

A Quantitative Analysis of Azerbaijan’s President’s Multilingual Discourse (2016–2026)

Mgr. Konstantin Startsev

Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava
Mlynská dolina F1, 842 48 Bratislava, Slovakia
startsev1@uniba.sk

1 Introduction

Dehumanisation (the denial of an individual’s essential humanity and identity) and toxification (the cognitive perception of the target group as fundamentally threatening to the perpetrators’ survival and society) are widely regarded as warning signs and catalysts of mass violence. A qualitative discourse analysis by Sahakyan (2023) documented systematic dehumanisation and toxification of Armenians in the speeches of the President of Azerbaijan in the context of the Second Karabakh War (27 September – 10 November 2020). She identified three narrative groups in President Ilham Aliyev’s rhetoric: (1) the identification of Armenians as the sole threat to Azerbaijan and Azerbaijanis; (2) their portrayal as non-human and barbaric in essence; and (3) the image of the Azerbaijani army. Sahakyan also observed (although she analysed only the English versions of the speeches) that Aliyev’s rhetoric is tailored to its audience: the international community receives a softened version.

We ask whether these findings can be confirmed quantitatively, and test three hypotheses: (i) whether dehumanising language can be detected and tracked within a large corpus; (ii) whether it intensifies in the run-up to war; and (iii) whether it varies systematically across the channels the regime uses for domestic and foreign audiences.

2 Methods

We compiled a trilingual corpus from the materials on the official presidential website *prezident.az* (2016–2026): 6,969 articles in Azerbaijani, Russian, and English ($\approx 20,892$ documents). Texts were assigned to four periods defined by the chronology of the conflict (peaceful, pre-war, war, post-war). To compare the versions addressed to different audiences, we built a trilingual dictionary of dehumanising terms and measured their frequency across the aligned versions. Content was then analysed with three methods: (1) **Distributional co-occurrence**: per-period Word2Vec spaces with frequency-weighted concept vectors (threat, vermin, moral disgust), following Caporusso et al. (2024); (2) **Supervised classification**: a TF-IDF + SGD linear classifier trained on several hundred manually annotated paragraphs following Shyrokykh et al. (2023); the linear model reached $F_1 \approx 0.82$ (five-fold cross-validation). (3) **BERT Transformer**: a fine-tuned mul-

tilingual transformer (XLM-RoBERTa) trained on the same data.

3 Results

The trilingual dictionary analysis confirms Sahakyan’s audience-tailoring observation. Across aligned articles, the English versions contain systematically fewer dehumanising and military-related terms than their Azerbaijani and Russian ones.

A covariance analysis shows that in every period references to Armenians are associated with dehumanizing language. The supervised classifier reveals a sharper temporal structure: the share of dehumanising paragraphs peaks in the pre-war period. The consistent linguistic gap indicates that the English version is consistently softer than the other languages.

4 Discussion

Our study quantitatively corroborate the qualitative mechanism described by Sahakyan (2023) now across three languages and 20,892 documents. Research into the language of dehumanisation and the tailoring of messages by autocracies should be further tested using data from other conflicts, as well as by exploring the potential of language monitoring to predict mass violence.

References

- Caporusso, J., Hoogland, D., Brglez, M., Koloski, B., Purver, M. and Pollak, S. (2024). A computational analysis of the dehumanisation of migrants from Syria and Ukraine in Slovene news media. Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S. and Xue, N. (ed.), In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 199–210. ELRA and ICCL.
- Sahakyan, N. (2023). The rhetorical face of enmity: the Nagorno-Karabakh conflict and the dehumanization of Armenians in the speeches of Ilham Aliyev. *South-east European and Black Sea Studies*, 23(4):863–882.
- Shyrokykh, K., Girnyk, M. and Dellmuth, L. (2023). Short text classification with machine learning in the social sciences: the case of climate change on Twitter. *PLoS ONE*, 18(9):e0290762.