The meaning of sentiment analysis of UN speeches on the Russia-Ukraine war: a comparative study using VADER and BERT NLP techniques

Chen Kertcher[1]* and Moti Zwilling[2]

[1]Department of Middle East-Political Science, Ariel University, Ariel, Israel, [2]Department of Economics and Business Administration, Ariel University, Ariel, Israel

The use of natural language processing (NLP) to analyze political speeches provides valuable insights into global politics. Political discourse shapes public opinion and international relations, particularly in forums like the United Nations (UN). Sentiment analysis, a key NLP technique, helps uncover the emotional and strategic aspects of such speeches. This study evaluates the sentiment in UN speeches on the Russia-Ukraine conflict using two advanced tools, VADER and BERT. A mixed-method approach compares their effectiveness and incorporates narrative analysis to explore context and implications. Results highlight a sentiment favoring global cooperation and Ukraine, challenging claims of a North–South divide, and demonstrate the benefits of a two-stage research approach.

1 Introduction

Natural language processing (NLP) has transformed the analysis of human communication, especially in politics, where speeches shape public opinion, influence policies, and impact international relations. The United Nations (UN) serves as a critical platform for such discourse, with leaders presenting their views on global issues, negotiating resolutions, and engaging in diplomacy on matters like security, economy, human rights, and the environment (Kentikelenis and Voeten, 2021; Kertcher, 2021; Parent and MacDonald, 2024).

Sentiment in political speeches can reveal foreign policy positions, stances on issues, and persuasive strategies. Analyzing this sentiment is essential for understanding implications on global diplomacy. While traditional analysis relied on human evaluation, the rise of advanced NLP tools meets the growing need to interpret sentiment accurately, driven by the volume and complexity of speeches in forums like the UN. This capability helps identify shifts in alignments, predict policy changes, and assess prospects for conflict or cooperation among nations.

This study addresses the need for sentiment analysis in political speech using NLP techniques. We apply a mixed-method approach to analyze the dominant sentiments in UN speeches related to the Russia-Ukraine war. By running Python code with two leading NLP tools (VADER and BERT) and incorporating a sample of narrative analysis, we gain deep insights into global political dynamics and the potential of using such techniques to understand the role of sentiment in international politics.

The remainder of the paper is structured as follows. Section 2 presents a contemporary literature review on the importance of examining emotions and sentiments in global politics,

along with an analysis of the potential and limitations of NLP techniques. Section 3 outlines the methodology, proposing a mixed-method approach to address current challenges. We explain how combining two advanced NLP techniques with sample narrative analysis can shed light on dominant global sentiments. Section 4 presents the main findings, explores biases within the NLP methods, and provides insights into the dominant and marginal sentiments in global political discourse. Finally, Section 5 offers a discussion and some conclusions.

2 Literature review

NLP is defined as a computational method aimed at enabling computers to process, interpret, and understand human languages through a variety of techniques that have increasingly been applied to the analysis of political speeches, a domain that has attracted significant scholarly attention in recent years (Mah and Song, 2024). One of the most widely used NLP techniques, as identified in the literature, is sentiment analysis, which involves assessing emotions conveyed in the text (Giatsoglou et al., 2017). This technique focuses on identifying and categorizing opinions expressed in texts to infer the writer's or speaker's stance on a given topic. The growing complexity and prevalence of political discourse, particularly in international settings such as the UN, underscores the need to explore how advanced NLP methods can yield valuable insights into global politics.

2.1 Sentiment analysis in political discourse and the Russia-Ukraine war

The emotional turn in international relations advances the constructivist approach, which emphasizes the role of ideas, norms, beliefs, and social identities in shaping political behavior (Wendt, 1992; Finnemore and Sikkink, 1998). Moreover, as Wiener (2018) demonstrates, norm contestation and interpretation through inclusive dialog among actors help sustain the legitimacy of international norms. Emotions, often seen as personal, are also social and cultural phenomena, playing a crucial role in politics. They play a critical role in international politics. Scholars recognize emotions as essential in justifying group actions and sustaining commitment to shared goals. Emotions influence decisions, including state actions (Krebs, 2015; Koschut, 2020; Wallis, 2024). Emotional processes are essential for groups to justify their actions and sustain long-term commitment to achieving shared goals (Niedenthal and Brauer, 2012, p. 269).

Scholars agree that emotions are an integral part of political discourse because language carries emotional meaning, and any analysis of emotions should consider how they mobilize communities for collective action (Bleiker and Hutchison, 2008; Koschut, 2018; Osnabrügge et al., 2021; Shah, 2024). States engage in war or humanitarian intervention not only for strategic or legal reasons but also driven by collective emotions that shape what is viewed as moral and desirable (Wallis, 2024). For instance, a sense of victimhood can fuel competition between groups over which has experienced greater suffering, as seen in the conflicts between Israelis and Palestinians or between Chinese and Japanese (Sasley, 2020). On the other hand, emotional discourse can promote selective empathy within groups, fostering sympathy and compassion for those who support their

policies while generating disgust and rejection toward those who oppose them (Head, 2020).

The full-scale war between Russia and Ukraine, which began on February 24, 2022, intensified global divisions and heightened negative emotions. It became the largest conflict in Europe since World War II, involving hundreds of thousands of soldiers, endangering millions of lives, and threatening to destabilize key global norms such as state sovereignty, territorial integrity, and peaceful conflict resolution. With Ukraine's nuclear reactors at risk and Russia's nuclear arsenal in play, the war holds significant potential for catastrophic consequences.

The conflict has reinforced alliances between the US, Europe, and other liberal democracies, which view Russia's aggression as a direct threat to international order and cooperation (Ellison et al., 2023). This is reflected in the unprecedented economic sanctions imposed on Russia and the increased military and financial aid provided to Ukraine. Western nations, seeking to preserve their global influence, argue that the invasion violates the core principles of international security and law, including sovereignty, rule of law, human rights, and territorial integrity.

In contrast to the Western condemnation of Russia in the name of international law, many studies highlight a more ambivalent stance from countries outside the West. This division can be partly attributed to long-standing grievances from nations in the Global South, which represent the majority of the world's population. Many of these countries, particularly in Asia, Africa, and Latin America, see the US as hypocritical and driven by self-interest and view Russia's actions as a challenge to Western dominance. Positive sentiment toward Russia is prevalent in countries like India and Pakistan, influenced by factors such as economic ties, dependence on food, energy, military supplies, and access to technologies (Ajala, 2022; Hartley, 2022; Mijares, 2022; Dempsey, 2023; Carrión-Vivar et al., 2024).

