A Corpus Assisted Discourse Study of President Volodymyr Zelenskyy's Wartime Addresses⋆

Svitlana Goloshchuk[1,*,†], Justyna Tomczak-Boczko[2,†]

[1] *Bratislava University of Economics and Business, Dolnozemská cesta 1, 852 35 Bratislava, Slovakia*

[2] *University of Szczecin, al. Piastów 40b, 71-064 Szczecin, Poland*

Abstract

This article reports a corpus-based comparative analysis of the English-language transcripts of President Volodymyr Zelenskyy's nightly video addresses during two one-month periods of the Russia–Ukraine war: 24 February–23 March 2022 and 24 February–23 March 2023. Two specialized corpora were compiled from the official presidential website (≈60,823 and ≈28,300 tokens, respectively) and processed in LancsBox. The study combines quantitative procedures—word frequency profiling, keyword analysis using simple maths and Cohen's d, concordance inspection, and collocation analysis (GraphColl)—with qualitative political discourse analysis to identify salient themes. Results indicate similarity in core narratives (national unity, defence, invasion) alongside marked shifts in lexical choice and stance. In 2022, keywords represent immediate defence and crisis (e.g., invaders, invasion, corridors, Mariupol, Melitopol), while in 2023 the discourse moves toward actors and coordination (e.g., Bakhmut, warriors, occupier/brigade, cooperation, justice, security). Collocational profiling of the node war shows 2022 patterns centred on against, after, Ukraine, day, world, whereas 2023 emphasises Russia, full-scale, during, suggesting a stronger attribution of agency and an increasingly international framing. Concordance choice further shows a move from early shock/negation toward more institutional, coalition-oriented messaging. Across both corpora, personalization via the lemma Putin is notably limited, which supports a strategic focus on broader geopolitical framing. The findings show both similarities and differences in President's crisis communication across the first and thirteenth months of the war.

1. Introduction

The Russian military's full-scale invasion of Ukraine in February 2022 led to a significant deterioration in the political rights and civil liberties enjoyed by Ukrainians. Since the beginning of the invasion, President Volodymyr Zelenskyy has been addressing Ukrainians daily to convey his position and call on the people of Ukraine to take decisive action. In addition, foreign governments, their members of parliaments, politicians, and ordinary citizens are also addressed intending to encourage them to actively support Ukraine against Russian aggression.

Zelenskyy addresses the nation in a video format daily late at night. During the day President also reports about the main activities he holds with national or foreign partners on his pages on social media. Although there are different approaches to defining features of political discourse (Filardo-Llamas et all, 2017; Zienkowski, 2018) in the present article we follow Rojo Martin and Teun van Dijk's definition that political discourse is "a special case of political action, [...] a functional or strategic part of the political process" [25]. Addresses delivered by Volodymyr Zelenskyy, in the context of the political situation in Ukraine, refer to a specific genre of political discourse where the president presents a formal speech to the nation as a whole [12]. They serve as a means for the leader to communicate the government's vision, policies, and plans to the public. These speeches are typically carefully crafted and prepared, combining persuasive language, rhetorical devices, and

---

emotional appeals to captivate and engage the audience. They are amplified through media channels, reaching wider audiences through live broadcasts, online streaming, or social media platforms. These platforms allow for greater dissemination and engagement with the address, enabling individuals to share their reactions and discuss the content in real-time.

In light of the full-scale war between Russia and Ukraine, there is a research interest to explore how the addresses are communicated by the President of Ukraine in his daily appeals to the nation, particularly through linguistic means. Specifically, it focuses on the linguistic means of representation of the war in the president's addresses during two periods: 1) 24.02.2022-23.03.2022, and 2) 24.02.2023-23.03.2023 marking the first months of the first and second year of the war conflict. The study utilizes corpus linguistics methods to analyze two specialized corpora compiled by the authors for the outlined objectives. The key research questions were formulated as follows:

RQ1: What are the key lexical features characterizing the addresses delivered by the President in the specified time frame?

RQ2: To what extent do the principal themes revealed in the addresses of both periods align or diverge?

The paper proceeds with the contextualization of the current study in the literature connected with the subject at issue. Subsequently, it describes the data collection procedure and the analysis methodology employed. Finally, the paper concludes by presenting an analysis and interpretation of its principal findings.

2. Literature review

Since the outburst of the war in Ukraine, several works on war discourses have been published. The majority of them analyses the means of war coverage in different media, for instance, Natalia Petiy investigating English-language media searches for cognitive metaphors utilized in depicting the war in Ukraine "The results of the analysis indicate that the war in Ukraine is frequently represented as a game or a battle between good and evil" [21]. Some authors have focused on specific verbal and nonverbal components of political addresses delivered by Volodymyr Zelenskyy [12]. Their findings demonstrate how these aspects communicate his leadership style from both internal and external perspectives during the conflict period.

