

UDC 81.322'272

## A COMPARATIVE CORPUS ANALYSIS OF THE RUSSIAN-UKRAINIAN WAR COVERAGE IN EASTERN AND WESTERN ENGLISH LANGUAGE NEWS RELEASES

*Marianna Lőrincz*

DSc (Pedagogical Sciences), Professor  
Philology Department

Ferenc Rákóczi II Transcarpathian Hungarian College of Higher Education

6, Kossuth Sq., Berehove, Transcarpathian region, 90202, Ukraine

lorinc.marianna@kmf.org.ua

ORCID: 0000-0002-2206-7113

This study presents the findings of a corpus-based comparative analysis of lexical characteristics in the prevalent discourses of the Russian-Ukrainian war coverage within Eastern and Western news releases. The analysis relied on two specialized corpora comprising articles from Eastern (475506 tokens) and Western (531973 tokens) news outlets compiled to serve the objectives of this study. To examine the textual data, keyword analysis, n-grams analysis, and collocation analysis were utilized.

The obtained findings revealed both similarities and divergences in terms of lexical characteristics and discourses in war coverage. Negative discourse prosodies related to the ongoing war were identified in both corpora, revealing a condemnation of the military intervention through their linguistic choices. However, notable differences were documented in the extent of emphasis on the war proceedings and related issues induced by war. The Eastern news sources displayed a degree of ambivalence in their position towards the conflict, with a tendency to present it as dependent on internal factors within Ukraine. On the other hand, the Western news sources extensively commented on the agents of the war and the war-related agenda, using emotionally-charged language to portray the war and its consequences. Furthermore, even within the context of war, the Eastern news outlets prioritized the themes of economic stability, Covid-19 pandemic, the need for dialogue, and international collaboration. In contrast, its Western counterpart paid more heed to the representation of the military intervention itself. The Eastern news coverage emphasized the region's policy of maintaining neutrality in the conflict, which contrasted with the West's unequivocal position in designating the aggressor side and placing stronger emphasis on the discussion of the military intervention. In sum, critically evaluating the messages encoded through linguistic means disseminated by the media can empower global readership to make more informed judgments in the discourse of war.

**Keywords:** corpus linguistics; discourse; Eastern; English language media; lexical features; Ukraine; war; Western

### *Introduction*

The role of media as a tool for shaping and channeling public opinion has been widely acknowledged [Shevchenko 2022]. It possesses the power to create and reinforce attitudes on socially sensitive issues. However, the media is not autonomous and can be influenced by the community itself, making it vulnerable to the detrimental impact of hidden forces. Consequently, it is frequently exploited to serve the interests of political entities and easily falls prey to propaganda. In the realm of conflict discourse, the media assumes a special status, as the depiction of a nation conveyed through news release can become a determining factor affecting its political relations on the international arena.

---

© 2023 M. Lőrincz; Published by the A. Yu. Krymskyi Institute of Oriental Studies, NAS of Ukraine on behalf of *The World of the Orient*. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>).

The media representation has important implications for a state's socio-political stability and cohesion [Lukin 2019]. Moreover, by serving the objectives of propaganda, the media can be instrumental in instilling distorted views and values, thereby exerting a negative impact on major decisions taken in society and politics [McLuhan 2005]. During times of conflict, news reports are notorious for disseminating narratives that rationalize or conceal acts of violence [Mercier 2005]. Therefore, understanding the prevailing discourses within the media during times of war becomes crucial.

In light of the full-scale war between Russia and Ukraine, there is a pressing need to explore how it is portrayed in news outlets, particularly through linguistic means. This study aims to provide a comparative analysis of the war's coverage in the media sources from Western and Eastern countries. Specifically, it focuses on the linguistic means of representation of the war in news releases from February 24, 2022, marking the commencement of the aggression, until the end of that year. The study utilizes corpus linguistics methods to analyze two specialized corpora compiled by the author for the outlined objectives. The key research questions were formulated as follows:

RQ1: What are the key lexical features characterizing the coverage of the Russian-Ukrainian war in Western and Eastern press releases?

