

*The use of machine learning provides a deeper understanding of how narratives are constructed and deployed*

PHOTO MIKE MACKENZIE

# Analysing Russian war narratives with machine learning

Drs. Pieter Jansen MSc\*

**The war between Russia and Ukraine, which escalated in February 2022, has not only caused extensive human suffering but highlighted the importance of information warfare as a modern battleground. The Russian Ministry of Defence (MoD) has strategically utilised daily briefings to shape the narrative surrounding the war, aiming to sway both domestic and international audiences. This article employs machine learning methods to dissect these briefings, uncovering the underlying themes and sentiments embedded within them. The study seeks to answer key questions: what are the dominant themes in the Russian MoD briefings? How are these themes deployed to influence perceptions? And does the emotional tone of these briefings fluctuate in response to the evolving war?**

To address these questions, I use machine learning to dissect the corpus of daily Russian MoD briefings, collected over a period of two and a half years (February 2022 until August 2024), shedding light on the intricate strategies of information warfare employed.<sup>1</sup> Two machine learning approaches are used: supervised and unsupervised learning. Supervised learning, which involves training a model on labelled data, is used to classify the briefings into specific categories based on key topics, such as nuclear, biological, and chemical (NBC) threats. This method also facilitates sentiment analysis, enabling the identification of the emotional tone – positive, negative, or neutral – expressed in the briefings. By contrast, unsupervised learning, particularly topic modelling, does not rely on labelled data. Instead, it identifies hidden patterns and structures within the data, allowing for the discovery of themes that are not predefined.

## Topic modelling

While supervised learning provides valuable insights into specific categories and sentiments, unsupervised learning methods like topic modelling are instrumental in uncovering themes within the briefings. Topic modelling decomposes the corpus into three components: words, topics, and briefings. It operates on the assumption that a briefing can be characterised by a mixture of topics, which are in turn represented by words drawn from a certain distribution.<sup>2</sup> In this analysis, topic modelling reveals the ten dominant themes in the Russian MoD briefings. These themes include military operations in key regions, nuclear and energy concerns, biological threats, and more. Each theme is represented by a set of keywords that frequently appear together in the briefings, providing a snapshot of the MoD's strategic priorities. For instance, one of the themes identified focuses on military operations in the Donetsk region, where the use of mechanised brigades and artillery is highlighted. Another theme revolves around nuclear energy, with particular emphasis on the Zaporozhye nuclear power plant and the associated safety concerns.

The consistent mention of biological and chemical weapons across multiple themes is a deliberate effort by the Russian MoD to shape the narrative around these topics, framing them as imminent threats posed by Ukrainian forces, and therefore a reason to fight.

## Detailed topic modelling insights

To further elaborate on the insights gained from topic modelling, let us explore some of the key themes identified:

1. **Military Operations:** this theme encompasses discussions around ongoing military operations, including troop movements, strategic engagements, and the deployment of weapon systems. The use of terms such as 'mechanised brigades', 'artillery', and 'Donetsk' indicates a focus on the tactical aspects of the war. The Russian MoD uses this theme to project a narrative of military strength and strategic superiority.
2. **Nuclear Energy:** the theme of nuclear energy is closely tied to incidents involving the Zaporozhye nuclear power plant. Keywords such as 'nuclear safety', 'Zaporozhye', and 'energy' highlight the MoD's emphasis on this issue. This theme serves to underscore the perceived threats posed by the war to civilian infrastructure and the broader implications of nuclear incidents.
3. **Defensive Measures and Counteroffensives:** another significant theme revolves around the defensive measures employed by Russian forces to repel Ukrainian counteroffensives. Terms like 'defensive measures', 'counter-offensive', and 'military hardware' are frequently associated with this theme. The narrative here is one of resilience and preparedness, portraying Russian forces as capable of withstanding and countering Ukrainian advances.

\* Drs. Pieter Jansen MSc works at the NATO Communications and Information Agency.

1 Official Russian MoD briefings in the English language collected from: [https://eng.mil.ru/en/special\\_operation.htm](https://eng.mil.ru/en/special_operation.htm).