Other authors have driven the further development of war discourse in Ukraine and the study carried out by Lőrincz [17] applies corpus analysis in Eastern and Western English language news releases to identify their main topics. The paper reveals that lemmas *Putin* and *Putin's* predominantly appear in the subject and attribute positions among the collocates of the term war, while lemma Zelenskyy is more commonly found in the object position. She concludes that the syntactic pattern suggests that lemma Putin is often associated with the agent or instigator of the action [17]. A recent study by Yiğit Salihoğlu and Çiğdem Karatepe similarly examined online news about the Ukraine-Russia war on BBC News and Al Jazeera English and came to some surprising conclusions. They suggest that "the illustration of news and the used language seemingly attempted to diminish the unpleasant aspects for Ukraine and its society" [26]. A more comprehensive description can be found in an analysis of online public sentiments in tweets related to the war. According to the study by Rahat Gulzar et al, it is revealed that most tweets are negative, and the total number of tweets declines over time [15]. Seminal contributions have been made by Baladrón-Pazos et al based on Spanish political tweets and their results show that a political party's communication generally meets the rules of political correctness and moderation [4].

There also have been several studies that investigate the ways how the president of Ukraine V. Zelenskyy and the president of Russia V. Putin are approached by media sources. The study by Raza and Malik utilises BBC as a data source to compare the media representation of Putin and Zelenskyy with the tools of multimodal CDA. They conclude that BBC's profiles are not neutral and "represent Putin as rigid, vastly powerful and a threat, whereas Zelensky is shown as amiable but dauntless in the face of war and political strife" [24]. From a different perspective, the paper by Innocent Chiluwa and Jurate Ruzaite compares the war rhetoric of Putin and Zelensky [8], whereas Gregić and Božić scrutinized crisis management through verbal and non-verbal communication of both presidents [14].

3. Methodology

To conduct this study, two specialized corpora of English language addresses delivered by President Volodymyr Zelenskyy were compiled, covering the period from February 24th, 2022, which marked the beginning of the Russian military assault, to March 23rd, 2022, and from February 24th, 2023 to March 23rd, 2023, thus covering the first months of each war year consequently. The first corpus (referred to as Corpus-22) consisted of about 60 addresses and speeches comprising a total of 60,823 words. The second corpus (referred to as Corpus-23) comprised around 37 files totalling 28,300 words. The referred video addresses, as well as transcripts of these speeches, are available in English on the official website of the President of Ukraine.

The present study covers two periods of one-month length within a year of the full-scale Russian war against Ukraine employing corpus linguistics methods. Corpus research is particularly advantageous for analyzing political discourse on war as, according to Paul Baker, it reduces researcher bias, facilitates the incremental effect of discourse and reveals resistant and changing discourses (2006, 10-15). This approach allows for a more robust and efficient study of linguistic data, enabling researchers to formulate more reliable interpretations and generalizations regarding prevalent discourses within a corpus [17]. McCarty discussing the issue of multi-word units in the scope of lexis emphasises that corpora reveal the regular, patterned preferences in given contexts, and "show how large numbers of users separated in time and space repeatedly orient towards the same language patterns when involved in comparable social activities" [19]. As a research tool, corpus analysis is useful in identifying themes and patterns across texts, and uncovering the underlying meanings prevalent in the text.

To process the target corpora, the Lancsbox software package was utilized. The objective was to conduct a comparative examination of the linguistic features of the president's addresses in each first month of the war. We employ different techniques on the corpora to ensure the quality of our analysis, as the methodological triangulation "facilitates validity checks of hypotheses, anchors findings in more robust interpretations and explanations, and allows the researcher to respond flexibly to unforeseen problems and aspects of the research" [1]. To achieve our task, we chose a combination of quantitative (frequency words analysis, keyword analysis, concordance analysis, collocation analysis) and qualitative measures (political discourse analysis for the identification of common themes) across language corpora. These measures guarantee that the words selected for our analysis are frequently used in a sample of texts and that the common themes identified on their basis with the exploratory, inductive approach to the empirically based study are compiled in a transparent and relevant way.

Frequency word analysis reveals the aboutness of the created corpora and as Paul Baker observes "frequency is one of the most central concepts underpinning the analysis of corpora" and also "one of the most oft-heard misconceptions of corpus linguistics" [3]. It comprises a list of all words in a corpus along with their frequencies and shows the percentage contribution that each word makes toward the corpus. It is a quantitative measure which "guarantees that the words selected are frequently used in a large number of texts and that the wordlist is compiled in a transparent and replicable way" [6]. Therefore, it is a quantitative methodology and may be reductive and generalizing, however, it helps to determine the focus of a text. "Frequency counts" as assures Paul Baker recalling the words of Stubbs: "No terms are neutral. Choice of words expresses an ideological position".

The next tool used is the keyword analysis which is employed to identify salient words in the corpus and the aboutness of the textual data. Baker and Ellece define keyness as "the relative frequency of a particular linguistic item in one text or corpus when compared against another text or corpus via statistical tests of significance (usually chi-squared or log-likelihood)"[3]. Identification of keywords is just the first stage, the next is a qualitative analysis to establish the use of the items in a context.

