Interpretable Models for Detecting Linguistic Variation in Russian Media: Towards Unveiling Propagandistic Strategies during the Russo-Ukrainian War

Anastasiia Vestel
Saarland University
Saarbruecken, Germany
anastasiia.vestel@uni-saarland.de

Stefania Degaetano-Ortlieb
Saarland University
Saarbruecken, Germany
s.degaetano@mx.uni-saarland.de

Abstract

With the start of the full-scale Russian invasion of Ukraine in February 2022, the spread of pro-Kremlin propaganda increased to justify the war, both in the official state media and social media. This position paper explores the theoretical background of propaganda detection in the given context and proposes a thorough methodology to investigate how language has been strategically manipulated to align with ideological goals and adapt to the changing narrative surrounding the invasion. Using the WarMM-2022 corpus, the study seeks to identify linguistic patterns across media types and their evolution over time. By doing so, we aim to enhance the understanding of the role of linguistic strategies in shaping propaganda narratives. The findings are intended to contribute to the broader discussion of information manipulation in politically sensitive contexts.

1 Introduction

The Russo-Ukrainian war has intensified the need to understand media manipulation and its societal impacts. There has been an increased number of endeavors for propaganda detection, in general and on the Russo-Ukrainian war specifically. Since language variation can be driven by external factors such as social, political, or cultural influences, studying linguistic change in the context of propaganda can help detect it more accurately. This argument is further supported by the fact that disinformation changes and evolves over time (Adriani, 2019), as is the case with Russian propaganda (Solopova et al., 2023a), which has been used by the government to justify the invasion and gain support from its population. Moreover, research has shown that linguistic change can occur not only diachronically, but also across diverse contexts, such as different political viewpoints (Azarbonyad et al., 2017; Ustyianovych and Barbosa, 2024). By comparing traditional mass media, i.e., press and TV, with social media in Russia, Alyukov et al. (2024) found that propaganda frames[1] differ between these two text types: state media are targeted at more passive audiences, whereas social media seek to convince those searching for alternative sources of information. This suggests that there might be fewer regime supporters on social media, and thus the political stance of the users might differ between the two text types.

This paper presents a research framework to analyze Russian state-controlled media and social media, which will allow us to answer the following research questions: (1) how language in these two text types linguistically differs and might reflect propaganda strategies (e.g., the use of euphemisms); (2) how it might have changed over time. As a result, we expect to see linguistic variations between the two media types, since they use distinct propaganda frames and strategies. Specifically, we might find a tendency towards euphemistic choices to prevail in state-controlled media texts in comparison to social media posts, such as by replacing *war* with *special military operation* (the former term is less likely to be propaganda, cf. Park et al., 2022; Solopova et al., 2023a). Additionally, by conducting our analysis, we anticipate to trace the diachronic evolution of Russian propaganda about the war in Ukraine.

Even though propaganda and disinformation detection is a common natural language processing (NLP) task, few studies have focused on linguistic change as a possible indicator of information manipulation. Furthermore, recent research relies on transformer-based architectures exploiting contextual embeddings for propaganda detection and classification into techniques (e.g., Hein, 2023).

[1] According to Entman (1993), to frame is to "select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described" (Entman, 1993, p. 52).

Proceedings of LaTeCH-CLfL 2025, pages 109–116
May 4, 2025 ©2025 Association for Computational Linguistics

While these methods perform relatively well, they heavily rely on annotated data and explainability remains a major issue, as they do not allow fully capturing on which basis the classification of propagandistic texts is motivated (cf. Da San Martino et al., 2021; Park et al., 2022). In this study, besides using word embeddings, we also propose interpretable methods applied to the analysis of language variation and change. Word embeddings (Mikolov et al., 2013; Hamilton et al., 2016) will allow us to model semantic shifts. Kullback-Leibler Divergence (KLD; Kullback and Leibler, 1951) is employed to detect and analyze features contributing to change across linguistic levels (Hughes et al., 2012; Bochkarev et al., 2014; Fankhauser et al., 2014; Klingenstein et al., 2014; Degaetano-Ortlieb and Teich, 2022). To capture more nuanced changes in the local linguistic context, we use surprisal, which models the (un)expectedness of words in particular contexts (Hale, 2001). The combination of these methods would help us detect distinctive features of linguistic change, providing highly interpretable results. Their detailed description is provided in Section 4.

