

**Borys Grinchenko Kyiv Metropolitan University
Faculty of Romance and Germanic Philology
Linguistics and Translation Department**

**U.S. PRESIDENTS' RHETORIC ON WAR IN UKRAINE:
A CORPUS-BASED COMPARISON OF JOE BIDEN'S AND
DONALD TRUMP'S SPEECHES**

MA Paper
Anastasiia Antoniuk
Group MLAm-1-25-2.0d

Research supervisor
Professor A.V. Chesnokova

Цим підписом засвідчую, що подані на захист
рукопис та електронний документ є ідентичні.
26.05.2026



Approved for defense at
the Linguistics and
Translation Department
26, May, 2026, Minutes 5
_____ Andrii KOZACHUK

Kyiv 2026

This paper presents the lexis used in Joe Biden's and Donald Trump's speeches on war in Ukraine, through a corpus analysis conducted using Voyant Tools software, which enables multilevel text analysis and Lingualyzer, a text analysis tool capable of evaluating hundreds of linguistic metrics. The paper is aimed to define lexical features and the frequency of lexical units in President Biden's and President Trump's speeches regarding the war in Ukraine. Furthermore, this study aims to assess how their specific linguistic choices measured against Douglas Biber's cross-linguistic limits shape political narratives and public perception of the Russian invasion. This study examines 8 speeches by President Biden (a total of 4,748 words) and 13 speeches by President Trump (a total of 4,971 words), as they reflect Americans' views on the events. The corpus is divided into two sub-corpora: speeches delivered by Joe Biden and Donald Trump, respectively. However, the Trump subcorpus is further divided into two sections: speeches delivered before he became president and speeches delivered after his election and inauguration. Additionally, a subcorpus from social media platform "X," consisting of four online posts, was compiled and analyzed. These findings point to the different rhetorical and narrative strategies employed by both leaders when addressing the conflict.

Keywords: *corpus analysis, President Biden, President Trump, Ukraine, war.*

TABLE OF CONTENTS

1. INTRODUCTION.....	4
2. THEORETICAL BACKGROUND.....	6
3. RESEARCH METHODS.....	12
3.1. Research design.....	13
3.2. Corpus architecture and speech selection.....	14
3.3. Voyant tools.....	14
3.4. Lingualyzer.....	15
3.5. Framework.....	16
3.6. The frequency of parts of speech in the English language by Douglas Biber.....	17
4. RESULTS.....	18
5. CONCLUSIONS AND DISCUSSION.....	47
6. REFERENCES.....	51

1. Introduction.

The foundations of corpus linguistics were established about a century ago. In the late 19th century, lexicographers also known as dictionary compilers collected examples of word usage to provide more precise definitions of terms. Before the advent of computers, researchers collected linguistic examples on small pieces of paper and stored them in boxes. The advent of computing technology paved the way for what researchers now call modern corpora. The first digital corpus, the “Brown Corpus,” appeared in 1961 and contained approximately one million words. Modern general corpora now contain hundreds of millions of words, and corpus linguistics makes a significant contribution to contemporary language research and teaching.

Corpus linguistics examines language through the analysis of a corpus (plural: corpora) (Hammond & Macken-Horarik, 1999). A corpus functions as an electronic database of authentic linguistic data that meets specific criteria. Essentially, corpus linguistics addresses two central questions:

- 1) What patterns emerge in relation to specific linguistic or lexical features?
- 2) What variations and registers shape these patterns?

Scholars must select data that precisely aligns with their core research interests to avoid conclusions that go beyond what their sample actually supports. Building a reliable corpus is a systematic, multi-stage process. This process requires researchers to develop a methodology, collect text samples, code data, and organize structural metadata (McEnery and Hardie 2012:241). At each stage of this process, researchers face a series of critical methodological decisions that ultimately determine the usefulness and validity of the final dataset. A research corpus is more than a random collection of texts; it is based on a specific digital framework. This term describes how the system conceptually separates and stores different layers of data, namely the main text and its general metadata. In basic configurations, structural markup simply isolates metadata and maintains the order of text arrays. However, as analytical demands grow, these structural frameworks form a complex hierarchy of information that can be computationally represented in several different ways, each with its own methodological trade-offs. Scholars such as Filter, Bieber, Johansson, Francis, Houston, Conrad, and McCarthy have contributed to the development of modern corpus linguistics, and their work continues to shape the field (Römer, 2011). John Sinclair is recognized as one of the most important and influential figures in this field of research (Stubbs, 2007).

Scholars argue that corpus studies go beyond the identification of quantitative patterns and include functional interpretations that explain the underlying causes of observed phenomena. For this reason, corpus-based research places significant emphasis on interpreting and illustrating quantitative

trends. Research in corpus linguistics typically combines both quantitative and qualitative methods, and researchers often move from quantitative results to qualitative conclusions (McIntyre & Walker, 2019). This structural study begins by considering corpora as organized sets of documents, often sorted into distinct subcategories and paired with corresponding metadata (such as author demographics, dates, or genre classifications). Within each individual document, the main challenge lies in how we represent the raw data itself. This includes standard text, transcribed dialogues, and multimodal data streams. Standard text models assume a clean, sequential chain of words, but real-world data often violates this assumption.

Tools such as Voyant Tools software are used for corpus analysis of texts. This open-source web-based text analysis environment supports the study of individual texts and corpora and serves researchers, students, and a broader audience. Users can explore online materials or upload their own texts, making the platform particularly useful for work in the field of digital humanities. Such analysis allows not only for refining and delving deeper into the findings of existing research but also for taking a fresh look at many seemingly traditional linguistic concepts and terms and redefining them.

The Lingualyzer program (Chesnokova & Zingier, 2024) was used to analyze the corpus and calculate the frequency of occurrence for each part-of-speech category. In accordance with standard linguistic models, parts of speech were divided into two functional groups:

Lexical (content) parts of speech: this category includes nouns, verbs, adjectives, and adverbs. These elements provide specific semantic information and the thematic essence of utterances.

Functional (grammatical) words: this category includes pronouns, prepositions, determiners, and conjunctions. These words serve as the structural “glue” of discourse.

FREQUENCY OF PARTS OF SPEECH			
		Conversation	Academic
Lexical	Nouns	15%	30%
	Verbs	12.5%	10%
	Adjectives	2.5%	10%
	Adverbs	5%	3%
Function	Pronouns	16.5%	4%
	Prepositions	5.5%	15%
	Determiners	4.5%	10%
	Conjunctions	4.5%	5%

Figure 1.1. The frequency of parts of speech in English language by Douglas Biber

The primary method for measuring lexical density in this study is determining the percentage ratio of parts of speech. Many lexical parts of speech, especially nouns and adjectives, are commonly found in the academic or “informative” register, which is typical of technical or legal reports. On the other hand, a high number of function words (such as *I*, *we*, and *they*) indicates that the register is conversational or “engaged,” which is typical of political speeches that seek to persuade or connect with people.

2. Theoretical background.

A variety of English language text corpora capturing the twentieth century have been compiled, primarily to facilitate grammatical research; however, these resources frequently present methodological limitations, as they are either restricted from public access or confined to a highly specific textual genre. For instance, the ARCHER corpus (A Representative Corpus of Historical English Registers) exemplifies a multi-genre database covering historical, medical, drama, and news reportage texts from 1650 to 1990 in fifty-year increments, yet its utility is constrained because it remains unavailable to the broader scientific community (Leech & Smith, 2005). Conversely, the Corpus of Late Modern English Prose is publicly accessible to researchers, but it is limited to a single genre – informal personal letters and its temporal focus aligns more closely with the nineteenth century than the twentieth (Denison, 1994). Other datasets sit at the intersection of both limitations, suffering from restricted access while simultaneously focusing on a narrow genre. This category includes the Corpus of English Newspaper Editorials (CENE), which samples institutional editorials from three British broadsheets such as *The Times*, *The Guardian*, and *The Daily Telegraph* – at decadal intervals across the twentieth century (Westin, 2002; Westin & Geisler, 2002; Leech & Smith, 2005), as well as Bauer’s corpus of *The Times*, which similarly isolates newspaper editorials sampled by decade (Bauer, 1994; Leech & Smith, 2005).

To overcome these barriers of restricted accessibility and genre isolation, the “Brown family” of corpora was developed, comprising five mutually comparable, multi-genre datasets that are publicly accessible through the ICAME Corpus Collection. This architectural framework is structurally divided into American and British components. The American sub-corpora include the Brown University Corpus of Written American English, representing language from 1961 (Francis, 1965), and the Freiburg-Brown Corpus of American English (Frown), capturing language from 1992 (Hundt et al., 1998). The British section consists of three sequential databases: the Lancaster 1931 corpus (BLOB), covering texts from 1931 - 3 years (Leech & Smith, 2005); the

Lancaster-Oslo/Bergen Corpus (LOB), tracking text from 1961 (Johansson et al., 1978); and the Freiburg-LOB Corpus of British English (FLOB), representing language from 1991 (Sand & Siemund, 1992). Across these five components, the texts are distributed equally among fifteen distinct communicative genres. Because all five databases were constructed using an identical sampling frame, topical distribution, and dimensional balance, they satisfy the necessary parameters of empirical representativeness and strict comparability required for diachronic research of twentieth-century written English.

The methodological precedent for this family was established with the publication of the original one-million-word Brown corpus in 1964, which was engineered to provide a standardized, common linguistic baseline upon which diverse, reproducible studies could be executed (Francis, 1965; Leech & Smith, 2005). Its design consists of 500 randomly extracted text segments of roughly 2,000 words each, systematically distributed across fifteen genres. To replicate this standard, the Lancaster-Oslo/Bergen corpus was compiled using an identical sampling framework, genre distribution, and chronological baseline of 1961, thereby enabling the first comprehensive, multi-genre synchronic comparison between British and American English (Johansson et al., 1978; Leech & Smith, 2005). During the 1990s, researchers at Freiburg University extended this design by compiling the Freiburg-LOB Corpus of British English and Frown corpora to represent written British and American English from 1991 and 1992, respectively. By meticulously mirroring the original Brown and Lancaster-Oslo/Bergen structural architectures, these new additions permitted scholars to model and compare short-term diachronic shifts across both primary national varieties (Leech & Smith, 2005). This model was subsequently pushed backward in time at Lancaster University with the compilation of the Lancaster 1931 corpus; this database targets a sampling point of 1931 to deliberately sustain the precise thirty-year temporal interval established between the 1961 databases and their 1990s counterparts.

Because the internal parameters of the entire Brown family are tightly controlled and mutually comparable, the collective framework facilitates several distinct dimensions of empirical inquiry. On a cross-sectional axis, it permits synchronic comparison between contemporary British and American English at two distinct historical junctures: 1961 and the early 1990s. On a longitudinal axis, the uniform architecture enables diachronic investigation across successive generations of language users, mapping the evolution of British English across a sixty-year span (1931, 1961, and 1991) and American English across a thirty-year timeline (1961 to 1992). Ultimately, by cross-referencing these axes, researchers can execute highly controlled comparative analyses to determine whether diachronic trajectories of language change operate uniformly or divergently across different geographical varieties of the same language during identical historical eras.

The methodological framework relies on corpus linguistics and the statistical analysis of keywords and collocations. Characterized by machine readability, authenticity, and representativeness, corpus linguistics allows critical discourse analysis to examine how specific subjects are discussed, investigate the idiolects of individual politicians, perform corpus comparisons, and isolate linguistic features such as keywords and contextual collocations (O’Keeffe & McCarthy, 2014).

The integration of CL and CDA can be operationalized through two approaches: a corpus-based CDA, which deductively uses the corpus to find linguistic evidence for a predetermined hypothesis, and a corpus-driven CDA, which inductively uses corpus data to generate new hypotheses (Baker, 2012). Specifically, a corpus-driven approach is used for the initial extraction of keywords and high-frequency lexical items, while a corpus-based approach allows for a detailed analysis of the collocational environments surrounding those keywords. Synthesizing these methods with the discourse-historical approach (DHA) minimizes researcher bias and subjectivity while providing a robust basis for qualitative and quantitative analysis of national identity discourse. This study draws its theoretical and methodological foundation from the work of Baker (2006, 2012), O’Keeffe (2006), McEnery and Hardie (2012), Haider (2017), Wodak and Meyer (2016), and Allan (2018).

The analytical focus of this investigation is confined strictly to the vocabulary level of linguistic variation, specifically tracking how lexical density and lexical richness fluctuated across the twentieth century. Lexical density stands as one of the most widely applied metrics for evaluating the diversity of a given vocabulary (Stamatatos et al., 2000), and it has been utilized in prior scholarship to assist in the chronological dating of literary works (Smith & Kelly, 2002). Mathematically, lexical density is operationalized as the ratio of unique word types to the total number of tokens within a text; consequently, a higher density value signifies a more expansive and varied vocabulary range.

However, a notable limitation of standard lexical density is that it treats distinct morphological variations of the same underlying lexeme as separate word types. To address this issue, Corpus Pastor et al. (2008) recommended the alternative application of lexical richness. This metric is calculated by dividing the total number of unique lemmas by the overall token count in a target text. By utilizing lemmatization, this second measurement effectively neutralizes the distorting effects of morphological inflection, ensuring that different grammatical variants of a single word are not erroneously counted as unique vocabulary items. For this reason, Corpus Pastor et al. (2008) maintained that lexical richness serves as a more accurate and reliable index for capturing the true extent of an author’s vocabulary variety.

In the comprehensive overview of this field, a corpus-based discourse analysis of Volodymyr Zelenskyy’s wartime rhetoric, Svitlana Goloshchuk and Justyna Tomczak-Boczko describe how the escalation of the conflict in Ukraine

has spurred the emergence of a substantial body of scholarly work devoted to the analysis of contemporary war discourse. A significant portion of this research focuses on the linguistic strategies employed in media coverage of events. For example, Natalia Petiy examined cognitive metaphors in English-language media, demonstrating that the conflict is often conceptualized either through the lens of a game or as a binary struggle between good and evil. In addition to general media reports, researchers have also closely examined the rhetoric of key political figures; specific verbal and nonverbal components of Volodymyr Zelenskyy's political speeches were analyzed to illustrate how his leadership style is projected to domestic and international audiences during the conflict.