Collocations, thought of as the "company that a word keeps", show some regularity in the appearance of words. Collocation studies show that a good deal of word combinations are "the

probabilistic outcomes of repeated combinations created and experienced by language users"[18]. The notion of collocation shifts the emphasis from a single word to pairs of words as integrated chunks of meaning. According to Stubbs, collocation analysis helps to "show the associations and connotations they have, and therefore the assumptions which they embody" [2] which implies, that collocations have ideological meaning, and thus, become an integral element of political discourse description and corpus linguistic tool.

The tools mentioned above are combined with the analysis of concordances, necessary to avoid making presuppositions about the ways that words are used within a text. As Baker states, a concordance analysis is one of the most effective techniques which allows researchers to carry out this sort of close examination [1]. According to the definition, a concordance is simply a list of all of the occurrences of a particular search term in a corpus, presented "in its context, it will not show you quite so easily whether the word occurs in little bursts or whether it is spread evenly throughout a text" [19].

The research structure includes the following stages: first, frequency words analysis is carried out and then positive and negative keywords are identified using the Words tool of the Lancsbox software. In particular, simple maths and Cohen's D statistics are employed to identify these keywords. Frequency cut-off points were not applied.

The top 100 keywords were examined. Function words were of no interest to the study as they primarily serve structural purposes rather than carrying substantive content. Consequently, they were disregarded in the analysis. For the analysis of collocates of the node "war", the following criteria were applied: Statistic: 01-Freq | Span: 5-5| | Statistic value threshold: 5.0. By adopting this default statistics approach to the study, reliability was ensured by determining the quantitative and qualitative features in the language of the corpora.

4. Results

Keyword statics (simple maths). To compare the most frequent words in the target corpora, the Words tool of the Lancsbox software was employed. Specifically, the keyword technique served to compare the lexical characteristics of the corpora. As a result, positive keywords, negative keywords, and lockwords were identified (Table 1). The positive keywords indicate the lexical choices that were more prevalent in Corpus-22 compared to Corpus-23. On the other hand, the negative keywords reveal the linguistic preferences of Corpus-23 when contrasted with Corpus-22. Lastly, the lockwords represent words that appeared with similar frequencies in both corpora. Because the keywords technique is a reliable method for uncovering the aboutness of the corpus, it was used to determine the principal discourses in the President's speeches during the first and thirteenth months of the war.

Table 1
Frequency of Special Characters

No.	Keywords+ (February-March 2022)	Lockwords	Keywords – (February-March 2023)
1.	invaders	14th	bakhmut
2.	2022	200	2023
3.	corridors	300	warriors
4.	It's	abandon	occupier
5.	mariupol	accordance	brigade
6.	seen	account	thank
7.	posthumously	across	separate
8.	lieutenant	adopted	justice
9.	melitopol	ambassadors	discussed
10.	moskow	ambulances	fellow
11.	planes	approach	results
12.	peaceful	association	legal
13.	vehicles	atlantic	frontline
14.	humanitarian	basement	year

15.	because	brussels	cooperation
16.	residents	burning	bravery
17.	look	busy	brigades
18.	invasion	captive	formula
19.	israel	cease	staff
20.	colonel	celebrate	helps

The tabulated data reveal that the President's addresses delivered in Corpus-22 primarily focus on the defence strategy of the invaded country, often referencing specific idea notions. Consequently, lemma invaders has the highest frequency, while corridors (combined with humanitarian) demonstrate the depth of the crisis within the country. Notably, lemmas Melitopol and Mariupol are also highly salient while they do not appear in Corpus-23. It is explained by the fact that these two cities were the first ones occupied and battled for during the first month of the war. Contrary, lemma Bakhmut ranks top in Corpus-23 as the situation in that area was disastrous at that period. The keyword analysis of Corpus-22 also includes lemmas posthumously, lieutenant, and colonel as part of the President's awareness of the individual warriors who die in the war. The focus of attention is given more to the internal crisis (lemmas invasion, invaders, residents) rather than the international problem (only lemma Moskow and Israel are used in the first target corpus). While lemma invader occupies the top position in Corpus-22, it ranks fourth (but is translated as occupier) among the keywords in Corpus-23 and immediately follows lemma warriors thus showing the interrelated process of the defence strategy. The keyword analysis of Corpus-22 points to closer attention to the lemmas plane, vehicle as they are regarded as the main instruments in the ongoing events. Additionally, Corpus-23 highlights lemmas warriors, occupier, brigade, fellow, brigades, and staff. This indicates that the target corpus follows the actors of the warfare more closely than its first counterpart. In contrast, the salience of lemmas seen, look, peaceful in Corpus-22 may suggest the President's position to solve the situation from the only perspective of settling peace in the country. The presence of lemmas thank, separate, justice, discussed, results, cooperation, formula, help in Corpus-23 also suggests that the President's policy changes in the direction of peace realisation and he aims at its multidimensional and international support. The presence of lemmas bravery and frontline also suggests that Corpus-23 is primarily concerned with the situation on the battlefield while seeking ways of resolving the conflict through dialogue and cooperation with partners.

