From *War* to *Special Military Operation*: Interpretable Detection of Linguistic Propaganda Framing in Russian Media

Anastasiia Vestel, Stefania Degaetano-Ortlieb

Saarland University
anastasiia.vestel@uni-saarland.de, s.degaetano@mx.uni-saarland.de

Abstract

This study investigates linguistic differences between Russian state-controlled media and social media reporting on the Russo-Ukrainian war, with a focus on propaganda frames and strategies. Using the WarMM-2022 corpus, we apply Kullback-Leibler Divergence (KLD) to identify distinctive linguistic features across media types. The use of KLD ensures the interpretability of the analysis, offering clear insights into the varying propaganda techniques, such as demobilization efforts on state media and more mobilizational rhetoric on social media. This method aligns with the principles of computational reproducibility, providing a transparent framework for linguistic analysis in political discourse. Our findings confirm previous research on Russian propaganda, underscoring the role of language in shaping propaganda strategies and highlighting the value of KLD as a method promoting accessibility and reusability in computational social science.

Introduction

The Russo-Ukrainian war has heightened concerns about media manipulation, leading to an increased focus on propaganda detection. Since linguistic variation is influenced by external social and political factors, we analyze semantic change, aiming at improving the automatic detection of propaganda. Research has shown that different political viewpoints contribute to linguistic shifts (Azarbonyad et al. 2017; Ustyianovych and Barbosa 2024). For example, Alyukov, Kunilovskaya, and Semenov (2024) found that propaganda frames (see Entman (1993) for a definition of framing) in Russian media varied between traditional state-controlled outlets and social media, with the former targeting passive audiences and the latter engaging users seeking alternative information. This suggests a potential difference in political alignment between audiences of these platforms. Our study focuses on linguistic differences across these two media types to draw possible connections to propaganda strategies. Our overarching research question is whether state and social media reveal linguistic differences and, by doing so, reflect differences in the propaganda frames and strategies employed in these media types. The motivation for addressing this issue lies in the fact that one should consider this variation when designing automatic propaganda detection systems and tailor them to the specific media type. In this paper, we move a step closer towards this goal, inspecting linguistic differences at the lexical level.

While natural language processing (NLP) methods for propaganda detection often rely on transformer-based models that require annotated data and lack transparency (Da San Martino et al. 2021; Park et al. 2022), we propose using Kullback-Leibler Divergence (KLD; Kullback and Leibler 1951) as an alternative. KLD has been widely applied to study linguistic variation in different scenarios (Hughes et al. 2012; Bochkarev, Solovyev, and Wichmann 2014; Fankhauser, Knappen, and Teich 2014; Klingenstein, Hitchcock, and DeDeo 2014; Degaetano-Ortlieb and Teich 2022) and offers interpretable and reproducible insights into language shifts, allowing us to identify distinctive linguistic features that signal propagandistic tendencies.

Related Work

NLP has been widely applied to investigate Russian propaganda following the full-scale invasion of Ukraine. One such effort, led by Alyukov, Kunilovskaya, and Semenov (2023), resulted in the creation of the Wartime Media Monitor (WarMM-2022) corpus, which comprises two sub-corpora: state-controlled media and social media. Building on this dataset, Alyukov, Kunilovskaya, and Semenov (2024) examined how propaganda frames differ across these media types. Their analysis revealed that traditional media, including television and newspapers, relied on normalization, a more demobilizational frame that emphasizes the return to normal life by downplaying the war's impact on society. Social media, on the other hand, predominantly employed the disinformation frame that labels pro-Ukrainian and pro-Western news as fake, which is a more mobilizational strategy. The distinction between these media types underscores their different functions: traditional outlets seek to maintain support among regime loyalists, while social media aim to activate and engage their users.

Other research has similarly examined linguistic patterns in Russian wartime media. Park et al. (2022) investigated differences between state-affiliated and independent news sources on social media platforms rather than comparing them with official media. Their study found that independent outlets consistently used the term *war*, while state-affiliated

sources preferred the euphemism *special military operation*, a pattern also noted by Ustyianovych and Barbosa (2024) in pro-Ukrainian versus pro-Russian Telegram channels. Additionally, Solopova, Benzmüller, and Landgraf (2023) trained machine learning models to detect pro-Kremlin rhetoric, revealing that the presence of the term *war* strongly correlated with a non-propagandistic classification, likely due to official censorship and the preference for state-approved terminology. Beyond individual word choices, Park et al. (2022) also highlighted broader framing differences between VKontakte and Twitter, supporting Azarbonyad et al. (2017)'s argument that semantic shifts occur not only over time but also across different political and social contexts. These findings reinforce the assumption that language use diverges between traditional and social media, reflecting the ideological positioning of their respective audiences.

Data

Our study uses the WarMM-2022 corpus (Alyukov, Kunilovskaya, and Semenov 2023), a dataset comprising 1.7 million posts related to the Russo-Ukrainian war. This corpus contains two distinct text types aimed at different audiences. The state-controlled media section includes 24.4 million tokens of press articles and 1.7 million tokens of TV transcripts, both exhibiting a more formal style. In contrast, the social media subcorpus, consisting of 268.4 million tokens, features a less formal register with limited government regulation. While state media primarily disseminate official propaganda, social media content represents a mix of regime-supporting narratives and opposition viewpoints, enabling an analysis of linguistic variation across different ideological contexts. The state media texts cover the period from February to September 2022, while social media posts span from July to September 2022.