RQ2: To what extent do the principal themes constructed by Western and Eastern news coverage align or diverge?

The paper proceeds with the contextualization of the current study in the pertinent literature. 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.

### *Literature review*

A growing body of literature has attempted to investigate the role of language in constructing diverse representations of reality [Ali and Omar 2016, 132]. Language has been pervasively described as ideology-laden [Lukin 2019], as manipulating language entails influencing public attitudes and responses to events. In fact, ideology is dependent on linguistic means, as they serve as potent vehicles for its expression. One powerful epitomizer of ideology is the media, which political forces employ for ideological purposes. In the literature, there is a general consensus on the critical link between media discourse, the construction of ideologies, and public perceptions [Ali and Omar 2016].

Specifically, media sources have been found to prepare society for the acceptance of military aggression as an inevitable and necessary solution to an impending crisis. Thus, Alzahrani et al. considered the role of media in manipulating public opinion by analyzing the framing shifts within pro-Russian propaganda discourse [Alzahrani et al. 2018]. The researchers concluded that Russian media employed information operations to shape public attitudes and foster acceptance of aggression and violence as an imminent outcome. Through their analysis of framing shifts, they uncovered that the Russian media had been laying the groundwork for the impending conflict since the annexation of Crimea in 2014. They suggest that the current military aggression could have been anticipated based on the examination of dominant narratives in the media.

In a similar vein, Kryzhanivska conducted a corpus analysis of war discourse based on news coverage from Ukrainian, Russian and British media outlets [Kryzhanivska 2022]. The author stressed that by examining the lexical properties of the corpus, it is possible to discern the media source's stance toward the war. She pointed out that Russia had been constructing a justification for this aggression through its media in advance. By portraying the Ukrainian government as fascist, the Russian media was preparing its population for the forthcoming war and achieved significant success in garnering widespread public support. The deliberate use of specific linguistic means served to create a desired ideological

meaning and legitimize the Russian invasion. The study highlighted the crucial role of the media in impacting public perceptions, underscoring the significance of studying of linguistic devices in the discourse surrounding the war.

The study of the role of language in disseminating ideology aligns closely with the perspective of critical discourse analysis, which centers on understanding the contribution and influence of language on social issues. Proponents of critical discourse analysis argue that linguistic and non-linguistic aspects are interconnected, and language itself contributes to social injustice [Blommaert and Bulcaen 2000]. A comparative analysis of war discourse using the framework of critical discourse analysis was attempted by Salihoglu and Karatepe [Salihoglu, Karatepe 2023]. Recognizing the significant role of media in propagating ideologies, the authors examined lexical choices in ongoing war coverage. They utilized the corpora from BBC News and AL Jazeera English to carry out a comparative analysis in linguistic patterning. While no statistically significant difference was detected between the two corpora, the main findings illustrated an overall backup of the Ukrainian side and conjuring the image of Russia as a global menace. True to the principles of critical discourse analysis, which seeks to expose social injustice, the researchers reflected on the impact of this ideology on the fate of individuals in Russia.

The use of linguistic means to represent and legitimize war has also been investigated through the lens of metaphor theory. In an attempt to understand the Russian-Ukrainian war discourse, Ilin and Nihmatova carried out a corpus-based analysis focusing on the “brotherly people” metaphor extensively used by Russian ideologists to justify the ongoing conflict [Ilin, Nihmatova 2023]. The study demonstrated how ideological content was conveyed through the above metaphor and emphasized the influential role of metaphors in molding political and academic discourses. It highlighted how metaphors express nuanced meanings that manipulate public attitudes. These findings underscore the crucial role of language in legitimizing war and emphasize the importance of analyzing the choice of linguistic resources in discourse.

With the aim of understanding how language can be used to manipulate public opinion for political reasons, Shevchenko et al. examined how Russia’s armed aggression against Ukraine was referenced in English-language media [Shevchenko et al. 2022]. By analyzing the semantic volume of lexemes associated with the concept of war, the researchers observed that the extensive use of references like *crisis*, *special operation* and *conflict* distorted the reality of war. The study, thus, underscored the influence of language on shaping public perception and emphasized the importance of unambiguous language use when reporting on warfare.