2 Related Work

Since the beginning of Russia's full-scale invasion of Ukraine, there have been a number of attempts to combat Russian propaganda with the help of NLP techniques. Some studies applied both traditional and neural machine learning to detect pro-Kremlin propaganda with promising results (Vanetik et al., 2023; Solopova et al., 2023b, 2024). Arguing that dehumanization leads to extreme violence, Burovova and Romanyshyn (2024) trained a few binary classifiers to detect dehumanizing language of Ukrainians on Russian social media, with the SpERT model outperforming the rest.

Other studies delved into the computational analysis of Russian propaganda about the war. For instance, Alyukov et al. (2023) created the Wartime Media Monitor (WarMM-2022) corpus, which includes publications on the Russo-Ukrainian war and consists of two parts: state media and social media, and used it to analyze major propaganda themes and strategies. In their later study, Alyukov et al. (2024), by working on the same dataset, explored the differences between propaganda frames representing diverse semantic entities in the two subcorpora. They identified the following frames: dependence (the narrative about Ukraine's dependence on the West), dehumanization (using dehumanizing language towards Ukrainians), normalization (downplaying the effects of the war on the everyday life in Russia), and disinformation (presenting news from Ukraine and the West as fake). The researchers found that the press and TV applied the dehumanization frame (which is in line with the results reported by Burovova and Romanyshyn, 2024), as well as dependence and normalization, while social media used the disinformation frame. These strategies, on the one hand, aimed to pacify the regime supporters who mostly consumed traditional media, and on the other hand, tried to mobilize the users of social media by employing the disinformation frame. The findings by Solopova et al. (2023a) also confirmed that the mobilization strategy was used by the government to target the Russian population. This indicates that the two text types are distinct from each other, as they are aimed at different audiences (at the semantic as well as other linguistic levels).

Similarly, Park et al. (2022) analyzed the media effects of Russian news about the war; however, instead of comparing press and TV with social media, they looked into state-affiliated and independent outlets on two online platforms, VKontakte and Twitter. They found that since the start of the full-scale invasion, independent media outlets have predominantly used the term *war*, while state-affiliated outlets have frequently opted for the euphemism (*special military*) *operation*. The same difference was observed by Ustyianovych and Barbosa (2024) between pro-Russian and pro-Ukrainian Telegram channels, indicating that political opinions might influence semantic choices and phrasing. Using the term *special military operation* was also given as an example of the normalization propaganda frame by Alyukov et al. (2024). This is in line with Solopova et al.'s (2023a) results, who trained two classifiers based on SVM and BERT to detect pro-Kremlin propaganda, and found that the word *war* was highly predictive for both models, meaning that a text containing it was more likely to be labeled as not propagandistic. The authors explain it by the fact that this term was deliberately avoided by government officials and even became illegal in Russia. Consequently, it rarely appeared in pro-Kremlin news, which relied on euphemisms instead.

Apart from linguistic variation between the state-affiliated and independent media, Park et al. (2022) also observed differences in the two platforms

(VKontakte vs. Twitter, particularly divergent framing strategies), as well as across time (before vs. after the beginning of the full-scale invasion, e.g., an increase of frequency of terms related to war). This confirms Azarbonyad et al.'s (2017) hypothesis that semantic change can occur both diachronically and in distinct contexts, such as divergent political viewpoints. Since traditional media and social media may reflect differences in the stance of the users, we presume that language might also vary between these two text types.