Methodological advances in this field often employ corpus linguistics to identify thematic and syntactic patterns in international media. Lőrincz used a corpus-based approach to compare Eastern and Western English-language news releases, highlighting their main thematic focuses. In particular, her syntactic analysis revealed that the lemmas *Putin* and *Putin's* predominantly function as subjects or attributes in combination with the term *war*, whereas *Zelensky* more often occupies the position of an object. This syntactic distribution implies that Putin is systematically presented as an agent or initiator of the conflict.

Other comparative studies have highlighted ideological nuances in media framing. Yiğit Salihoğlu and Çiğdem Karatepe examined the online coverage of the war by BBC News and Al Jazeera English, noting a tendency in both visual and textual language to soften the most painful realities experienced by Ukraine and its population. Expanding their analytical scope to social media corpora, Rahat Gulzar et al. tracked public sentiment on Twitter (now X), finding a predominantly negative tone across the dataset alongside a steady decline in the total volume of tweets over time. Similarly, Baladrón-Pazos et al. compiled a corpus of Spanish political tweets, finding that party communication regarding the war generally adhered to standards of moderation and political correctness.

Finally, a separate line of research focuses on comparative media coverage and the direct rhetoric of opposing heads of state. Using multimodal critical discourse analysis (CDA), Raza and Malik examined the portrayals of both leaders on the BBC and concluded that the coverage lacks neutrality; instead, the broadcaster constructs a dichotomy that presents Putin as an inflexible, extremely powerful threat, while portraying Zelenskyy as an approachable yet resolute leader in the midst of a geopolitical crisis. Using a more direct rhetorical approach, Innocent Chiluwa and Jurate Ruzaitė compared Putin's and Zelenskyy's war speeches, while Gregić and Božić assessed their crisis management strategies through a combined analysis of their verbal and nonverbal communication.

The cognitive metaphor theory, according to Petiy, pioneered by Lakoff and Johnson (1980), posits that human thought processes are fundamentally metaphorical, wherein abstract conceptual domains are systematically mapped onto and understood through more concrete physical experiences. While

cognitive metaphor theory historically relied on intuitive or introspective linguistic examples, contemporary discourse analysis increasingly demands empirical validation. Consequently, cross-disciplinary research at the intersection of linguistics, psychology, and cognitive science has shifted toward identifying these conceptual mappings within naturally occurring, large-scale language data (e.g., Kövecses, 2002; Thibodeau et al., 2019).

When applied to political discourse and geopolitical conflicts, cognitive metaphor theory provides a vital framework for decoding how military actions are morally justified, structurally framed, and ideologically packaged for the public (Lakoff, 1992; Sabbah, 2009). However, to systematically capture how these mental models operate across entire media landscapes such as the representation of the Syrian refugee crisis (Salem et al., 2022) or the war in Iraq scholars require a methodology that scales beyond isolated textual examples. This is where the tools of corpus linguistics become indispensable. In contemporary media analysis, corpus linguistics serves as an empirical bridge to cognitive metaphor theory. While cognitive theory defines what mental metaphors are, corpus analysis provides a suite of statistical tools such as collocations, frequency counts, and keyword-in-context concordance to demonstrate how frequently and in which precise lexical environments these metaphors are deployed.

The quantitative textual analysis for this study was conducted using two primary computational tools: Voyant Tools and Lingualyzer .

To complement the linguistic profiling, Voyant Tools was employed as an analytical resource. Voyant Tools is a web-based, open-source text analysis environment designed to facilitate the macro-level interpretation of texts and digital corpora for academic researchers, students, and the general public. The platform enables users to either input custom datasets or analyze online textual content directly, making it a widely recognized utility within digital humanities research for extracting corpus statistics, word frequencies, and distributional patterns.

Lingualyzer functions as an automated quantitative content analysis platform engineered to evaluate textual data across hundreds of distinct linguistic metrics at multiple textual levels. Capable of processing 41 different languages, the software is frequently utilized for a broad spectrum of linguistic investigations, including cross-linguistic and comparative discourse studies. The platform was developed and designed by Guido M. Linders and Max M. Louwse, with additional development input from Kirill O. Mitev (Tilburg University). The computational engineering and design of the tool were funded through research grants awarded to Max M. Louwse, supported by the European Union, OP Zuid, the Dutch Ministry of Economic Affairs, the Province of Noord-Brabant, and the Municipality of Tilburg.

Scholarly analysis of presidential rhetoric frequently utilizes classical linguistic frameworks to decode political intent. Yang et al. investigated the

inaugural address of President Rodrigo Duterte through Searle's Speech Act Theory to evaluate how language is deployed to persuade, inform, and establish commitments. Their framework relied on five speech act classifications: assertives (affirming facts), directives (eliciting action), commissives (promising future action), expressives (conveying emotional states), and declarations (executing institutional change). Utilizing the official transcript, the authors manually coded each clause, finding that commissives were the most prevalent (45%), followed by assertives (27.5%) and directives (15%). While these proportions align with the standard expectations of an inaugural address, the study was constrained by the subjectivity inherent in manual coding, as the results were subject to personal interpretation. An equivalent speech act analysis targeting historical US presidential speeches was executed in.

Other lines of research leverage digital corpora to identify long-term thematic and emotional shifts in executive communication. Nicole et al. examined the historical evolution of US presidential language by compiling a corpus from the University of Virginia's Miller Center. Following text vectorization, they applied topic modeling and sentiment analysis, observing clear chronological patterns, such as the emergence of slavery as a central topic during the Civil War era. Their sentiment analysis revealed that President Dwight Eisenhower achieved the highest positivity metrics, while presidents with lower public approval ratings systematically registered lower sentiment scores. Furthermore, comparative rhetorical strategies among contemporary political figures have been highlighted by Jiménez-Preciado et al. Their analysis of the campaign rhetoric of Donald Trump, Joe Biden, and Kamala Harris demonstrated that Trump utilized significantly more emotionally charged language and exhibited lower lexical diversity compared to his political opponents.

Recent scholarship has also explored the capacities and boundaries of Natural Language Processing in measuring political tone and polarization. Nemeth et al. provided a scoping review on the application of NLP tools to assess language polarization, noting that a speaker's political stance is heavily contingent not only on individual authorship but also on systemic variables, such as whether their political party is currently in power. They concluded that integrating multi-type datasets serves as an effective strategy to mitigate these analytical limitations. Conversely, Rydeen et al. encountered distinct methodological boundaries when evaluating the rhetorical tone of presidential State of the Union addresses. Although they anticipated a direct correlation between a president's political party and their tone on specific subjects, their analysis was hindered by the complex nature of human emotion and textual data. A primary limitation was the structural homogeneity of the dataset itself; because State of the Union addresses are delivered annually according to predictable institutional standards, the authors lacked the diverse contextual data necessary to build individualized profiles for specific presidents.

To counter these quantitative limitations, some researchers adopt specialized qualitative models. Haiyue et al. evaluated Joe Biden's State of the Union addresses through the framework of spatial and temporal proximization theories. They noted that Biden operationalized spatial proximization by introducing external threats to instigate anxiety among the public, while using temporal proximization by referencing specific past or imminent occurrences to reinforce this emotional effect. However, because this study was strictly qualitative rather than quantitative, the authors acknowledged that the empirical foundation of the data was somewhat restricted.

While prior scholarship has extensively investigated historical presidential discourse or analyzed general media framing of modern conflicts, there remains a critical empirical void regarding the direct, comparative rhetoric of active heads of state during contemporary warfare. My study directly addresses this gap by presenting an unprecedented, comparative corpus-linguistic analysis of the wartime speeches of President Joe Biden and President Donald Trump. Consequently, my paper offers a first-of-its-kind, data-driven framework for understanding how language, ideology, and cognitive framing are weaponized at the highest institutional levels during the ongoing war in Ukraine.

3. Research Methods.

This research employs a corpus-assisted interpretive approach, bridging the empirical strength of corpus linguistics with the contextual awareness of political discourse analysis. Instead of viewing corpus tools simply as objective counting mechanisms, this study treats them as heuristic guides for systematically analyzing the emotional undertones of political language (Baker, 2006; McEnery & Hardie, 2012). This methodology is useful for exploring emotionally charged political rhetoric, where affective meaning is embedded within widespread, repetitive lexical patterns rather than isolated instances of figures of speech.

This study utilizes a mixed-methods research design, integrating quantitative corpus metrics with qualitative political discourse analysis to examine the thematic frameworks and rhetorical positioning in the speeches of President Joe Biden and President Donald Trump regarding the war in Ukraine. The quantitative phase leverages Voyant Tools, a web-based digital humanities platform, to calculate word frequencies, generate word clouds, illustrative trends and map the chronological distribution of key lexical items across both text corpora. These computational outputs serve as an empirical foundation to isolate dominant themes, ideological markers, and key points within their respective political rhetoric. Conversely, the qualitative phase involves a close contextual and critical interpretation of these visual and statistical patterns, evaluating how they reflect competing political agendas and foreign policy doctrines. By utilizing Voyant Tools' interactive suite, which extends beyond basic frequency

tallies to encompass trend graphs and contextual concordances, this approach enables a systematic, empirical exploration of the linguistic structures shaping contemporary geopolitical discourse. Another micro-quantitative analysis is based on *Lingualyzer*, a computational linguistics platform designed for quantitative text analysis across hundreds of metrics. Built specifically for behavioral and cognitive science researchers, it facilitates multilingual and cross-linguistic studies across 41 languages without requiring advanced programming expertise.

3.1 Research Design.

This research adopts a mixed-methods, corpus-assisted discourse analysis design where quantitative patterns serve as the foundation for qualitative interpretation. The empirical basis of this study relies on two self-compiled corpora containing the speeches of Joe Biden and Donald Trump regarding the war in Ukraine. Adopting this hybrid paradigm addresses the inherent limitations of single-method approaches: purely qualitative analyses frequently suffer from impressionism, while strictly quantitative studies risk separating data from its context (Charteris-Black, 2011; Partington, Duguid, & Taylor, 2013).

During the initial stage of the data collection process, a corpus of speeches delivered by Joe Biden was compiled, hereafter designated as Corpus 1 (a total of 4,748 words). The primary source for these texts was the Miller Center archive, which hosts an extensive repository of presidential transcripts. During the text preparation stage, the corpus was meticulously cleaned to eliminate formatting inconsistencies, editorial notes, and irrelevant metadata, ensuring the integrity and accuracy of the subsequent analysis. From this database, eight speeches were selected using purposive sampling, spanning from the onset of the full-scale invasion of Ukraine in February 2022 until Biden's final address as president in September 2024. Specifically, the material comprises speeches delivered on February 24, 2022; March 1, 2022; March 26, 2022; September 21, 2022; February 21, 2023; October 20, 2023; March 7, 2024; and September 24, 2024. These selected texts serve as the primary empirical material for this component of the study.

Conversely, Corpus 2 encompasses the public discourse of Donald Trump, totaling 4,971 words. Given Trump's extensive reliance on digital media as a primary communicative channel, this corpus is structured into three distinct sub-corpora to capture platform-specific and temporal shifts: the Pre-presidential (Pre-POTUS) period, the Presidential (POTUS-Era) period, and a dedicated microblogging sub-corpus consisting of posts extracted from X (formerly Twitter). In contrast to the uniform nature of Corpus 1 (or the Joe Biden corpus), Trump's communicative material exhibits substantial stylistic, structural, and delivery variation, thereby providing a distinct analytical framework. This categorization allows for the identification of lexical changes, thematic accents,

and stylistic shifts across different time periods and communication platforms. Because political rhetoric is heavily influenced by evolving social, political, and technological contexts, this tri-part division helps isolate key linguistic transformations and trace variations in word frequency across nearly half a century of the Presidents' public discourse. The analysis will focus on adapting discursive strategies over decades, linguistic patterns, and word frequency changes. This division provides a plainer and cleaner picture of how the American politician's communication style with society has changed over his extensive public career. Furthermore, this division highlights the continuity of the Presidents' political discourse and its transformation. This data collection approach allows the further analyses of the political discourse transformation and focuses on maximum precision.

The sub-corpora are arranged in strict chronological progression. The Pre-presidential (Pre-POTUS) sub-corpus captures the period when Joe Biden was the incumbent executive and comprises three interviews delivered on March 7, September 23, and September 26, 2024, totaling 936 words. The Presidential (POTUS-Era) sub-corpus begins following Trump's electoral victory in November 2024 and subsequent inauguration, spanning ten speeches delivered between January 23, 2025, and March 17, 2025. Finally, the digital communications sub-corpus consists of four targeted posts published on X between February 19 and March 7, 2025.

3.2 Corpus architecture and speech selection.

The dataset foundation of this study is comprised of three distinct datasets structured to capture the rhetorical and platform-specific nuances of contemporary American political discourse. This includes a primary corpus of Joe Biden's presidential speeches, a corpus of Donald Trump's political speeches (subdivided into pre-POTUS and POTUS periods), and a dedicated sub-corpus of Trump's X (formerly Twitter) data. Rather than relying on random sampling, the texts within these three datasets were gathered through a targeted, purposive sampling strategy based entirely on their direct relevance to the research objectives. The primary inclusion criterion dictated that the texts must explicitly mention or address the war in Ukraine. This strict topical focus ensures that the compiled corpora qualifies as a specialized corpus which, as argued by Flowerdew (2012), is exceptionally effective for uncovering dense rhetorical and emotional patterns in political language. The limited and highly focused scope of the data is therefore a deliberate analytical choice, ensuring that the resulting datasets contain the necessary lexical density required for a fine-grained exploration of emotionally and ideologically salient lexis across different timelines and digital mediums.

3.3 Voyant Tools.

Initially, Voyant Tools is utilized to construct global lexical profiles for each speech within the corpora. Methodological precedents indicate that visualization-oriented platforms are highly efficient at isolating prevailing affective clusters and thematic densities within political communication (Sinclair, 2004; Rockwell & Sinclair, 2016). Within the framework of this research, Voyant Tools serves a threefold exploratory purpose: it facilitates the detection of emotionally charged, high-frequency lexical items; maps the spatial concentration and dispersion of keywords across the texts; and allows for a preliminary delineation between optimism-oriented and resistance-oriented vocabularies. Consequently, this initial exploratory phase serves a heuristic function, generating baseline analytical hypotheses that are subsequently scrutinized using more precise corpus linguistic techniques.

