УДК 81.11

THE LANGUAGE OF WAR: CORPUS ANALYSIS OF JOE BIDEN AND DONALD TRUMP'S SPEECHES ON UKRAINE

Antoniuk Anastasiia
6th year student of the Faculty of Romance and Germanic Philology
Borys Grinchenko Kyiv Metropolitan University
arantoniuk.frgf24m@kubg.edu.ua

РИТОРИКА ВІЙНИ: КОРПУСНЕ ПОРІВНЯННЯ ПРОМОВ ДЖО БАЙДЕНА І ДОНАЛЬДА ТРАМПА ЩОДО ВІЙНИ В УКРАЇНІ

Антонюк Анастасія
студентка 6 курсу Факультету романо-германської філології
Київський столичний університет імені Бориса Грінченка
arantoniuk.frgf24m@kubg.edu.ua

This article presents the lexis used in Joe Biden and Donald Trump's speeches on Ukraine through corpus analysis conducted with the Voyant Tools software, which enables multilevel text analysis, including word-frequency and collocation analysis as well as the generation of word clouds and trend visualizations, and reveals the evolution of their rhetorical strategies. The paper is aimed to define the lexical features and the frequency of use of lexical items in the speeches of President Biden and President Trump on the war in Ukraine. Additionally, this research aims to understand how their linguistic choices influence public perception of the Russian invasion and to what extent these linguistic choices shape political reactions. This research looks at 8 speeches by President Biden (4,748 words total) and 13 speeches by President Trump (4,971 words total) in that they reflect American views on the events. The corpus is subdivided into two sub-corpora: the speeches delivered by Joe Biden and Donald Trump, respectively. However, the Trump sub-corpus is further divided into two sections: speeches delivered prior to becoming president, and addresses following his election and inauguration. Additionally, a social media network "X" sub-corpus consisting of four online statements was compiled and analysed. The results indicate that President Biden predominantly mentioned "Ukraine", "people", and "freedom" in his speeches. Alternatively, President Trump's most-used words are "Ukraine", "war", and "Russia". These findings suggest differing rhetorical and narrative strategies used by the two leaders when addressing the conflict.

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

Стаття присвячена дослідженню лексики, що використовується в промовах Джо Байдена та Дональда Трампа щодо подій в Україні, на основі корпусного аналізу, проведеного за допомогою програмного забезпечення Voyant Tools, яке забезпечує багаторівневий аналіз тексту, включаючи аналіз частотності слів і словосполучень, а також створення хмар слів і візуалізацію тенденцій, і розкриває еволюцію їхніх риторичних стратегій. Метою цієї роботи є вивчення лексичних особливостей та частоти вживання лексичних одиниць у промовах президента Байдена та президента Трампа щодо війни в Україні. До того ж це дослідження має на меті зрозуміти, як їхній лінгвістичний вибір впливає на сприйняття російського вторгнення громадськістю та в якій мірі цей лінгвістичний вибір формує політичні реакції. У цьому дослідженні розглядаються 8 промов президента Байдена (загалом 4748 слів) та 13 промов президента Трампа (загалом 4971 слово), оскільки вони відображають американський погляд на ці події. Корпус поділений на два підкорпуси: промови Джо Байдена та Дональда

Трампа відповідно. Втім, підкорпус Трампа надалі поділяється на дві частини: промови, виголошені до того, як він став президентом, та промови після його обрання та інавгурації. Крім того, було складено та проаналізовано підкорпус соціальної мережі «X», що складається з чотирьох онлайн-заяв. Результати показують, що президент Байден у своїх промовах переважно згадував «Україну», «людей» та «свободу». Натомість найчастіше вживаними словами президента Трампа є «Україна», «війна» та «Росія». Ці висновки свідчать про різні риторичні та наративні стратегії, які використовують два лідери, коли говорять про конфлікт.

Ключові слова: *корпусний аналіз, президент Байден, президент Трамп, Україна, війна.*

Introduction. Foundational work in corpus linguistics began roughly a century ago. By the late nineteenth century, lexicographers, also known as lexicon compilers, collected examples of language use to define terms with greater precision. Before the development of computers, researchers gathered linguistic samples on small slips of paper and stored them in pigeonholes. The arrival of computing technology created the conditions for what researchers now describe as modern corpora. The first digital corpus, the Brown Corpus, appeared in 1961 with approximately one million words. Contemporary general corpora now contain hundreds of millions of words, and corpus linguistics contributes substantially to contemporary language research and instruction.

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

1. Which patterns occur in relation to particular linguistic or lexical features?
2. Which distinctions and registers shape these patterns?

Such scholars as Filter, Biber, Johansson, Francis, Houston, Conrad, and McCarthy advanced the development of modern corpus linguistics, and their work continues to shape the field (Römer, 2011). John Sinclair is acknowledged as one of the most significant and influential figures in this area of study (Stubbs, 2007).

