and proofreading. Innovative algorithms for the automatic balancing of grammatical gender forms in Polish will enable the improvement and standardization of the language and help build bridges between different cultures and languages.

BIBLIOGRAPHY:

1. Gonen, H., Kementchedjheva, Y., & Goldberg, Y. (2019). How Does Grammatical Gender Affect Noun Representations in Gender-Marking Languages? In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL) (pp. 463-471). Association for Computational Linguistics.
2. Kaczmarek, K., Pokrywka, J., & Gralinski, F. (2022). Using Transformer models for gender attribution in Polish. 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS), 73-77.
3. Palczynski, M., & Pechmann, P. (2001). Application of semantic analysis of Polish speech for purpose of automatic gender identification. International Conference on Advanced Computer Systems.

4. Formanowicz, M., Cislak, A., Horvath, L.K., & Sczesny, S. (2015). Capturing socially motivated linguistic change: how the use of gender-fair language affects support for social initiatives in Austria and Poland. *Frontiers in Psychology*, 6.
5. Nowosad-Bakalarczyk, M. (2020). On the Category of Sex and Grammatical Gender in the Polish Language.
6. Krajewski, G. (2005). The role of grammatical gender in the acquisition of noun inflection in Polish. *Psychology of Language and Communication*.
7. Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. (2011). Author Gender Identification from Text. *Digital Investigation*, 8(1), 78-88.
8. Alekseev, A., & Nikolenko, S. (2020). Recognizing Preferred Grammatical Gender in Russian Anonymous Online Confessions. In P. Sojka, I. Kopeček, K. Pala, & A. Horák (Eds.), *Text, Speech, and Dialogue. TSD 2020* (pp. 32-40). Springer, Cham.
9. Zhou, P., Zhang, W., Peng, H., & Zhao, J. (2019). Examining Gender Bias in Languages with Grammatical Gender. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 5276-5284).
10. McCarthy, A. D., Martin, A., Hewlett, W., & Adelani, D. (2020). Measuring the Similarity of Grammatical Gender Systems by Comparing Partitions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 5664-5675).