The results of the keyword analysis obtained using Cohen's D test are presented in Table 2.

Table 2
Keywords statistics (Cohen's D)

No.	Keywords + (2022)	Keywords − (2023)
1.	2022	2023
2.	they	thank
3.	not	health
4.	war	fellow
5.	because	glory
6.	us	and
7.	invaders	bakhmut
8.	but	warriors
9.	what	terrorist
10.	invasion	helps
11.	know	occupier
12.	if	held
13.	have	course
14.	It's	separate
15.	so	international
16.	people	brigades
17.	mariupol	who
18.	say	cooperation
19.	civilians	security
20.	corridors	good

Keywords statistics (Cohen's D). By applying the Cohen's D statistics, we compared the obtained data and found that Corpus-22 (Table 2, column 1) describes the main agents of the warfare through *lemmas they, us, invaders, people* and *civilians.* Notably, lemma *war* occupies only fourth place in Corpus-22 while in the next one, it is not indicated at all. The attempt to explain the situation is achieved by employing the lexemes with informative meaning (e.g., lemmas *because, but, what, if, so*). The only lexeme denoting particular actions to be done is lemma *corridors* which appear the last one in our list. On the other hand, the keywords from Corpus-23 (Table 2, column 2) primarily focus on achieving the main concepts of peace (e.g., lemmas *health, glory, separate, international, cooperation, security, good*) as well as on the agents of the conflict (e.g., *lemmas fellow, warrior, brigade vs occupier, terrorist*). We may also find some common verbs which demonstrate a particular activity done to change the situation (e.g., lemmas *thank, helps, held*) while in Corpus-22 (Table 2, column 1) there are ones of unspecified meaning (e.g., lemmas *know, have, say*).

We may also find quite a lot of keywords in Corpus-22 which show the determination and clear vision of the current situation (e.g., lemmas *not, but, what, if, so*) as in the following examples: *If we protect Kyiv, we will protect the state;* or *And even if you destroy all our cathedrals and churches, you will not destroy our sincere faith in God, in Ukraine.* The keyword if (98 hits, 16.11 rel.freq.) is used together with the present tense verbs which means that the President is fully aware of the near future: *if someone thinks, if they can accumulate, if everyone leaves, if you try to hide, if anyone still doubts, if the invasion continues.* Lemma *not* has 584 hits (96.02 rel.freq.) in Corpus-22 while only 100 hits (35.34 rel.freq.) in Corpus-23. Some examples of Corpus-22: *Even the border between them was only conditional, only on maps, but not in the soul. Not in the soul; This is not to be changed by missiles; They are not afraid of even that. The lower frequency of the negative sentences in the second portion of the corpus* (February-March 2023) also shows that the whole rhetoric of the President's speeches changes and becomes more impersonal and cooperative. We can find the lemmas *and* (977 hits, 345.23 rel.freq.), *who* (292 hits, 103.18 rel.freq.). The sentences from the corpus to illustrate the use of conjunction and are the following: *The Staff considered the production and supply of ammunition and weapons; Separately and in great detail... This is a very valuable and value-based cooperation.* Lemma *who* is found in various contexts and refers to multiple actors of the discourse: *Glory to each and everyone who is in combat; Thank you to everyone who helps us; I want to address those who are still waiting.*

Concordance. A concordance is a comprehensive index of the words used in a text or a body of text. Ordinarily, it will not only index but also cite all passages in which a given word occurs. To verify the reliability of the inferences derived from the quantitative data, further investigation was conducted by analysing the concordance lines, which provide a comprehensive index of lexical units used in Corpus-22 and Corpus-23.

The analysis was conducted using KWIK technique of the LancsBox software. Studying concordance lines assists in clarifying the instances in which a particular word or phrase appears in context within a text corpus. It also provides valuable insights into the usage, meaning, and collocation of words or phrases within a given corpus formulated based on numeric findings.

Keywords put in context demonstrate a profound explanatory basis and lead the researcher towards objective conclusions. A subset of results received from the KWIK analysis is presented in Table 3.

Table 3
Concordance lines sourced from Corpus-22

Left context	Node	Right context
forget. This attack on Kharkiv is a	war	crime. This is state terrorism of the
City Military Administration for the period of	war.	To guarantee the defence of the city.
Head of the Military Administration. After the	war,	we will return everything to its place

strangers here. This is what the people's	war	is. This is what the people of
"united". The seventh day of this terrible	war	began. A war we feel the same
were used. Take them home. Ukrainians! Every	invader	should know: they will not get anything
will make our cities destroyed by the	invader	better than any city in Russia. Enerhodar.
many other towns and villages, which the	invader	managed to capture, are being held hostage.
of our people who will persecute the	invader	to the end. On the third day,
their land and will not give the	invader	a single piece of land. Not a
victory. Except for the truth. Except for	peace.	Except for the tranquillity we want to
world for the sake of Ukraine and	peace	even more. Neutral Switzerland has supported EU
be destroyed everywhere. They will not have	peace.	They will have no food. They will
live. The war must be stopped and	peace	restored as soon as possible. Today was,
does not know how to live in	peace.	Today I held a meeting with the