### **Methodology**

#### *Corpora*

To conduct this study, two specialized corpora of English language news articles were compiled, covering the period from February 24<sup>th</sup>, 2022, which marked the beginning of the Russian military assault, to January 1<sup>st</sup>, 2023. The first corpus (referred to as the Eastern Corpus – EC) consisted of about 3000 article excerpts comprising a total of 475 506 words. These excerpts were sourced from Eastern news outlets, including *Asia News Monitor*, *South China Morning Post*, and *Hindustan Times*. The inclusion of *Asia News Monitor* was based on its provision of news coverage from various Asian countries such as Japan, South Korea, China, and India. *South China Morning Post* reported on current affairs from Hong Kong and China, while *Hindustan Times* served as an English-language daily newspaper based in India.

The second corpus (the Western corpus – WC) comprised around 3000 article excerpts totaling 531 972 words. These excerpts were obtained from Western newspapers, namely *The Guardian*, *The Daily Mail and Mail on Sunday* and *The New York Times*. *The Guardian* reports on news from Britain and around the world. As a broadsheet, it is associated

with in-depth journalism and formal writing style. On the other hand, *The Daily Mail and Mail on Sunday*, classified as a tabloid, also covered news from the UK. Lastly, *The New York Times* is based in the USA and is renowned for high-quality journalism. It is considered to be one of the most prestigious newspapers globally.

The articles were collected with the assistance of a news aggregator service LexisNexis. The principle search term employed to identify relevant articles was *war in Ukraine*, chosen for its clarity and unambiguous reference. Articles that contained the search term either in their title or text were selected and included in the designed corpora.

#### *Procedure*

The present study covers an almost yearlong period of the full-scale Russian war against Ukraine and its representation in both Western and Eastern media, employing corpus linguistics methods. Corpus research is particularly advantageous for analyzing media discourse on war across a vast number of texts and identifying patterns in the language that is laden with ideology [O'Halloran 2010]. This approach allows for more robust and efficient study of linguistic data, enabling researchers to formulate more reliable interpretations and generalizations regarding prevalent discourses within a corpus. As a research tool, corpus analysis is useful in identifying trends and patterns across texts, uncovering the underlying meanings inherent in the corpus [Orpin 2005].

To process the target corpora, the Lancsbox software package was utilized. The objective was to conduct a comparative examination of the linguistic features of war coverage in the Eastern and Western media. To achieve this, keyword analysis, analysis of n-grams, and collocation analysis were employed. Specifically, keyword analysis was selected as it is a suitable method for identifying salient words in the corpus and the aboutness of the textual data [Gabrielatos 2018].

A text's keyword profile is instrumental in discerning its dominant narratives and reflecting prominent socio-cultural meanings [O'Halloran 2010; Rayson 2019]. In addition, since multiword expressions such as n-grams and collocations are key in language description, the n-grams and collocates of the lexeme "war" were analyzed. As argued by Brezina et al., texts are organized into lexical patterns or networks that complement each other. Through the study of collocations, one can gain insight into the relationship between the lexis and the text, as well as the underlying connection between the texts and the speakers' mindset [Brezina et al. 2015, 139]. Therefore, by examining keywords and multiword expressions, the prevailing discourses and discourse prosodies can be highlighted in the target corpora. Discourse prosodies, as described in the literature, refer to attitudinal meanings expressed within the textual data [Bednarek 2009, 131]. Lastly, to assess the reliability of interpreting the obtained quantitative findings, the KWIK tool of Lancsbox was utilized to examine the use of keywords in context. This tool displays all instances of the query term in the form of a concordance [Brezina et al. 2021, 18].

The research procedure was structured as follows: first, positive and negative keywords were identified using the Words tool of the Lancsbox software. In particular, simple maths and Cohen's D statistics were employed to determine 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: 05 - MI3 | Span: 1-1 | Collocation frequency threshold: 10.0 | Statistic value threshold: 10.0 | CPN: 05 - MI3 (10.0) / L1-R1 / C: 10.0-NC: 10.0. These specific criteria were chosen to address the issue of an overpopulated graph. Overall, by adopting this approach to the study, reliability was ensured by determining the quantitative and qualitative features in the language of the corpora. Consequently, this approach helped mitigate the problem of subjectivity that underlies human interpretation.