2 K. Benoit et al, 'Quanteda: an R package for the quantitative analysis of textual data', *The Journal of open Source Software* 3 (2018) (30) 774. See: <https://www.theoj.org/joss-papers/joss.00774/10.21105.joss.00774.pdf>.



Figure 1 Topic modelling reveals ten dominant themes

- 4. **Biological and Chemical Weapons:** the focus on biological and chemical weapons is a recurring theme in the Russian MoD briefings. The consistent use of terms such as ‘biological weapons’, ‘chemical threats’, and ‘laboratories’ show a strategic effort to frame Ukraine as a potential aggressor in this domain. This theme is particularly significant given the ethical, psychological, and geopolitical implications of such narratives.
- 5. **International Involvement and Biological Research:** a theme that intersects with the biological and chemical weapons narrative is the involvement of international actors, particularly Western powers, in biological research. Keywords like ‘biological research’, ‘international involvement’, and ‘Western powers’ indicate a narrative that seeks to implicate foreign nations in the war. This theme is used to justify Russian actions and to portray the war as having global dimensions.
- 6. **Psychological Operations and Strategic Communications:** the Russian MoD’s use of language in the briefings also reflects a broader strategy of psychological operations. Keywords such as ‘information warfare’, ‘strategic communications’, and ‘propaganda’ are associated with this theme. The goal here is to shape public perception and control the narrative around the war, both domestically and internationally.

By decomposing the briefings into these themes, topic modelling provides a clearer understanding of the Russian MoD’s strategic communication efforts. The consistent focus on NBC threats, the portrayal of military strength, and the emphasis on international involvement all serve to reinforce the narratives that the MoD seeks to propagate.