The obtained findings align with the results taken from the quantitative analysis. As expected, the President stresses the illegal and terroristic nature of the Russian invasion and fully supports Ukrainian defenders and civilians. He encourages people to protect their homeland and truly believe in justice and the upcoming victory. Furthermore, the enemy soldiers are addressed directly without any rude rhetoric but only with a disapproving tone. The theme of peace is highlighted to be the highest moral and stability value which is consistent with the concept of truth, EU values, and prosperity. Overall, the analysis suggests that Zelenskyy's addresses are characterised by a strong sense of inspiration, strategic communication, and support of Ukrainian citizens, portraying them as heroes fighting for their homeland.

Collocations. As our final approach to the analysis of our corpora, we utilized the GraphColl tool to study the collocates of the lexeme war and provide an explanation of the main themes available in the target corpora. Analysis of collocates identifies the meaning they are put in. The results of the collocation analysis are shown in Figure 1 and Figure 2, which visualize the collocates:

Figure 1: Collocates of lemma war in Corpus-22.

Figure 2: Collocates of lemma war in Corpus-23.
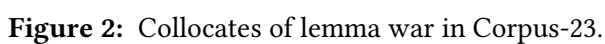
The findings reveal several similarities in the lexical patterns of both corpora. The frequency of the node war comprises 313 hits, and it collocates with 113 lemmas in Corpus-22, while Corpus-23 shows the frequency of 72 hits and collocation with only 27 lemmas.

In Corpus-22, the immediate left-hand collocates of the node war include terms like *this, after, day, world, end, days*. Among the right-hand collocates are lemmas like *we, against, our, Ukraine, they, people*. In Corpus-23, the left-hand collocates of war include terms such as *russia, full-scale, russian, during, first, year, crimes*, and the right-hand collocates contain the lemmas *we, all, Ukraine, who*.

The most statistically significant collocates in Corpus-22 include *against, after, Ukraine, day, world*. To show more statistical values, we include a short description of these lemmas. The node *war* combines with lemma *against* 44 times out of 313 (7.23) in 32 texts out of 60, e.g., *Another night of Russian's full-scale war against us, against the people, has passed.* The lemma *after* shows its presence with the node war 29 times out of 313 in 22 texts out of 60, e.g., *My dears, the time will come when we will be able to sleep. But it will be after the war. After the victory.* The lemma *Ukraine* is used with the war node 26 times out of 313 (4.27) in 20 texts: *Ukraine never wanted this dreadful war. And Ukraine doesn't want it* (Table 4):

Table 4
Collocates of the node "war" in Corpus-22

Position	Collocate	Statistics	Freq (coll.)	Freq (corpus)
L	of	134.0	134	2110
L	this	114.0	114	618
L	to	73.0	73	1904
M	is	68.0	68	925
R	and	62.0	62	1569
R	we	61.0	61	971
R	for	59.0	59	953
R	in	57.0	57	1066
R	against	52.0	52	194
R	will	45.0	45	554
R	not	41.0	41	584
R	that	41.0	41	727
-	war	38.0	38	313
R	our	37.0	37	1008
L	after	34.0	34	63
R	ukraine	27.0	27	551
R	it	26.0	26	531
L	be	25.0	25	339
L	day	25.0	25	133
L	world	24.0	24	191
R	are	23.0	23	719
R	they	22.0	22	498
L	end	20.0	20	35
R	people	20.0	20	500
R	us	20.0	20	279
L	days	17.0	17	67

In Corpus-23, we may find lemmas *russia, full-scale, against, russian, during, Ukraine*. The lemma *russia* occurs 11 times out of 72 (3.89) and in 9 texts out of 37 as in one of the examples: *Russia started a full-scale war against us.* The node *full*-scale occurs 10 times out of 72 (3.53), and in 8 texts out of 37. It is illustrated in the following sentence*: For the fact that during a year of full-scale war, Spain has stood with us in defense against Russian terror.* And the node *war* collocates with the lemma against 10

times out of 72 (3.18) in 9 texts out of 37 ones, e.g., *There is an obvious war of tyranny against freedom* (Table 5):

Table 5
Collocates of the node "war" in Corpus-23

Position	Collocate	Statistics	Freq (coll.)	Freq (corpus)
L	of	37.0	37	1234
L	this	23.0	23	246
R	and	19.0	19	977
R	we	18.0	18	416
L	in	17.0	17	551
L	to	15.0	15	886
L	for	13.0	13	487
L	russia	11.0	11	86
R	all	10.0	10	265
L	full-scale	10.0	10	13
R	is	10.0	10	372
R	against	9.0	9	78
R	our	9.0	9	604
L	russian	9.0	9	169
R	will	9.0	9	261
R	are	7.0	7	264
L	by	7.0	7	128
L	during	7.0	7	12
L	first	7.0	7	40
R	ukraine	6.0	6	345
L	were	6.0	6	68
L	year	6.0	6	48
L	crimes	5.0	5	20
R	that	5.0	5	267
R	who	5.0	5	292
L	of	37.0	37	1234