Methodology

Measuring Variation Between Text Types

To assess the linguistic differences between state and social media in the WarMM-2022 corpus, we employ KLD (Kullback and Leibler 1951). KLD quantifies divergence between two probability distributions, helping identify features that distinguish these text types. We apply KLD to content words, comparing the probability distributions of linguistic features in state-controlled (*State*) and social media (*Social*) texts. KLD measures the additional bits needed to encode one distribution using another, revealing the extent to which specific features contribute to linguistic divergence. Unlike simple frequency-based methods, KLD captures even low-frequency yet distinctive linguistic variations (Degaetano-Ortlieb, Säily, and Bizzoni 2021). KLD calculation for *State* given *Social* is formalized as:

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

This methodological approach supports transparency and interpretability by allowing a direct inspection of linguistic features that drive divergence between text types. In contrast to opaque neural models, KLD provides reproducible outputs based on clearly defined probability distributions. Although neural models tend to perform well in terms of classification accuracy, they rely heavily on annotated datasets and often struggle to generalize across different domains. In comparison, KLD adheres to the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, and facilitates verification and extension of results in computational social science.

Data Pre-Processing

As almost 50% of the texts in the social media subcorpus are duplicates (Alyukov, Kunilovskaya, and Semenov 2023), we decided to remove them before applying KLD. This decision was motivated by the fact that word frequencies can significantly affect KLD results, which would show bias towards duplicated content rather than reflecting the significant linguistic features of individual texts. Even though one could argue that keeping the duplicates would allow for a more representative analysis, we believe that deleting them would provide for a better comparison between the language of social and state media. Note that in a follow-up step, we are considering keeping duplicates to quantify differences in propaganda frames and strategies. Following the approach of Alyukov, Kunilovskaya, and Semenov (2023), duplicates were detected by matching the first 20 tokens of a text, and only the texts with the earliest publication date were kept as they are the original posts.

Next, we limited the scope of linguistic features only to content words, namely nouns, verbs, adjectives, adverbs, and proper nouns, as these parts of speech are the most likely to contribute significantly to the divergence between text types. Also, as Russian is a highly inflected language, we included only the lemmas of each word rather than all word forms. Then, we selected 15 dates in July, August, and September 2022 (as only these three months contain texts from both state and social media) for our analysis, on which important events of the Russo-Ukrainian war took place[1]. We then queried the corpus to match the words with content POS tags and the specified dates. The resulting data, which amounted to approx. 21.8 million words, was divided into five subsamples, each containing texts from three dates, with an average of 4.4 million words per subsample. After this, the hapax legomena of each of the five subsamples of the corpus were removed to reduce noise. Finally, we run the KLD script (Degaetano-Ortlieb and Teich 2022), which is publicly available on GitHub, on each of the pre-processed subsamples of the data.

Results

Overall KLD Values

Table 1 shows overall KLD values for social and state media, as well as the difference between them in each subsample. It also includes average KLD values of both media types

[1]The dates were chosen based on the information from the timeline of the war by Fakty ICTV, a Ukrainian news agency. For a full list of the selected dates and the description of the corresponding events, see Appendix A.

Subsample #	Media type	KLD value	KLD difference
1	Social	0.863	0.386
	State	0.477	
2	Social	0.761	0.297
	State	0.464	
3	Social	0.783	0.303
	State	0.48	
4	Social	0.787	0.281
	State	0.506	
5	Social	0.802	0.3
	State	0.502	
Average	Social	0.799	0.313
	State	0.486	

Table 1: Overall KLD values for different media types in the five subsamples.

Figure 1: Top 50 words with the highest KLD values for social media (left) and state media (right). Size accounts for KLD values and shade indicates frequencies from higher (black) to lower (light gray).

and the difference between them. We can see that KLD values for social media are always higher than for state media: on average, 0.313 more bits are needed to model the former subcorpus with the latter. This could mean that the social media subcorpus has a larger vocabulary size compared to state media. An explanation for this could be that social media texts include a wider variety of registers (e.g., both formal and informal styles), while official media are mostly characterized by formal language.

Comparison of Media Types

To assess the correlation of KLD results between the five subsamples, we calculated the percentage of overlapping words among the 50 most distinctive words for both media types. There are 52.48% and 64.84% overlapping words for social and state media, respectively. This suggests that KLD is quite robust across different subsamples of the data. Therefore, we calculated the average KLD values of the words and selected the top 50 words with the highest KLD values for a more in-depth comparative analysis of text types.

Figure 1 shows word clouds of linguistic features with the highest average KLD values for social and state media among the subsamples (for more details, see Appendix C,

Topic	Social media	State media
(dis)information	6%	0%
events	0%	8%
geography	6%	26%
life	4%	0%
nation & ethnicity	4%	4%
necessity	4%	0%
news	0%	4%
noise	24%	6%
other	24%	16%
political figures	0%	8%
slogans & ideology	4%	0%
statements	0%	16%
time	2%	6%
USSR	4%	0%
war	18%	6%

Table 2: Percentages of words with the top highest KLD values for social and state media grouped by topic.