Due to the war's profound impact on global security, emergency debates were quickly convened at the UN. The UN Security Council, responsible for maintaining international peace and security, was unable to pass a resolution due to Russia's veto (UN SC/14808). As in previous crises where the Security Council failed to act, the UN General Assembly held an emergency session, allowing all member states to express their views on the war. On March 2, 2022, following six debates, the first resolution was adopted (UN A/RES/ES-11/1), with 141 votes in favor, 5 against, and 35 abstentions. The resolution condemned Russia's invasion, demanded the withdrawal of its forces, called for the reversal of its recognition of Donetsk and Luhansk, and included condemnation of Belarus.

Despite the emotional division between states in the Global North and South, the war has generally fostered negative global sentiment and emotions. This scenario presents an excellent opportunity to test the effectiveness of NLP methods in assessing global sentiment trends on such a significant issue.

2.2 Challenges in analyzing political speeches

Sentiment analysis, using NLP and machine learning, automatically classifies subjective information as positive, negative, or neutral. It is valuable for analyzing political speeches, particularly in

global politics, but faces unique challenges with speeches by heads of state and diplomats (Miranda and Bringula, 2021).

Sentiment analysis provides insights into public discourse but struggles with the nuanced, ambiguous language of diplomatic speeches, especially at the UN. Diplomatic language often uses ambiguity, irony, and polite hedging, which can mislead tools designed for straightforward language (Semenov and Tsvyk, 2021).

Additionally, diplomatic speeches are often deeply rooted in historical and cultural contexts. Sentiment analysis tools may struggle to capture these subtleties (Liu and Wang, 2020). Moreover, while sentiment analysis can provide an overall positive or negative score, it may miss the specific goals, criticisms, and variations within a speech (Surowiec and Miles, 2021).

In recent years, the use of NLP methods for sentiment analysis has seen significant growth (Wankhade et al., 2022; Jim et al., 2024). Sentiment analysis in the context of public diplomacy is a growing field. Several studies have focused on politicians' use of Twitter (Antonakaki et al., 2021) and sentiment analysis of official media statements (Fisher et al., 2022). However, these approaches have not yet fully addressed the complexities of UN diplomatic speeches. For instance, Widmann and Wich tackled cultural sensitivity by developing a sentiment analysis tool specifically for German political discourse, training a model on over 10,000 German-language phrases (Widmann and Wich, 2023). Their research demonstrated that NLP models can effectively automate text analysis and accurately measure emotional language within brief political statements on such platforms as Twitter and Facebook. This success was also demonstrated in studies of political sentiment on television and other media (Atmaja and Sasou, 2022), but it may fail with longer speeches such as UN diplomatic speeches.

Widmann and Wich's models present some limitations their models require substantial processing power and time, especially when applied to large datasets. Additionally, their approach may limit its generalizability across different domains or languages without additional customization.

Other studies that dealt with the analysis of speeches in this domain have been critiqued for relying too heavily on traditional, static frameworks in an increasingly dynamic research environment (Kentikelenis and Voeten, 2021; Fisher et al., 2022; Jim et al., 2024). One major criticism is that these studies may have overemphasized quantitative methods without sufficiently integrating qualitative data, which could capture the nuance of human behavior and motivation. While quantitative approaches offer broad insights, they often fail to explain underlying drivers, such as social norms, cultural pressures, or individual psychological states, resulting in oversimplified conclusions.

Additionally, other methods introduced in the literature may lack adaptability to rapidly changing sociopolitical landscapes. For instance, Kentikelenis and Voeten (2021) may not fully account for the swift shifts in global governance and political dynamics, which could render their conclusions quickly outdated.

Given all of the above, the application of sentiment analysis to UN speeches remains underexplored. While some studies have examined narratives in UN forums through human coding (Kertcher, 2021, 2024), others have employed mixed methods to analyze sentiment in annual leaders' speeches from 1970 to 2018 regarding exits from the global economic order. However, these studies still relied on human coding (Kentikelenis and Voeten, 2021). Although

sentiment analysis is a valuable tool for understanding public discourse, its limitations in diplomatic settings must be considered. Human expertise remains essential for interpreting the nuances, context, and intentions embedded in diplomatic language. Nonetheless, there is a growing need to leverage the most advanced NLP techniques and explore how combining them with human insight can yield relevant and meaningful findings for the study of global politics.

2.3 Advancements in NLP techniques for sentiment analysis

In comparison to existing literature, our study employs methods that offer several notable advantages, enhancing both effectiveness and applicability. First, VADER and BERT are specifically designed to handle informal language and shorter text formats, which are frequently encountered in political discourse, especially on social media platforms. Models like VADER and TextBlob can analyze short texts (around 500 words), while more sophisticated models like BERT are better suited for longer speeches (1,000+ words) for more detailed fine-tuning and pattern recognition. This adaptability enables our study to capture subtle distinctions in sentiment that may be overlooked by more complex models reliant on structured speech formats, such as those found in UN speeches.

Furthermore, the simplicity of VADER and BERT allows for rapid analysis. By focusing primarily on sentiment polarity and emotional tones, our approach reduces the risk of overcomplicating the analysis, leading to clearer and more actionable insights into public opinion. This streamlined methodology not only enhances the clarity of the analysis but also aligns with the practical needs of policymakers and analysts, who often require timely and relevant information.

3 Methodology

3.1 Research questions

The current study addresses several research questions and their derived hypotheses:

RQ1: *Do different NLP techniques vary in their ability to identify sentiments in diplomatic speeches?*

H1: Keyword-based techniques (such as VADER) will perform less effectively compared to more advanced NLP techniques, leading to varying results that raise questions about which NLP technique is most suitable for sentiment analysis of diplomatic speeches.

RQ2: *Is there a relationship between the diplomatic topic and the anticipated sentiment, as analyzed through NLP techniques?*

H2: We anticipate that NLP analysis will predominantly highlight negative sentiments in speeches concerning the Russia-Ukraine war due to the conflict's profound impact on international peace and security.

RQ3: *Are there common patterns in the narrative derived from sentiment analysis of UN speeches related to the Russia-Ukraine conflict?*

H3: The sample narrative analysis will succeed in illuminating the meaning of positive and negative results. We hypothesize that there will be four main narratives that represent two contradictory axes. On the first axis will be a negative narrative that is anti-Russian and a positive sentiment that is pro-Ukrainian. On the second axis will be the potential for a negative narrative that is anti-Western and anti-Ukraine and a positive narrative that emphasizes support for Russia. We assume that the second axis will be marginal and perhaps will not be identified by our method.