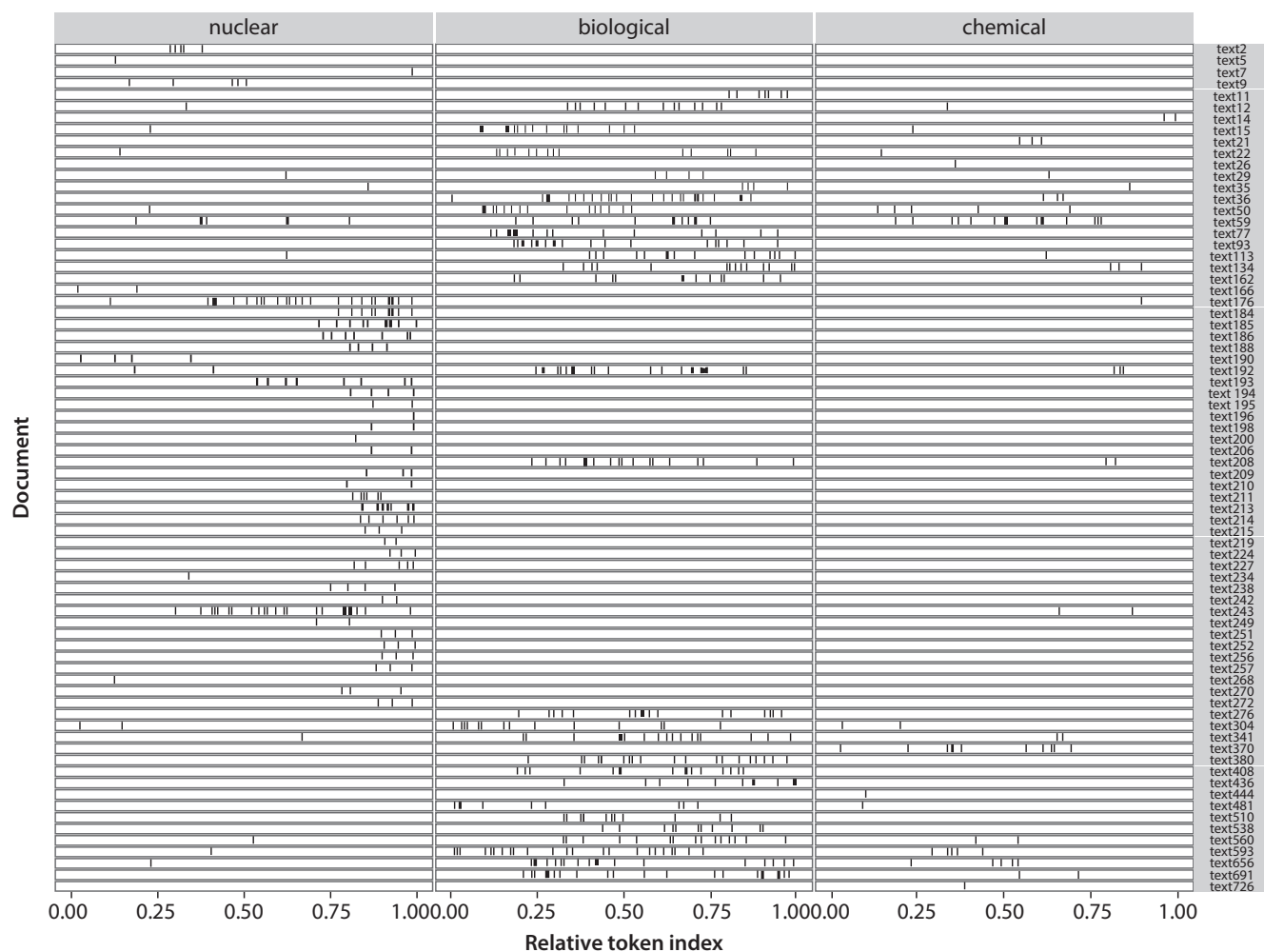


Figure 2 Lexical dispersion plot – nuclear, biological and chemical

## Lexical dispersion of NBC

To complement the insights gained from topic modelling, a lexical dispersion analysis<sup>3</sup> was conducted to examine the frequency and distribution of specific keywords related to the NBC theme. The choice to focus on NBC – nuclear, biological, and chemical – threats stems from their importance in the MoD briefings. NBC threats are consistently emphasised in the narratives, reflecting the deliberate effort to frame Ukraine as a significant and continuous source of danger. By analysing NBC themes through Lexical Dispersion and Sentiment Analysis, the study highlights how these threats are used to shape public opinion, stoke

fear, and maintain a sense of urgency and legitimacy around the conflict.

Lexical dispersion plots (Figures 2-4) visually represent where and how often certain words appear in the text, providing clues about the contexts in which these words are used. In this analysis, the keywords ‘nuclear’, ‘chemical’, and ‘biological’ were plotted across the corpus of Russian MoD briefings. The results reveal interesting patterns in the use of these terms. The term ‘nuclear’ shows concentrated usage during specific periods, notably at the beginning