which contains the list of these words ranked by their KLD values). To conduct a more in-depth qualitative analysis, we grouped the words from the feature list ranked by KLD values into topics (Table 2 indicates the percentages of words in each topic for both media types). Given that KLD allows us to inspect which features contribute most to the divergence between the two media types (cf. Degaetano-Ortlieb and Teich 2022), we can observe some interesting patterns between state and social media. For instance, words related to geographical entities are much more distinctive for newspapers and TV (26%), especially the names of Ukrainian territories that are or were occupied by Russia at the time, such as *Donbas*, *DPR* (an abbreviation of *Donetsk People's Republic*), *LPR* (an abbreviation of *Luhansk People's Republic*)[2], as well as the words used in the full names of these and other territories invaded by Russia, namely *Donetsk*, *Luhansk*, *Zaporizhzhia* and *Kherson [Oblasts]*; *republic* and *oblast*. This could indicate the government's efforts to emphasize their military successes and establish authority over these regions. This propaganda strategy could be related to the normalization frame, which was found to be prominent in state media (Alyukov, Kunilovskaya, and Semenov 2024): Russia claims the regions of Ukraine it occupied as inherently Russian territories.

Another interesting distinction between the language of the two media types is that the words *Russia* and *Russian Federation* are more prominent for state media, while *Ukraine* (both capitalized and in lowercase) is more characteristic of social media. This might point to differences in propaganda framing: on the one hand, state media focuses on Russia, while also trying to avoid naming the opposing side, thus creating an illusion that the war is far away and does not affect the life in Russia; social media, on the other

[2]DPR and LPR are self-proclaimed, Russian-backed territories in eastern Ukraine that were recognized by Russia in 2022 and later used to justify the annexation of parts of Donetsk and Luhansk Oblasts.

hand, is clear in naming the enemy, relying on a more mobilizational approach, which is in line with the findings by Alyukov, Kunilovskaya, and Semenov (2024). However, one should not rule out the possibility of anti-government users on social media who speak up against the war in Ukraine, which could also increase the frequency of the name of the country, thus affecting KLD values.

Verbs that indicate statements, such as *to state*, *to communicate*, *to point out*, *to emphasize*, *to add*, etc., are highly characteristic of official media (16%), which makes sense given the specificities of this text type: politicians are often quoted on TV and in newspaper articles. For the same reason, the names and positions of political figures are also distinctive for the language of state media (8%): *president*, *head* (as in "head of state"), *Vladimir* (Putin's first name), *minister*, and so on. The higher contribution of the president's first name (Vladimir) as opposed to his last name (Putin) could be explained by the fact that the more formal register of state media uses the combination of the politicians' first and last names rather than only last names, as it is more common in the informal social media texts.

It is worth noting that words describing certain political events, such as elections, are distinctive for state media (8%), including *referendum*, *voting*, and *to vote*. They are mostly mentioned in the context of the sham referendums on Russia's annexation of Ukraine's occupied territories conducted in late September 2022, as they often co-occur with the names of these regions in the corpus (e.g., *DPR, LPR, Zaporizhzhya Oblast and Kherson Oblast are preparing to hold a referendum on joining Russia*). Their high contribution to the state media language could indicate the promotion of the narrative that the local population of the Ukrainian territories occupied by Russia presumably wants to become part of Russia. This aligns with the propaganda about territorial control attested by the above-mentioned finding regarding the names of geographical regions.

The language of social media is highly characterized by words related to war (18%), e.g., *to fight*, *front*, *army*, *soldier*, an abbreviation of *Armed Forces of the Russian Federation*, *to kill*, etc. Only three words that belong to this topic are among those with the highest KLD values for state media: *military*, *situation* and *special*, all of which are (parts of) euphemisms that refer to Russia's invasion of Ukraine. As expected, the direct term *war* has the highest contribution to the language of social media. This variation in terminology to refer to war was attested in previous research, although in slightly different contrasting contexts (cf. Park et al. 2022; Solopova, Benzmüller, and Landgraf 2023; Ustyianovych and Barbosa 2024). Notably, this does not necessarily mean that social media users mostly include anti-government or anti-war voices, as propaganda is still highly widespread in this media type. Instead, the direct term *war* and other words related to this topic might indicate that pro-government users are trying to mobilize the Russian population and encourage it to fight against Ukraine, a strategy which is more common on social media (Alyukov, Kunilovskaya, and Semenov 2024). On the other hand, the euphemism *special military operation* also appears in the form of an abbreviation as one of the most distinctive words for social media. This finding is not surprising, as abbreviations are more common in informal text types compared to the official language of state media.

In addition, the high contribution of words such as *propaganda*, *fable* and *truth* to the language of social media, which together constitute 6% of words in this text type, might point to the prevalent usage of the disinformation frame in this text type, as influencers are trying to convince the users of which news should be trusted and which should be discarded as pro-Ukrainian or pro-Western propaganda. This is in line with the findings by Alyukov, Kunilovskaya, and Semenov (2024), which suggest that the disinformation frame is typical of social media as opposed to state media.

An interesting distinction between state and social media is that two contrasting words that describe Russians are characteristic of each text type. In state media, the adjective that encompasses the Russian population overall, which would be transcribed as [rossiyskiy], has a higher contribution. On the other hand, the term that refers only to ethnic Russians, [russkiy], excluding the numerous minorities that live in the country, is more prominent in social media. This variation could indicate different propaganda strategies in the media: social media users appeal to the idea of ethnic Russians, who need to "defend" their "Slavic brothers" (i.e., the Russian-speaking population in Ukraine), while state media emphasize the unity of all the republics in Russia. More specifically, the word [russkiy] is very often encountered in the phrase *the Russian world*, a concept used to promote the idea of a shared cultural, spiritual, and political space uniting ethnic Russians and Russian speakers, often to justify Russia's influence over post-Soviet states. This expression, which also has a mobilizational character, is encountered about ten times more frequently on social media than on state media.