Frequency data is instrumental in assessing the relative prominence of linguistic features within a specific discourse domain, as well as enabling cross-corpus comparisons. A case in point is illustrated by comparing the 5 most frequent nouns from two distinct corpora, each containing 8 and 13 speeches by Joe Biden and Donald Trump. Juxtaposing these ranked lists uncovers predictable lexical overlaps (such as *people, war, Ukraine*), while simultaneously exposing the unique rhetorical priorities of each speaker. Biden's list underscores themes of the countries (*Ukraine, Russia, world*), war, and references to the rival (*Putin*). Conversely, Trump's vocabulary heavily features Ukraine's national security (*war, Ukraine*), and a notable tendency to use verbs (*think, say*).

While these raw frequencies representing the absolute counts of items within the data allow for straightforward ranking and preliminary insights, they possess inherent limitations when comparing datasets of unequal volumes. To accurately evaluate whether a linguistic feature is statistically more or less prevalent across texts of differing sizes, normalising the frequency counts becomes imperative. Significantly, tracing the emergence and obsolescence of specific lexical items serves as a metric for assessing shifting thematic priorities over time, illustrating how certain issues gain prominence while others recede from public discourse. By leveraging the analytical capabilities of Voyant Tools – specifically its trend mapping, word cloud generation, and collocation graphs – this study extracts actionable visual data. Ultimately, these corpus linguistics visualizations offer a macro-level perspective on the rhetorical trajectory characterizing the Biden and Trump datasets.

3.4 Lingualyzer.

Lingualyzer (Linders & Louwse, 2023) is a web-based, freely accessible computational linguistics application designed to support multilingual and

multidimensional quantitative text analysis. Developed at the intersection of cognitive science and artificial intelligence, the tool serves a vital heuristic function in the digital humanities. It bridges the gap between complex computational text-processing pipelines and qualitative text hermeneutics by automatically converting unstructured, unprocessed text into quantifiable, theoretically motivated variables.

Structurally, Lingualyzer operates across a broad array of linguistic dimensions and levels of textual analysis. Unlike traditional corpus tools that focus narrowly on syntax or raw word counts, Lingualyzer extracts measures across three core strata simultaneously: the lexical, the syntactic, and the semantic. Furthermore, the application processes data dynamically, computing these multidimensional metrics at the individual sentence level, the paragraph level, and the global document level. This layered architecture allows researchers to systematically map subtle shifts in a text's rhetorical density, grammatical complexity, and thematic trajectories.

The primary method for measuring lexical density in this study is to determine the percentages of parts of speech. Many lexical parts of speech, especially nouns and adjectives, are typically found in an academic or "informational" register, which is common in technical or legalistic reporting. On the other hand, a lot of function words (like *I*, *we*, and *they*) indicate that the register is conversational or "involved," which is common in political speeches that try to persuade or connect with people.

3.5 Framework.

The analytical framework adopted in this study operates on the principle that emotional meaning within political discourse is co-constructed through the intersection of linguistic structures.

The dimension establishes the study's quantitative foundation by extracting recurring lexical configurations across the three compiled datasets (Biden's speeches, Trump's multi-era speeches, and Trump's X sub-corpus). Raw frequency metrics isolate sustained affective keywords, while Key-Word-in-Context and collocation analyses expose how emotional meanings are forged via predictable lexical co-occurrences rather than isolated word choices. Building upon the corpus-derived data, the tier provides an interpretive mechanism to evaluate textual patterns based on affective and moral alignment. Affect analysis categorizes positive or negative connotations, evaluating how sentiment is sustained or regulated across formal speeches and microblogging platforms. Judgment and moral evaluation analyses unpack how social factors and geopolitical behaviors are evaluated, revealing how specific vocabulary is mobilized to legitimize foreign policy positions, construct ethical hierarchies, and engineer audience alignment.

By synthesizing macro-level data visualization through word clouds, temporal keyword trends, this methodology demonstrates how nuanced frameworks can be empirically mapped without reducing complex political rhetoric. Furthermore, this study breaks new ground by incorporating Biber's (1999) foundational framework on the grammatical distribution and frequency of parts of speech in the English language. Utilizing Biber's multidimensional approach allows the research to systematically measure grammatical variations across genres, providing an empirical benchmark to contrast the formal, structural parameters of presidential addresses against the highly colloquial, unmediated nature of digital microblogging. This hybrid design positions the study within the expanding domain of corpus-assisted emotion analysis. By applying this platform-sensitive lens to three distinct datasets (Biden's addresses, Trump's multi-period speeches, and Trump's X sub-corpus), the research responds directly to ongoing calls for more methodologically transparent and theoretically informed corpus designs within political linguistics (Partington et al., 2013; Baker & McEnery, 2015).

3.6 The frequency of parts of speech in the English language by Douglas Biber.

To establish a structural comparison between the compiled datasets, this study utilizes the foundational part-of-speech framework established by Biber et al. (1999). Importantly, Biber et al.'s baseline findings are derived from the 40 million-word Longman Spoken and Written English (LSWE) corpus, a comprehensive empirical repository comprising 37,244 distinct texts balanced across four core registers of approximately 5 million words each: conversation, fiction, news, and academic prose. For my research 'academic' is the closest register to 'formal', which Biber does not have. Besides, 'conversation' and 'academic' are two extremes, so my analysis neatly situates each of my corpora in between the two. To ensure mathematical comparability across asymmetrical datasets, all findings within the LSWE framework are normalized to a standard frequency of occurrence per 1 million words. For the purposes of this study, the analysis isolates and focuses exclusively on two of these core registers: conversation and academic prose, the latter of which is operationalized within this research as the baseline for a formal written register.

Within the context of this study, Biber et al.'s reference norms for academic prose are utilized to represent the structural behavior of formal, informational written texts, while their conversation metrics represent spontaneous, interactive oral production. Biber et al.'s empirical standards demonstrate that specific grammatical distributions serve as reliable indicators of these registers; formal academic prose heavily favors nominal density, with nouns comprising approximately 30% of the text and verbs compressing to roughly 10%. Conversely, conversational registers favor interactive engagement,

marked by a lower noun density (15%), heightened verb frequencies (12.5%), and a significant reliance on pronouns (16.5%) to facilitate real-time communication.

The discourse is composed of content words, which include nouns, verbs, adjectives, and adverbs. Their arrangement in this speech demonstrates a calculated balance between action-oriented oratory and factual clarity. To identify the specific register of this discourse, a mathematical analysis of nominal density was performed by dividing the frequency of nouns by the total word count. To ensure analytical precision across the extensive dataset, artificial intelligence (specifically ChatGPT) was utilized to convert raw part-of-speech frequencies into standardized percentages and map these findings against established linguistic benchmarks. The specific computational instruction executed was: *'divide the cumulative sum of each part of speech by the total token count of the text to calculate the final relative percentages.'* Following the initial extraction of grammatical metrics from Lingualyzer, each part-of-speech variable was processed through the AI interface and systematically cross-examined in alignment with Biber's foundational part-of-speech framework.

4. Results.

Voyant Tools analysis of Joe Biden's speeches .

At the initial stage of the research, speeches given by the former President of the United States, Joe Biden, were collected, which in what follows is labelled as Corpus 1. The Miller Center contains a large collection of meeting transcripts. 8 speeches dated from the beginning of the full-scale invasion, February 2022, until Biden's last speech in the position of a country leader, September 2024, were selected. The speeches delivered on February 24, 2022; March 1, 2022; March 26, 2022; September 21, 2022; February 21, 2023; October 20, 2023; March 7, 2024; and September 24, 2024, will serve as the primary material for this study.

Word cloud.

The word cloud aggregates the most commonly utilized language within the corpus and facilitates the recognition of persistent trends in lexical selection. The words such as *Ukraine (130)*, *people (38)*, *Russia (35)*, *world (33)*, and *Putin (31)* are the most often occurring ones in the corpus. *Ukraine (130)* is the most common lexeme, as can be seen in Figure 3.1 below, reflecting both the geopolitical gravity of the conflict and the speeches' primary subject. Other frequently occurring items are *people (38)*, *Russia (35)*, *the world (33)*, and *Putin (31)*. This tendency shows consistent focus on the social aspect, the main players in the war, and the wider global aspects.



Figure 4.1. Word cloud of Joe Biden's speeches.

Terms.

A foundational component of this corpus-based analysis is the Terms tool, which offers a structured, quantitative breakdown of lexical distribution regarding the war in Ukraine across the datasets. While a word cloud relies on spatial aesthetics to highlight prominent vocabulary, the Terms tool delivers the identical underlying data through a systematic, tabular matrix that pairs every conflict-related lexical item with its absolute frequency count. By utilizing this structured matrix, researchers can conduct rigorous cross-corpus comparisons to isolate recurring thematic threads, stylistic shifts, and rhetorical trends distinguishing Biden's and Trump's foreign policy discourse. Ultimately, tracking these quantified linguistic patterns over time makes it possible to detect evolving geopolitical preoccupations and capture how both figures strategically frame the Ukraine war to the public.

Key words are *Ukraine* (130), *people* (38), *Russia* (35), *world* (33), *Putin* (31), *Ukrainian* (26), *united* (23), *war* (22), *states* (17), *Russian* (16).

In conclusion, the empirical data extracted from the Biden sub-corpus reveals a deeply strategic, highly structured foreign policy narrative. By utilizing the specific word frequency trends identified through Voyant Tools, this analysis demonstrates how Biden relies on clear linguistic asymmetries to center the victim state, hyper-personalization to isolate Vladimir Putin as the sole aggressor, and humanitarian framing to focus on the resilience of the ordinary populace.

			Term	Count
<input type="checkbox"/>	<input type="checkbox"/>	1	ukraine	130
<input type="checkbox"/>	<input type="checkbox"/>	2	people	38
<input type="checkbox"/>	<input type="checkbox"/>	3	russia	35
<input type="checkbox"/>	<input type="checkbox"/>	4	world	33
<input type="checkbox"/>	<input type="checkbox"/>	5	putin	31
<input type="checkbox"/>	<input type="checkbox"/>	6	ukrainian	26
<input type="checkbox"/>	<input type="checkbox"/>	7	united	23
<input type="checkbox"/>	<input type="checkbox"/>	8	war	22
<input type="checkbox"/>	<input type="checkbox"/>	9	states	17
<input type="checkbox"/>	<input type="checkbox"/>	10	russian	16

Figure 4.2. Terms of Joe Biden's speeches.

Trends.

The trend visualization depicts a line graph that charts the relative frequencies of chosen phrases throughout the corpus. A description at the top of the graph indicates the color associated with each term, and each line represents a particular word and appears in a different color. Variations in the frequencies of terms like world, Ukraine, Russia, Putin, and people show how the corpus's thematic emphasis has changed over time. These shifts are a result of shifting geopolitical circumstances, changes in the conflict's public and diplomatic discourse, and events on the battlefield. These phrases' movement over the timeline shows how the speeches' main points change to reflect new developments and shifting political environments. This feature in the "Trends" tool is seen in Figure 4.3 below.

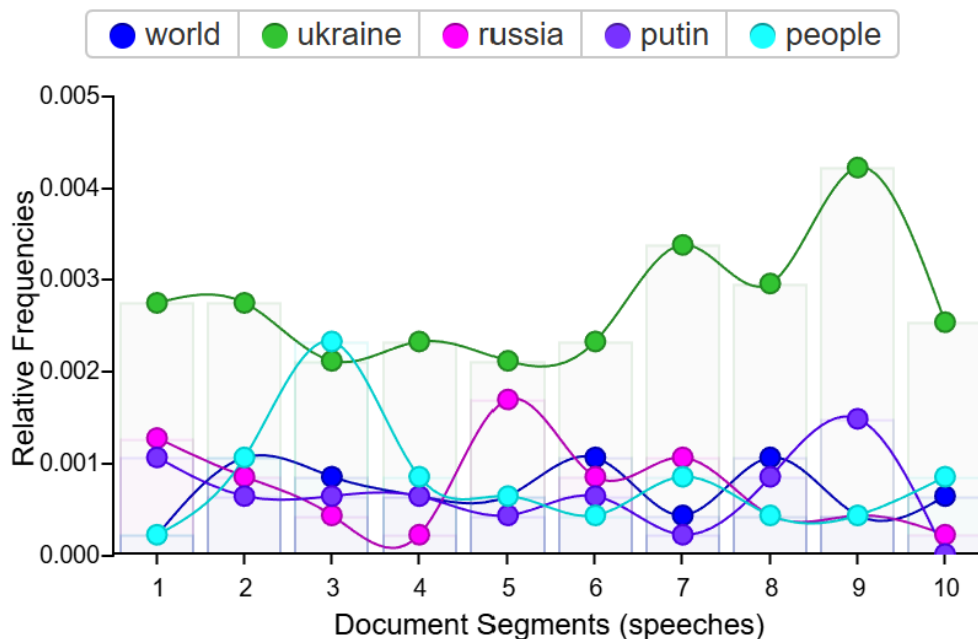


Figure 4.3. Trends of Joe Biden's speeches.

The graph demonstrates that the lexeme *Ukraine* (green) displays the greatest frequency throughout the corpus. The word *people* (light blue) ranks second in prominence. The remaining words, *Putin*, *Russia*, and the *world*, show substantial shifts over time and occur at somewhat lower rates.

Analysis of Donald Trump's speeches.

Corpus 2, which consists of material from Donald Trump's public statements (4,971 words total), was organized into three sub-corpora: the Pre-POTUS period (three texts), the POTUS-Era period (ten texts), and a set of X posts (four texts), given Trump's extensive use of social media as a communicative platform. In contrast to Corpus 1, Trump's speeches display considerable variation in structure and delivery, which creates a distinct analytical environment.