The presence of the collocates such as *against, Ukraine* indicates a high level of convergence between the target corpora. Both corpora include several collocates revealing negative discourse prosodies, with a more pronounced attitude observed in Corpus-23. Notably, the current war is identified as *terrorist, dreadful, brutal, terrible, shameful, the worst war since World War II* in Corpus-22. The use of adjectives with highly disapproving meanings demonstrates a more determined stance expressed in Corpus-22 towards the issue of war compared to the previous findings obtained through the keyword analysis. Thus, both target corpora include lexical items that point to the brutality of the war, its scale, the tragedy of the people, and the fact that it is illegal and terroristic.

Dissimilarities can be detected in the usage of the lemmas *russia, russian* which are among the immediate collocates of the node war in Corpus-23 whereas they do not appear in Corpus-22 among the first 25 collocates (actually they appear on the 31 and 37 positions). Furthermore, lemmas Russia Russian are left-hand collocates having the subject and attribute position in the sentence structure. This suggests the attacking country is presented as the main doer of the conflict and the one to be responsible for it. Similarly, we do not find particle not in any war collocation in Corpus-23 while in the 2022 one, it is quite frequent (6.25). It might be explained by the conflict abruptness during its first month and rejection and disagreement with the situation in general. Similarly, the lemma end (both as a noun and as a verb) is not found in Corpus-23 while in the first war month, it is quite frequent

(3.29). The president uses it to show the positive war outcome: *The war must end; ... to put pressure on Russia to end this war.* Noteworthy, lemma *Putin* does not feature in any of these corpora.

Common themes. The synthesis of the keyword analysis and collocated data of the node "war" revealed common themes that were emphasized in the target corpora (Table 6). Since many narratives intertwine, the division of linguistic data into distinct thematic groups is not straightforward.

Table 6
Common themes and lexical choices

Themes 2022	Lexical choices	Themes 2023	Lexical choices
National unity	our, we, you, Ukraine, us, your, Ukrainians, Ukrainian, I	National unity	our, we, Ukraine, you thank, Ukrainian, Ukrainians, your, us
Country defence	not, war, people, Mariupol, civilians, corridors, Melitopol, humanitarian, residents, against military (actions), troops	Country defence	Bakhmut, warrior, brigades, frontline, against, defence, war, brigade, fighting, soldiers, sanctions, forces, lives, life, march
Country invasion	they, invaders, invasion, Moskow, Russia, Russian	Country invasion	terrorists, occupier, Russia, Russian, enemy, evil aggression, terror
Defenders	posthumously, lieutenant, colonel	Peace concepts	justice, legal, bravery, security, freedom, grateful, peace, free, together, law, good, power
Military aid	planes, vehicles	International aid	discussed, results, cooperation, formula, helps

The tabulated data reveal both points of convergence and divergence in the war discourses of the contrasted corpora. Despite both corpora having similar discourse prosodies regarding the ongoing military invasion, differences are observed in several aspects. While in Corpus-22 as well as in Corpus-23 the themes of national unity, country defence and invasion are similar, still they include different keywords to approach these topics. A constitutive difference is observed through the shift from emphasising country defenders and military aid to introducing the issues associated with peace concepts and the importance of international support to resolve the conflict.

The tabulated data reveal both points of convergence and divergence in the war discourses of the contrasted corpora. Despite both corpora having similar discourse prosodies regarding the ongoing military invasion, differences are observed in several aspects. While in Corpus-22 as well as in Corpus-23 the themes of national unity, country defence and invasion are similar, still they include different keywords to approach these topics. A constitutive difference is observed through the shift from emphasising country defenders and military aid to introducing the issues associated with peace concepts and the importance of international support to resolve the conflict.

5. Conclusion

This study presents the findings of a corpus-based comparative analysis of lexical characteristics of the political address delivered in a video format by President V. Zelenskyy since the start of the Russian full-scale invasion into the territory of Ukraine. We use a combination of quantitative (frequency words analysis, keyword analysis, concordance analysis, collocation analysis) and qualitative measures (political discourse analysis for the identification of common themes) across

language corpora. These measures guarantee that the words selected for the identification of corresponding themes are frequently used in a chosen number of texts.

Both similarities and differences were detected in the coverage of lexical features of the address within the target corpora, highlighting general trends in political discourse. By applying Simple maths and Cohen's D statistics we found that both corpora attributed to the defence strategy of the invaded country employing for its coverage different lexemes. In Corpus-22 we could find the items *not, war, people, Mariupol, civilians, corridors, Melitopol, humanitarian, residents, against military (actions), troops,* but Corpus-23 is characterised by the concentration of lemmas *Bakhmut, warrior, brigades, frontline, against, defense, war, brigade, fighting, soldiers, sanctions, forces, lives, life, march.* In contrast, the keywords from Corpus-23 primarily focus on achieving the main concepts of peace (e.g., lemmas *health, glory, separate, international, cooperation, security, good*) as well as on the agents of the conflict (e.g., lemmas *fellow, warrior, brigade vs occupier, terrorist*).