Homoglyphs as a Potential Propaganda Strategy

The "noise" category of the features ranked by KLD contains many words that do not necessarily contribute to the divergence between the media types because they appeared accidentally in the corpus, e.g., *text* and *image* on social media (as it is often encountered in the phrase *Text on image*, which accompanies a certain picture) or an abbreviation for *TV host* on state media, since this word is frequently used in TV transcripts, as in *HOST:* [Quote]. However, this category also includes some unexpected results we found by looking at the most distinctive words for social media. Specifically, some of them were quite rare, or it was not clear why they appeared among the most distinctive words for this text type. On the other hand, they had rather low frequencies compared to the rest in the word cloud (as they were in light gray). Examples include a non-existent word, [smolodu], due to an error resulting from automated POS tagging (the original word means *from a young age*); an archaic word for *again*; and the female name *Eva*. In order to find an explanation for the high KLD values of these words, we queried the corpus to see the context surrounding them. To our surprise, the reason behind the high KLD values for these words was that some duplicates were not removed during the automated deduplication process. This is in turn explained by the fact that many

words contained homoglyphs, i.e., characters that look visually similar or identical but originate from different scripts (these could be, for example, Cyrillic letters replaced with Latin characters).

What is more interesting, the words with homoglyphs differ among the duplicated posts. If the words containing homoglyphs were the same, the deduplication script would delete them as well; however, since homoglyphs appear sporadically, the script failed to detect these texts as duplicates. As a result, many of the words from these posts had high frequencies in social media compared to the state media, which explains their high KLD values. An example of a duplicated post with varying homoglyphs is provided in Appendix B.

Given that the number of duplicated posts that were not removed automatically is high enough to affect the KLD results, we presume that inserting Latin characters instead of Cyrillic letters in different words of the same texts was not an incidental error, but rather an information manipulation technique. Homoglyphs are often used to avoid plagiarism detection (Alvi, Stevenson, and Clough 2017) and for other malicious purposes, such as phishing attacks (Lee, Yam, and Chan 2020). As it is common for social media platforms to have automatic filters that detect and block duplicate content to prevent spamming and the spread of disinformation, some users might employ homoglyphs to bypass these filters. It is unclear whether this is done manually or through an automated process such as a computer program that inserts homoglyphs in various words of the same text. Either way, this seems to be a strategy to disseminate propaganda by circumventing filters on social media platforms.

Preliminary Diachronic Analysis

Since we selected 15 dates in July, August, and September 2022 and split the texts into five parts, we decided to analyze the differences between KLD results for each subsample to see whether specific events correlate with shifts in the words that contribute most significantly to the divergence. By comparing unique words for each subsample (i.e., those that were among the top 50 words with the highest KLD values), we could examine how temporal dynamics evolve differently in the two media types. For instance, for the first subsample of July 3, 27, and 29, the word *victory* was unique for social media, which could be tied to Russia's capture of Lysychansk in Luhansk Oblast; in state media, unique words include *Lavrov* (Russia's Minister of Foreign Affairs) and *negotiations*, which were probably discussed in the context of the Black Sea Grain Initiative. For the second subsample, the word *battle* is distinctive on social media, unlike for other periods, which could be explained by the discussion of the Ukrainian counteroffensive in Kharkiv Oblast that began on September 6. The word *mobilization* is characteristic of the state media language in late September (the third subsample), which makes sense, since this is when a partial mobilization started in Russia. Not surprisingly, the word *to vote* and the words that make up *polling station* are distinctive for state media from September 25 to 27, as this is when the sham referendums were conducted on the occupied territories. Finally, *KharkivOblast* (likely used as a hashtag) is prominent on social media between September 28-30, during the last phase of the Kharkiv counteroffensive.

These findings already show us that KLD can capture not only divergence between text types but also differences between periods, since changes in the top KLD words are associated strongly with particular events of the war. However, for a more in-depth diachronic analysis, we plan to apply KLD to different periods, which would show us more meaningful distinctions.

Conclusion

Our study examined linguistic variation in Russian state-controlled and social media to understand their role in propagandistic communication during the Russo-Ukrainian war. Using KLD, we identified key lexical differences between the two media types, revealing distinct propaganda strategies. State-controlled media favored euphemistic terminology and references to geopolitical control, aligning with the normalization frame to demobilize the population. In contrast, social media exhibited a mobilizational approach, employing direct war-related vocabulary and disinformation framing. In addition, while state media promote the idea of unity of all Russian citizens, social media emphasize the importance of Russian ethnicity in the current war. We have also found that homoglyphs are sometimes employed to spread propaganda by duplicating content multiple times on social media. Finally, a preliminary temporal analysis shows that KLD is a robust measure not only for detecting text type variation but also tracing linguistic changes through time.

Overall, this work contributes to reproducible and interpretable social science research by employing transparent methods, well-documented data processing, and an analysis framework aligned with the FAIR principles. Theoretical insights from our findings can inform future studies on media influence and help policymakers, journalists, and platform moderators better understand and respond to evolving propaganda strategies in digital communication.