The sub-corpora were arranged chronologically. The Pre-POTUS one is compiled of speeches delivered on March 7, 2024; September 23, 2024, and September 26, 2024. November 18, 2024 (the day President Trump won the elections) marks the beginning of the POTUS Era, which happened on January 23, 2025; February 17, 2025; February 18, 2025; February 18, 2025; February 26, 2025; February 28, 2025; March 4, 2025; March 5, 2025; March 8, 2025; and so far the last speech that was analyzed is spoken on March 17, 2025. X posts were published on February 19, 2025; February 24, 2025; February 28, 2025; March 7, 2025.

The first stage of the analysis focuses on the Pre-POTUS era, when Joe Biden was president. During this period, Donald Trump gave three interviews totaling 936 words. These interviews were conducted on March 7, 2024, September 23, 2024, and September 26, 2024.

			Term	Count
<input type="checkbox"/>	<input type="checkbox"/>	1	it's	12
<input type="checkbox"/>	<input type="checkbox"/>	2	ukraine	10
<input type="checkbox"/>	<input type="checkbox"/>	3	think	9
<input type="checkbox"/>	<input type="checkbox"/>	4	say	8
<input type="checkbox"/>	<input type="checkbox"/>	5	war	7
<input type="checkbox"/>	<input type="checkbox"/>	6	country	7
<input type="checkbox"/>	<input type="checkbox"/>	7	people	5
<input type="checkbox"/>	<input type="checkbox"/>	8	know	5
<input type="checkbox"/>	<input type="checkbox"/>	9	i'm	5
<input type="checkbox"/>	<input type="checkbox"/>	10	russia	4

Figure 4.5. Terms of Donald Trump's pre-POTUS speeches.

Trends.

1. The analysis of individual lexical items in the Pre-POTUS sub-corpus reveals notable patterns. The first item is not technically a word. Still, a contraction *it's* appears 12 times, typically in sentences such as its [*not Ukraine anymore*] and in combination with various adjectives, including *it's* [*sad, terrible, gone, happened, destroyed, tough, locked, believable, and fault*], each of which occurs at least once or twice.

2. The most frequently occurring term in the sub-corpus is *Ukraine*, mentioned 10 times, often in contexts emphasizing territorial loss or the ongoing conflict.

3. The verb *think* frequently occurs in reference to Trump's reflections on former political opponents, specifically Joe Biden and Barack Obama, in relation to the war and associated events.

4. Similarly, the verb *say* appears in limited collocations such as [they] *say* and in statements indicating hypothetical or rhetorical possibilities, including claims of resolving the conflict rapidly.

5. The term *war* features prominently and carries complex connotations. For instance, it occurs alongside easily in statements suggesting the prevention of a broader conflict, while later in the corpus, the United States is held responsible for the ongoing war in Ukraine. Overall, the Pre-POTUS sub-corpus demonstrates extensive use of adjectives and strategic lexical choices that reflect both evaluative and political dimensions of discourse.

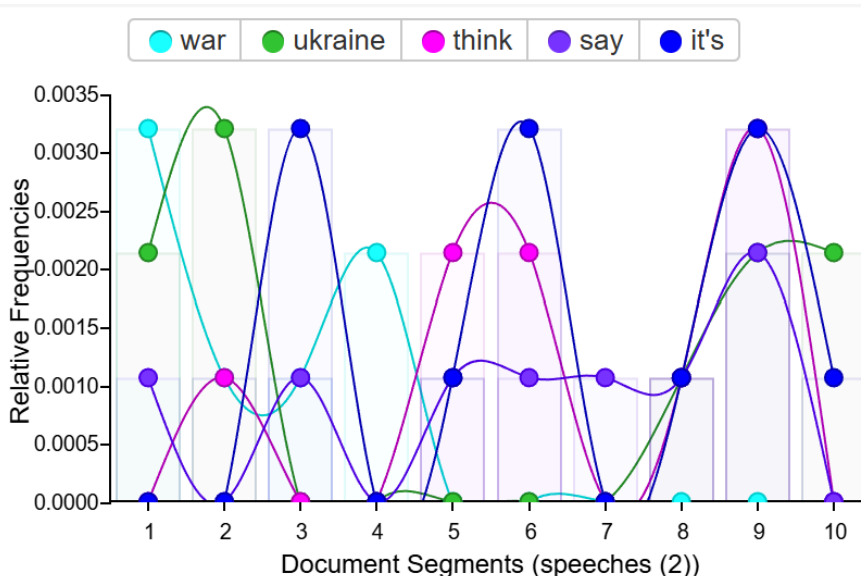


Figure 4.6. Trends of Donald Trump’s pre-POTUS speeches.

The period of analysis when Donald Trump won the elections for the second time and became the 47th president of the United States is labelled in our research as the POTUS Era (3,456 words total).



Figure 4.7. Word cloud of Donald Trump’s POTUS speeches.

As can be seen from Figure 4.7 above, the top 5 words are *Ukraine* (41), *think* (30), *war* (26), *president* (22), and *lot* (21).

Terms.

The key terms are Ukraine (41), think (30), war (26), president (22), lot (21), going (20), want (16), doing (16), Putin (15), people (15).

The data from Trump's presidential era highlights a shift toward a highly personalized and transactional communication style.

			Term	Count
+	<input type="checkbox"/>	1	ukraine	41
+	<input type="checkbox"/>	2	think	30
+	<input type="checkbox"/>	3	war	26
+	<input type="checkbox"/>	4	president	22
+	<input type="checkbox"/>	5	lot	21
+	<input type="checkbox"/>	6	going	20
+	<input type="checkbox"/>	7	want	16
+	<input type="checkbox"/>	8	doing	16
+	<input type="checkbox"/>	9	putin	15
+	<input type="checkbox"/>	10	people	15

Figure 4.8. Terms of Donald Trump's POTUS speeches.

In conclusion, the POTUS section of the corpus shows a propensity to assign credit to Ukraine, frequently using unfavorable or simplistic word choices. The study shows that references to war are relatively less prominent in these speeches, with Ukraine serving as the primary emphasis. Throughout the talks, personal opinion appears as a crucial rhetorical component.

Finally, the X sub-corpus contains a total of 579 words.

Terms.

The key words are *war* (8), *Zelenskyy* (7), *Russia* (7), *united* (6), *Ukraine* (6), *Trump* (6), *states* (6), *president* (6), *peace* (5), *2025* (4).

The empirical data from Trump's X (Twitter) sub-corpus reveals a highly politically calculated rhetorical framework. The prominence of *war* (8), *Zelenskyy* (7), and *Russia* (7) establishes a direct focus on the primary factors and the conflict itself, yet this is strictly filtered through a lens of American political leadership, as evidenced by the recurring clusters of *united*, *states*, *president*, and *Trump* (each 6). Notably, the presence of *peace* (5) in green which stands for positive connotation combined with the specific temporal marker *2025* (4) underscores a forward-looking, goal-oriented narrative. Ultimately, this microblogging discourse strips away dense diplomatic jargon to deliver a highly

condensed, urgent message to the public, framing the crisis as a problem centered around key leaders that requires immediate, decisive American intervention to achieve a swift resolution.

			Term	Count
<input checked="" type="checkbox"/>	<input type="checkbox"/>	1	war	8
<input checked="" type="checkbox"/>	<input type="checkbox"/>	2	zelenskyy	7
<input checked="" type="checkbox"/>	<input type="checkbox"/>	3	russia	7
<input checked="" type="checkbox"/>	<input type="checkbox"/>	4	united	6
<input checked="" type="checkbox"/>	<input type="checkbox"/>	5	ukraine	6
<input checked="" type="checkbox"/>	<input type="checkbox"/>	6	trump	6
<input checked="" type="checkbox"/>	<input type="checkbox"/>	7	states	6
<input checked="" type="checkbox"/>	<input type="checkbox"/>	8	president	6
<input checked="" type="checkbox"/>	<input type="checkbox"/>	9	peace	5
<input checked="" type="checkbox"/>	<input type="checkbox"/>	10	2025	4

Figure 4.9. Terms of Donald Trump's X speeches.

As demonstrated in Figure 4.10 below, the top 5 words are *war* (8), *Zelenskyy* (7), *Russia* (7), *United* (6), *Ukraine* (6).

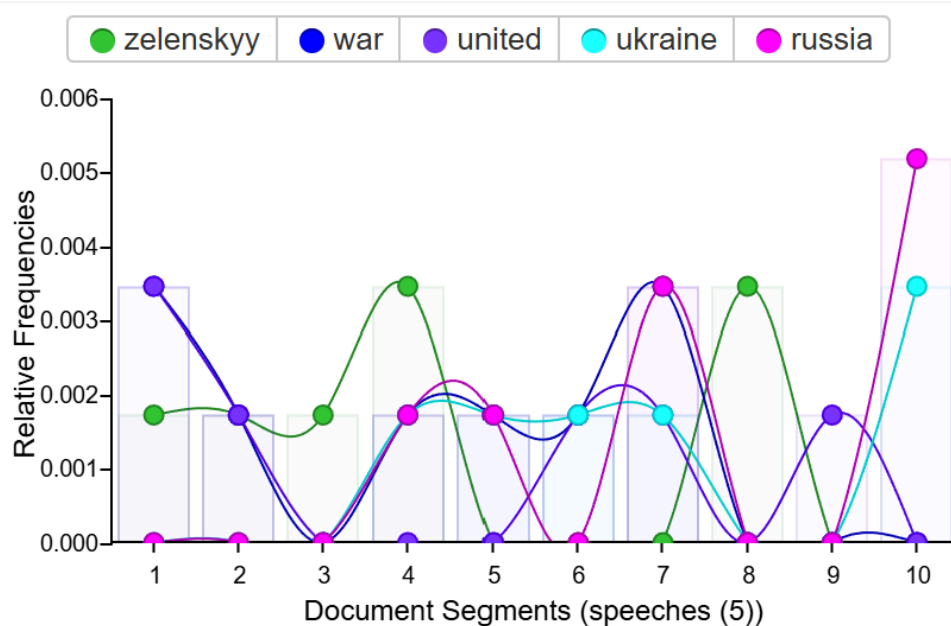


Figure 4.10. Trend of Donald Trump's X speeches.

In summary, the most common term in this section of the corpus is *Russia*, which is followed in decreasing order of frequency by *war*, *Zelenskyy*, *Ukraine*, and *united*.

Lingualyzer analysis of Joe Biden’s speeches.

At the initial stage of the research, speeches given by the former President of the United States, Joe Biden, were collected and in what follows, are labelled as Corpus 1. The Miller Center contains a large collection of meeting transcripts. 8 speeches dated from the beginning of the full-scale invasion, February 2022, until Biden’s last speech in the position of a country leader, September 2024, were selected. The speeches delivered on February 24, 2022; March 1, 2022; March 26, 2022; September 21, 2022; February 21, 2023; October 20, 2023; March 7, 2024; and September 24, 2024, will serve as the material for this study (Antoniuk A.R. (2025). THE LANGUAGE OF WAR: CORPUS ANALYSIS OF JOE BIDEN AND DONALD TRUMP’S SPEECHES ON UKRAINE. MUNDUS PHILOLOGIAE, 5). To illustrate the evidence, here is the table of dates, occasions, and titles to support my research paper:

Date	Occasion	Title
February 24, 2022	President Joe Biden delivers his first speech on the day of the Great Invasion	Joe Biden Speech Transcript: Russia Invades Ukraine
March 1, 2022	President Joe Biden delivers his first State of the Union address. He begins with a pledge to defend the country of Ukraine from the Russian invasion.	State of the Union Address
March 26, 2022	At the Royal Castle in Warsaw, Poland, President Joe Biden talks about Russian aggression and the war in Ukraine.	Remarks in Support of the People of Ukraine
September 21, 2022	President Joe Biden addresses the 77th session of the United Nations General Assembly. He	Speech before the 77th Session of the United Nations General Assembly

	speaks out against Russia and its invasion of Ukraine.	
February 21, 2023	Speaking at the Royal Castle in Warsaw, Poland, President Biden talks about the first year of the war in Ukraine after the Russian invasion.	Remarks on the One-Year Anniversary of the Ukraine War
October 20, 2023	President Biden addresses the American people to make the case for why US support for Ukraine and Israel is crucial.	Remarks on the US Response in Support of Israel and Ukraine
March 7, 2024	President Biden urges Congress to send him a bipartisan national security bill to help fund Ukraine.	State of the Union Address
September 24, 2024	President Biden highlighted the major international conflicts the world is facing: the Russian-Ukrainian War.	Address before the 79th United Nations General Assembly

Table 4. 11. The table of speeches delivered by President Joe Biden

The initial focus of this corpus analysis focuses on the speech delivered on February 24, 2022, the first day of the full-scale invasion. This speech contains 735 words, including 146 nouns and 155 verbs. Additionally, the lexical profile includes 65 adjectives and 27 adverbs. From a structural perspective pronouns occur 62 times, while determiners account for 58 instances, and conjunctions total 50.

Again, we return to the parts of speech illustration to understand the style of the language and vocabulary used.

The discourse is composed of content words, which include nouns, verbs, adjectives, and adverbs. Their arrangement in this speech demonstrates a calculated balance between action-oriented oratory and factual clarity. To identify the specific register of this discourse, a mathematical analysis of

nominal density was performed by dividing the frequency of nouns by the total word count. AI was used to transform raw part-of-speech counts into standardized percentages and connect these results with recognized linguistic norms in order to preserve accuracy throughout the extensive data collection.

There are 146 noun words in the President's speech such as *Ukraine, Putin, Russia, allies, people* as well as *assault, provocation, attack, troops, border, sanctions, consequences*. Normally, there must be either 15% or 30%, depending on the style: either conversation or academic. The nominal density (19.8%) is above the typical conversational norm of 15%. It is still far below the academic discourse requirement of thirty percent, though. This implies that although the speech is quite informative, it does not employ the "nominal heavy" complexity typical of scientific or technical reporting favoring an understandable language.

There are 155 verbs in the text, which is about twice as many (21%) as the academic style standard of 10%. An "action-oriented" discourse is characterized by a high verb frequency. This is very consistent with a conversational register that emphasizes actions and processes over static descriptions. Verbs such as *invaded, planning, moved, rejected, authorizing* are predominantly active and agentive, assigning unambiguous responsibility to Putin while positioning Biden as a reactive but decisive authority (*authorizing, designed, assured*).