3 K. Watanabe and S. Muller, ‘Quanteda tutorials’. See: <https://tutorials.quanteda.io>.



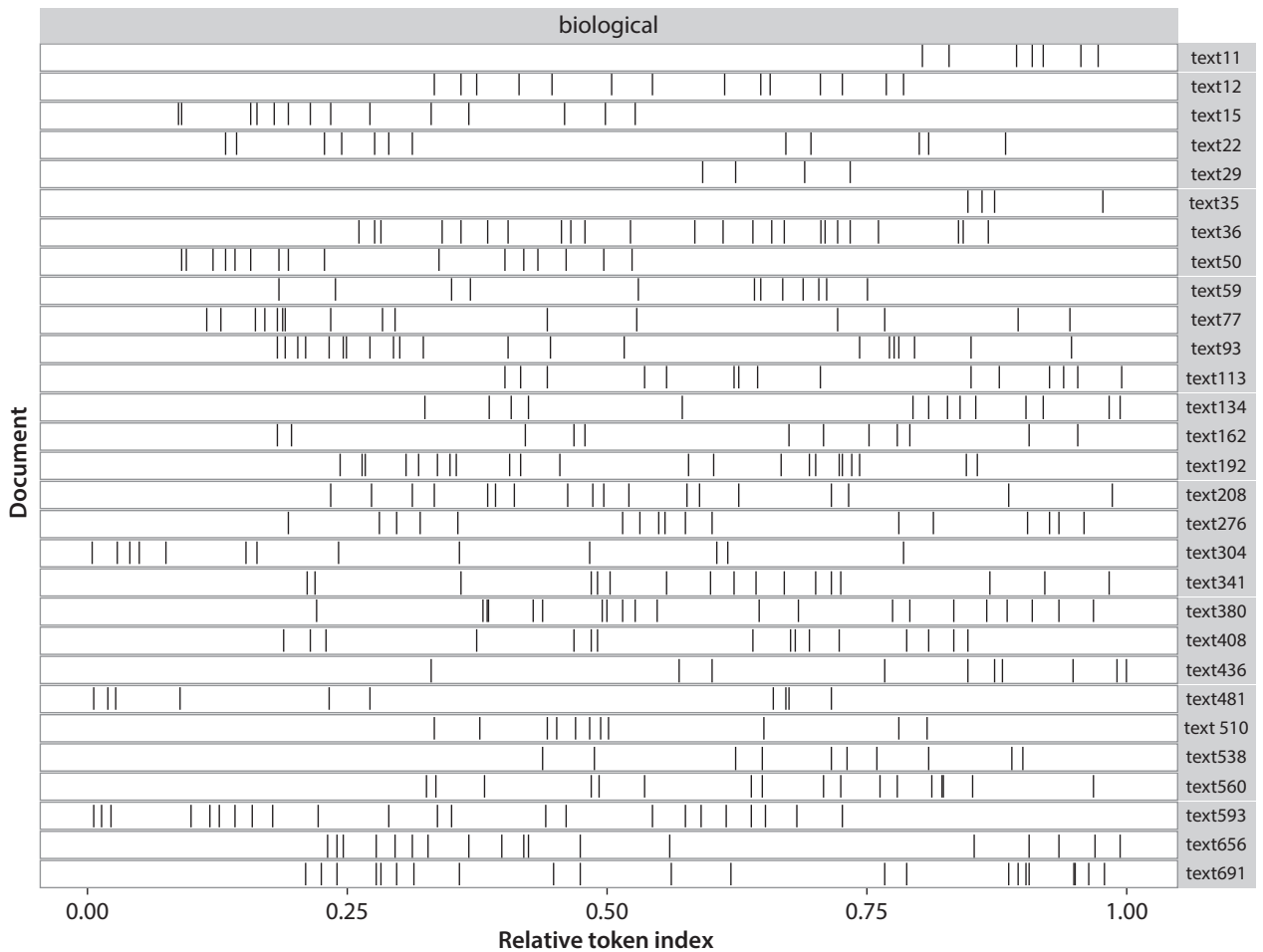


Figure 3 Lexical dispersion plot – biological

of the war in Chernobyl and during the Zaporozhye nuclear plant incidents. These peaks correspond to critical moments when nuclear safety was at the forefront of international concern.

The term 'biological' is used more consistently throughout the briefings, suggesting a sustained narrative around biological threats. Unlike 'nuclear' and 'chemical', which show peaks at specific moments, 'biological' appears to be a recurring theme in the Russian MoD's communication strategy until December 2023. This consistency indicates a deliberate effort to keep the narrative of biological threats in the public discourse, framing Ukraine as a continuous

source of danger in this domain. The term 'chemical', finally, appears with less frequency than 'biological' but follows a similar consistent pattern with a spike on 28 May 2024. 'Chemical' is often associated with allegations against Ukrainian forces, framing them as potential users of chemical weapons. This narrative serves to justify Russian military actions and to stoke fear of chemical threats.

