

Lenin in Translation: The Global Journey of Revolutionary Discourse

Mikhail Mikhailov & Research Group

1. Background and significance

1.1. Significance of research project in relation to current knowledge, research-based starting points

Vladimir Lenin (1870–1924) died 100 years ago. Yet, this person is still the subject of lively debate: in some countries, statues of Lenin are removed and he is blamed for all current problems, while elsewhere streets, ships and children are named in his honour. There exist a wealth of very contradictory publications about Lenin and still we know surprisingly little about this person.

Vladimir Lenin's own works and the mythology around his personality, the “Leniniana”, had substantial influence on the development of the present-day Russia, at least because many of Russia's current elite belong to the so-called “last Soviet generation” that read in their youth more works by Lenin than by any other author (Yurchak 2006). Subsequently, if one wishes to understand contemporary Russian society and thinking, it is necessary to understand Lenin's discourse. This research project focuses on gaining renewed understanding of Lenin's historical heritage, by studying how it has affected the formation of political discourse in Russia, Finland, and other countries.

Lenin is probably one of the most studied thinkers of the 20th century, but the focus has been on his political theory and his political activities. In linguistics, Lenin's works have been studied mainly from the point of view of language and style, and the translations of Lenin's works have been studied even less, although Lenin is probably the world's most translated political writer: he occupies the 7th place in Index Translationum database (<https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=5&nTyp=min&topN=50&lg=0>), and his works were translated into 125 languages. Besides, these studies often served the purpose of establishing Lenin's Cult (Tumarkin 1983).

Therefore, a multidisciplinary study of Lenin's works, their translations (as well as of the process of dissemination of Marxism-Leninism worldwide), using modern methods of data collection, processing and analysis will give a clue to a better understanding of contemporary Russia and CIS area. To facilitate research and make it more convincing, new distant reading tools will be developed for detecting relevant features of text and their visualisation, i.e., ‘distant reading’ and ‘distant drawing’. Finally, a conversational agent based on Lenin's texts (“Lenin-bot”) will be developed.

1.2. Research premise, aims and objectives

The aim of this project is to initiate a systematic corpus-based study of Lenin's works and their translations. The project team is interdisciplinary, it consists of linguists, translation scholars, and data scientists. We study original texts, translations and the history of their making, publication and distribution. Research is performed in cooperation with libraries, archives, and museums. The electronic text corpora compiled during the project, as well as other materials will be made publicly available. At the current stage of research, the focus is on original works in Russian, as well as Russian-Finnish, Russian-English, and Russian-Spanish data, but in the long term we are planning to create an international network and expand the research to other languages: German, Dutch, Slovene, etc. The new IT-tools for distant reading and visualisation will be created and finally a multilingual conversational agent based on Lenin's works and their translations ("Lenin-bot") will be developed.

1.3. Research questions and/or hypotheses

The goals of the project are to find answers to the following research questions:

- What were the main features of Lenin's discourse, and were they also transferred to the translations?
- When did the translation of Lenin's works start? How were they published? Who were the translators? Why were many of the works retranslated?
- Is there a difference between the early translations of Lenin's works and those published in the times of Lenin's cult?
- How Lenin's heritage is reflected in political discourses in Russia, Finland, U.K., U.S.A. and other countries in different time periods?

The hypothesis is that publishing and translating of Lenin's works was a part of the creation of Lenin's cult, therefore only features supporting the image of "the most human of all humans" were transferred to the translations. This tendency can be found both in publishing of original works and in translations, but in translations it is more visible. For the same reason the influence of Leninist discourse is more likely to be found in the leftist discourse, although some maxims might be smuggled to the broader political circles.

2. Impact

2.1. Outcomes and impacts within academia

The project will develop new methods of analysis of multilingual political discourse that will be applicable to research of current political discourse.

The findings of the project will activate research on the topic of using translation in political propaganda and provide a rich dataset on Leninist discourse.

New IT-tools will help to explore multilingual data in digital humanities and social sciences.

The data collected will be valuable in itself. Parallel corpora is difficult to collect and most of the large parallel corpora are institutional texts (UN, European Parliament, European Commission), religious texts (the Bible) or technical documentation for a limited number of language pairs. Parallel corpora of Marxist texts can become new large multipurpose datasets with many rare language pairs included (languages of Russia, Caucasus, Central Asia etc.) that will provide linguists with data for contrastive and typological research.

2.2. Effects, impact and interaction beyond academia

The project will help to digitise some of the materials from libraries, archives and museums. The cooperation with museums of Tampere will provide new exponents, a Lenin conversational agent among them.

The digital corpora and analytical tools will be a useful source of information for educational organisations and the bodies of power.

The parallel corpora will become a useful dataset for training machine translation systems and other AI applications, as well as for online dictionary services.