The text contains 65 adjectives, which makes it 8.8%. Adjectives such as *brutal, premeditated, naked, unjustifiable* function as moral verdicts, pre-empting any ambiguity about the nature of the attack. This is close to the 10% academic requirement and is quite a difference from the 2.5% found in simple conversation. This suggests a significant degree of qualitative evaluation and descriptive detail about the invasion.

With 27 occurrences, the frequency of adverbs is marginally higher (3.6%) than the academic criterion of 3% but lower than the 5% observed in conversation. The adverbs *purposefully, strategically, unilaterally* are notably process-oriented.

Now, the next step is the function parts of speech. The 62 pronouns utilized greatly surpass the academic limit of 4% bringing it 8.4%. The conversational style is characterized by high pronoun usage, which creates a direct connection between the speaker, the audience, and the subject. Pronouns reveal a sharp us/him opposition: *he/his* is consistently attached to Putin as sole aggressor, while *we/our* constructs American-allied solidarity.

The usage of determiners (such as "the," "this," and "those") at 58 instances is between (7.8%) the academic standard of 10% and the conversational benchmark of 4.5%. This suggests that the items referenced in the speech were identified with a moderate degree of specificity.

There are 50 conjunctions (6.8%), exceeding the academic and conversational requirements of 5% and 4.5%, respectively. The conjunction *although* appears significantly, e.g. *Although [we provided \$650 million...our*

forces will not engage] marking a concessive structure that simultaneously asserts support and limits.

A "hybrid" rhetorical profile is revealed by the mathematical data. The President conveys the seriousness of the invasion by using a high density of adjectives and a moderate increase in nouns (parts of an academic or informational style). Nonetheless, a majority of pronouns (8.4%) and verbs (21%) serve as the main syntactic engine, solidifying the address in a conversational register.

This corpus analysis's next phase centers on a crucial speech made by President Biden in March 2022. With a total word count of 416, this particular corpus is shorter than the first speech. The distribution of syntactic categories offers crucial information on the development of the administration's wartime register, despite its shorter length.

There are 76 nouns, which makes it 18%; the nominal density is in line with earlier results, falling between the academic benchmark of 30% and the conversational benchmark of 15%. One week later, the noun profile shifts markedly from military-legal to moral-emotional vocabulary: *fearlessness, courage, determination, homeland, unity, souls, hearts, bond*. This reflects the State of the Union's performative function and Biden is now speaking to Congress and the American people, not documenting a crisis.

With 83 verbs, the text maintains a high frequency of 20% that almost doubles the academic standard of 10%. Verbs like *inspires, stand, inflicting, choosing, rise* are present-tense and dynamic, creating immediacy and collective momentum.

This text contains 28 adjectives, which is 6.7%. This is still much higher than the 2.5% conversational average, even though it is less than the 8.8% reported in the February speech. The adjective cluster such as *premeditated, unprovoked, unmistakable, iron, unified* combines factual judgment with inspirational framing.

The greater number of pronouns (43 instances) made 10.3% in this sample is a noteworthy discovery. This 10% density, which significantly exceeds the academic limit of 4%, indicates a very "involved" interpersonal register. One clear indicator of the conversational style is the frequent use of deictic markers (such "I," "we," and "you"), which are meant to build rapport and a sense of shared responsibility. Pronouns are overwhelmingly collective: *we, our, us* dominate, with *he* appearing only in relation to Putin. The conjunction *while* introduces contrast: *while [it shouldn't have taken something so terrible...]*, signalling rhetorical concession before affirmation – a classic oratorical move.

This text's 23 conjunctions somewhat surpass the academic norm of 5%, due to it being 5.5%. This guarantees the structural coherence needed for oral oratory, enabling the speaker to make connections between difficult concepts in a way that is nonetheless understandable to a broad audience.

To conclude that analysis, it is important to note that the syntactic hybridity pattern is supported by the study of the March 2022 address. The speech is clearly classified as conversational due to the high density of pronouns (10.3%) and the overwhelming abundance of verbs (20%), even though the noun and adjective counts offer enough lexical "weight" to retain an authoritative tone. Adverb *never* appears for the first time here, introducing the absolute-negative framing that will become a Biden signature throughout the corpus.

The speech given in Warsaw on March 26, 2022, is the third important piece of information in that analysis. With 757 words in total, this speech is a crucial rhetorical moment meant for both a local European audience and a worldwide audience. Additional proof of the "hybridized" register used in the first month of the conflict may be seen in the syntactic distribution taken from this corpus.

This address's speech implies a deliberate increase in nominal and descriptive mass.

With 154 occurrences in nouns, this speech has the highest nominal density in the corpus to date, with a total number of 20.3%, coming very near to the academic standard of 30%. The noun profile here becomes explicitly ideological: *freedom, democracy, struggle, nation, frontlines, empire, liberty, darkness*. Delivered at the Royal Castle in Warsaw, the speech situates the Ukraine war within a grand historical narrative of democratic resistance, and the nouns encode this directly.

The transcript contains 145 verbs (19%), almost twice as many as the academic standard of 10%. The high frequency of verbs guarantees that the language stays action-oriented even though the number of nouns has increased. By giving active processes precedence over static analysis, this preserves the speech's conversational engine. Verbs such as *sapping, replenish, erase, refuse* introduce new transitivity patterns that Russia is now an entity that *saps* and *erases*, while Ukraine *refuses* and *fight*s.

This study's 83 pronouns (11%) indicate a high level of interpersonal participation, significantly higher than the academic standard of 4%. In order to establish a shared emotional and political space between the speaker and the *people of Ukraine*, this 11% density is a clear indicator of the conversational style. It does this by using deictic references.

With 36 occurrences, the conjunction frequency (4.7%) is nearly in line with the 5% Academic and 4.5% conversational criteria. This suggests a well-balanced syntactic structure that preserves the flow required for a significant public speech while offering the logical "connectors" for complicated concepts.

A precise rhetorical balance is revealed when analyzing the data from March 26. The overwhelming number of verbs (19%) and the high frequency of pronouns (11%) guarantee that the vocabulary stays essentially conversational,

even while the noun density (20.3%) and adjective count (7.4%) offer the "academic" weight necessary for a significant foreign policy address. The adjectives *sacred*, *perennial*, *cynical*, *obscene*, *collective* are among the most evaluatively loaded in the entire corpus, signalling Biden's highest rhetorical register.

This longitudinal study's fourth phase looks at President Biden's September 21, 2022, speech to the UN General Assembly. This 507-word speech reflects a high-level diplomatic setting where a major shift toward a formal register might be anticipated. Nonetheless, the evidence indicates that a hybridized syntactic framework is still being followed.

There are 91 noun occurrences and 18%, the nominal density is in line with earlier wartime addresses, falling between the academic benchmark of 30% and the conversational benchmark of 15%. The sub-20% noun count indicates that the speaker valued straightforward communication above the intricate nominalization typical of multilateral diplomatic texts, although a UN address typically requires a high informative density. By September 2022 the noun inventory shifts to institutional-legal vocabulary: *referenda*, *violation*, *charter*, *statehood*, *culture*, *sovereignty*, *principles*, *resolve*. This register shift is conditioned by the UN context and Biden is now speaking to a multilateral body and must frame the war in terms of international law rather than moral emotion.

There are 107 verbs in the text, which totals into 21%, which is much more than the academic standard of 10%. Verbs such as *organizing*, *annex*, *asserted*, *extinguishing*, *condemned*, *marshaled*, *chose* are notably more formal and legalistic, with *extinguishing* being the most semantically powerful: it frames the war as an attempt to eradicate Ukraine's ontological existence as a nation

The text's 38 adjectives and its being 5% surpass the conversational average of 2.5%. Although it is still somewhat below the 10% academic criterion, this frequency corresponds with a more formal tone and permits the accurate qualitative characterisation of international law and sovereign boundaries.

The adverbial density is managed with 20 occurrences and 4%, just above the academic limit of 3%. This demonstrates a careful application of modifiers to set the urgency and mode of international cooperation in context. Adverbs in this speech are unusually precise: *explicitly*, *unequivocally*, *perfectly* are all intensifiers of institutional certainty, signalling Biden is addressing skeptical or non-aligned states.

Compared to the academic standard of 4%, the 54 pronouns employed in this sample represent a high density 10.7%. A clear indicator of the conversational style is this 10.7% frequency. Pronouns introduce *you* for the first time in the corpus, directed at the General Assembly, creating a dialogic dimension absent from earlier speeches.

With 25 occurrences and 5%, the conjunction frequency perfectly matches the Academic criterion of 5%.

A summary of the data from September 21 shows that the President's "UN register" does not completely change into an academic style. The predominance of verbs (21%) and pronouns (10.7%) guarantees that the communication stays interpersonal and action-oriented.

The memorial speech given in Warsaw on February 21, 2023, the first anniversary of the war, is the subject of the study's fifth analytical milestone. This speech offers a crucial longitudinal perspective on the stability of the administration's rhetorical register over twelve months of continuous warfare, using an adjusted total word count of roughly 950.

The nominal quantity of nouns is impressively constant (19.4%) with the baseline (February 2022) and following addresses, with 184 occurrences. This indicates a "sustained informational plateau," as it is much above the 15% conversational benchmark but well below the 30% academic criterion. The one-year mark produces the richest and most varied parts of speech profile in the corpus. Nouns span three semantic fields simultaneously: geopolitical (*coalition, territory, land, power*), moral-atrocity (*depravities, crimes, humanity, destruction, atrocities*), and resilience (*cornerstone, stability, prosperity, heroes, commitment*). This tripartite structure reflects the speech's complex rhetorical task such as accounting for one year of war, condemning Russian crimes, and renewing commitment.

There are 185 verbs in the text (19.4%), which is almost twice as many as the academic standard of 10%. For example verbs like *unleashed, shattered, committed, targeted, stolen* documents Russian atrocities; present tense (*doubts, stands, leads*) asserts Ukraine's survival; future (*will prevail, will not waver*) projects conviction.

The text's 70 adjectives maintain the high qualitative precision trend with its 7%. This number is close to the 10% academic norm and almost triples the 2.5% conversational average. The adjective *craven* applied to Putin's "*lust for land and power*" is the most semantically intense evaluative term across all eight speeches.

The adverbial density approximately matches the 5% conversational threshold with 53 occurrences and 5.5% itself. Compared to the more restrained UN address, this shows a little more expressive use of modifiers, allowing for a stronger focus on the manner and persistence of Ukraine's resistance.

Compared to the 4% academic limit, the 99 pronouns (10.4%) used in this sample represent a high density. This 10.4% frequency maintains a high level of "speaker-audience involvement," a key feature of the conversational register, making it an effective interpersonal tool. Pronouns like *we* and *you* are crucial in this situation for fostering a feeling of global togetherness.

At 62 occurrences (6.5%), the conjunction frequency exceeds the academic criteria of 5% and the conversational benchmark of 4.5%. During a

high-profile broadcast, this higher usage promotes the "oral flow" of the oratory, making it possible to integrate complex geopolitical concepts into a story that is nevertheless understandable to a wide global audience.

An exceptional level of register stability is shown by the study of the February 2023 data. After a year of the war, the President's rhetoric preserves a deliberate syntactic hybridity.

This corpus study's sixth speech examines President Biden's October 20, 2023, broadcast speech, a pivotal rhetorical moment in which he connected support for Israel and Ukraine to a single national security framework. This 746-word sub-corpus offers 134 instances of nouns (17.96%). While this remains substantially above the 15% conversational average, it shows a slight decrease in nominal weight compared to the Warsaw address (20%). This speech, delivered alongside the Gaza crisis, shows a notable shift in noun semantics toward security-strategic framing: *security, appetite, control, leadership, alliances, values, capability, bravery*. The word *appetite* in [*Putin's*] *appetite [for power and control]* is the speech's most striking nominalization, reducing Russian expansionism to an animalistic drive.

There are 161 verbs (22%) in the text, which is more than twice as many as the academic standard of 10%. Verbs like *echoes, emboldened, watching, erase* reflect an anticipatory threat register and Biden is now arguing forward, warning of global consequences if Ukraine falls.

The transcript's 46 adjectives (6%) are slightly lower than the academic's 10%, but they still surpass the conversational average of 2.5%. The adjective *vital* appears here with high collocational frequency (*vital for America's national security*), marking a shift from altruistic to self-interest framing of US involvement which is a rhetorical adjustment likely responding to domestic political pressure.

With 38 occurrences (5%), the adverbial density is exactly in line with the conversational guideline of 5%. This demonstrates a more expressive, forceful style of oratory.

This speech's 80 pronouns considerably surpass the 4% academic restriction while maintaining the high-involvement trend. A defining feature of the conversational style, its 11% density serves to provide a direct line of communication between the Commander-in-Chief and the domestic audience while highlighting group responsibility through the frequent use of pronouns.

The frequency of conjunctions was above the academic norm of 5% in 46 (6%) cases. Conjunctions peak here particularly *if* and *when*, constructing extended conditional arguments about global security consequences.

The next empirical data point in this longitudinal study is the State of the Union Address delivered on March 7, 2024. This corpus consists of 517 words. To begin with, nouns' 84 occurrences (16.25%) – this address's nominal density is slightly higher than the 15% conversational average. Nouns such as *march, chaos, leadership, risk, harm* are abstract and threat-oriented.

There are 107 verbs (21%) in the text, which is more than twice as many as the academic norm of 10%, with action verbs *invading, sowing, stop, defend, blocked, emboldening* packed tightly – suggesting compressed argumentative urgency.

This transcript contains 28 adjectives. This is a considerable drop from the prior wartime baseline (about 9%), even though it is higher than the 2.5% conversational average. With 26 occurrences, there are 5% of adverbs.

The September speech given during the 79th UN General Assembly serves as Joe Biden's corpus's last empirical data stage. This 351-word transcript summarizes the administration's high diplomatic manner.

With 65 occurrences of nouns, the nominal density is consistent with the 18% range, making it conversational with its 15%. The final speech shows a legacy and vindication noun profile: *survival, challenges, crisis, democracy, choice, peace, weary* and Biden is now framing the war as a test that has been substantially met.