The spike on May 28, 2024 highlights allegations regarding the use of chemical weapons by Ukrainian forces. The Russian MoD claims that Ukrainian forces have employed toxic chemicals, including BZ-type agents, and have used UAVs to deliver these chemicals. The MoD also accuses

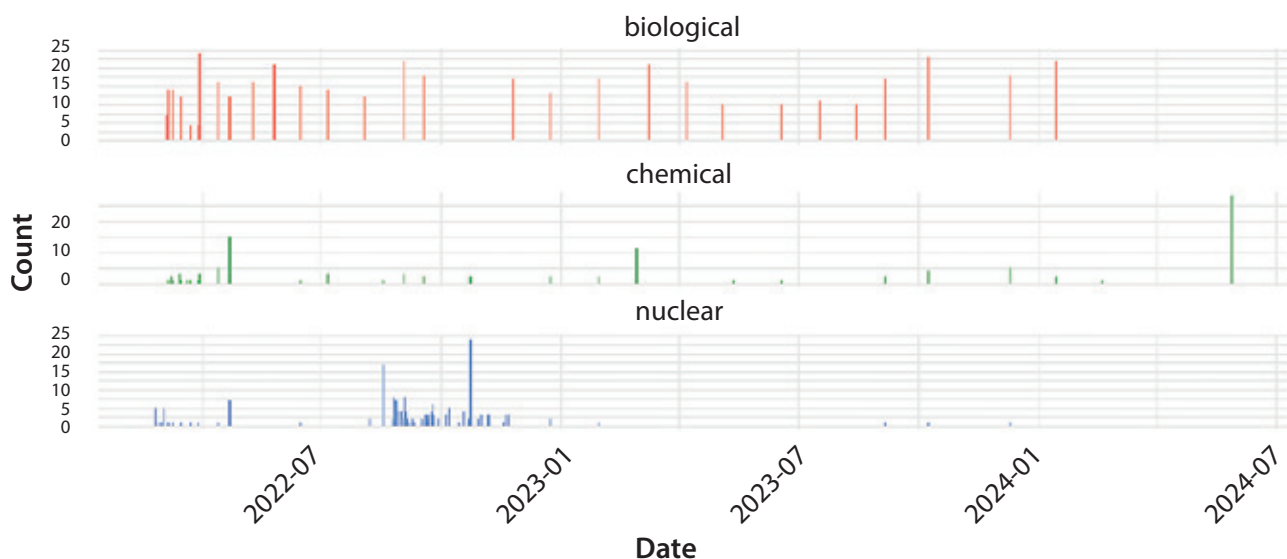


Figure 4 NBC daily count

Western countries of controlling the Organisation for the Prohibition of Chemical Weapons (OPCW) and using it for political purposes. The briefing emphasises that these allegations have gained traction in foreign media and among experts, sparking international debate and scrutiny over Ukraine's actions and compliance with chemical weapons conventions.

The lexical dispersion analysis further reinforces the insights gained from topic modelling, highlighting the strategic use of language by the Russian MoD. By consistently referring to NBC threats, the MoD seeks to maintain a narrative of imminent danger, thereby legitimising its military actions and influencing both domestic and international audiences. An interesting result of the topic modelling and lexical dispersion analysis is that the military is obsessed with biological and chemical weapons and nuclear safety. By contrast, the Russian leadership dwells on nuclear weapons.<sup>4</sup>

## Sentiment analysis

Sentiment analysis<sup>5</sup> reveals the emotional tone of the presented information, showing how the narrative shifts throughout different stages of

the war. The sentiment plot (see Figure 5), which represents the average sentiment scores, is segmented into distinct periods. These periods align with significant phases of the war and exhibit fluctuations in sentiment, likely driven by key events on the battlefield.

**February 2022 – May 2022: initial decline in sentiment.** The onset of the invasion in February 2022 corresponds with a sharp decline in sentiment, as indicated by the downward slope of the red moving average line. This period covers the initial invasion and the intense fighting that followed, characterised by high uncertainty and significant challenges for Russian forces. The negative sentiment reflects

- 4 Articles in *The New York Times* in which Russian President Vladimir Putin discussed nuclear weapons: 'Russia claims its Sarmat intercontinental missile is on 'combat duty'' (1 September 2023); 'Kim Jong-un inspects missiles and nuclear bombers in Russia' (16 September 2023); 'Russia may be planning to test a nuclear-powered missile' (2 October 2023); 'U.S. deflects Putin's nuclear alert as another effort at escalation' (27 February 2022); 'Why Putin went straight for the nuclear threat' (1 April 2022); 'How seriously should we take Putin's nuclear threat in Ukraine?' (24 September 2022); 'Russia's advances on space-based nuclear weapon draw U.S. concerns' (February 2024); 'Biden warned of a nuclear Armageddon. How likely is a nuclear conflict with Russia?' (10 October 2022); 'In Washington, Putin's nuclear threats stir growing alarm' (1 October 2022); 'Putin is brandishing the nuclear option. How serious is the threat?' (3 March 2022).
- 5 Watanabe and Muller, 'Quanteda tutorials'.