Limitations and Future Work

This study presents a preliminary analysis of linguistic divergence between text types in Russian media, focusing only on individual words. In the future, we plan to apply KLD to other linguistic features (e.g., parts of speech and syntactic patterns) to explore linguistic strategies of propaganda. Finally, even though our current work includes an exploratory temporal analysis of the Russian propaganda, we aim to apply KLD to different periods of the war to study the diachronic evolution of pro-Kremlin narratives.

Another limitation of our work is that during the analysis stage, we discovered that not all duplicated posts were removed from the social media subcorpus due to the presence of homoglyphs, which affected the KLD results. To address this issue, we plan to assess the scale of duplicates, delete them, and re-run the experiments on the cleaned version of the corpus. In addition, it would also be interesting to apply KLD to the raw corpus that contains duplicates, as the results would be more representative of the social media language, and compare them to our findings from the KLD analysis of the cleaned corpus.

Ethics Statement

The WarMM-2022 dataset was provided to us by the authors of the original paper (Alyukov, Kunilovskaya, and Semenov 2023), and the permission to use it was obtained from them. The metadata containing personally identifiable information was not used in our research. It is worth noting that the text data may include offensive content, such as dehumanizing language towards Ukrainians (Alyukov, Kunilovskaya, and Semenov 2024); however, none of it was provided in the paper for illustration purposes.

A note has to be made regarding the potential negative societal impact or misuse of our research, since it touches on propaganda, a highly sensitive and contentious topic. In their paper, Park et al. (2022) discuss three media effects models of propaganda: agenda setting, framing, and priming. Studying agenda setting means focusing on what topics are discussed in the media, whereas research on framing emphasizes how the communication of those topics can shape the audience's perception and interpretation of them (Scheufele and Tewksbury 2007). Finally, priming describes the influence that framing and agenda setting have on public opinion (Entman 2007). Park et al. (2022) argue that in contrast to agenda setting and framing, research on priming can be misused by malicious actors. In our study, we only investigated agenda setting and framing (e.g., by showing that geographical entities are distinctive for state media and discussing how the variation in the war terminology may be used to target different audiences on state vs. social media). Studying priming, on the other hand, would involve examining public engagement in the media with regard to certain propaganda strategies, which would give us an intuition as to what strategies are effective. Since this is not part of our research, we believe our study does not present any risks of negative societal impact or misuse.

To mitigate any other potential negative outcomes, the study uses existing non-personal textual data (as mentioned above), applies reproducible and interpretable methods, and documents the preprocessing steps and analytical setup to ensure transparency and responsible research practices.

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, as well as Sergei Bagdasarov, Stefan Fischer, and Luigi Talamo for offering technical support.

References

Alvi, F.; Stevenson, M.; and Clough, P. 2017. Plagiarism Detection in Texts Obfuscated with Homoglyphs. In Jose, J. M.; Hauff, C.; Altıngovde, I. S.; Song, D.; Albakour, D.; Watt, S.; and Tait, J., eds., *Advances in Information Retrieval*, 669–675. Cham: Springer International Publishing. ISBN 978-3-319-56608-5.

Alyukov, M.; Kunilovskaya, M.; and Semenov, A. 2023. Wartime Media Monitor (WarMM-2022): A Study of Information Manipulation on Russian Social Media during the Russia-Ukraine War. In Degaetano-Ortlieb, S.; Kazantseva, A.; Reiter, N.; and Szpakowicz, S., eds., *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 152–161. Dubrovnik, Croatia: Association for Computational Linguistics.

Alyukov, M.; Kunilovskaya, M.; and Semenov, A. 2024. Confuse and Normalise: Authoritarian Propaganda in a High-Choice Media Environment and Russia's Invasion of Ukraine. In Goode, P., ed., *Russian propaganda today: Challenges, effectiveness, and resistance*, in print. University of Michigan press, University of Manchester Press.

Azarbonyad, H.; Dehghani, M.; Beelen, K.; Arkut, A.; Marx, M.; and Kamps, J. 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, 1509–1518. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-4918-5.

Bochkarev, V.; Solovyev, V.; and Wichmann, S. 2014. Universals versus Historical Contingencies in Lexical Evolution. *Journal of The Royal Society Interface*, 11(101): 20140841.

Da San Martino, G.; Cresci, S.; Barrón-Cedeño, A.; Yu, S.; Di Pietro, R.; and Nakov, P. 2021. A Survey on Computational Propaganda Detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI'20, 4826–4832. Yokohama, Japan. ISBN 978-0-9992411-6-5.

Degaetano-Ortlieb, S.; Säily, T.; and Bizzoni, Y. 2021. Registerial Adaptation vs. Innovation Across Situational Contexts: 18th Century Women in Transition. *Frontiers in Artificial Intelligence*, 4: 609970.

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

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

Entman, R. M. 2007. Framing Bias: Media in the Distribution of Power. *Journal of Communication*, 57(1): 163–173.

Fankhauser, P.; Knappen, J.; and Teich, E. 2014. Exploring and Visualizing Variation in Language Resources. In Calzolari, N.; Choukri, K.; Declerck, T.; Loftsson, H.; Maegaard, B.; Mariani, J.; Moreno, A.; Odijk, J.; and Piperidis, S., eds., *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 4125–4128. Reykjavik, Iceland: European Language Resources Association (ELRA).

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

Klingenstein, S.; Hitchcock, T.; and DeDeo, S. 2014. The Civilizing Process in London's Old Bailey. *Proceedings of the National Academy of Sciences of the United States of America*, 111(26): 9419–9424.