Diachronic variation of war narratives was also analyzed by Solopova et al. (2023a), who looked at the evolution of pro-Kremlin propaganda within the first year of the full-scale invasion. Compared to the beginning of 2022, they found an increase in the use of the term *Kyiv Regime*, claims, assertive words, adverbs, and other high-modality words, as well as the mention of the West and negotiations on Russian Telegram at the start of 2023. In contrast, *special military operation*, negotiations, sanctions, genocide, fake news, and Belarus were discussed less frequently in the Russian state-run media in 2023 in comparison with 2022. In a similar vein, Burovova and Romanyshyn (2024) observed varying temporal dynamics of the dehumanization rhetoric, whose changes coincided with important events before or after the start of the full-scale invasion. In particular, they found that certain types of dehumanization began to rise shortly before the invasion and declined at its onset, suggesting that dehumanization plays a preparatory role in legitimizing acts of genocide. These developments reveal important shifts in propagandistic narratives over time.

3 Data

For our pilot study, we are using the WarMM-2022 corpus (Alyukov et al., 2023), which is a collection of 1.7M posts on the Russo-Ukrainian war. Our motivation for choosing it is two-fold. Firstly, the corpus includes two text types targeting different audiences. The state-controlled mass media include 24.4M tokens of press and 1.7M tokens of TV transcripts, and their style is more formal. Social media posts consist of 268.4M tokens and are characterized by limited governmental control and less formal register. Whereas the former text type promotes state-imposed propaganda, the latter includes both publications by regime supporters and anti-government voices. These differences between the text types would allow us to study linguistic variation in divergent contexts. Secondly, the WarMM-2022 corpus is diachronic: the state media subcorpus covers the period from February until September 2022, whereas the posts on social media span from July to September 2022, which is useful for analyzing linguistic change over time.

4 Proposed Methodology

4.1 Measuring Divergence Between Media Types and Time

To measure how much the two text types of the WarMM-2022 corpus (state vs. social media) differ from each other and by which linguistic features, we use KLD (Kullback and Leibler, 1951). KLD is used to quantify the divergence between two probability distributions of linguistic features. Using the whole lexicon to depict the lexical level, as well as vocabulary subsets such as content words, part-of-speech tags, etc. to represent more abstract linguistic levels, we implement KLD on the two probability distributions: *State* (for state media) and *Social* (for social media).

We apply KLD to the WarMM-2022 corpus comparing probability distributions of text types and diachronically by using various linguistic features. The probability distribution is based on the unigram probability of a linguistic feature (e.g., a word) to occur in one or the other sub-corpus. In general, KLD measures the number of additional bits needed to encode one distribution with the other distribution. For example, KLD of *State* given *Social* is measured as:

$$D(\text{State} \parallel \text{Social}) = \sum_i p(\text{feature}_i \mid \text{State}) \log_2 \frac{p(\text{feature}_i \mid \text{State})}{p(\text{feature}_i \mid \text{Social})}$$

In this equation, $p(\text{feature}_i \mid \text{State})$ stands for the *i*-th linguistic feature in the *State* distribution and $p(\text{feature}_i \mid \text{Social})$ for the *i*-th feature in the *Social* distribution. As the overall divergence is a sum of the individual divergences of each feature, we get to know how much linguistic features contribute to divergence-revealing features that are disproportionately emphasized in one corpus relative to the other. In comparison to using mere frequency, with KLD we are also able to detect low-frequency but distinctive features of variation (cf. Degaetano-Ortlieb et al., 2021).

Previous studies have demonstrated KLD's utility in analyzing linguistic variation and change, enabling comparisons of linguistic features across registers (Fankhauser et al., 2014), styles (Hughes et al., 2012), social variables and combinations of these (Degaetano-Ortlieb et al., 2021) as well as linguistic differences in criminal trials (Klingenstein et al., 2014), and word frequency shifts across languages (Bochkarev et al., 2014).

By applying KLD to the WarMM-2022 corpus, we expect to see some differences between the two text types. Furthermore, KLD can be applied to investigate diachronic linguistic change. For instance, Degaetano-Ortlieb and Teich (2022), who explored the evolution of scientific English, showed that external factors such as new scientific discoveries influenced the vocabulary of the language, which was reflected by peaks in KLD. Therefore, this method can help us study how linguistic strategies of propaganda shifted over time. Overall, KLD will offer us a nuanced perspective on how narratives adapt to audience and platform constraints and evolve diachronically.