There are 67 verbs in the text, which is almost twice as many (19% as the academic standard of 10%). Verbs in the past tense (*defended, ensured, understood, stepped, stood*) dominate the early portion, constructing a narrative of completed collective achievement. This is a marked contrast to the very first speech performed on the day of the Great Invasion, where verbs were predominantly present or future.

Twenty adjectives (8%) are used, making it academic. Adjectives *bigger, stronger, more united* that were applied to NATO encode the vindication thesis directly.

With 21 occurrences (6%), the adverbial density is exactly in line with the conversational criterion of 5%. The adverb *merely* [*we could have stood by and*] *merely* [*protested*] is the speech's most rhetorically loaded word, framing inaction as morally inadequate through minimization.

The 31 number of pronouns (8%) is greatly above the academic limit of 4%.

With 29 occurrences (8%), this conjunction frequency is the highest in the corpus, surpassing the academic and conversational criteria of 5% and 4.5%, respectively. Conjunctions reach its highest frequency, reflecting the most complex argumentative structure in the corpus: Biden is synthesising two years of policy, constructing causal chains, and presenting a choice to the international community.

President Biden's rhetorical framework is a deliberate syntactic mix, according to this longitudinal examination of eight speeches. It is evident by comparing linguistic density to predetermined standards that the administration blends academic and conversational styles to achieve both authority and accessibility. This analysis reveals a clear rhetorical arc across the eight speeches. Nouns shift from evidentiary-military (2022), moral-ideological (early 2022) institutional-legal (mid 2022), atrocity-documentation (2023),

strategic-security (2023–2024) to legacy-vindication (2024). Verbs move from present-urgent to past-accomplished. Adjectives become progressively more evaluatively intense before moderating in the final speech. The adverb *never* functions as a consistent semantic anchor across the entire corpus, appearing in all eight speeches as a device of absolute negation against Russian success.

Analysis of Donald Trump’s speeches.

Corpus 2, which consists of material from Donald Trump’s public statements (4,971 words total), was organized into three sub-corpora: the Pre-POTUS period (three texts), the POTUS-Era period (ten texts), and a set of X posts (four texts), given Trump’s extensive use of social media as a communicative platform. In contrast to Corpus 1, Trump’s speeches display considerable variation in structure and delivery, which creates a distinct analytical environment.

The sub-corpora were arranged chronologically. The Pre-POTUS one is compiled of speeches delivered on March 7, 2024; September 23, 2024; and September 26, 2024. November 18, 2024 (the day President Trump won the elections) marks the beginning of the POTUS Era, which happened on January 23, 2025; February 17, 2025; February 18, 2025; February 26, 2025; February 28, 2025; March 4, 2025; March 5, 2025; March 8, 2025; and so far the last speech that was analyzed is spoken on March 17, 2025. X posts were published on February 19, 2025; February 24, 2025; February 28, 2025; March 7, 2025.

Date	Occasion	Title
March 7, 2024	First comment on the war	Trump says he can end Ukraine war in a day
September 23, 2024	To blame U.S. leadership for the war	During long, winding speech, Trump blames U.S. for war in Ukraine
September 26, 2024	To state and argue that any deal would have gone better	'Ukraine is gone, it's not Ukraine anymore,' Trump says during campaign stop
November 18, 2024	To call for immediate de-escalation and warn that Biden’s policies are leading to "nuclear World War III".	Donald Trump calls for "De-escalation and Peace between Ukraine and Russia" 47th President Agenda

January 23, 2025	The war would never have started if the 2020 election hadn't been "rigged".	The war in Ukraine would have never happened if I were president
February 17, 2025	To state that Putin "should make a deal" because the war is destroying Russia.	Trump says Putin should make deal to end Ukraine war
February 18, 2025	To criticize Ukraine in the war	Trump says Ukraine 'should have never started' the war with Russia
February 26, 2025	To criticize Ukraine for wasting money	REMARKS BY PRESIDENT TRUMP BEFORE CABINET MEETING
February 28, 2025	To conduct negotiations on a ceasefire	Heated exchange between Trump, Zelenskyy, Vance at Oval Office
March 4, 2025	To state that Ukraine is ready to make peace	Address to a Joint Session of Congress
March 5, 2025	To confirm that Ukraine is ready for dialogue	Trump says he received letter from Zelenskyy saying Ukraine ready for dialogue
March 8, 2025	To defend Putin	Trump says Putin launching massive strike on Ukraine is 'what anybody would do.'
March 17, 2025	To announce an upcoming call with Putin	Trump says he will talk to Putin on Tuesday as he pushes for an end to the Ukraine war

Table 4.12. The table of speeches delivered by President Donald Trump

The Trump corpus analysis begins with his first comments on the beginning of the Great Invasion on March 7, 2024, with a short 78-word transcript, where extreme interpersonal interaction is prioritized over conventional informational structures, according to the quantitative data from this micro-text, which shows a clear shift in register.

With just four occurrences (5%), the nominal density is much below the 15% conversational norm and about six times lower than the 30% academic benchmark and are generic: *war, world, office, order, day*.

With 20 verbs (25%) used in the text, more than one in four words is a verb; the academic standard of 10% is more than doubled by this, with future-tense constructions occupying every agentive slot: *will prevent, will have settled, will get solved, will take*.

The adjective count is at its lowest point for conversational style, with just two occurrences and 3%. The adjective *disastrous* is the only evaluative term applied to the war itself.

With 12 occurrences (15%), the adverbial density is almost three times higher than the conversational average of 5%, e.g. *very easily, quickly, exactly, rapidly*.

Compared to the academic maximum of 4%, the 14 pronouns indicate a huge density with its 18%, overwhelmingly first-person singular *I* appears 7 times in 78 words, an extraordinary density that mirrors the speech's central rhetorical claim.

The five conjunctions (6%) made it typically academic.

The next speech consists of 693-word shows a visible difference between kinetic action and static information.

The nominal density is significantly low at 66 occurrences (10%). It is still about 5% below the 15% conversational average and a startling 20% below the 30% academic criterion, although being higher than the prior micro-transcript.

There are 205 verbs in the text, which means that action-oriented words make up 30% of the total speech. This is three times the academic standard of 10%, and are predominantly conversational present tense: *say, call, think, want, hear, come, walk*, these are action verbs of everyday speech rather than political governance.

There are 31 adjectives (4%), which is higher than the conversational average of 2.5% but still well below the academic criterion of 10%. The adjective *greatest* applied to Zelenskyy [*the*] *greatest* [*salesman in history*] is deeply ironic, deploying superlative praise as covert insult, reframing Ukrainian diplomacy as manipulation.

The adverbial density surpasses the 3% academic and 5% conversational marks with 52 occurrences, and its 8%.

The 151 pronouns used, which make up 22% of the full speech, are an important finding. This is much higher than typical conversational levels and

more than five times the academic 4% limitation, driven by dense first-person singular usage (*I* appears over 20 times) alongside frequent *he/they* in reference to Zelenskyy, Biden, and Obama.

The 46 conjunctions surpass the academic standard of 5% with its 7%. The conjunction *because* appears with notable frequency, constructing causal chains that assign responsibility: [*it's our fault*] *because* [*it's a war that should never have been allowed to happen*].

The last pre-POTUS speech, given on September 26, 2024, contains 106 words.

With 13 occurrences (12%), the nominal density is slightly higher than in earlier speeches, but it is still below the conversational average of 15%. Nouns are concrete and grief-coded: *people, cities, towns, dead, building, tower* which is a notable departure from the abstract institutional nouns of Biden's UN speeches.

There are 34 verbs in the text, which means that action words make up almost one-third (32%) of the speech. This exceeds the academic style of 10% by more than three times.

The frequency of the 15 pronouns (14%) is more than three times the academic maximum of 4%.

With seven occurrences and 7%, the conjunction frequency is higher than the academic criterion of 5%.

The day President Trump won the elections on the 18th of November, 2025, the victory speech's word count of 167 illustrates the quick change from contender to president-elect.

The nominal density is greater than in earlier samples, with 22 occurrences and 13% total. This shows a steady rise in "informational anchors", even if it is still below the 15% conversational average and the 30% academic limit.

There are 38 verbs in the text, and 23%, which is much more than the academic standard of 10%. It is nevertheless high enough to maintain an "action-oriented" conversational register.

With 15 adjectives and 9%, the academic level of 10% is almost reached.

The adverbial density surpasses the 3% academic and 5% conversational criteria with 12 occurrences and 7%.

The decline in pronoun density to 8% is an important discovery, even if it is still double the 4% academic limit.

With 14 occurrences and 8%, the conjunction frequency is significantly higher than the academic criterion of 5%.

The POTUS era started with a 41-word short speech. Delivered just after winning the election, this speech sits at the boundary between candidate and president-elect.

With six occurrences (15%), the nominal density is nearly exactly in line with the conversational average of 15%. Key nouns such as *situation, brink,*

catastrophe, possibility, war, peace are abstract threat nouns that frame the conflict as an existential emergency requiring immediate intervention.

Verbs make up an impressive 39% of this content (16 occurrences). This is over four times the academic criterion of 10%, driven almost entirely by modal and conditional constructions: *should have never started, wouldn't have happened, would have never happened, couldn't happen*. This dense modal cluster performs retrospective delegitimation: every verb denies the war's necessity or inevitability under Trump.

Descriptive density is 2%. Adjectives include *dangerous, explosive, nuclear, horrific, ridiculous* and is a concentrated cluster of intensified negatives that construct urgency.

The two adverbs employed are precisely at the 5% conversational marker. The adverb *never* appears twice in counterfactual constructions (*would have never happened if I was in the White House*), introducing the counterfactual blame frame that will dominate the entire POTUS sub-corpus.

There are five pronouns (12%), which is three times the academic restriction of 4%. With four occurrences and 10%, the conjunction frequency is almost twice as high as the academic criterion of 5%.

The next POTUS speech was conducted on February 17, 2025. This 77-word speech transcript demonstrates an excessive concentration of action. According to the data, the aggressive style got even more syntactically forceful as the presidency went on.

The nominal density is remarkably low (8%), with only 6 occurrences, nearly half of the 15% conversational average.

Verbs make up an impressive 43% of the text (33 times). This exceeds the academic standard of 10% by more than four times.

With 14 pronouns, the density is huge (19%), due to the fact that it is more than four times the academic maximum of 4%.

The frequency of conjunctions is below the 5% academic and 4.5% conversational criteria, making it 4%.

The presidential speech delivered on February 18, 2025 consists of a 700-word sub-corpus and is an important analytical unit that offers a reliable dataset for assessing the speaker's rhetorical development throughout his presidency.

With 78 occurrences and 11%, the nominal density is still much lower than the academic limit of 30% and the conversational average of 15%.

There are 216 verbs in the text, and 31%, which is three times the academic standard of 10%, with an extraordinary concentration of past counterfactual modals: *would have never happened, could have been settled, could have made, would have given*, grammatically encoding an entire alternative history in which Trump's hypothetical dealmaking prevents all losses.

There are 29 adjectives and 4%, which is significantly below the 10% academic mark but above the 2.5% conversational norm. The adjectives

pathetic, incompetent, half-baked target Biden exclusively, while *magnificent, golden* are applied to Ukrainian cultural heritage is a rare moment of aesthetic evaluation in Trump's corpus.

With 63 occurrences and 9%, the adverbial density is almost twice as high as the conversational standard of 5%.

There were 123 pronouns used (made it 18%), which is more than four times the academic maximum of 4%, are dominated by *I* in agentive positions (*I could have made, I have the power, I told him*) and *he* for both Putin and Biden as contrasting figures where Putin is controllable, Biden is incompetent.

The conjunction frequency stands at the academic threshold of 5% at 46 occurrences and 7%.

The speech given on February 26, 2025, consists of 739 words.

With 91 occurrences and 12%, the nominal density is slightly higher than in previous micro-transcripts, but it is still much lower than the 15% conversational average and the 30% academic standard.

With 199 verbs, the text maintains a frequency of over 27%. This exceeds the academic benchmark of 10% by more than 2.5 times.

There are 44 adjectives, which is about twice as many as the conversational average of 2.5%, with it being 6%. With positive evaluatives directed at Zelenskyy personally: *great, sterling, tremendous, fantastic, talented, phenomenal*. This adjectival generosity is contextually strategic where Trump is performing diplomatic cordiality while simultaneously reassigning blame for Crimea (*you lost Crimea during the Obama administration*)

With 63 occurrences and 9%, the adverbial density is still strong and well above the conversational standard of 5%, dominated by intensifiers: *very, really, absolutely, strategically, phenomenally*.

The 20% pronoun density (148 instances) in this speech is a crucial finding, due to the fact that it is five times the academic maximum of 4%.

With 62 occurrences and 8%, the conjunction is strong, surpassing the 5% academic and 4.5% conversational limitations.

On March 4, 2025 President Donald J. Trump gave his first address to the Joint Houses of Congress during his second term with 657 words.

With 79 occurrences, the nominal density stays within the 12% range that has been seen during the presidency, which falls below the conversational average of 15%. With financial and institutional vocabulary dominating: *dollars, billions, defense, security, conflict, ocean, table, letter, sovereignty, independence*.

There are 207 verbs in the text, which is an exceptionally high frequency that accounts for over one-third of the entire address, with 31%. This is three times the academic standard of 10%.

The presence of 36 adjectives is higher (5%) than the conversational average of 2.5%.

Adverbs (5.78%): With 38 occurrences and 6%, the adverbial density is very similar to the conversational limit of 5%. The adverbs *tirelessly*, *needlessly*, *sadly* are the most formally elaborate in the POTUS corpus, functioning as markers of effort, regret, and criticism respectively.

The phrase "*like taking candy from a baby*" is the only idiomatic expression in either corpus used to describe financial policy and is a notable register anomaly, briefly collapsing the formal congressional register into populist metaphor.

The 21% pronoun density (141 instances) exceeds the academic restriction of 4% by more than five times.

With 27 occurrences (4%), the conjunction frequency falls short of the 4.5% conversational limit.

On March 5, 2025, President Trump gave a short 139-word speech.