Scholars argue that corpus research extends beyond the identification of quantitative patterns and includes functional interpretations that explain the underlying reasons for observable phenomena. For this reason, corpus-based investigations devote substantial attention to interpreting and illustrating quantitative tendencies. Studies in corpus linguistics commonly integrate both quantitative and qualitative methods, and researchers frequently proceed from quantitative outcomes toward qualitative insight (McIntyre & Walker, 2019).

For corpus analysis of texts, such tools as Voyant Tools software are used. This web-based, open-source text analysis environment supports the exploration of individual texts and corpora and serves researchers, students, and broader audiences. Users can examine online materials or upload their own texts, which makes the platform particularly useful for work in the digital humanities. Such an analysis allows not only to improve and further explore the conclusions of already conducted studies but also to look at and define many seemingly traditional linguistic concepts and terms in a new way.

It is also important that corpus analysis of the language of war in Ukraine (2022-2025 at the moment of analysing) will allow us to demonstrate the divergence in rhetorical strategies between President Joe Biden and President Donald Trump. Such analysis of the corpus, which comprises 8 speeches by President Biden (4,748 words total) and 13 speeches by President Trump (4,971 words total), makes it possible to systematically track the shift in their linguistic preferences and ideological framing of the conflict.

It is possible to see how the Presidents' contrasting social narratives, emotional appeals, and core vocabulary surrounding the invasion changed and differed across their respective sub-corpora, particularly when distinguishing between Trump's pre- and post-presidency addresses. This comparison provides a crucial opportunity to analyze the extent to which U.S. political rhetoric on a major global crisis is shaped by prevailing political registers and personal style.

Most importantly, no such systematic corpus analysis has previously been conducted that directly contrasts the specific lexical and thematic patterns in the speeches of both Presidents Biden and Trump dedicated exclusively to the war in Ukraine.

Methods. 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 3.1. Word cloud 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 3.2.
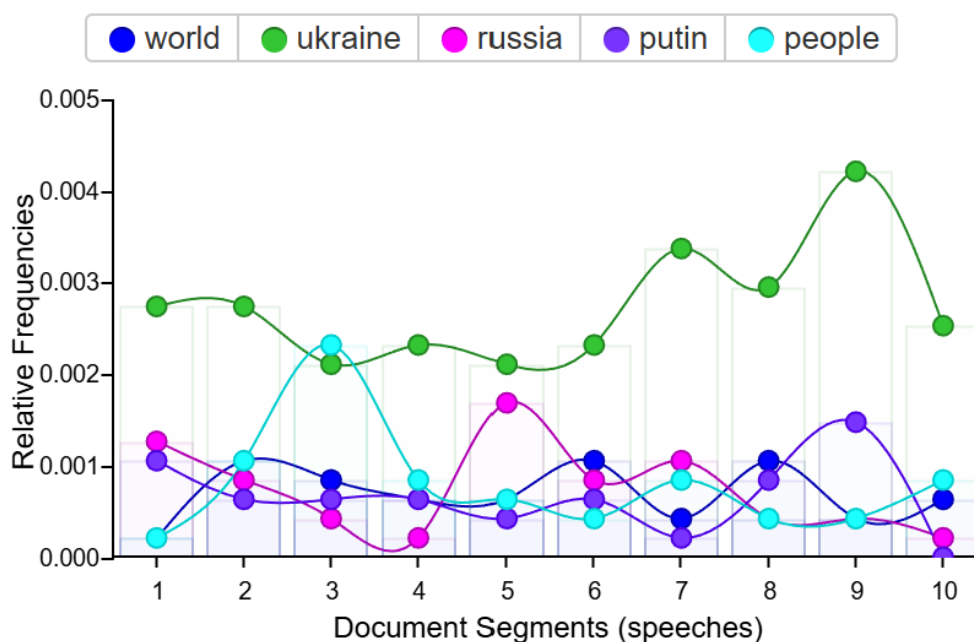
Figure 3.2. Trends of Joe Biden's speeches.

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.

Word cloud. Such words as *It's (12), Ukraine (10), think (9), say (8), and war (7)* are the most frequently occurring ones in the corpus.