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

Lee, J. S.; Yam, G. P. D.; and Chan, J. H. 2020. PhishGAN: Data Augmentation and Identification of Homoglyph Attacks. *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 1–6.

Park, C. Y.; Mendelsohn, J.; Field, A.; and Tsvetkov, Y. 2022. Challenges and Opportunities in Information Manipulation Detection: An Examination of Wartime Russian Media. In Goldberg, Y.; Kozareva, Z.; and Zhang, Y., eds., *Findings of the Association for Computational Linguistics: EMNLP 2022*, 5209–5235. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics.

Scheufele, D. A.; and Tewksbury, D. 2007. Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models: Models of Media Effects. *Journal of Communication*, 57(1): 9–20.

Solopova, V.; Benzmüller, C.; and Landgraf, T. 2023. The Evolution of Pro-Kremlin Propaganda From a Machine Learning and Linguistics Perspective. In Romanyshyn, M., ed., *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, 40–48. Dubrovnik, Croatia: Association for Computational Linguistics.

Ustyianovych, T.; and Barbosa, D. 2024. Instant Messaging Platforms News Multi-Task Classification for Stance, Sentiment, and Discrimination Detection. In Romanyshyn, M.; Romanyshyn, N.; Hlybovets, A.; and Ignatenko, O., eds., *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING 2024*, 30–40. Torino, Italia: ELRA and ICCL.

Ethics Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, because this study uses available textual data, respects privacy norms, avoids profiling individuals, and focuses on media discourse rather than personal content, ensuring ethical integrity.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the main claims in the abstract and introduction align closely with the paper's methods, findings, and overall scope.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, the paper explains how KLD is suited for analyzing linguistic divergence related to propaganda strategies.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, the paper clearly discusses potential artifacts (duplicates and homoglyphs) and their effects on the analysis.
   (e) Did you describe the limitations of your work? Yes, see Limitations and Future Work.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see Ethics Statement.
   (g) Did you discuss any potential misuse of your work? Yes, see Ethics Statement.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see Ethics Statement.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, the paper conforms to the relevant ethics review guidelines.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, the assumptions underlying the methodological approach are stated clearly. The paper identifies the assumption that distinct linguistic features reflect underlying propaganda strategies and that KLD is a valid metric for detecting these differences.
   (b) Have you provided justifications for all theoretical results? Yes, the theoretical justification for using KLD, including its ability to capture both high- and low-frequency features, is supported by references to previous research (see Introduction and Measuring Variation Between Text Types).
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, the Introdcution includes discussions of alternative models (e.g., transformer-based techniques) and previous analyses of propaganda, offering context for the chosen approach.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, alternative mechanisms are taken into account. We consider that factors such as duplicates and homoglyphs might account for observed differences, acknowledging that not all high-KLD contributions are propagandistic per se (see Homoglyphs as a Potential Propaganda Strategy).
   (e) Did you address potential biases or limitations in your theoretical framework? Yes, potential biases and limitations are discussed, such as biases stemming from data artifacts.
   (f) Have you related your theoretical results to the existing literature in social science? Yes, the theoretical results are well connected to existing social science literature (see Comparison of Media Types).
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the

social science domain? Yes, even though the paper is mainly an analytical contribution, it discusses implications for further research and, by extension, potential policy and practice considerations (see Conclusion).

3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? Yes, the key assumptions for the method are provided, such as the use of lemmas to look at change and the use of probability distributions for comparison.
   (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Yes, sufficient details are provided to reproduce the main experimental result such as the dataset, the processing steps and the analytical method (see Data and Methods).
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes, the necessary experimental details are specified, such as the selection of subsamples, the choice of time periods, and the parameters for the KLD analysis (see Data Pre-Processing).
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes, we address robustness through overlap metrics rather than conventional error bars (see Overall KLD Values).
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? No, because the experiments were run on a local device.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes, the evaluation approach is well justified, as we validate across multiple subsamples.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, see Data (for the creators of the corpus) and Data Pre-Processing (for the authors of the code).
   (b) Did you mention the license of the assets? Yes, see Data and Data Pre-Processing.
   (c) Did you include any new assets in the supplemental material or as a URL? Yes, we include tables and appendencies.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, see Ethics Statement.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, see Ethics Statement.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? NA
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
   (d) Did you discuss how data is stored, shared, and deidentified? NA

Appendix A. Selected dates for the analysis

Table 3 shows an overview of the analyzed subsamples, which were selected based on important events of the Russo-Ukrainian war in July-September 2022.

Appendix B. Example of a duplicated post with homoglyphs

Table 4 shows two visually identical posts where the homoglyphs (Latin characters used instead of Cyrillic letters) are underlined.

Appendix C. Words contributing most to the divergence between the media types

Table 5 shows the top 50 words with the highest average KLD values for social media and state media, together with their frequencies (probabilities).