The nominal density is precisely at the 15% conversational average with 21 occurrences and 15%. Key words are *peace*, *sovereignty*, *independence*, *agreement*, *minerals*, *security*, *table*.

There are 34 verbs in the text (made 24%), which is about 2.5 times the academic standard of 10%.

This transcript's 12% adjective density (16 times) is a noteworthy discovery. This surpasses even the academic limit of 10%. The adjective *beautiful* is applied twice to the prospect of Russian peace signals (*wouldn't that be beautiful*).

With eight occurrences and 8%, the adverbial density corresponds to the 5% conversational marker.

The 18 pronouns and 13% used are three times the academic maximum of four percent.

With seven occurrences and 5%, the conjunction frequency is identical to the academic norm of 5%.

On the 8th of March, President Donald Trump gave a 146-word speech. With just 9 occurrences and 6%, the nominal density is remarkably low, falling over one-third of the 15% conversational average.

Verbs make up a staggering 36% of the text (53 occurrences). This exceeds the academic standard of 10% by more than three times.

Adjectives (4.11%): The six adjectives with 4% are above the conversational average of 2.5%.

With 12 occurrences and 8%, the adverbial density is over twice as high as the conversational guideline of 5%.

The 24% pronoun density (35 instances) is six times the 4% academic limit and one of the highest findings in the research. In this improvised press exchange, with *I*, *he*, *they*, *it* cycling rapidly across referents. The statement that Russia bombing Ukraine is "*what anybody would do*" is grammatically encoded through the pronoun *anybody* which is a universalising indefinite that normalises Russian aggression by generalising it to hypothetical universal behaviour.

In 11 cases there are 8%, the frequency of conjunctions is higher than the academic norm of 5%.

The last but not least 214-word presidential speech was given on March 17, 2025.

With 30 occurrences (14%), the nominal density is a bit lower than the conversational average of 15%.

There are 67 verbs in the text, which made up 31%, which is more than three times the academic standard of 10%.

The adjective density is still modest (3%), just above the conversational average of 2.5%, with only 7 occurrences.

Adverbs (5.14%): With 11 occurrences and 5%, the adverbial density is exactly in line with the 5% conversational standard.

The 20% pronoun density (43 times) is five times the academic cap of 4% and are split between *we* (Trump's negotiating team) and *I*, with *you* now absent perhaps due to the fact that Zelenskyy is no longer being addressed but discussed in the third person, a subtle demotion from interlocutor to subject.

The conjunction frequency above the academic norm of 5% in 12 cases and 6%.

To conclude, a very consistent and unique language architecture is shown by the long-term corpus analysis of Donald Trump's speeches from the pre-presidential era to the mature POTUS period.

Finally, moving to the X sub-corpus, there are 4 posts from the POTUS era, which I analyzed altogether due to the article word limitation.

An important look at this 580-word corpus of social media content. The limitations of the platform usually encourage conciseness on X (previously Twitter), but the syntactic distribution reveals a distinctive linguistic profile that increases descriptive intensity while reflecting his spoken oratory.

Nouns are less than the 15% conversational average (11%), with 65 occurrences. It demonstrates how the discourse avoids the thick "nominal packing" present in the 30% academic standard, even when written

The 122 verbs (21%) continue to have a high frequency, almost twice as high as the academic standard of 10%.

Adjectives (7%) are far greater than the conversational average of 2.5%, with 42 occurrences. The 34 adverbs (6%), which function as tonal intensifiers in place of a live speech's vocal accent, are in line with the 5% conversational criterion.

With 48 occurrences (8%), pronouns double the 4% academic limit.

With 37 conjunction occurrences (6%), this frequency exceeds the 4.5% conversational norm.

To support the findings, Table 4.13 below demonstrates a comparison of parts of speech data in the two corpora.

Part of speech	Biber et al. (1999) Conversational	Biber et al. (1999) academic	Biden corpus (average)	Trump sub-corpus (average)	Trump X sub-corpus
Noun	15%	30%	17-20%	8-14%	11%
Verb	12.5%	10%	19-20%	27-31%	21%
Adjective	2.5%	10%	6-9%	4-9%	7%
Adverb	5%	3%	5-6%	7-9%	6%
Pronoun	16.5%	4%	8-11%	17-24%	8%
Conjunction	4.5%	5%	8%	6-8%	6%

Table 4.13. *Parts of speech across corpora.*

A central objective of corpus linguistics is to identify how changes in the communicative channel or medium affect linguistic choices. When analyzing the political discourse of Donald Trump, comparing his spoken oratory to his written communications on X (formerly Twitter) reveals a sharp stylistic divergence. While his general sub-corpus exhibits an exceptionally high pronoun frequency of 17–24% aligning closely with or even exceeding standard conversational registers (16.5%) this figure drops dramatically to just 8% within the Trump X sub-corpus. In the context of register and discourse analysis, this shift is not merely an accidental variation; it represents a fundamental transition in text structure and communicative intent.

Traditional political speechmaking allows for a loose, repetitive, and narrative-heavy syntax. Spoken registers are naturally dense with pronouns because speakers constantly rely on immediate personal references (*"I will prevent war, we had a debate, and they would call me..."*) to maintain a conversational flow.

On X, however, the text is constrained by strict character limits and a visually fast-paced environment. To adapt, the grammar undergoes a process of dense condensation. The reduced pronominal frequency observed in Trump's Twitter sub-corpus (8%) relative to his spoken sub-corpus (17–24%) constitutes one of the most analytically significant findings of the present study. At first glance, this reduction appears paradoxical: social media platforms are conventionally associated with informal, interactive discourse, which Biber et al. (1999) identify as a primary conditioning environment for elevated pronoun use.

This statistical variation reflects a deeper shift in the communicative function of the discourse. In his spoken speeches, Trump's high pronoun density

(17–24%) creates an intensely interpersonal, dynamic relationship between the sender and the receiver.

However, a closer examination of the pronominal patterns within the Twitter sub-corpus reveals that this reduction is not indicative of reduced affective engagement, but rather reflects a fundamental shift in discourse mode from dialogic to proclamatory that systematically suppresses interactional pronoun categories while foregrounding agentive first-person reference.

A quantitative survey of the four tweets comprising the Twitter sub-corpus identifies approximately 25 pronominal tokens in total, distributed across three dominant categories: first-person singular (*I*: 7 occurrences), first-person plural (*we/our*: 6 occurrences), and third-person singular masculine (*he/his*: 7 occurrences). Crucially, second-person pronouns (*you/your*) are entirely absent from the sub-corpus. This zero-frequency finding is particularly telling within Biber's framework: the second person is the prototypical marker of interactional involvement, indexing direct audience address and dialogic reciprocity. Its complete absence in Trump's Twitter sub-corpus indicates that, despite the interactive affordances of the platform, the texts function as unidirectional proclamations rather than conversational exchanges. This aligns with Biber et al.'s (1999) observation that pronoun frequency is conditioned not merely by medium but by the degree of genuine interactional involvement encoded in the discourse.

Further insight is gained by examining the functional distribution of the pronouns that do occur. Third-person masculine pronouns (*he/his*) appear almost exclusively in reference to Ukrainian President Volodymyr Zelenskyy, functioning as referential rather than dialogic tokens. Their role is evaluative and distancing – constructing Zelenskyy as an object of scrutiny rather than a participant in discourse. This referential use contributes minimally to interactional density and is more consistent with the informational pole of Biber's continuum, where third-person reference dominates over audience-directed address.

The distribution of first-person pronouns reveals an equally significant pattern. While political speech conventionally employs the first-person plural to construct national solidarity and collective agency, a feature well-documented in presidential rhetoric, Trump's Twitter sub-corpus exhibits a marked asymmetry between *I* and *we*. First-person singular forms consistently occupy syntactically agentive positions: *I love Ukraine, I want PEACE, I am strongly considering, I emphasized, I have determined*. By contrast, *we* appears in contexts of ambiguous reference – sometimes indexing the United States as a collective entity [*we are successfully negotiating*], sometimes the Trump administration specifically – but never functioning as an inclusive pronoun that incorporates the audience. This asymmetry constructs the speaking subject as a singular, volitional agent whose authority is individual rather than institutional or collective.

This pronominal self-centering is reinforced by the transitivity patterns that characterize the sub-corpus. A systematic analysis of subject-verb-object constructions across the four tweets reveals a consistent asymmetry in the distribution of agency. Trump is lexicalized as the subject of active, volitional processes [*negotiating, considering, determining, emphasizing*]. Other political factors, by contrast, are consistently assigned passive or negative agentive roles: Biden [*never tried*]; Europe [*has failed to bring Peace*]; Zelenskyy [*talked*] the United States into expenditure is a causative construction that frames him as a manipulative rather than legitimate agent. Russia, notably, appears as an aggressor only within a distancing quotation (*'pounding'*), a graphological strategy that will be addressed in the following section. Trump thus emerges as the sole consistently active, competent agent across the corpus – a pattern that constitutes what may be termed a heroic self-framing strategy, in which the speaking subject is constructed in implicit contrast to a field of failure and incompetence.

This transitivity asymmetry maps directly onto the pronominal findings: the dominance of *I* in agentive positions is not merely a stylistic preference but a grammatically encoded ideological stance. It is worth noting that this pattern diverges markedly from Biden's sub-corpus, in which collective factors as Ukraine, NATO allies, the democratic world are more frequently foregrounded as agents, and the presidential *I* is more typically embedded within inclusive *we* constructions.

A further paralinguistic feature unique to the Twitter sub-corpus warrants attention in relation to register and stance. Trump employs graphological scare quotes on different occasions across the four texts: "*TRUMP*," "*MISSING*," "*like a fiddle*," "*gravy train*," "*Economic Partnership*," "*Critical Minerals and Rare-Earths Deal*," and '*pounding*' that are identified stance markers as a key feature of written registers in which the producer cannot rely on prosody or gesture to signal epistemic positioning. Scare quotes function here as written stance devices, performing one of two operations: epistemic distancing, in which the quoted term is implicitly contested or ironized ("*gravy train*," "*MISSING*"), or reframing, in which a term is resignified through its graphic isolation. The self-quotation of "*TRUMP*" is particularly anomalous rather than a personal pronoun or proper name used referentially, it functions as a brand token, performing institutional rather than personal identity and further suppressing the interactional register that high pronoun use would otherwise index.

Taken together, these findings support the interpretation that Trump's Twitter sub-corpus occupies a hybrid position on Biber's register continuum. Nominally interactive in medium, it is structurally informational in form: pronoun density is suppressed by the absence of second-person address; what pronouns do occur are predominantly referential or agentively self-centered; and paralinguistic devices substitute for the affective functions that interactional pronouns would otherwise serve in spoken discourse. The reduction from

17–24% in the spoken sub-corpus to 8% in the Twitter sub-corpus thus reflects not a diminution of rhetorical intent, but a register-conditioned redistribution of the linguistic means by which that intent is expressed.

5. Conclusions and discussion.

The comprehensive corpus analysis of President Joe Biden's and Donald Trump's public discourse regarding the war in Ukraine yields critical insights into the linguistic architecture, thematic emphases, and structural choices defining modern executive communication during geopolitical crises. Spanning a corpus of eight Biden speeches (February 2022 – September 2024), three Trump pre-presidential speeches (March–September 2024), and ten Trump POTUS-era speeches and statements (November 2024 – March 2025), the analysis reveals not merely stylistic differences between two individual speakers, but structurally distinct and ideologically coherent discourse systems, each grammatically encoding a fundamentally different construction of the war, its causes, its responsibilities, and its resolution. By integrating macro-level lexical frequency tracking via Voyant Tools with micro-level Part-of-Speech density metrics extracted through Lingualyzer, this investigation demonstrates how ideological positioning, platform-specific constraints, and rhetorical goals manifest systematically in text. The empirical findings reveal that both figures utilized language strategically to frame the conflict, assign accountability, and shape global public opinion, though they operated through fundamentally distinct rhetorical paradigms.

Crucially, the rigor, methodology, and preliminary outcomes of this research framework have already undergone formal academic peer review and dissemination. The structural concepts and data collection methods utilized throughout this study were compiled and published across two distinct articles within *Mundus Philologiae*. Furthermore, the empirical findings regarding executive rhetorical adaptation were presented and defended before the scholarly community during a formal university conference (Філологічні перлини, the 16th of May, 2025). These academic milestones not only validate the analytical framework utilized herein but also emphasize the contemporary relevance of computational linguistics in evaluating political discourse during ongoing international conflicts.

At the macro-linguistic level, both leaders maintained a core focus on the primary geopolitical, humanitarian, and strategic realities of the crisis. This thematic concentration is empirically supported by the consistent frequency distributions of centralized keywords across both datasets, notably: *Ukraine, Russia, people, world, Putin*. Trump's verbal profile is characterised by three dominant patterns absent or marginal in Biden's corpus: future-tense self-projection (*will prevent, will have settled, will get solved*), past counterfactual construction (*would have never happened, could have been*

settled, would have given), and present-tense cognitive hedging (I think, I don't know, I find it difficult). Each pattern encodes a distinct rhetorical stance: the future constructions assert personal omnipotence; the counterfactuals construct an alternative history in which Trump's hypothetical actions prevent all suffering; the hedging verbs perform neutrality in contexts where condemnation would be expected. Biden's verb profile, by contrast, is institutionally distributed across collective subjects – *we have assembled, we will not tire, we stand in solidarity* – and spans a wider tense range that, as the diachronic analysis demonstrates, shifts systematically from crisis-urgency in 2022 to vindication-retrospection in 2024.

However, variations in word usage among the analyzed sub-corpora illuminate how rhetorical methods systematically adjust to shifting settings, including official presidential speeches, pre-POTUS commentary, and immediate social media communications.

While a traditional surface-level analysis outlines *what* subjects dominate the discourse, the algorithmic parts of speech tracking provided by Lingualyzer proves that syntactic construction is just as crucial to strategic communication as lexical choice. By examining the density of lexical words versus function words, this study moves beyond mere frequency counts to reveal two essentially distinct models of political authority and communication design.