We argue that KLD offers interpretability advantages over more opaque machine learning methods in detecting divergent language use which can be mapped to propaganda techniques and provides a deeper understanding of how these techniques are linguistically construed and evolve over time. While neural models achieve high accuracy, their reliance on labeled data and challenges in domain transfer limits adaptability to novel datasets and hardly allows analyzing linguistic choices. In contrast, KLD's reliance on probability distributions aligns with the FAIR (Findable, Accessible, Interoperable and Reusable) principles, enabling reproducibility and transparency in computational linguistics research.

4.2 Surprisal

According to information theory, information is defined as unpredictability within a given context, often described as surprisal (Hale, 2001). Surprisal quantifies the degree of unexpectedness of a unit, such as a word in a sequence, based on its preceding context. It is expressed in bits, with higher values indicating greater unpredictability and lower values reflecting higher predictability. For instance, in the context of Russian propaganda, the surprisal of the word *operation* given *special military* would be measured as follows:

$$S(\text{operation}) = -\log_2 p(\text{operation} \mid \text{special military})$$

Since the term *special military operation* was introduced at the beginning of the full-scale invasion, we hypothesize that the surprisal of the word *operation* in the given context will be higher in February 2022, but it will drop in the following months, indicating the conventionalized usage of this term in state-imposed propaganda.

Surprisal has been applied in a number of studies on language change, e.g., to trace the evolution of scientific English (Teich et al., 2021; Degaetano-Ortlieb and Teich, 2022; Steuer et al., 2024) and to analyze linguistic variation in Early Modern English (Gergel et al., 2017), suggesting the validity of this method for this task.

4.3 Word Embeddings

In distributional semantics, words are represented as vectors in a space based on their co-occurrence patterns, allowing their representations to be compared across different periods (Hamilton et al., 2016). Word embeddings are a commonly used method to study semantic change (Hamilton et al., 2016; Bizzoni et al., 2020; Giulianelli et al., 2020; Montariol et al., 2021). It has also been applied to examine linguistic variation in political and social contexts (Azarbonyad et al., 2017; Garg et al., 2018; Wevers, 2019; Marjanen et al., 2019; Tripodi et al., 2019), including the Russo-Ukrainian war (Ustyianovych and Barbosa, 2024).

We also believe that word embeddings are useful for investigating semantic shifts that might reveal propaganda strategies. For example, Russia has been using the narrative of "Nazi Ukraine" to justify its invasion, claiming that the current Ukrainian government commits genocide against Russians (Fortuin, 2022). By visualizing the word *Nazi* in the semantic space, we anticipate that it will be closer to words related to Nazi Germany and World War II before or at the very beginning of the full-scale invasion, but afterward, this word will probably be more strongly associated with Ukraine, its government and people.

5 Preliminary and Expected Results

Drawing from Alyukov et al.'s (2024) work, we anticipate finding differences and/or similarities between state and social media, as well as tracing the evolution of Russian propaganda over time by applying the above-mentioned methods. This would allow us to study linguistic change both diachronically and across media types. We might also gain insights into the interplay between the text types. Specifically, narratives that originate in the official media might influence social media discourse. This could happen through the repetition and reinforcement of state-approved messages by pro-government social media users and the dissemination of mainstream propaganda by bots or paid commentators (Alyukov et al., 2023).

As the first step of our pilot study, we conducted some experiments by applying KLD to a small subset of the WarMM-2022 corpus, and we could already see some of the results we expected. Specifically, we compared the usage of nouns in social and state media posts from July 30 and 31, 2022 (approx. 2 million nouns). While the direct term война[2] is the most distinctive noun for social media, state media mostly uses opaque euphemisms like спецоперация[3], ситуация[4] and демилитаризация[5]. This is in line with previous studies, which showed a clear distinction between the words denoting the war used in propagandistic or non-propagandistic texts (Solopova et al., 2023a), pro-Russian or pro-Ukrainian news (Ustyianovych and Barbosa, 2024) and state-affiliated or independent outlets (Park et al., 2022). Another interesting observation is that there is a high contribution of words such as правда[6] and факт[7] to the language of social media, as opposed to that of press and TV. This could indicate the government's efforts to employ the disinformation frame, which, as was shown by Alyukov et al. (2024), is predominant on social media as a means to discourage users from seeking out other sources of news that contradict the pro-Kremlin narratives.