3. Research data and material, methods and research environment

3.1. Research data

Lenin's works can be found on many Internet sites, both original works and their translations into many languages (e.g. leninism.su, marxists.org), still, many translations are not yet digitized or are available only as scanned pdf files. We are compiling several digital corpora that will be used as sources of data for our research and will be made publicly available after the conclusion of the project. These are:

- LenCor, a corpus of Complete works by Lenin. Current version: 2,360 documents, 4.5 M running words).
- RuDiRe, a corpus of Russian political discourse of the time of revolutions (1890s-1920s). A reference corpus compiled of works by prominent politicians, philosophers, writers, scholars, artists, and journalists of that time. Currently contains 181 documents, 2 M running words.
- ParVLen, a multilingual parallel corpus of Lenin's works. The goal is to include in the corpora at least the most important of Lenin's works. The Russian-Finnish data is collected with an emphasis on retranslation, i.e., the works that were translated several times will receive special attention. At the moment of submitting the application the sizes of the parallel corpora were as follows:
Russian-English: 39 text pairs, about 600,000 running words per language,
Russian-Finnish: 153 text pairs, about 1.2 M running words per language,
Russian-Spanish: 12 text pairs, about 250,000 running words per language.

The corpora are all full-text corpora (no IPR issues; both the original works and all translations are copyright-free). The texts are processed with Universal Dependencies parsers, parallel texts are

aligned at sentence level. The main corpus management tool is NoSketch Engine, parts of the data will be also processed with Sketch Engine.

3.2. Research methods

In the study of text materials, the methods of corpus-based discourse analysis are used. We will find typical for Lenin's texts means of influencing the audience: metaphors and metaphoric models, irony, sarcasm, expressive language. The study will combine quantitative data and usage examples. Parallel corpora will be used to study how the features of the original works are (or are not) transferred into translations. The translations will be studied both in relation to original texts and to each other and to other target languages.

To illustrate the overall nature of Lenin's *persuasive* discourse, for example, we will first order the target texts along a temporal timeline (or other relevant contextual dimension, e.g., intended target audience) and then detect the densities (and density changes) of manipulative expressions within these texts. Well known tools of persuasion include the use of pronouns "we" and "you" to create an in- and outgroup; extreme case formulations ('of course', 'no doubt', 'know perfectly well', etc.); grammatical presupposition triggers, e.g. questions; and citing sources using mocking/glorifying stance. We will first collect lists of potentially persuasive expressions in target languages, and then visualise the corresponding trends along the timeline in texts ("stock market curves").

Using the methods of digital humanism, we will find out whether Lenin's ideas had any impact on Finnish political discourses in different time periods by comparing Finnish translations with parliamentary minutes and newspapers. We will use existing online full-text databases (e.g. ParlamenttiSampo, Kansalliskirjaston sanoma- ja aikakauslehtikokoelman suomenkielinen osakorpus). The study of text recycling has been an active research area in recent years, and we can apply ready-made methods to our materials (see e.g. Salmi et al. 2020). So far, the recycling of texts has been studied by focusing on one source material at a time, but the project develops methods that can be used to trace the movement of texts between several different source materials.

To aid serendipitous discoveries regarding Lenin's use of language (e.g., persuasion), we will design easy-to-use computerised tools which allow observing relevant textual features ("distant reading") in corpora, and visualising them automatically ("distant drawing"), e.g., so show linguistic features of specific style appearing across different contexts and timelines.

In this project, we will design, develop, and evaluate a conversational agent called the "Lenin Bot," leveraging large language models (LLM) such as OpenAI's GPT-4 and its future iterations. The capabilities of these LLMs will be enhanced by incorporating an extensive and unique corpus, inaccessible to existing language models, through Retrieval-Augmented-Generation (RAG) technologies. We will employ collaborative fine-tuning processes, involving translation studies students, to improve the agent's performance in its specific domain. This entails versioning and customizing the bot's stylistic tone and personality, as well as enabling its operation across various languages, and cultural, historical, and temporal contexts. These approaches aim to significantly mitigate "hallucinations", a common issue in current language models.

The results obtained with the corpus tools are supplemented by means of the sociology of translation, and digital humanism. The social and political context of Finnish translations made in different time periods, in different contexts, and in different countries, the goals, agency and institutional status of translators and publishers (see e.g. Krekola 2006) are studied both by combining existing research data and by doing archival work with primary materials.

3.3. Environment

The research is being carried out at the faculty of Information Technology and Communication Sciences (ITC) of the Tampere University, the members of the team are from two units of the faculty: Languages Unit and Computing Sciences Unit.

The project has two virtual servers to host the data. The corpora are processed with the help of Universal Grammar parsers and the set of corpus tools No Sketch Engine.