RQ4: *In what manner does sample narrative analysis reveal the limitations of specific NLP methods, such as VADER and BERT?*

H4: Sample narrative analysis can illuminate the discrepancies in results between NLP methods. It will identify the topics and issues in which NLP's methods are less accurate. Therefore, it can identify the best NLP method to analyze diplomatic speeches quickly and effectively.

3.2 Two-stage approach: NLP and sample narrative analysis

This paper adopts a mixed-method approach consisting of two stages. The first applies quantitative NLP techniques, and the second centers on qualitative narrative analysis. Together, these stages offer a more holistic understanding of language, sentiment, and the narratives within the text.

3.3 Stage 1: data sampling, acquisition, and analysis

The data was analyzed using two widely used techniques for NLP sentiment analysis: BERT (Bidirectional Encoder Representations from Transformers) and VADER (Valence Aware Dictionary for Sentiment Reasoning). These methods were chosen for their distinct strengths in handling sentiment analysis of political discourse.

VADER is defined as a rule-based method that calculates four sentiment scores: Negative (Neg), Neutral (Neu), Positive (Pos), and Compound. The Compound score is a weighted composite that ranges from −1 (most negative) to +1 (most positive). It is computed by summing the valence scores of each word in the lexicon, adjusting for heuristics, and normalizing the result. Sentiment classification is then determined by the Compound score: a score ≥ 0.05 is classified as positive, ≤ −0.05 as negative, and between −0.05 and 0.05 as neutral.

BERT, in contrast, is a transformer-based machine learning model fine-tuned for sentiment analysis. It classifies text into multiple sentiment categories: Very Negative, Negative, Neutral, Positive, and Very Positive. BERT uses logit values to predict sentiment, where the final classification is based on the highest logit value, which is then processed through a SoftMax function to generate probabilities. The sentiment label is determined by the highest probability.

Sentiment analysis within the realm of diplomatic discourse poses unique challenges owing to the implicit and strategically nuanced characteristics inherent in political language. The VADER model is particularly designed for short, informal texts, such as social media communications, utilizing pre-established sentiment lexicons and heuristic rules to categorize sentiment based on explicit emotional vocabulary, punctuation, and intensity modifiers. However, its effectiveness diminishes when applied to formal, structured speeches, where sentiment is frequently expressed indirectly rather than through overt emotional language.

Conversely, BERT, a transformer-based deep learning model, employs contextual embeddings to understand word relationships in specific linguistic contexts, thereby enabling it to capture intricate variations in sentiment within lengthier and more structured texts. Although BERT offers enhanced flexibility and accuracy in the analysis of diplomatic communications, it is also vulnerable to biases present in its training data, which may lead to misclassifications of culturally nuanced expressions, diplomatic hedging, or strategic ambiguity.

Given these complementary strengths and limitations of these methods, the current study employed both VADER and BERT to leverage their distinct capabilities, VADER's proficiency in detecting explicit sentiment markers and BERT's ability to interpret context-dependent sentiment, allowing for a more comprehensive and nuanced analysis of sentiment in diplomatic speeches.

Specifically, in the current study, we applied both methods to analyze speeches delivered during the first six meetings of the UN General Assembly (UNGA) emergency session, held between February 28 and March 2, 2022. This session culminated in the adoption of the first resolution on the Russia-Ukraine conflict after the UN Security Council failed to pass any resolution on the issue. The study focuses on the UNGA emergency session for several reasons: it provides a global forum where all states can participate, it operates without veto power (unlike the Security Council), and it allows for the discussion of any topic, making it an ideal setting for sentiment analysis.

To ensure a structured and representative subset for human narrative analysis, a targeted sampling approach was employed as described in 3.4, focusing on sentiment classification results from VADER and BERT. The selection criteria included (1) high-convergence cases (12 top speeches), where both models classified speeches similarly as either positive or negative; (2) high-divergence cases, where the models produced conflicting sentiment classifications; (3) geopolitical diversity, ensuring representation from different diplomatic perspectives, and (4) To enhance the efficacy of the comparison, we assigned VADER's neutral score as positive in our sample analysis, given that BERT lacks a neutral category. This assignment was based on the observation that neutral diplomatic speeches typically support conformity, are less inflammatory, and lack accusatory or aggressive language. This approach ensured a systematic and reproducible methodology, facilitating a balanced exploration of both aligned and conflicting sentiment interpretations.

We analyzed a total of 161 speeches, comprising 4,005 sentences and 90,698 words, which led to a largely one-sided resolution against Russia. This dataset offers a robust opportunity to evaluate the use of sentiment in diplomatic debate, despite the inherent challenges of applying NLP methods to professional and formal language. For each state, we calculated an overall sentiment score, known as the

compound score, using both VADER and BERT. We then compared the results to identify patterns of sentiment (positive or negative) and assessed the convergence and divergence between the methods. This comparison allowed us to identify trends in sentiment that align with our initial hypotheses.

3.4 Stage 2: sample narrative analysis

The second stage of the analysis employs a sample narrative analysis. Narrative analysis is a well-established method for examining international politics and the declared policies of states. It posits that speeches function as stories, creating sequences of events that connect characters and scenes through causal relationships, ultimately imbuing them with meaning. These narratives often grapple with moral and normative dilemmas, sometimes offering solutions (Krebs, 2015; Shenhav, 2015; Opperman and Spencer, 2023; Kertcher, 2024; Miskimmon et al., 2024). While narrative analysis can be subject to the influence of subjective interpretation and can be slow to process large volumes of data, it offers a complementary perspective to NLP techniques.

Unlike other studies that rely solely on narrative analysis, we conducted targeted sampling to explore the meanings of positive and negative sentiments identified by NLP methods. A human narrative analysis was performed on a carefully selected subset of 12 speeches comprising 537 sentences and 12,038 words. This sample was chosen to facilitate a comparison of eight pairs of speeches: two with the highest positive VADER scores that also excelled in BERT, two with strong positive BERT scores that performed well in VADER, two speeches where both methods classified the sentiment as negative, and two where

both identified it as positive. Furthermore, we included two speeches that received a high positive VADER score but a negative BERT score and two with a high positive BERT score but a negative VADER score.