Based on the analysis conducted using the KWIC technique within the LancsBox software, it can be concluded that the method provides valuable insights into the lexical patterns and usage within Corpus-22 and Corpus-23. Nevertheless, the contextualized presentation of keywords offers a strong explanatory framework, affirming that President Zelenskyy emphasizes the illegal and terroristic nature of the Russian invasion while offering unwavering support to Ukrainian defenders and civilians. His rhetoric encourages citizens to protect their homeland and maintain a steadfast belief in justice and eventual victory. Notably, Zelenskyy addresses enemy soldiers with a disapproving tone, avoiding offensive language, which underscores a measured and strategic communication style.

In the final stage of the analysis, the GraphColl tool was employed to examine the collocates of the lexeme war and to identify the primary themes present in the target corpora. The use of adjectives with highly disapproving meanings (*terrorist, dreadful, brutal, terrible, shameful, the worst war since World War II*) demonstrates a more determined stance expressed in Corpus-22 towards the issue of war compared to the previous findings obtained through the keyword analysis. Thus, both target corpora include lexical items that point to the brutality of the war, its scale, the tragedy of the people, and the fact that it is illegal and terroristic. However, it also reveals notable dissimilarities in the use of specific lemmas between Corpus-22 and Corpus-23. Interestingly, in Corpus-23, there is a stronger emphasis on attributing responsibility for the conflict to Russia, as reflected by the prominence of the lemmas *russia* and *russian.* This shift suggests a deliberate framing of the aggressor as the primary actor in the narrative. Additionally, the absence of certain terms, such as not and end in Corpus-23, compared to their notable presence in Corpus-22, indicates a change in tone and focus. Early discourse appears to reflect shock, rejection, and a hopeful emphasis on resolving the conflict, while later discourse shifts to a more assertive and accusatory narrative. The exclusion of the lemma Putin from both corpora further suggests a strategic choice to de-personalize the conflict and focus on the broader geopolitical context.

The theme of peace is presented as a paramount moral and stabilizing value, consistent with the principles of truth, European Union values, and prosperity. Overall, the analysis highlights that Zelenskyy's addresses are characterized by inspirational messaging, strategic communication, and an unwavering emphasis on the resilience and heroism of Ukrainian citizens in their struggle to defend their homeland.

To gain deeper insights into the linguistic features of political address across different war periods, it would be valuable to analyse corpora comprising each war year separately with a larger number of data. Furthermore, conducting a comparative analysis of addresses from each period, including the application of corpus linguistic tools and the analysis of political discourse would be highly informative. This would provide a more comprehensive picture of the linguistic choices and discourses employed by the President regarding the studied topic.

Acknowledgements

Declaration on Generative AI

The authors have not employed any Generative AI tools.

6. References

[1] Baker, Paul, and Jesse Edgber. "Introduction." Triangulating Methodological Approaches in Corpus-Linguistic Research, edited by Paul Baker and Jesse Edber. Routledge, 2016.

[2] Baker, Paul, and S. Ellece. Key Terms in Discourse Analysis. Continuum, 2011.

[3] Baker, Paul. Using Corpora in Discourse Analysis. Continuum, 2006.

[4] Baladrón-Pazos, Antonio José, Beatriz Correyero-Ruiz, and Benjamín Manchado-Pérez. "Spanish Political Communication and Hate Speech on Twitter During the Russian Invasion of Ukraine." Politics and Governance 11, no. 6 (2023). https://doi.org/10.17645/pag.v11i2.6328

[5] Braithwaite, Sharon. "'I Need Ammunition, Not a Ride': Zelensky Refuses US Offer to Evacuate," The Journal Times, February 26, 2022. https://journaltimes.com/news/world/i-need-ammunition-not-aride-zelensky-refuses-us-offer-to-evacuate/article_d7d67dca-0547-5a5e-a301-2a80aebd566d.html,

[6] Brezina, Vaclav, and Dana Gablasova. "Is There a Core General Vocabulary? Introducing the New General Service List." Applied Linguistics (2013): 1–23. https://doi:10.1093/applin/amt018

[7] Camargo, Fernández Laura, and Urbán Crespo Miguel. "Retórica, propaganda e identidad en la invasión de Ucrania. El antifascismo como argumento en los discursos de Putin y Zelenski." [Rhetoric, propaganda and identity in the invasion of Ukraine. Anti-fascism as an argument in Putin's and Zelenski's speeches]. Refracción 6 (2022): 283-312.

[8] Chiluwa, Innocent, and Jurate Ruzaite. "Analysing the language of political conflict: a study of war rhetoric of Vladimir Putin and Volodymyr Zelensky." Critical Discourse Studies (2024). https://doi.org/10.1080/17405904.2024.2331186

[9] Clark, Catherine, and Katie McQuade. "The Making of Zelensky's Leadership. The Man or the Crisis?" in Leadership During a Crisis. A Focus on Leadership Development, edited by Christian Harrison. Routledge, 2024.