Biden's discourse is defined by high noun density and a structured frequency of conjunctions. This grammatical architecture imbues his rhetoric with the logical coherence, structural stability, and informational weight expected of institutional leadership. By grounding his addresses in substantive nouns and connectives, Biden establishes an administrative, analytical framework that presents international policy as a matter of universal principles, institutional frameworks, and collective alliances.

Conversely, Trump's rhetorical style is driven by a saturation of pronouns and high-frequency verbs. This structural pattern frames the geopolitical crisis not as an institutional dilemma, but as a dynamic sequence of urgent, individualized actions and interpersonal negotiations. This verbal orientation creates an active, immediate, and highly personalized narrative of authority.

Furthermore, the cross-platform analysis demonstrates that these grammatical profiles shift dramatically depending on the medium of delivery. The examination of Trump's "X" (formerly Twitter) sub-corpus underscores how rhetorical techniques adapt to digital platforms. On social media, the primary communicative objective shifts away from describing policy details toward "branding" the conflict through powerful emotional qualifiers, immediate personal pronouns, and condensed, high-impact phrasing (McIntyre & Walker, 2019).

In contrast, President Biden's institutional speeches consistently prioritize collective action, international cooperation, and the humanitarian cost of the conflict, maintaining an informational density that mirrors traditional diplomatic

registers. Ultimately, Trump's remarks exhibit a more critical, adversarial, and opinion-driven rhetorical approach – particularly visible in his pre-POTUS interviews and social media posts – whereas Biden relies on an institutional syntax to project stability and multilateral alignment.

To expand upon the empirical foundations established in this thesis, several clear pathways for future academic inquiry are proposed. These prospects aim to deepen the field of computational political discourse analysis by extending the temporal, cross-cultural, and multi-modal scope of the current corpus. Future studies could significantly broaden this corpus-based approach by incorporating a larger volume of speeches, interviews, and digital communications from both figures over an extended timeframe. Conducting a longitudinal study that systematically compares language use at various successive stages of the war would offer a more granular understanding of strategic communication and rhetorical adaptation. Tracking these shifts would reveal how linguistic framing, thematic emphasis, and grammatical complexity evolve in direct response to changing material conditions on the battlefield and shifting domestic political climates.

A crucial next step involves shifting the analytical lens from text production to text reception. It is vital to examine, in an evidence-based way (Zyngier, Chesnokova and Viana, 2007; Zyngier et al., 2008; van Peer and Chesnokova, 2019), the actual reactions of audiences to these political speeches. This approach would clarify the specific verbal triggers that stimulate public engagement, emotional responses, or political alignment, bridging the gap between computational text metrics and empirical cognitive outcomes.

Building on the principles of empirical aesthetics and reader-response metrics, it is worthwhile to investigate whether the reception of this political discourse varies across distinct cultural cohorts. Specifically, future research should check, in an evidence-based way (van Peer and Chesnokova, 2019; van Peer and Chesnokova, 2025), whether the reaction of audiences in two fundamentally different cultures namely American and Ukrainian differs when exposed to the same analytical texts. Such an inquiry would reveal how domestic political positions versus direct exposure to the realities of conflict influence the cognitive and emotional processing of executive rhetoric.

The present study is not without limitations, which simultaneously point toward productive avenues for future research. Most notably, the analysis of Donald Trump's linguistic output is divided between a pre-presidential sub-corpus of three campaign-era speeches (March–September 2024) and a POTUS sub-corpus of ten speeches and statements from his second term (November 2024 – March 2025). While this division enables a limited comparison across political contexts, it excludes a substantial body of relevant material produced during Trump's first presidential term (2017–2021) – the period during which he was actively managing US-Ukraine relations as well as his post-presidential period (2021–2024), during which he commented on the

war from outside institutional power with fewer constraints on rhetorical register.

Finally, future research should explore how the physical modality of language delivery impacts human perception and evaluation. It would be highly valuable to verify experimentally if speeches heard as spoken audio or viewed in video format are assessed in the exact same manner as their written scripts read in text form (van Peer and Chesnokova, 2024). This direction would isolate the persuasive power of structural syntax and lexical choices from prosodic elements like intonation, pacing, and visual delivery, offering a comprehensive view of how strategic communication functions across different sensory channels in the modern political landscape.

REFERENCES

1. Anadolu Agency. (n.d.). Anadolu Agency. <https://www.aa.com.tr/en>
2. Anthony, L. (2015). TagAnt (Version 1.2.0). Tokyo: Waseda University. <http://www.laurenceanthony.net/software/tagant/>. Accessed 25 May 2019.
3. Anthony, L. (2016). AntConc (Version 3.4.4). Tokyo: Waseda University. <http://www.laurenceanthony.net/software/antconc/>. Accessed 25 May 2019.
4. Antoniuk A.R. (2025). THE LANGUAGE OF WAR: CORPUS ANALYSIS OF JOE BIDEN AND DONALD TRUMP'S SPEECHES ON UKRAINE. *MUNDUS PHILOLOGIAE*, 5. <https://mundphil.kubg.edu.ua/index.php/journal/article/view/76>
5. Archer, D., Rayson, P., Piao, S., & McEnery, T. (2004). Comparing the UCREL Semantic Annotation Scheme with lexicographical taxonomies. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX (European Association for Lexicography) International Congress (Euralex 2004) (Vol. III, pp. 817–827)*. Lorient: Université de Bretagne Sud.
6. Archives | International Premier Journal of Languages & Literature. (2026). Ipjll.com. <https://ipjll.com/ipjll/index.php/journal/issue/archive>
7. Arif, M., Anwar, B., & Shahid, U. (2026). S Functional variation in lexical bundles: A corpus-based comparative analysis of native and non-native web-based English. *Social Sciences Spectrum*, 5(1), 409–429. <https://doi.org/10.71085/sss.05.01.506>
8. Aston, G., & Burnard, L. (1998). *The BNC handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
9. Atwell, E. (2008). Tokenizing and part-of-speech tagging. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook. Volume 1 (pp. 527–551)*. Berlin/New York: Walter de Gruyter.
10. Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
11. Bennett, G. (2010). Using Corpora in the Language Learning Classroom. <https://doi.org/10.3998/mpub.371534>
12. Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.
13. Biber, D., Gray, B., & Poonpon, K. (2011). Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development? *TESOL Quarterly*, 45(1), 5–35. <https://doi.org/10.5054/tq.2011.244483>

14. Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
15. Biber, D., Reppen, R., & Conrad, S. (2004). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
16. Bird, S., Loper, E., & Klein, E. (2009). *Natural language processing with Python*. Sebastopol: O'Reilly Media.
17. (Charteris-Black, 2011; Partington, Duguid, & Taylor, 2013).
18. Chesnokova, A. and S. Zyngier. (2024). Literary Awareness and the Poetic Experience: An Evidence-Based Assessment. *The Modern Higher Education Review* 9, 84–100. <https://edreview.kubg.edu.ua/index.php/edreview/article/view/178/202>
19. Crasborn, O. (2010). What does 'informed consent' mean in the internet age? Publishing sign language corpora as open content. *Sign Language Studies*, 10(2), 276–290.
20. Crasto, E. P., Phadke, G. G., Mohammed, A., & Sharma, R. (2025). Echoes of Leadership : A Natural Language Processing Analysis of the Presidents of The United States of America. *Proceedings of the 2025 18th International Conference on Computer Science and Information Technology*, 326–336. <https://doi.org/10.1145/3783862.3783904>
21. Douglas, F. M. (2003). The Scottish corpus of texts and speech: Problems of corpus design. *Literary and Linguistic Computing*, 18(1), 23–37.
22. Fabiszak, M., & Kaszubski, P. (2005). *A Corpus Based Study of War Metaphors*. Pragmatics Today.
23. Francis, W. N., & Kucera, H. (1964/1979). *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Department of Linguistics, Brown University. <http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM>. Accessed 24 May 2019.
24. Garside, R., Leech, G., & Sampson, G. (Eds.). (1987). *The computational analysis of English: A Corpus-based approach*. London: Longman.
25. Ghodke, S., & Bird, S. (2010). Fast query for large Treebanks, 267–275. *Proceedings of NAACL 2010*. Los Angeles, CA.
26. Green, L. J. (2002). *African American English: A linguistic introduction*. Cambridge: Cambridge University Press.
27. Greenbaum, S. (Ed.). (1996). *Comparing English worldwide: The international Corpus of English*. Oxford: Clarendon Press.
28. Hammond, J., & Macken-Horarik, M. (1999). *Critical Literacy: Challenges and Questions for ESL Classrooms*.
29. Hardie, A. (2014). Modest XML for Corpora: Not a standard, but a suggestion. *ICAME Journal*, 38, 72–103.

30. Hermann Moisl. (2007). Creating and Digitizing Language Corpora, Volume 1: Synchronic Databases. Academia.edu. https://www.academia.edu/71812922/Creating_and_Digitizing_Language_Corpora_Volume_1_Synchronic_Databases
31. Hope vs. Defiance: A Comparative Corpus Analysis of Emotional Lexis in MLK's I Have a Dream and Churchill's We Shall Fight on the Beaches Azmat Batool¹ , Dr. M. Faisal Hanif² , Yasmin Saleh³ , Mussarat Qamar⁴ , *Faiza Altaf
32. Jaworska, S. (2016). A comparative corpus-assisted discourse study of the representations of hosts in promotional tourism discourse. *Corpora*, 11(1), 83–111. <https://doi.org/10.3366/cor.2016.0086>
33. Johansson, S., Leech, G., & Goodluck, H. (1978). Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers. Department of English, University of Oslo. <http://clu.uni.no/icame/manuals/LOB/INDEX.HTM>. Accessed 24 May 2019.
34. Linders, G. M., & Louwse, M. M. (2023). Lingualyzer: A computational linguistic tool for multilingual and multidimensional text analysis. *Behavior Research Methods*, 1-28. <https://link.springer.com/content/pdf/10.3758/s13428-023-02284-1.pdf>
35. McIntyre, D., & Walker, B. (2019). 1 Combining corpus linguistics and stylistics. In Edinburgh University Press eBooks (pp. 1–24). <https://doi.org/10.1515/9781474413220-005>
36. Mikros, G. (2013). *Corpus Manager: A Tool for Multilingual Corpus Analysis*. Cambridge University Press.
37. Mitkov, R. (n.d.). Using Comparable Corpora to Track Diachronic and Synchronic Changes in Lexical Density and Lexical Richness. Retrieved May 26, 2026, from <https://mt-archive.net/BUCC-2012-Stajner.pdf>
38. Petiy, N. (2023). CONCEPTUAL METAPHORS VERBALIZING WAR IN UKRAINE IN MEDIA DISCOURSE. *Věda a Perspektivy*, 2(21). [https://doi.org/10.52058/2695-1592-2023-2\(21\)-203-214](https://doi.org/10.52058/2695-1592-2023-2(21)-203-214)
39. PhD, D. B. (2026, April 13). Claude AI (Anthropic)¹, Duncan Borthwick² ¹ Anthropic, San Francisco, CA, USA · ² Labcoats and Blackboards Ltd. Published: April 2026 · DOI: none ABSTRACT This paper demonstrates the application of large language model AI — specifically Claude (Anthropic) — to the synthesis and quantitative analysis. <https://www.linkedin.com/pulse/quantitative-linguistic-analysis-us-presidential-duncan-borthwick-phd-bjkpe>
40. Presidential speeches | Miller Center. (2016). Miller Center. <https://millercenter.org/the-presidency/presidential-speeches>

41. Romāne-Kalniņa, L. (2020). KEYWORD ANALYSIS IN PRESIDENTIAL RHETORIC THROUGHOUT HUNDRED YEARS OF INDEPENDENCE OF ESTONIA, LATVIA AND LITHUANIA: CORPUS LINGUISTIC TOOLS IN CRITICAL DISCOURSE ANALYSIS. *Language for International Communication: Linking Interdisciplinary Perspectives*. Volume 3, 95–109. <https://doi.org/10.22364/lincs.2020.08>
42. Römer, U. (2011). Corpus Research Applications in second Language teaching. *Annual Review of Applied Linguistics*, 31, 205–225. <https://doi.org/10.1017/s0267190511000055>
43. Stubbs, M. A tribute to John McHardy Sinclair (14 June 1933-13 March 2007). *The Phraseological View of Language: A Tribute to John Sinclair*, edited by Thomas Herbst, Susen Faulhaber and Peter Uhrig, Berlin, Boston: De Gruyter Mouton, 2012, pp. 1-16. <https://doi.org/10.1515/9783110257014.1>
44. Svitlana Goloshchuk, & Tomczak-Boczko, J. (n.d.). A Corpus Assisted Discourse Study of President Volodymyr Zelenskyy's Wartime Addresses. <https://ceur-ws.org/Vol-4107/paper7.pdf>
45. Traugott, E. C. (2012). Geoffrey Leech, Marianne Hundt, Christian Mair and Nicholas Smith, *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press, 2009. <https://doi.org/10.1017/s1360674311000359>
46. van Peer, W. and A. Chesnokova (2019). What Literature Does to Our Emotions, and How Do We Know? Empirical Studies Will Tell. *Синопис: text, контекст, медіа. Synopsis: Text, Context, Media*, 25(1), 1–10.
47. van Peer, W. and A. Chesnokova (2024). “The Sound Must Seem an Echo to the Sense”: Experiencing Oral and Silent Reading of Poetry. In L. Pillière and S. Sorlin (eds), *Style and Sense(s)*. Cham: Palgrave Macmillan, pp. 235–257.
48. van Peer, W. and A. Chesnokova. (2025). How do you feel? Literature's Contribution to Learning the Granularity of Emotions. In C.-H. Mayer and E. Vanderheiden (eds), *International Handbook of Emotions. Resourceful Cultural Perspectives*, Vol. 1. Cham: Springer, pp. 253–261.
49. Zyngier, S., A. Chesnokova and V. Viana (eds.) (2007). *Acting & Connecting. Cultural Approaches to Language and Literature*. Munster: LIT Verlag.
50. Zyngier, S., Bortolussi, M., Chesnokova, A. and Auracher, J. (eds). (2008). *Directions in Empirical Literary Studies: In honor of Willie van Peer*. Amsterdam / Philadelphia: John Benjamins.