First, we compared the divergence and convergence of the compound scores from each NLP sentence. To allow for a full comparison with BERT, which only classifies sentences as positive or negative, we treated VADER's neutral sentiment as positive, as such statements are not hostile to any position. A summary of these findings is presented in Figure 1.

Second, we performed a human assessment of the positive or negative sentiments of the different sentences in our sample group of speeches and compared these assessments to the results from VADER and BERT. This approach enable us to evaluate the differences between the two NLP techniques and explain the meanings of positive or negative sentiments. Moreover, the method ultimately aids in determining which of the NLP methods was more precise in assessing the sentiments within diplomatic speeches, particularly in the context of the Russia-Ukraine conflict.

4 Results/findings

RQ1: *Consistency and Consensus of NLP Findings.*

We hypothesized that, given the context of the war, the majority of speeches would be characterized as predominantly negative. However, both techniques indicated that most speeches were, in fact, more positive than negative, with VADER reporting 73.2% positive and BERT indicating 77.6%. Additionally, 60% of the results from both NLP methods aligned in identifying speeches as
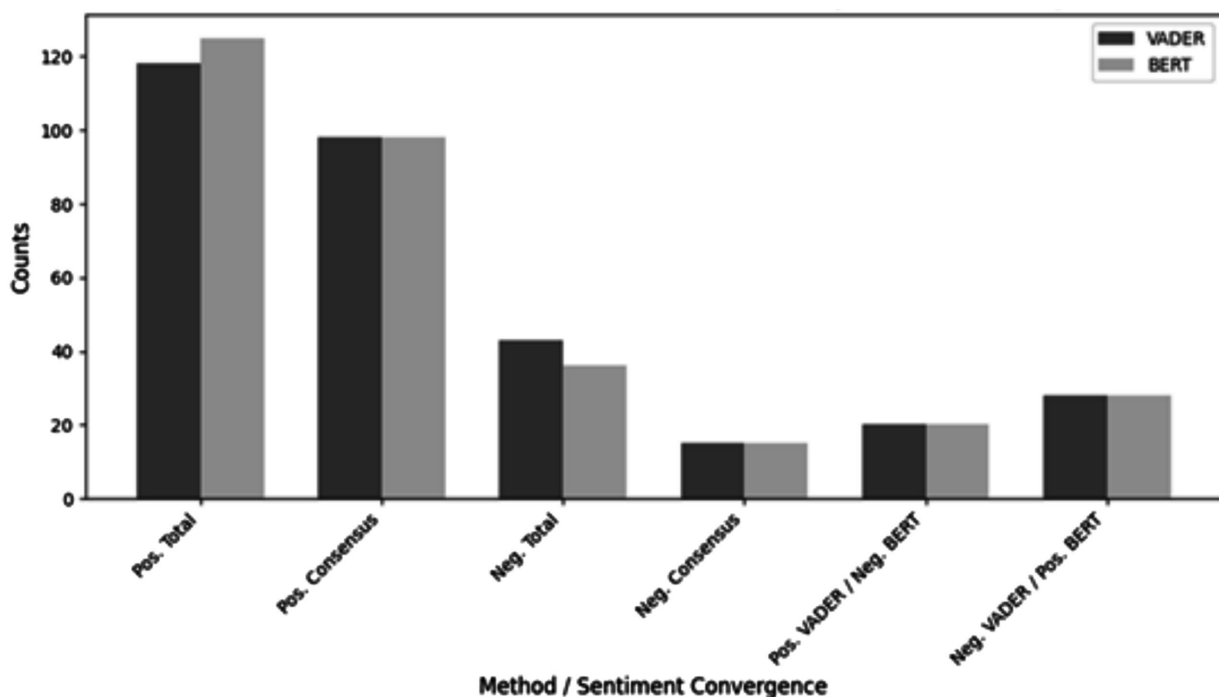

FIGURE 1
Frequency comparison of sentiment analysis scores (VADER vs. BERT).

TABLE 1 Sample of narrative analysis by country group and convergence percentage (Vader's neutral is calculated as positive as explained in the methodology).

Country	Sentences	Words	Group Polarity	Convergence (%)
Belize	18	443	Negative (B-/V-)	61.0
Colombia	40	1736	Negative (B-/V-)	57.5
Canada	97	1775	Negative (B-/V-)	59.0
Costa Rica	33	849	Negative (B-/V-)	63.6
EU+	58	1,268	Positive (B+/V+)	77.0
Myanmar	24	470	Positive (B+/V+)	87.5
Papua New Guinea	21	588	Positive (B+/V+)	61.9
Luxembourg	45	1,012	Positive (B+/V+)	75.0
Haiti	18	406	Mixed (V+/B-)	83.3
Ecuador	21	663	Mixed (V+/B-)	61.9
Denmark	71	1,128	Mixed (V-/B+)	67.6
Ukraine	91	1700	Mixed (V-/B+)	62.6
Group total (Neg)	188	-	Negative Group	60.0
Group total (Pos)	148	-	Positive Group	76.0
Group total (Mixed)	201	-	Mixed Group	66.0
Grand total	537	-	All Groups	66.8

positive. According to VADER, 26.7% of all state speeches were classified as negative, which is quite similar to the 22.3% identified by BERT (Figure 1).

RQ2: *The Ratio of Positive to Negative Sentiments Using NLP Methods.*

Another hypothesis was that the one-sided and straightforward resolution adopted during the UNGA emergency session would lead to greater alignment in the results between both NLP techniques. We found that both methods reached a consensus on the sentiment of 113 speeches, identifying 98 as positive and 15 as negative, resulting in an agreement rate of 70% (Figure 1).

Therefore, we can assume that the UN speeches had a strong positive sentiment and not a negative one. This finding contradicts our hypothesis, which was based on the severity of the international crisis and the UNGA resolution's strong condemnation of Russia.

RQ3: *How Does Sample Narrative Analysis Illuminate the Meaning of Positive and Negative Sentiments?*

While the quantitative results indicate sentiment trends, they do not provide an explanation of their meanings. Criticism regarding the potential of diplomatic speeches highlights the difficulties of using NLP techniques, particularly in interpreting meaning and establishing consistency. However, our narrative analysis of the highest-scoring positive and negative speeches from the BERT and VADER lists reveals two prominent trends of positive and negative sentiments. Despite variations in positive and negative scores, all 12 speeches in our sample expressed criticism of the war, positioning themselves along a primary axis that condemned Russia and supported positive actions.

Our sample narrative analysis method aimed to assess the convergence and divergence of results between the NLP techniques.