In the future, we plan to do a more comprehensive KLD analysis comparing state and social media posts from the whole WarMM-2022 corpus, as well as studying diachronic linguistic change in the context of propaganda and applying other methods mentioned in Section 4, namely surprisal and word embeddings. As a more ambitious goal, we hope that our work will contribute towards combating disinformation, specifically in war contexts. In terms of practical applications of our methodology, we expect it could be employed in studying other political or historical events.

[2] [voyna] — *war*.

[3] [spetsoperatsiya] — an abbreviation from "special [military] operation".

[4] [situatsiya] — *situation*, as in "situation in Ukraine".

[5] [demilitarizatsiya] — *demilitarization*, a term used by the Russian government to justify its invasion of Ukraine.

[6] [pravda] — *truth*.

[7] [fakt] — *fact*.

6 Conclusion

This work underscores the potential of open, transparent methodologies to democratize access to knowledge and foster resilience against disinformation. By leveraging interpretable methods such as KLD, surprisal, and word embeddings, our study aims to provide a robust framework for detecting and analyzing propaganda strategies in Russian state-controlled and social media.

By systematically examining linguistic change both across text types and over time, our study contributes to a deeper understanding of propaganda mechanisms and their societal implications. It also highlights the importance of combining interpretability and reproducibility in computational linguistics research, particularly in political contexts.

In addition to its academic contributions, this research has significant practical implications. It equips researchers, policymakers, and media analysts with tools to critically examine information landscapes and identify deliberate attempts to influence public opinion. Ultimately, by demonstrating how linguistic change can be an indicator of propagandistic strategies, we aim to advance efforts to counteract disinformation and enhance media literacy.

7 Future Work

We use KLD, surprisal, and word embeddings for a preliminary analysis of propagandistic narratives, which would reveal certain linguistic features that drive change in this domain. In future studies, we might also use graph neural networks, as they have been shown to provide promising and interpretable results in semantic change (Chen et al., 2023) and disinformation detection (Panayotov et al., 2022). We also plan to consider a combination of these methods as a complementary means to transformer-based approaches, specifically, by using machine learning methods to detect propaganda. Possible directions include classifying news into fake or real (as in Solopova et al., 2024), pro- or anti-regime (similar to Ustyianovych and Barbosa, 2024), and according to propaganda frames (following the work by Alyukov et al., 2024). Potentially, we might extend our research and analyze not only linguistic change of propaganda across time and text types, but also how narratives about the war differ between languages such as Russian, Ukrainian, and English, representing another dimension of linguistic variation. Finally, we could also investigate pro-Kremlin propaganda that preceded the full-scale invasion of Ukraine in 2022, e.g., since the start of the war in Donbas in 2014.

8 Challenges and Limitations

Propaganda detection is a complicated task not only for computers but even for humans, as many people fall victim to information manipulation in today's enormous influx of news in media. First of all, there is no single definition of propaganda in general or a single framework for detecting it with NLP techniques. We aim to address these challenges by providing a working definition of propaganda based on previous research in the field, as well as proposing a thorough methodology for tackling it computationally. Secondly, propaganda identification can be biased, as it depends on the political stance of the researcher. To eliminate any possible bias, we again plan to rely on related work and use data-driven approaches to detect propaganda, which were described in Section 4.

9 Ethical Considerations

Propaganda and war are highly sensitive topics. However, since we are using an already available corpus of news on the Russo-Ukrainian war (WarMM-2022), our research does not involve human participants (e.g., to annotate texts as propaganda or not), thus eliminating any ethical concerns in this regard. In the future, we might also use other datasets that were employed in previous research on the topic of the Russian invasion of Ukraine.

10 Acknowledgments

Funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

We sincerely thank Maria Kunilovskaya for providing the WarMM-2022 dataset, which has been invaluable for our research.