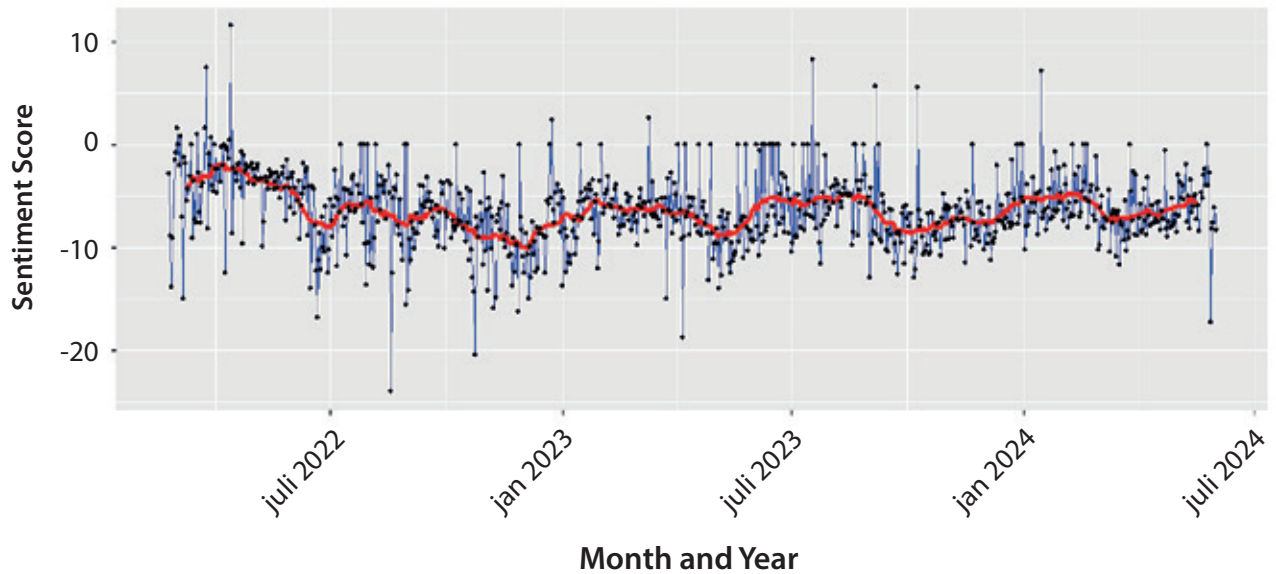


Figure 5 Daily sentiment analysis

the difficulties faced by the Russian military, including unexpected resistance, logistical issues, and the international community's swift condemnation and sanctions. Key events during this time include the start of the full-scale invasion of Ukraine in February 2022 and the battles for key cities like Kyiv and Kharkiv in March 2022, alongside the imposition of international sanctions on Russia.

**June 2022 – December 2022: gradual recovery and fluctuations.** Following the initial downturn, there is a period of gradual sentiment recovery, particularly noticeable from mid-2022 onward. The sentiment fluctuates but shows an overall upward trend as the year progresses. This period corresponds to some operational successes, such as the capture of key territories in eastern Ukraine, and the consolidation of Russian forces in strategically important areas. Noteworthy events include Russia's capture of Severodonetsk and Lysychansk in June 2022, strengthening its control in the Donbas region.