Subsample #	Size (in tokens)	Date	Event
1	4,844,884	03.07.2022	Russian armed forces captured Lysychansk, a city in Luhansk Oblast.
		22.07.2022	Russia and Ukraine signed a deal to lift Russia's blockade on grain exports from Ukraine's Black Sea ports.
		29.07.2022	Russia's mass murder of Ukrainian prisoners of war in Olenivka, Donetsk Oblast.
2	4,364,781	09.08.2022	Explosions at the Russian Saky military airbase in Crimea.
		06.09.2022	Beginning of Ukraine's counteroffensive in Kharkiv Oblast.
		15.09.2022	A mass burial of Ukrainian civilians was revealed in Izium, a city in Kharkiv Oblast that was recently liberated by Ukrainian forces from the Russian occupation.
3	4,261,001	21.09.2022	Beginning of a partial mobilization in Russia; the biggest prisoner exchange between Ukraine and Russia.
		23.09.2022	Russia's attempts to annex Donetsk, Luhansk, Kherson and Zaporizhzhia Oblasts by conducting sham referendums on the occupied territories.
		24.09.2022	
4	4,426,534	25.09.2022	
		26.09.2022	
		27.09.2022	
5	3,883,186	28.09.2022	
		29.09.2022	
		30.09.2022	

Table 3: Corpus statistics on analyzed subsamples

Excerpt of post #1	Excerpt of post #2	Transcription	Translation
Мужской выбор Русская пословица гласит: «Береги платье снову, а честь – смолоду». Можно сбежать от мобилизации, но как после этого смотреть в глаза тем, кто не струсил? #Россия #Мобилизация #Донбасс #СВО #Война #Мужчина #Воин #Мотивация #Украина	Мужской выбор Русская пословица гласит: «Береги платье снову, а честь – смолоду». Можно сбежать от мобилизации, но как после этого смотреть в глаза тем, кто не струсил? #Россия #Мобилизация #Донбасс #СВО #Война #Мужчина #Воин #Мотивация #Украина	Muzhskoy vybor Russkaya poslovitsa glasit: «Beregi plat'ye snovu, a chest' – smolodu». Mozhno sbezhat' ot mobilizatsii, no kak posle etogo smotret' v glaza tem, kto ne strusil? #Rossiya #Mobilizatsiya #Donbass #SVO #Voyna #Muzhchina #Voin #Motivatsiya #Ukraina	A man's choice A Russian proverb says: "Take care of your clothes when they're new, but your honor from a tender age." You can run away from mobilization, but how can you look into the eyes of those who did not chicken out? #Russia #Mobilization #Donbas #Special Military Operation #War #Man #Warrior #Motivation #Ukraine

Table 4: Example of a duplicated post with homoglyphs.

words	kld1	kld2	p1	p2
война	0.01889599713	-0.004192441836	0.00871236896	0.001950530525
id	0.009952021185	-4.08E-05	0.001254726564	5.15E-06
украина	0.005534982873	-3.48E-05	0.0007545204392	4.80E-06
Украина	0.005327380528	-0.003729465072	0.01229520147	0.00919239308
россия	0.003375514101	-1.94E-05	0.0004494196159	2.66E-06
быть	0.00333772958	-0.002186061158	0.00554928044	0.003663399033
так	0.003150903865	-0.001847413739	0.004210732272	0.002519757808
русский	0.003100236833	-0.001432250096	0.002950147001	0.001456437128
народ	0.002809849396	-0.001005788289	0.001919536409	0.0007045476768
СВО	0.002400526851	-0.000873991321	0.001663276028	0.0006178550212
изображение	0.00237544162	-0.0001169942535	0.00054255105	2.81E-05
воевать	0.002234637552	-0.0005666856008	0.001139418628	0.0002995930383
человек	0.002133267103	-0.001544510731	0.004969710762	0.003707071924
как	0.002023741188	-0.001070944864	0.002209584286	0.001171967432
|	0.001844343704	-1.05E-05	0.0002423604757	1.50E-06
СССР	0.001821054386	-0.0004754090724	0.000944100599	0.0002508780661
год	0.001800345247	-0.001478979918	0.006939377887	0.0058067856
знать	0.001792123222	-0.0007988678797	0.001544319622	0.0006925538974
нет	0.001683859701	-0.0009224960043	0.001978443373	0.00110217946
текст	0.001590656068	-0.0002952521175	0.0006569867266	0.0001259232286
мир	0.001569846452	-0.0009572430934	0.002289075043	0.001431443573
враг	0.001565912753	-0.0004078842041	0.0008082310471	0.0002120268908
эксклюзив	0.001534397244	-9.67E-06	0.0002072756659	1.36E-06
изощряться	0.001504113066	-5.99E-06	0.0001886518956	7.51E-07
там	0.001459443567	-0.0008621290006	0.001966333708	0.001180382812
жить	0.001452463093	-0.0006071206224	0.001180858496	0.0005109799915
почему	0.001440041129	-0.0005657403208	0.00107452852	0.0004264135976
СвоихНеБросаем	0.001427647018	-9.60E-06	0.0001929036283	1.43E-06
смолодить	0.001420405904	-5.65E-06	0.0001781530075	7.09E-07
фильм	0.001404527478	-0.0002701705647	0.0006014267232	0.0001297692053
жизнь	0.001393324568	-0.0007412128203	0.001587242591	0.000872003048
фронт	0.001345785954	-0.0006068861524	0.001182648214	0.0005400996228
пропаганда	0.001336698742	-5.32E-06	0.0001676541194	6.67E-07
нужный	0.001318386877	-0.0007879817635	0.001797964302	0.001083885449
снову	0.001310540254	-5.22E-06	0.0001643732168	6.54E-07
армия	0.001310009475	-0.0007916715481	0.00188248594	0.00116989478
небылица	0.001299825213	-2.04E-05	0.000216870269	3.40E-06
ребенок	0.001283675502	-0.0007117687146	0.001515999377	0.0008439361935
солдат	0.001263107412	-0.0004915870696	0.0009459876329	0.0003805440723
надо	0.001199953623	-0.0005882686822	0.001182625866	0.0005883298806
земля	0.00113705402	-0.000449983041	0.0008528507367	0.0003393197494
ВСРФ	0.001120739208	-4.47E-06	0.0001406308758	5.61E-07
Денацификация	0.001115436651	-4.55E-06	0.0001404854932	5.73E-07
друг	0.001110516588	-0.0004999940618	0.001001491932	0.0004723906508
Ева	0.001101288714	-2.14E-05	0.000183850886	4.48E-06
Бог	0.001060342965	-0.000125387156	0.0003440563387	4.15E-05
советский	0.001053420575	-0.0004909767119	0.001021137294	0.0005120671016
VS	0.001051170282	-8.30E-06	0.0001444736653	1.35E-06
убивать	0.001026291025	-0.0002143197361	0.0004566272501	9.90E-05
правда	0.001021779707	-0.0003493608484	0.0007174945835	0.0002890110279