First, the method demonstrated convergence of positive sentiment, convergence of negative sentiment, and divergence among the NLP techniques, exhibiting consistent results. Second, the positive sentiment group was the most aligned among both NLP techniques and human analysis, with Papua New Guinea being the only exception, showing a convergence rate of only 0.619% in sentiment identification. Examining the other three countries within this group, only 39 out of 127 sentences were not in agreement, representing about one-third of the contested sentences, which is considered a high score in the literature. Third, the two groups focused on negative sentiment or a lack of consensus among the techniques exhibited significantly lower convergence rates. Interestingly, the sentences classified as consensually negative received, on average, less agreement overall (Table 1).

Each group was analyzed separately based on four narrative criteria. First, the 'setting' or context of the conflict/war was examined in relation to international norms, laws, and institutions. The second narrative perspective focused on assessing the actions of the key actors, primarily Russia and Ukraine, while also mentioning other entities like UN agencies, EU initiatives, and international legal bodies such as the International Court of Justice (ICJ) and International Criminal Court (ICC). Third, the emplotment could have taken various forms, from downplaying the negative impacts of Russian actions to outright condemnation or from expressing general support for Ukraine to sharing distressing images and stories of tragedy. Finally, the proposed solutions to the conflict encompassed a range of approaches, from advocating for a general end to hostilities or encouraging peace to more aggressive measures such as prosecuting Russian leadership at the ICC and imposing strict economic sanctions on Russia. The relevance of historical analogies—often cited as a reason not to employ NLP techniques—was minimal in the broader context of the speeches, having little effect on the overall number of words and sentences used.

4.1 Positive speeches

Positive speeches represent the most consensual sentiment in the General Assembly discussion. Both NLP techniques agreed that approximately three quarters of the speeches are positive a relatively high score which stand against our preliminary hypothesis. What defined a speech as positive? We used the narrative analysis identifying the setting, the actors, and the emplotment that predominantly conveyed a positive sentiment. Our sample examined the speeches of four actors: Luxembourg, Myanmar, Papua New Guinea, and the EU representative along with other states that endorsed the official statement identified as EU+ in the text.

First, all four actors emphasized a setting centered around international law, diplomacy, and a sense of global community. Quotes included the preamble of the UN Charter: "We the peoples of the United Nations, determined to save succeeding generations from the scourge of war" (Myanmar), "respect for the rules-based international order under international law, including the Charter" (Papua New Guinea), and "The EU and its member states will continue to do their utmost to protect the integrity of the rules-based international system" (EU+).

Second, positive speeches advocated for principles such as sovereignty and territorial integrity for Ukraine. For instance "we commend the courage of the President of Ukraine and the Government and the people of Ukraine" (Luxembourg), and "expressing the European Union's full solidarity with and support for Ukraine and its people" (EU+).

Russia was urged to recognize that "the General Assembly has the responsibility to send a strong and united message on the need for Russia to immediately and unconditionally cease its military operations, withdraw its troops, and opt for genuine dialogue" (EU+). Additionally, "We condemn the involvement of Belarus in the aggression against Ukraine" (Luxembourg).

These speeches promoted global political cooperation through the UNGA or other actors, with statements like "thank all Member States supporting Ukraine in every possible way" (Myanmar). They also supported aid for Ukraine, "commend[ing] Ukraine's neighbors for their generosity" (Luxembourg), and "thank[ing] Ukraine's neighbors for opening their borders and hearts to receive Ukrainians" (Papua New Guinea).

Finally, actions proposed to resolve the crisis reflected a positive sentiment, such as calling for Russian withdrawal: "We urge Russia to immediately and unconditionally withdraw from Ukraine" (Papua New Guinea); minimizing casualties: "the parties must take every measure to protect civilians, including children and humanitarian workers, and civilian infrastructure" (Luxembourg); and promoting diplomatic dialogue and negotiation for peace under UN auspices: "encourage both parties to use peaceful dialogue through diplomacy" (Papua New Guinea), and "We further call on Russia to engage earnestly in dialogue with a view to a political and diplomatic solution" (EU+).

In summary, consensus on positive sentiment does not entail ignoring the seriousness of the international crisis nor adopting a pro-Russian stance. Instead, it reflects how dominant narratives express a positive sentiment aligned with shared global norms and values, a preference for collective action as a global community, and support for Ukraine, all while refraining from a strong anti-Russian sentiment or limiting its expression in the speeches. This is accomplished by incorporating some negative elements that do not overshadow the overall sentiment of the speeches.

4.2 Negative speeches

Negative speeches represent a different category. First, it should be noted that NLP methods struggle to reach an agreement on negative diplomatic speeches. Even when both techniques agree that certain speeches are negative, they often disagree on which specific sentences are more negative or positive. In our sample, they reached a consensus on approximately 60% of all sentences, a relatively low score for what was expected to be a consensual issue according to our preliminary hypothesis (see Table 1). Thus, what characterized a diplomatic speech as negative?

Initially, the setting of negative speeches painted a grim picture of the contemporary international crisis, highlighting the direct connection between Russian actions and the erosion of various aspects of international norms and order. Statements such as "Unacceptable violation of Ukraine's sovereignty and territorial integrity and in contravention of Article 2 of the Charter and the norms and principles of international law" (Belize) and "illegal aggression against a neighboring State" (Costa Rica), or "violation of international humanitarian law and international human rights law" (Costa Rica) were common. Additionally, remarks like "Russia's offensive… violates the principle that no State should be subjected to the use or threat of use of force against its sovereignty, political independence, or territorial integrity" (Colombia) and "Russia has launched an illegal and unjustifiable war" (Canada) were noted, alongside claims that "Belarus has also violated its obligations under the United Nations Charter and international law" (Canada).

Negative speeches often detailed the suffering of the Ukrainian people due to the war, expressing sentiments like "seeing the loss of life, disruption of livelihoods, destruction of property, and terror" (Belize). They also conveyed tragic updates, such as "this morning we learned of the deaths of dozens of Ukrainian children" (Colombia) and "they are looting and forcibly displacing civilians" (Canada).