References

Roberto Adriani. 2019. The Evolution of Fake News and the Abuse of Emerging Technologies. *European Journal of Social Sciences*, 2(1):18–24.

Maxim Alyukov, Maria Kunilovskaya, and Andrei Semenov. 2023. Wartime Media Monitor (WarMM-2022): A Study of Information Manipulation on Russian Social Media during the Russia-Ukraine War. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 152–161, Dubrovnik, Croatia. Association for Computational Linguistics.

Maxim Alyukov, Maria Kunilovskaya, and Andrei Semenov. 2024. Confuse and Normalise: Authoritarian Propaganda in a High-Choice Media Environment and Russia's Invasion of Ukraine. In Paul Goode, editor, *Russian Propaganda Today: Challenges, Effectiveness, and Resistance*, page in print. University of Michigan press, University of Manchester Press.

Hosein Azarbonyad, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. 2017. Words are Malleable: Computing Semantic Shifts in Political and Media Discourse. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 1509–1518, New York, NY, USA. Association for Computing Machinery.

Yuri Bizzoni, Stefania Degaetano-Ortlieb, Peter Fankhauser, and Elke Teich. 2020. Linguistic Variation and Change in 250 Years of English Scientific Writing: A Data-Driven Approach. *Frontiers in Artificial Intelligence*, 3:73.

Vladimir Bochkarev, Valery D. Solovyev, and Søren Wichmann. 2014. Universals versus Historical Contingencies in Lexical Evolution. *Journal of The Royal Society Interface*, 11(101):20140841.

Kateryna Burovova and Mariana Romanyshyn. 2024. Computational Analysis of Dehumanization of Ukrainians on Russian Social Media. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, pages 28–39, St. Julians, Malta. Association for Computational Linguistics.

Jing Chen, Emmanuele Chersoni, Dominik Schlechtweg, Jelena Prokic, and Chu-Ren Huang. 2023. ChiWUG: A Graph-based Evaluation Dataset for Chinese Lexical Semantic Change Detection. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, pages 93–99, Singapore. Association for Computational Linguistics.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2021. A Survey on Computational Propaganda Detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI'20, pages 4826–4832, Yokohama, Yokohama, Japan.

Stefania Degaetano-Ortlieb, Tanja Säily, and Yuri Bizzoni. 2021. Registerial Adaptation vs. Innovation

Across Situational Contexts: 18th Century Women in Transition. *Frontiers in Artificial Intelligence*, 4:609970.

Stefania Degaetano-Ortlieb and Elke Teich. 2022. Toward an Optimal Code for Communication: The Case of Scientific English. *Corpus Linguistics and Linguistic Theory*, 18(1):175–207.

Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

Peter Fankhauser, Jörg Knappen, and Elke Teich. 2014. Exploring and Visualizing Variation in Language Resources. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4125–4128, Reykjavik, Iceland. European Language Resources Association (ELRA).

Egbert Fortuin. 2022. "Ukraine Commits Genocide on Russians": The Term "Genocide" in Russian Propaganda. *Russian Linguistics*, 46(3):313–347.

Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644.

Remus Gergel, Martin Kopf-Giammanco, and Julia Masloh. 2017. Surprisal and Satisfaction: Towards an Information-theoretic Characterization of Presuppositions with a Diachronic Application. In *IWCS 2017 - 12th International Conference on Computational Semantics - Short Papers, Montpellier, France, September 19 - 22, 2017*. The Association for Computer Linguistics.

Mario Giulianelli, Marco Del Tredici, and Raquel Fernández. 2020. Analysing Lexical Semantic Change with Contextualised Word Representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3960–3973, Online. Association for Computational Linguistics.

John Hale. 2001. A Probabilistic Earley Parser as a Psycholinguistic Model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, NAACL '01, pages 1–8, USA. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Vitalij Hein. 2023. *Propaganda Detection in Russian and American News Coverage about the War in Ukraine through Text Classification*. Thesis, Technische Universität Wien.

James M. Hughes, Nicholas J. Foti, David C. Krakauer, and Daniel N. Rockmore. 2012. Quantitative Patterns of Stylistic Influence in the Evolution of Literature. *Proceedings of the National Academy of Sciences*, 109(20):7682–7686.