words	kld1	kld2	p1	p2
заявить	-0.002023797521	0.005983783304	0.001296657653	0.003828294658
президент	-0.002061475203	0.005678735039	0.001413362383	0.003887424554
сентябрь	-0.002100076551	0.005149730325	0.001693741403	0.004047391259
глава	-0.001578210393	0.004894032583	0.0009890273918	0.003016930812
сообщить	-0.001231245703	0.004767869602	0.0006370001727	0.002446044679
Владимир	-0.001585069246	0.004416607413	0.001073559777	0.002988701927
ранее	-0.0009589673178	0.004308960256	0.0004447132792	0.001989217745
КОР	-5.00E-06	0.004280864519	5.19E-07	0.0004373534142
референдум	-0.001537738333	0.00388329632	0.001235452438	0.002984855597
отметить	-0.0009202435629	0.003851484918	0.0004561556578	0.001872240038
Россия	-0.002918533184	0.003554737936	0.01077573879	0.013011643
республика	-0.001108501195	0.003479210768	0.0006848863251	0.002119671118
область	-0.00178164713	0.00334440838	0.00230954737	0.004025804674
РФ	-0.001931102922	0.00333157185	0.002475433067	0.004255130058
также	-0.001367990876	0.003164561885	0.001150755599	0.002634522008
частичный	-0.001403063886	0.00306291719	0.001255689803	0.002728956444
запорожский	-0.0009832598854	0.003056568347	0.0006253110341	0.001889181166
февраль	-0.0008916649068	0.002936513905	0.0005306337379	0.001717886229
регион	-0.001083495979	0.002899095081	0.000787399733	0.002064265753
военный	-0.002074296712	0.002779008651	0.005573662361	0.007242522175
ДНР	-0.001078452874	0.002730516398	0.0008338818983	0.002064695948
херсонский	-0.0008426640262	0.00262055837	0.000549849098	0.001632432114
Донбасс	-0.00140050452	0.002595643161	0.001636367285	0.00297788359
подчеркнуть	-0.0005202051634	0.002533684333	0.0002283773819	0.001109635015
Известия	-5.60E-05	0.002401784767	1.10E-05	0.000436925549
слово	-0.001306215875	0.002384870577	0.001523856288	0.002765486043
состав	-0.001139273472	0.00231795437	0.001112815974	0.002263008674
российский	-0.001674528431	0.002200351834	0.004324556265	0.005659089295
ситуация	-0.001037750508	0.00212037021	0.001014633566	0.00206457509
защита	-0.000749038653	0.00209908455	0.0005127441398	0.001420050121
народный	-0.0008956029435	0.00207989421	0.0007774582472	0.001749894727
министр	-0.0006358508302	0.002075427915	0.0003846380004	0.001226277818
решение	-0.001045236626	0.002044944604	0.001102296419	0.002134254087
представитель	-0.0007056175564	0.002005230181	0.0004712286603	0.001333410604
санкция	-0.0008446439841	0.001959620836	0.0007060182381	0.00162383485
ЛНР	-0.0008958964762	0.001952566276	0.0008798986092	0.001817771966
специальный	-0.001002310946	0.001879936618	0.001126556889	0.002093308784
объявить	-0.0008093631031	0.001877201561	0.0007021511385	0.001580024318
голосование	-0.0004202566978	0.001874500944	0.0002056513091	0.0008771985354
безопасность	-0.0007044430793	0.001858021374	0.0005155001192	0.001339143831
луганский	-0.0005208256874	0.001800713523	0.0003096003667	0.001022454732
голосовать	-0.0004427602148	0.001793470557	0.0002254123723	0.0008930088593
РИА	-0.0004281580575	0.001777968489	0.000210285782	0.0008669015912
корр	-7.59E-06	0.001771872707	9.75E-07	0.000224222651
сообщать	-0.0006533241676	0.001766780583	0.000464064574	0.001239032997
добавить	-0.0004833693214	0.001742307303	0.0002638432828	0.0009438800014
страна	-0.001329911785	0.001708492085	0.004016524285	0.005063663769
проведение	-0.0007168315227	0.001658856605	0.0006286324589	0.001405354029
донецкий	-0.0006367071658	0.001627444305	0.000483413005	0.001214302437
отмечать	-0.0004297267859	0.00158821702	0.0002319348337	0.0008452842149

Table 5: Top 50 words with the highest KLD values for social media (left) and state media (right), together with their respective probabilities ("p"); "kld1" and "p1" are used for social media, while "kld2" and "p2" represent values for state media.