Russian actions were met with strong condemnation. Regarding diplomatic manipulation, one speaker stated, "Whatever lies are spoken here today trying to justify the unjustifiable or to explain the inexplicable, it is President Putin's war of choice that is making their lives much more difficult" (Canada). The nuclear threat was addressed with statements like "condemning the Russian Federation's decision to put its nuclear forces on alert" (Costa Rica), along with warnings that "threatening Ukraine, Europe, and the entire world with the use of nuclear weapons" (Canada) constitutes a serious issue. There were also concerns regarding the attack on fundamental rights: "including extrajudicial killings, abductions, forced disappearances, politically motivated prosecutions… among other reported abuses of fundamental freedoms" (Costa Rica). Furthermore, the impact of aggression was discussed in terms of global security: "Russia's unilateral and unjustified actions… could result in a new global migration crisis that would put the stability and peace of Europe and the world in serious jeopardy. […] the situation has caused economic panic all over the world" (Colombia).

In contrast to the positive narratives, negative speeches contained much less discussion about global cooperation to assist

the Ukrainian people or other forms of international collaboration. This may partly explain the more pronounced negative scores assigned. Like the analysis of predominantly positive sentiments, NLP techniques also identified positive labels within the negative sentiments, with some statements being contested between the techniques. For instance, "to the Russian protesters bravely demonstrating against the war…" (Canada) illustrates this point. Another example involves broader global goals: "of humankind, we must pursue our agenda addressing climate change, gender equity, energy transitions, reducing global hunger, combating poverty, and achieving all the Sustainable Development Goals" (Colombia).

RQ4: *Explaining Divergences in Results.*

Understanding the divergences in the results of NLP techniques is crucial for the evolution of AI. In this study, we identified that 48 out of 161 speeches, or approximately 30% of all speeches, received opposite scores from different techniques. We identified four states with the highest divergence scores and conducted a human narrative analysis to explain these discrepancies and better understand the contrasting results reflected in the techniques' compound score.

Despite reaching an agreement on over 60% of sentiments' meanings—83% in the case of Haiti—the NLP methods were unable to determine whether the overall sentiment score should be classified as positive or negative (see Figure 2). This inconsistency raises concerns and highlights the caution needed when applying NLP techniques to diplomatic speeches. However, narrative analysis can clarify these scores.

When the VADER score is positive and BERT is negative, the analysis reveals several insights. Ecuador's and Haiti's speeches followed a similar narrative structure emphasizing positive sentiment, focusing on international norms and global crises. They also articulated the need for various actions to address the situation. Additionally, both speeches contained negative sentences that likely influenced BERT more than VADER. Haiti and Ecuador explicitly condemned Russia with such statements as "Ecuador learned with shock and disappointment that a military invasion of Ukraine had begun" (Ecuador) and "following the decision by Moscow to raise its nuclear alert level" (Haiti). They also used stark descriptions of the suffering endured by Ukrainians, such as "massive flow of people fleeing the battlefield" (Haiti) and "when considering the situation in the temporarily occupied territories of Ukraine" (Ecuador).

Conversely, when BERT scores are positive and VADER scores are negative, such as for Ukraine's and Denmark's speeches, the findings indicate that they utilized a similar narrative structure characterized by negative sentiment, with a detailed portrayal of Russian aggression and its impact on Ukraine. Moreover, these speeches included positive statements that likely affected BERT more than VADER. Both also highlighted positive elements like international norms and specific calls to action involving the ICJ, ICC, and support for refugees. While we cannot definitively explain the discrepancies, it is noteworthy that Ukraine's speech referenced historical analogies, such as, "it was chosen by someone who is now sitting in a bunker. We know what happened to the person who sat in the bunker in Berlin in May 1945," along with a poignant personal story about a Russian soldier who died in the war. This narrative style, which was unique among the 11 other speeches examined in detail, may have influenced the overall score. In
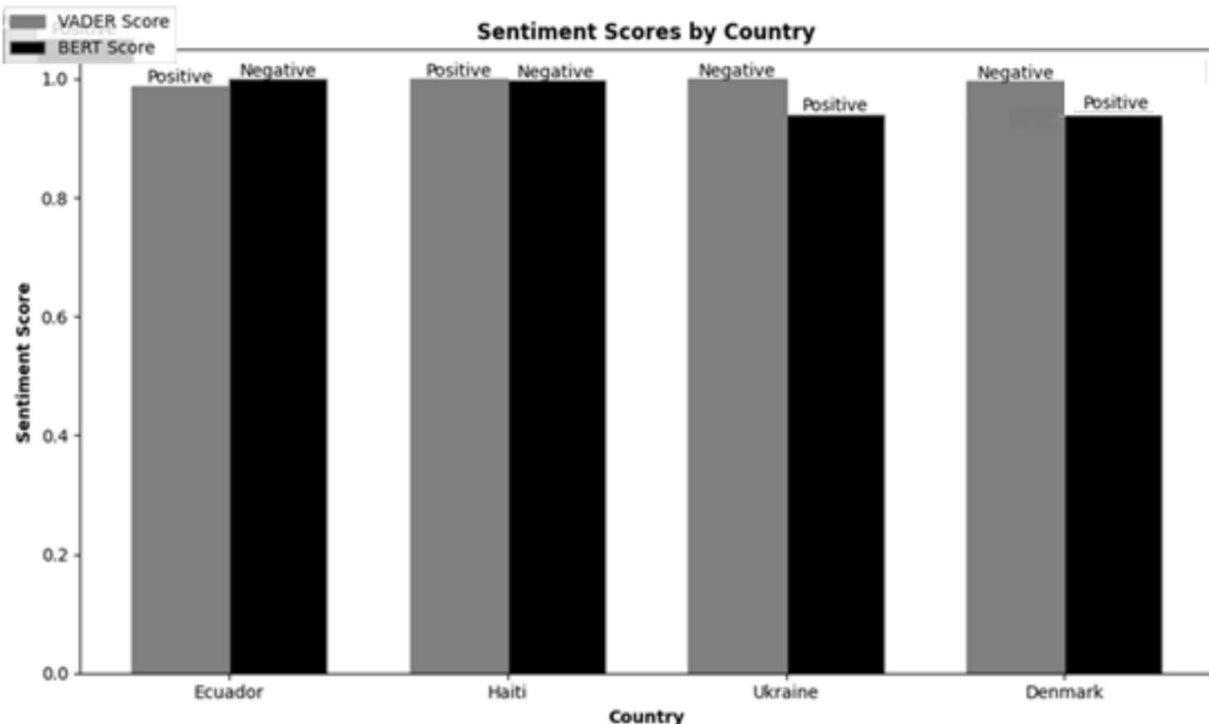

FIGURE 2
Most highly sentiment discrepancies sentiment scores by country.

contrast, Denmark's speech, while generally aligned with negative speech characteristics, included numerous statements about positive actions taken by the international community.