**January 2023 – August 2023: stabilisation with volatility.** In early 2023, the sentiment shows a more stable trend, with the red moving average line hovering near neutral. However, volatility

persists, reflecting the ongoing ebb and flow of the war. This stabilisation indicates a period of entrenched positions, with fewer significant shifts on the battlefield, leading to a more measured tone in the briefings. Significant events during this period include the continuation of intense fighting in the Donbas region, particularly around Bakhmut, during January and February 2023, and the start of Ukraine's major counteroffensive operations in the south and east in June 2023.

**September 2023 – February 2024: sentiment rebound.** This period is marked by an upward trend in sentiment, due to favourable developments for Russia. These include repelling Ukrainian counteroffensives and stabilising frontlines. Key events in this period include reports from October and November 2023 of Russian forces successfully defending against Ukrainian advances in key contested regions, and renewed Russian offensives.

**March 2024 – August 2024: stabilisation at higher sentiment.** In the final months of the plot, the sentiment appears to stabilise at a higher level than earlier in the war. The average line flattens, indicating that the positive

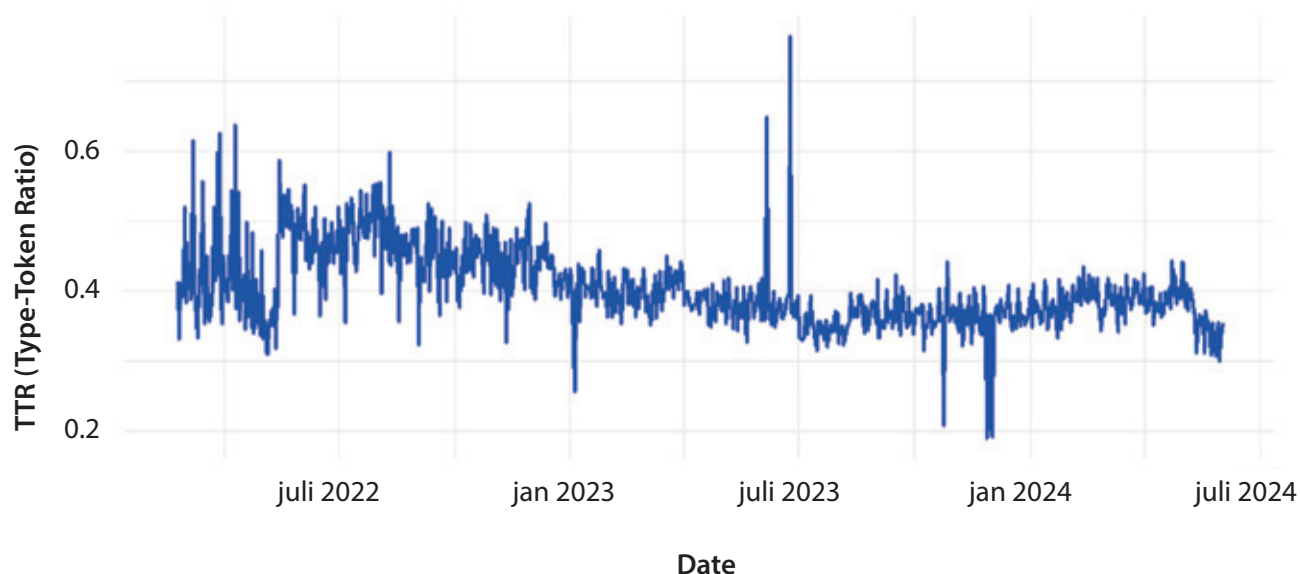


Figure 6 Type-Token Ratio over time

narrative has been maintained. This period reflect a relative stabilisation of the war, with the MoD sustaining its optimistic messaging. During this time, continued military operations focused on fortifying gains and projecting a sense of control, despite ongoing war dynamics, particularly from March to June 2024.

Overall, the sentiment plot provides a dynamic view of the Russian MoD's briefings over the course of the Ukraine war, with fluctuations in sentiment closely tied to key events on the battlefield. The initial decline reflects the challenges of the invasion, while periods of recovery and stabilisation align with strategic gains or efforts to manage public perception. The significant rebound from late 2023 to early 2024 shows a more positive narrative, in response to perceived or real successes.