Sara Klingenstein, Tim Hitchcock, and Simon DeDeo. 2014. The Civilizing Process in London's Old Bailey. *Proceedings of the National Academy of Sciences*, 111(26):9419–9424.

Solomon Kullback and Richard A. Leibler. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Jani Marjanen, Lidia Pivovarova, Elaine Zosa, and Jussi Kurunmäki. 2019. Clustering Ideological Terms in Historical Newspaper Data with Diachronic Word Embeddings: HistoInformatics2019 - the 5th International Workshop on Computational History. *HistoInformatics 2019 : International Workshop on Computational History 2019*, pages 21–29.

Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*.

Syrielle Montariol, Matej Martinc, and Lidia Pivovarova. 2021. Scalable and Interpretable Semantic Change Detection. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4642–4652, Online. Association for Computational Linguistics.

Panayot Panayotov, Utsav Shukla, Husrev Taha Sencar, Mohamed Nabeel, and Preslav Nakov. 2022. GREENER: Graph Neural Networks for News Media Profiling. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7470–7480, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Chan Young Park, Julia Mendelsohn, Anjalie Field, and Yulia Tsvetkov. 2022. Challenges and Opportunities in Information Manipulation Detection: An Examination of Wartime Russian Media. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5209–5235, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Veronika Solopova, Christoph Benzmüller, and Tim Landgraf. 2023a. The Evolution of Pro-Kremlin Propaganda From a Machine Learning and Linguistics Perspective. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 40–48, Dubrovnik, Croatia. Association for Computational Linguistics.

Veronika Solopova, Viktoriia Herman, Christoph Benzmüller, and Tim Landgraf. 2024. Check News in One Click: NLP-Empowered Pro-Kremlin Propaganda Detection. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 44–51, St. Julians, Malta. Association for Computational Linguistics.

Veronika Solopova, Oana-Iuliana Popescu, Christoph Benzmüller, and Tim Landgraf. 2023b. Automated Multilingual Detection of Pro-Kremlin Propaganda in Newspapers and Telegram Posts. *Datenbank-Spektrum*, 23(1):5–14.

Julius Steuer, Marie-Pauline Krielke, Stefan Fischer, Stefania Degaetano-Ortlieb, Marius Mosbach, and Dietrich Klakow. 2024. Modeling Diachronic Change in English Scientific Writing over 300+ Years with Transformer-based Language Model Surprisal. In *Proceedings of the 17th Workshop on Building and Using Comparable Corpora (BUCC) @ LREC-COLING 2024*, pages 12–23, Torino, Italia. ELRA and ICCL.

Elke Teich, Peter Fankhauser, Stefania Degaetano-Ortlieb, and Yuri Bizzoni. 2021. Less is More/More Diverse: On The Communicative Utility of Linguistic Conventionalization. *Frontiers in Communication*, 5. Publisher: Frontiers.

Rocco Tripodi, Massimo Warglien, Simon Levis Sullam, and Deborah Paci. 2019. Tracing Antisemitic Language Through Diachronic Embedding Projections: France 1789-1914. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 115–125, Florence, Italy. Association for Computational Linguistics.

Taras Ustyianovych and Denilson Barbosa. 2024. Instant Messaging Platforms News Multi-Task Classification for Stance, Sentiment, and Discrimination Detection. In *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING 2024*, pages 30–40, Torino, Italia. ELRA and ICCL.

Natalia Vanetik, Marina Litvak, Egor Reviakin, and Margarita Tyamanova. 2023. Propaganda Detection in Russian Telegram Posts in the Scope of the Russian Invasion of Ukraine. In *Proceedings of the Conference Recent Advances in Natural Language Processing - Large Language Models for Natural Language Processings*, pages 1162–1170. INCOMA Ltd., Shoumen, BULGARIA.

Melvin Wevers. 2019. Using Word Embeddings to Examine Gender Bias in Dutch Newspapers, 1950-1990. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 92–97, Florence, Italy. Association for Computational Linguistics.