5 Discussion and conclusion

This study explored how global sentiment can be interpreted through a mixed-method approach that combines innovative NLP techniques with narrative analysis. Diplomatic speeches present unique challenges for NLP methods due to their complexity, reluctance to take a definitive stance, and extensive use of historical and cultural context. The study revealed several insights through this mixed-method approach.

First, it highlighted the tendency for diplomatic speeches to convey clear narratives featuring specific settings, actors, plotlines, and recommendations. Second, it demonstrated that speeches embedded in historical contexts, such as the Ukrainian speech, can significantly influence BERT scores; however, such instances are more the exception than the norm. Third, NLP methods uncovered two competing sentiments, which narrative analysis effectively explained by overcoming the complexities involved and clarifying the goals.

Fourth, the meaning of these sentiments plays a crucial role in elucidating the global use of emotional aspects. Positive sentiment speeches generally emphasize international values, norms, institutions, and actors that support Ukraine, along with various solutions to the conflict. In contrast, negative sentiment speeches focus more on detailed accounts of Ukrainian suffering and Russian aggression, including occupation, civilian bombings, and nuclear threats. These findings contribute to our understanding of how constructivist theories and strategic narrative studies can benefit from a mixed-method approach. As demonstrated in this paper, despite challenges to the global order and economic or political ties with Russia, the majority of actors supported the condemnation of Russia and called for adherence to accepted norms, such as state sovereignty, global institutions, the avoidance of international war, and the peaceful resolution of conflicts.

Fifth, while both sentiments contain elements of the other, their scope differs. Contrary to our preliminary hypothesis, positive speeches advocate for a global commitment to the core values of the international community, disregarding the Global North–South divide. Ultimately, we argue that the invasion of Ukraine initially fostered a sense of global crisis that transcended North–South sentiments. This was the prevailing narrative during the UNGA, where countries united to support Ukraine, called for collaborative efforts, and advocated for various measures to end the conflict, all while maintaining a firm stance against Russia without resorting to estranging rhetoric.

Sixth, although VADER employs a predefined lexicon accompanied by heuristic rules, it may encounter difficulties in capturing the subtle nuances, indirect phrasing, and context-dependent meanings that frequently characterize diplomatic discourse. Conversely, BERT, which possesses a superior capacity for understanding surrounding context, still misinterprets historically or culturally significant references within the international dialog. For instance, a statement such as "certain actors have historically disregarded territorial integrity" may be interpreted as neutral by VADER due to the lack of overtly negative terminology. In contrast,

BERT, acknowledging the historical context of aggression, might classify it as negative. Additionally, the use of cultural politeness strategies such as hedging can result in divergences in sentiment classification between these two analytical approaches.

In addition, both VADER and BERT depend on pre-defined lexicons and training datasets that may inadequately capture the subtleties of diplomatic discourse. For example: as diplomacy often utilizes hedging, strategic ambiguity, and indirect criticism, it can result in misclassification by these models.

Lastly, it is essential to acknowledge the challenges posed by NLP methods. Typically, studies employ a single technique; however, we aimed to demonstrate that using multiple techniques can impact results. Although the techniques largely concurred on the prevailing sentiment, we observed that the code used for BERT was less accurate than VADER in evaluating the sentiment of diplomatic speeches despite being regarded as more advanced in NLP analysis. This observation highlights the necessity of critically assessing state-of-the-art NLP methods and their outcomes instead of adopting them unquestioningly.

6 Limitations and recommendations for future research

This study underscores the evolving role of NLP in political science but does not assert that the methods employed are definitive or exhaustive. Continuous development in NLP necessitates ongoing evaluation of methodologies and their inherent biases. Expanding datasets especially to include variations in sentence length could enhance findings and offer insight into global sentiment shifts during conflicts. Such approaches can deepen understanding of how diplomatic speeches shape or challenge global norms, legitimacy, and international alignments.

While NLP effectively captures general sentiment, it may overlook nuanced narratives, necessitating complementary qualitative methods. Exploring correlations between advanced NLP techniques and sentiment analysis could provide a more refined understanding of the emotional undercurrents influencing global politics. Applying this approach to issues like war, economic changes, and climate crises could reveal regional trends and contentious debates. Such approaches, among others, can deepen our understanding of how diplomatic speeches both construct and deconstruct the meaning of global norms, the legitimacy of certain actions, and potentially reveal tensions between realist explanations for particular actions, contrasting them with broader frameworks. These new studies could also shed light on how state blocs respond to the evolving global order, identifying the issues that promote conformity versus those that encourage deviance from accepted norms.

To strengthen validity, future research should involve independent coders and measure inter-coder reliability (e.g., Cohen's kappa). Implementing automated algorithms with majority voting mechanisms that consolidate outputs from multiple NLP models, such as VADER and BERT, may further enhance robustness and reduce model specific biases. This multidimensional framework, benchmarked against human coders, aims to ensure consistent, interpretable sentiment analyses in diplomatic discourse.

A limitation of this study lies in its reliance on VADER and BERT for narrative analysis without extensive statistical validation. Future

studies should compute inter-method agreement and correlation measures to increase analytical rigor. Additionally, addressing multilingual content and translation biases remains crucial. Although this study segmented and tokenized texts, future work should integrate language-specific calibrations and consistency checks to improve accuracy. Finally, refining context-aware embeddings and domain-specific sentiment lexicons can enhance the precision of diplomatic discourse analysis. While VADER excels with short social media texts, BERT offers more nuanced classifications for complex diplomatic language.

In conclusion, this paper's mixed-method approach—combining NLP techniques with sample narrative analysis—holds potential for application in other fields of global politics. For instance, it can be used to explore how identities, norms, and meanings are constructed around issues beyond war, such as global warming, the global economic order, and human rights. Such studies could provide valuable insights into how diplomatic rhetoric co-constructs international relations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

CK: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. MZ: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ajala, O. (2022). The case for neutrality: understanding African stances on the Russia-Ukraine conflict. *J. Mil. Strateg. Stud.* 22, 130–153.