## Type-Token Ratio

In addition to topic modelling, lexical dispersion, and sentiment analysis, the Type-Token Ratio (TTR) is used as a linguistic metric<sup>6</sup> to measure the lexical diversity within the MoD briefings. TTR is calculated as the number of unique words

(types) divided by the total number of words (tokens) in a given text. A higher TTR indicates a more diverse vocabulary, while a lower TTR signals repetition and limited lexical variety.

TTR functions as an indicator of lexical richness in a text. Evidently, if a briefing has a higher TTR, a more diverse vocabulary is utilised, which means greater usage of different words. On the other hand, if it has a lower TTR, there is more repetition of words, and probably a limited range of vocabulary. Given this is an analysis of military briefings, one would not expect a varied use of language. As an analogy, think of listening to the daily weather forecast on television, which is not the most enlightening prose. Yet, in cases where there is an increase in TTR, there must be a good reason. The plot shows two significant peaks in TTR: one around June 6, 2023, and another on June 24, 2023. These peaks show the MoD uses a more diverse vocabulary during critical events to convey information.

<sup>6</sup> Ibidem.



## The Russian MoD adjusts its messaging to reflect the evolving dynamics of the war

The first peak in TTR, around June 6, 2023, coincides with the start of Ukraine's counter-offensive. This period is marked by an increase in lexical diversity, as the Russian MoD employs a more varied vocabulary to describe the complex and evolving situation on the ground. The need for precision and detail in communicating the events of the counteroffensive likely drives the increase in TTR.

Almost three weeks later, on June 24, the second TTR peak occurs. This was during the 'Prigozhin's raid on Moscow'. This event, which involved significant internal turmoil within Russia, necessitated a diverse and specific vocabulary to capture the complexity of the situation. The use of terms like 'mobilised', 'destabilised', 'tactics', and 'strategies' reflects the MoD's effort to convey the gravity of the situation while managing public perception.

When critical events occur, the MoD is compelled to use more specific and diverse language to describe the complexities of these situations. This results in a noticeable spike in TTR, inadvertently highlighting these events,

even if the MoD aims to downplay or obscure their significance. Thus, TTR becomes a subtle yet powerful indicator of moments that the MoD may prefer not to spotlight, offering deeper insights into the underlying dynamics of its briefings.

## Conclusion

This comprehensive analysis of the daily Russian MoD briefings over nearly two and a half years reveals several key themes and strategies employed in the MoD's strategic communication efforts. Through the application of machine learning techniques – supervised learning, topic modelling, lexical dispersion, sentiment analysis, and Type-Token Ratio – the study uncovers the intricate ways in which the MoD uses language to shape public opinion, justify military actions, and influence both domestic and international audiences. The research questions guiding this analysis have been addressed as follows.

First, dominant themes: the analysis identifies ten dominant themes in the Russian MoD briefings, including military operations, NBC threats, and international involvement in biological weapons research. These themes reflect the MoD's strategic priorities and its efforts to control the narrative around the war.

Second, strategic deployment of themes: the consistent focus on NBC threats, particularly biological and chemical weapons, is a deliberate effort by the Russian MoD to frame Ukraine as a continuous source of danger. This narrative is strategically deployed to justify Russian military actions and to stoke fear both domestically and internationally.

The sentiment analysis reveals fluctuations in the emotional tone, finally, of the briefings. Peaks and troughs correspond to key events in the war. The MoD adjusts its messaging to reflect the evolving dynamics of the war, using positive sentiment to project strength and resilience, and negative sentiment to prepare the audience for challenging times.



PHOTO: MINISTRY OF DEFENCE OF THE RUSSIAN FEDERATION

*Russian Forces in North Eastern Ukraine. Russian military briefings are an orchestrated operation to influence narratives and sentiments about the military reality*

The briefings are not merely a record of military events but an orchestrated operation. The use of machine learning in this analysis provides a deeper understanding of how narratives are constructed and deployed. In an era where information is as crucial as physical might, the ability to analyse and interpret the strategic use of narratives becomes vital. Battles are fought not just on the ground but in the minds and perceptions of the global audience as well. ■