[10] Demczuk, Andrzej. "Idealized Influence of President Volodymyr Zelensky." Politeja 5, no. 86 (2023): 329-34. https://doi.org/10.12797/Politeja.20.2023.86.15.

[11] Donaj, Łukasz, and Marcin Wochelski. "Ewolucja przemówień Wołodymyra Zełenskiego jako prezydenta Ukrainy przed 24 lutego 2022 roku." [Evolution of Volodymyr Zelensky's Speeches as President of Ukraine before 24 February 2022]. Šowoeuropejskie Studia Polityczne, August (2022): 145-62.

[12] Goloshchuk, Svitlana. "Communicative and Structural Analysis of the Political Address (Based on Volodymyr Zelenskyy's Addresses)." Закарпатські філологічні студії, vol. 1, 2024, pp. 71-75.

[13] Goloshchuk, Svitlana, and Justyna Tomczak-Boczko. "Verbal and Nonverbal Components of Political Addresses by President Volodymyr Zelenskyy (Wartime Period)." Jazyk a politika. Na pomedzi lingvistiky a politologie IX: Zborník príspevkov z 9. ročníka medzinárodnej vedeckej konferencie, edited by Štefančík R., Bratislava, 2024, pp. 75-86.

[14] Gregić, Margareta, and Jelena Božić. "Crisis Management through Verbal and Non-Verbal Communication – Zelensky vs Putin." National Security and the Future 3, no. 24 (2023): 98-130. https://doi.org/ 10.37458/nstf.24.3.5.

[15] Gulzar, Rahat, Sumeer Gul, Manoj Kumar Verma, Mushtaq Ahmad Darzi and Farzana Gulzar, Sheikh Shueb. "Analyzing the Online Public Sentiments Related to Russia-Ukraine War over Twitter". Global Knowledge, Memory and Communication, (2023). https://doi.org/10.1108/GKMC-03-2023-0106

[16] LancsBox. "Lancaster University Corpus Tool Box." Accessed September 12, 2024. http://corpora.lancs.ac.uk/lancsbox.

[17] Lőrincz, Marianna. "A Comparartive Corpus Analysis of the Russian-Ukrainian War Coverage in Eastern and Western English Language News Releases." Shìdnij svìt 3 (2023): 115–30. https://doi.org/10.15407/orientw2023.03.115

[18] McCarthy, Michael. Explorations in Corpus Linguistics. Cambridge University Press, 2006.

[19] McEnery, Tony, and Andrew Wilson. Corpus Linguistics. An Introduction. Edinburgh University Press, 2005.

[20] Olcott, Don. "The Zelensky Files: Leadership Strategies and Practices for University Leaders." Ukrainian Journal of Educational Studies and Information Technology 11, no. 3 (2023): 136 –55. https://doi.org/10.32919/uesit.2023.03.01.

[21] Petiy, Natalia. "Conceptual Metaphors Verbalizing War in Ukraine in Media Discourse." Věda a perspektivy 30 (2023): 203-214. https://doi.org/10.52058/2695-1592-2023-2(21)-203-214

[22] Pfleger, Sabine. "La construcción simbólico-discursiva de Volodymir Zelensky: la actualización del mito de héroe." [Volodymir Zelensky's symbolic-discursive construction: the actualisation of the hero myth]. Lengua y Sociedad. Revista de Lingüística Teórica y Aplicada, 21, no.° 1, (2022): 53-72. https://doi.org/10.15381/lengsoc.v21i1.22584

[23] President of Ukraine. Volodymyr Zelenskyy. "Official Website." Accessed September 8, 2024. https://www.president.gov.ua/en.

[24] Raza, Amber, and Sadia Malik. "The Ideological Media Representation of Putin and Zelensky in BBC Profiles: A Multimodal Critical Discourse Analysis." Pak. Journal of Media Science 3, no. 2 (2022): 148-68.

[25] Rojo, Luisa Martín, and Teun A. van Dijk. "'There was a Problem, and it was Solved!': Legitimating the Expulsion of `Illegal' Migrants in Spanish Parliamentary Discourse." Discourse & Society 8, no 4 (1997): 523-566. https://doi.org/10.1177/0957926597008004005

[26] Salihoğlu, Yiğit, and Karatepe, Çiğdem. "Investigation of Online News about Ukraine-Russia War: A Corpus-Based Critical Discourse Analysis." RumeliDE Journal of Language and Literature Studies, 33 (2023): 1171-86.

[27] Shuster, Simon. "2022 Person of the year. Volodymyr Zelensky." Time, December 7, 2022. https://time.com/person-of-the-year-2022-volodymyr-zelensky/

[28] Yakymchuk, V., & Lopatiuk, N. (2024). Communication strategies and tactics in political speeches of V. Zelenskyy. Germanic Philology. Journal of Yuriy Fedkovych Chernivtsi National University, (835–836), 127–135. https://doi.org/10.31861/gph2022.835-836.127-134