The project will collaborate with FIN-CLARIAH and the Bank of Languages of Finland, as well as with OPUS project (Open Source Corpus). The finalised corpora will be made publicly available at the online services of these infrastructures (kielipankki.fi, corp.csc.fi, <https://opus.nlpl.eu/>).

The project will also use various machines and tools for scanning paper documents, optical character recognition, aligning parallel texts etc. Most of the tools are available at the University, some will be purchased (e.g. commercial aligning tools, commercial version of LLMs for aligning and developing the conversation agent).

4. Current stage of the project

4.1. Conference presentations

- Mikhail Mikhailov & Julia Souma. "Poliittisten tekstien kääntäminen: yhteys lukijoihin vai alkuperäistekstiin? Tapaus Lenin." KäTu symposium, oral presentation. Tampere University April 2024.
- Mikhail Mikhailov, Julia Souma, Nelli Meer, Petra Miettunen, Elizaveta Sidorova. "Monta kertaa käännetty Lenin. Leninin rinnakkaiskorpus, venäjä-suomi -osio." KäTu symposium, poster. Tampere University April 2024 (Best poster award).
- Mikhail Mikhailov. "From Ulyanov to Lenin: a corpus-based discourse analysis of Vladimir Lenin's work." CILC-2024, oral presentation. Las Palmas, May 2024.
- Mikhail Mikhailov & Kaisa Koskinen. "Retranslating Lenin." Retranslation in Context VI, oral presentation. Izmir, October 2024.
- Mikhail Mikhailov. "Lenin en español. ¿Menos arrogante?" GranaSlavic2025 (University of Granada). June 2-4, 2025.
- Mikhail Mikhailov, Satu Valkiainen - The Roads We Take. The Metaphor of a Journey in Lenin's Works and Their Finnish Translations. University of Bergamo, June 2025.

- Mikhail Mikhailov, Daria Dayter, Julia Souma – Abominable, Disgusting, Repugnant and Vile. Strong language in Lenin’s works and their translations. SKY 2025, Tampere University, October 2025.

4.2. Publications

- Mikhail Mikhailov, Julia Souma, Satu Valkiainen. 2025 “*Hännystely, häntyyryys, perässälaahustaminen vai hvostismi?* Leninin uudelleensuomennoksia vertailemassa” [=Hännystely, häntyyryys, perässälaahustaminen vai hvostismi? Studying retranslations of Lenin’s works]. *MikaEl* 18:2.
- Mikhail Mikhailov and Kaisa Koskinen. 2026 (forthcoming) “Translating Lenin as Institutional Translation: From Ideal to Cult.” *Translation Studies*.
- Mikhail Mikhailov. “Opyt korpusnogo analiza rabot V.I. Lenina”. [=Corpus-based analysis of Lenin’s works]. Article submitted to the journal *Scando-Slavica* (JuFo-2), January, 2025.

References

- Arvola, Paavo, Tuulikki Alamettälä (2022): IRVILAB: Gamified Searching on Multilingual Wikipedia. *SIGIR 2022*: 3329-3333.
- Baker, Paul 2006. *Using Corpora in Discourse Analysis*. Continuum Discourse Series. London and New York: Continuum.
- Brooks, Jeffrey & Georgiy Chernyavskiy 2007. *Lenin and the Making of the Soviet State. A Brief History with Documents*. New York: Palgrave Macmillan.
- Charteris-Black, Jonathan 2018. *Analysing Political Speeches: Rhetoric, Discourse and Metaphor*. 2nd edition. London: Palgrave.
- Dayter, Daria & Sofia Rüdiger 2022. *The Language of Pick-Up Artists. Online Discourses of the Seduction Industry*. London: Routledge.
- Dayter, Daria, Miriam A. Locher & Thomas C. Messerli 2023. *Pragmatics in Translation. Mediality, Participation and Relational Work*. Cambridge: Cambridge UP.
- Ennker, Benno 1997. Die Anfänge des Leninkults in der Sowjetunion. Köln: Böhlau. / Ennker, Benno 2011. Formirovanije kulta Lenina v Sovetskom Sojuze. Moskva: ROSSPEN.
- Ertürk, Nergis & Özge Serin 2016. Marxism, Communism, and Translation: An Introduction. *Boundary 2*, 43(3), 1–26.
- Fairclough, Norman 2013. *Critical Discourse Analysis: the Critical Study of Language*. 2nd edition. London: Routledge.
- Gillings, Mathew, Gerlinde Mautner & Paul Baker. 2023. *Corpus-assisted Discourse Studies*. Cambridge University Press.
- Keskustalo, Heikki, Laura Korkeamäki, Selja Vanamo, Kimmo Kettunen, Sanna Kumpulainen (2023): Analyzing gender clues in war-time letters. *Digit. Scholarsh. Humanit.* 38(1): 209-223.
- Koskinen Kaisa & Outi Paloposki 2019. New directions for retranslation research: Lessons learned from the archaeology of retranslations in the Finnish literary system. *Cadernos de Tradução* 39 (1), 23-44.
- Koskinen, Kaisa 2018. Revising and Retranslating. In: Kelly Washbourne & Ben Van Wyke (eds.) *Routledge Handbook of Literary Translation*. Routledge, 315–324.