Figure 3.3. Word cloud 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.
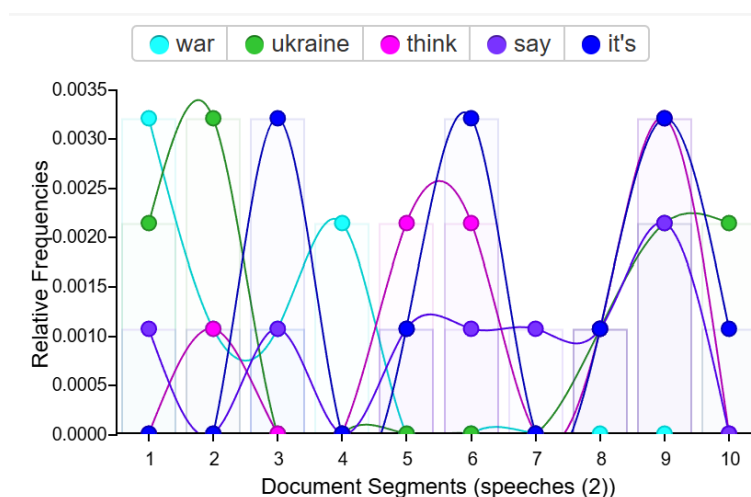
Figure 3.4. Trends of Donald Trump's pre-POTUS speeches.

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 3.5. Word cloud of Donald Trump's pre-POTUS speeches.

As can be seen from Figure 3.5 above, the top 5 words are *Ukraine (41), think (30), war (26), president (22), and lot (21).*

In conclusion, the Pre-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. As demonstrated in Figure 3.6 below, the top 5 words are *war (8), Zelenskyy (7), Russia (7), united (6), Ukraine (6).*
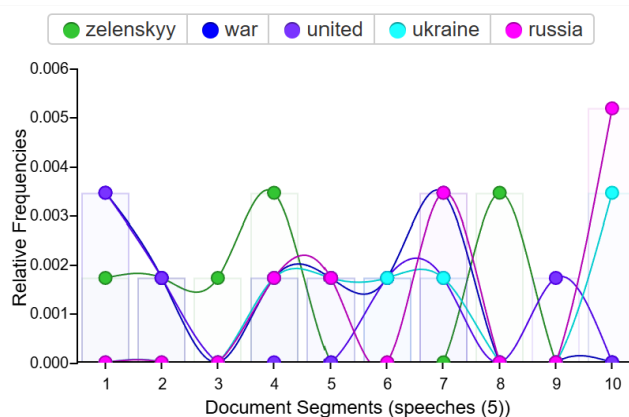

Figure 3.6. Trends of Donald Trump's speeches.

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.*

Conclusions and further research prospects. Corpus analysis of President Trump's and President Biden's speeches on the war in Ukraine has provided insights regarding thematic emphases, linguistic choices, and frequency patterns during the crisis. Both presidents' primary attention on the geopolitical, humanitarian, and strategic aspects of the conflict is reflected in the analysis, which

identifies consistent patterns in vocabulary, such as the frequent use of Ukraine, Russia, people, world, and Putin. Word usage variations among sub-corpora show how rhetorical methods adjust to shifting settings, such as social media communication, presidential speeches, and pre-POTUS comments. These results highlight how both presidents used language strategically to define the debate, assign criticism, and shape public opinion. While Trump's remarks show a more critical and opinion-driven rhetorical approach, especially in pre-POTUS interviews and X posts, President Biden's speeches often highlight group action, international cooperation, and the humanitarian cost of the conflict.

By looking at more speeches from both presidents over time, future studies could broaden this corpus-based approach and find patterns, emphasis shifts, or framing changes as the conflict progresses. Another possibility would be to examine, in an evidence-based way (Zyngier, Chesnokova and Viana, 2007; Zyngier et al., 2008; van Peer and Chesnokova, 2019), the reaction of audience to such speeches in that verbal triggers of this reaction are made clear. A deeper understanding of strategic communication and rhetorical adaptation in the context of ongoing geopolitical events may be obtained through a longitudinal study that compares language use at various stages of the war.

REFERENCES

1. Biber, D., Reppen, R., & Conrad, S. (2004). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
2. Mikros, G. (2013). *Corpus Manager: A Tool for Multilingual Corpus Analysis.* Cambridge University Press.
3. Fabiszak, M., & Kaszubski, P. (2005). *A Corpus Based Study of War Metaphors*. Pragmatics Today.
4. 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
5. Bennett, G. (2010). Using Corpora in the Language Learning Classroom. https://doi.org/10.3998/mpub.371534
6. Hammond, J., & Macken-Horarik, M. (1999). Critical Literacy: Challenges and Questions for ESL Classrooms.
7. 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
8. Römer, U. (2011). Corpus Research Applications in second Language teaching. *Annual Review of Applied Linguistics*, *31*, 205–225. https://doi.org/10.1017/s0267190511000055
9. van Peer, W. and A. Chesnokova (2019). What Literature Does to Our Emotions, and How Do We Know? Empirical Studies Will Tell. *Synopsis: Text, Context, Media,* 25(1), 1–10.
10. Zyngier, S., A. Chesnokova and V. Viana (eds.) (2007). *Acting & Connecting. Cultural Approaches to Language and Literature*. Munster: LIT Verlag.
11. 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.
12. 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
13. Presidential speeches | Miller Center. (2016). Miller Center. https://millercenter.org/the-presidency/presidential-speeches