Antonakaki, D., Fregopoulou, P., and Ioannidis, S. (2021). A survey of twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Syst. Appl.* 164:114006. doi: 10.1016/j.eswa.2020.114006

Atmaja, B. T., and Sasou, A. (2022). Sentiment analysis and emotion recognition from speech using universal speech representations. *Sensors* 22:6369. doi: 10.3390/s22176369

Bleiker, R., and Hutchison, E. (2008). Fear no more: emotions and world politics. *Rev. Int. Stud.* 34, 115–135. doi: 10.1017/S0260210508007821

Carrión-Vivar, K. D., Jima-González, A., and Alcántara-Lizárraga, J. Á. (2024). Between tradition and pragmatism: challenges for Latin America amid the Russia–Ukraine war. *Bull. Lat. Am. Res.*:3571. doi: 10.1111/blar.13571

Dempsey, J. (2023). Judy asks: is Russia's war on Ukraine a global war? *Commentary*.

Ellison, J., Cox, M., Hanhimäki, J. M., Harrison, H. M., Ludlow, N. P., Romano, A., et al. (2023). The war in Ukraine. *Cold War History* 23, 121–206. doi: 10.1080/14682745.2023.2162329

Finnemore, M., and Sikkink, K. (1998). International norm dynamics and political change. *Int. Organ.* 52, 887–917. doi: 10.1162/002081898550789

Fisher, S., Klein, G. R., and Codjo, J. (2022). Focusdata: foreign policy through language and sentiment. *Foreign Policy Anal.* 18:002. doi: 10.1093/fpa/orac002

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., and Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Syst. Appl.* 69, 214–224. doi: 10.1016/j.eswa.2016.10.043

Hartley, A. (2022) 'Why so many African leaders support Putin', *The Spectator*. Available at: https://www.spectator.co.uk/article/why-so-many-african-leaders-support-putin/.

Head, N. (2020). "Contesting emotional governance – empathy under fire in the Israeli public sphere during operation protective edge" in The power of emotions in world politics (Abingdon: Routledge), 113–129.

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., and Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: a state-of-the-art review. *Nat. Lang. Process. J.* 6:100059. doi: 10.1016/j.nlp.2024.100059

Kentikelenis, A., and Voeten, E. (2021). Legitimacy challenges to the liberal world order: Evidence from United Nations speeches, 1970–2018. *Rev. Int. Organ.* 16, 721–754. doi: 10.1007/s11558-020-09404-y

Kertcher, C. (2021). The dispositive of terrorism during the war on terrorism: the UNSC's approach to concrete terror emergencies in the Middle East. *Crit. Stud. Terror.* 14, 555–577. doi: 10.1080/17539153.2021.1989130

Kertcher, C. (2024). A thin debate: comparing the conflicting political narratives of the UN security council and general assembly on the Gaza strip from 2006 to 2018. *Int. Politics*. doi: 10.1057/s41311-023-00549-2

Koschut, S. (2018). The power of (emotion) words: on the importance of emotions for social constructivist discourse analysis in IR. *J. Int. Relat. Dev.* 21, 495–522. doi: 10.1057/s41268-017-0086-0

Koschut, S. (2020). The power of emotions in world politics. London, New York: Routledge.

Krebs, R. R. (2015). Narrative and the making of US national security. Cambridge: Cambridge University Press, 138.

Liu, W., and Wang, Y. (2020). The role of offensive metaphors in Chinese diplomatic discourse. *Discourse Context Media* 37:100418. doi: 10.1016/j.dcm.2020.100418

Mah, A., and Song, E. (2024). Elite speech about climate change: analysis of sentiment from the United Nations conference of parties, 1995–2021. *Sustain. For.* 16:2779. doi: 10.3390/su16072779

Mijares, V. M. (2022). The war in Ukraine and Latin America: reluctant support. *GIGA Focus Latin America* 2. doi: 10.57671/GFLA-22022

Miranda, J. P. P., and Bringula, R. P. (2021). Exploring Philippine presidents' speeches: a sentiment analysis and topic modeling approach. *Cogent Soc. Sci.* 7:1932030. doi: 10.1080/23311886.2021.1932030

Miskimmon, A., O'Loughlin, B., and Roselle, L. (2024). "Strategic narratives, soft power, and foreign policy" in The Oxford handbook of foreign policy analysis. eds. J. Kaarbo and C. G. Thies. *1st* ed (Oxford University Press), 445–461.

Niedenthal, P. M., and Brauer, M. (2012). Social functionality of human emotion. *Annu. Rev. Psychol.* 63, 259–285. doi: 10.1146/annurev.psych.121208.131605

Opperman, K., and Spencer, A. (2023). "Narrative analysis" in Routledge handbook of foreign policy analysis methods (London, New York: Routledge), 117–132.

Osnabrügge, M., Hobolt, S. B., and Rodon, T. (2021). Playing to the gallery: emotive rhetoric in parliaments. *Am. Polit. Sci. Rev.* 115, 885–899. doi: 10.1017/S0003055421000356

Parent, J. M., and MacDonald, P. K. (2024). In search of status: China and the USA in United Nations speeches, 1970–2020. *Chin. J. Int. Politics* 17, 1–27. doi: 10.1093/cjip/poae001

Sasley, B. (2020). "Victimhood as power in international conflict" in The power of emotions in world politics (Abingdon: Routledge), 48–64.

Semenov, A., and Tsvyk, A. (2021). The approach to the Chinese diplomatic discourse. *Fudan J. Humanit. Soc. Sci.* 14, 565–586. doi: 10.1007/s40647-021-00321-x

Shah, T. M. (2024). Emotions in politics: a review of contemporary perspectives and trends. *Int. Polit. Sci. Abstracts* 74, 1–14. doi: 10.1177/00208345241232769

Shenhav, S. R. (2015). Analyzing social narratives. *1st* Edn: Routledge.

Surowiec, P., and Miles, C. (2021). The populist style and public diplomacy: kayfabe as performative agonism in Trump's twitter posts. *Public Relations Inquiry* 10, 5–30. doi: 10.1177/2046147X20979294

Wallis, J. (2024). How do the emotional and embodied experiences of international interveners influence their understanding and practice of peacebuilding? *Coop. Confl.* 59, 209–227. doi: 10.1177/00108367231184724

Wankhade, M., Rao, A. C. S., and Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* 55, 5731–5780. doi: 10.1007/s10462-022-10144-1

Wendt, A. (1992). Anarchy is what states make of it. The social construction of power politics. *Int. Organ.* 46, 391–425.

Widmann, T., and Wich, M. (2023). Creating and comparing dictionary, word embedding, and transformer-based models to measure discrete emotions in German political text. *Polit. Anal.* 31, 626–641. doi: 10.1017/pan.2022.15

Wiener, A. (2018). Contestation and constitution of norms in global international relations: Cambridge University Press.