- Koskinen Kaisa & Outi Paloposki 2015. *Sata kirjaa, tuhat suomennosta. Kaunokirjallisuuden uudelleenkäyttäminen*. Helsinki, SKS.
- Krekola, Joni 2006. *Stalinismin lyhyt kurssi. Suomalaiset Moskovan Lenin-koulussa 1926–1938*. Helsinki: SKS.
- Krekola, Joni 2008. Lenin Lives in Finland. In: Juhana Aunesluoma & Pauli Kettunen (eds.) *The Cold War and the Politics of History*. Helsinki: Edita, 107–125.
- Mikhailov, Mikhail 2022. Corpus-based Analysis of Russian Translations of Animal Farm by George Orwell. In: Meng Ji and Michael P. Oakes (eds.) *Corpus Exploration of Lexis and Discourse in Translation*. London: Routledge, 56–82.
- Mikhailov, Mikhail 2021. God, the Devil, and Christ: A Corpus Study of Russian Syntactic Idioms and Their English and Finnish Translation Correspondences. In: Aleksandar Trklja and Łukasz Grabowski (eds.) *Formulaic Language: Theories and Methods*. Berlin: Language Science Press, 191–223. <https://langsci-press.org/catalog/view/304/3060/2282-1>.
- Mikhailov Mikhail & Robert Cooper 2016. *Corpus Linguistics for Translation and Contrastive Studies: a guide for research*. London and New York: Routledge.
- Mikhail Mikhailov, Miia Santalahti and Julia Souma 2019. PEST: A parallel electronic corpus of state treaties. In: Irene Doval, Maria Teresa Sánchez Nieto (eds.) *Parallel Corpora: Creation and Applications*. Benjamins, 183–195.
- Munday, Jeremy & María Calzada Pérez (eds.) 2020. New Contexts in Discourse Analysis for Translation and Interpretation. *Special issue of Meta* 65 (1).
- Numminen, Jaakko (ed.) 1987. *Lenin ja Suomi*. Vol. 1. Helsinki: Opetusministeriö.
- Numminen, Jaakko (ed.) 1989. *Lenin ja Suomi*. Vol. 2. Helsinki: Opetusministeriö.
- Numminen, Jaakko (ed.) 1990. *Lenin ja Suomi*. Vol. 3. Helsinki: Opetusministeriö.
- Rauhala, Paula 2023. *Marx in the West and in the East: Reading Capital in the Divided Germany*. Ph.D. dissertation: Tampere University. <https://trepo.tuni.fi/handle/10024/146256>
- Salmi, Hannu et al. 2020. The Reuse of Texts in Finnish Newspapers and Journals, 1771–1920: A Digital Humanities Perspective. *Historical Methods. A Journal of Quantitative and Interdisciplinary History*, 54(1), 14–28.
- Service, Robert 1985. *Lenin: A political life. Volume 1. The Strengths of Contradiction*. London: Macmillan.
- Service, Robert 1991. *Lenin: A political life. Volume 2. Worlds in Collision*. London: Macmillan.
- Service, Robert 1995. *Lenin: A political life. Volume 3. The Iron Ring*. London: Macmillan.
- Souma, Julia 2023. *Diplomatičeskoe «razrešaetsâ» : kommunikativnye akty razrešeniâ i zapreta v finlândsko-rossijskih meždunarodnyh soglašeniâh*. Ph.D. Dissertation, Tampere University. <https://urn.fi/URN:ISBN:978-952-03-3042-2>
- Turunen, Risto 2021. *Shades of Red: Evolution of the Political Language of Finnish Socialism from the 19th Century until the Civil War of 1918*. Helsinki: The Finnish Society for Labour History. Ph.D. dissertation: Tampere University. <https://urn.fi/URN:ISBN:978-952-03-1906-9>.
- Velikanova, O. 1996. *Making of an Idol: on Uses of Lenin*. Göttingen.
- Ventsel, Andreas 2011. “Lenin is the Stalin of Today”: A Deictic Approach to the Cult of the Leader. *Russian Journal of Communication* 4(1–2), 38–52, DOI: 10.1080/19409419.2011.10756789.
- Yurchak, Alexei 2006. *Everything was Forever, Until it Was no More: the Last Soviet Generation*. Princeton: Princeton University Press.