**Results**

To compute 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 in Table 1 indicate the lexical choices that were more prevalent in the EC compared to the WC. On the other hand, the negative keywords reveal the linguistic preferences of the WC when contrasted with the EC. Lastly, the lockwords represent words that appeared with similar frequencies in both corpora. Because keywords technique is a reliable method for uncovering the aboutness of the corpus, it was used to determine the principle discourses in the media’s coverage of the war.

**Table 1. Keywords statistics (simple maths)**

No.	Keywords + (EC)	Lockwords	Keywords – (WC)
1.	india	blitzkrieg	zelenskyy
2.	beijing	criminalize	putin’s
3.	delhi	dealings	moscow’s
4.	zelenskyy	depth	ukrainian’s
5.	china	distraction	kremlin
6.	covid-19	doomed	china’s
7.	ongoing	dreadful	retreat
8.	taiwan	employee	hospital
9.	dialogue	exceed	russians
10.	cooperation	excess	soldiers
11.	war-torn	fascism	victory
12.	asia	fiercely	chilling
13.	trade	frightened	family
14.	bilateral	genuine	horrors
15.	moscow	gloom	biden’s
16.	relations	hamper	brexit
17.	abstain	immigration	battle
18.	resolution	inaction	donbass
19.	war-hit	intimidation	kill
20.	condemn	justifying	occupy

The tabulated data reveal that the Eastern news coverage of the war primarily focuses on the internal affairs of the respective countries, often referencing specific place names. Consequently, lemmas such as *India*, *Beijing*, *Delhi*, *China*, *Taiwan*, and *Asia* have the highest frequencies. Notably, lemmas *Zelenskyy* and *Covid-19* are also highly salient. While lemma *Zelenskyy* ranks fourth among the keywords in the EC, it occupies the top position in the WC. Lemma *Covid-19*, on the other hand, does not appear among the keywords in the WC. The keyword analysis of the WC points to closer attention to lemmas *Putin*, *Moscow*, *Kremlin*, *China*, and *Russians* as they possess the highest level of salience. Hence, they are posed as the principal actors in the ongoing events. Additionally, the WC highlights lemmas *retreat*, *hospital*, *soldiers*, *victory*, *family*, *horrors*, *battle*, *kill* and *occupy*. This indicates that the Western news coverage follows on the details and developments of the warfare more closely than its Eastern counterpart. In contrast, the salience of lemmas *dialogue*, *cooperation*, *trade*, *bilateral*, *relations*, *abstain*, *condemn* in the EC may suggest a reluctance to take a firm stance in the conflict and a promotion of the idea of dialogue and cooperation. The presence of lemmas *trade* and *relations* also suggests that the Eastern side is primarily concerned with maintaining economic and political stability in the region, while condemning a military resolution to the conflict.

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 + (EC)	Keywords – (WC)
1.	ukraine	*ukraine
2.	country	*war
3.	russia	*winning
4.	president	dangerous
5.	china	defend
6.	us	defy
7.	global	demilitarise
8.	military	denazify
9.	crisis	depend
10.	india	deplorable
11.	food	desperate
12.	conflict	despicable
13.	moscow	destroy
14.	sanction	disappointed
15.	security	disgusting
16.	supply	dislocation
17.	economic	dismayed
18.	economy	distorted
19.	nato	drag
20.	biden	madness

The comparison of the data obtained reveals that the Eastern news release (Table 2, column 2) primarily focuses on the main agents of the events (e.g., lemmas *Ukraine, country, Russia, president, China, US, India, Moscow, NATO, Biden*) as well as on the economic affairs (e.g., lemmas *food, sanction, supply, economic, economy*). On the other hand, the keywords from the WC are mostly descriptors of the warfare itself (e.g., lemmas *war, winning, defend, defy, demilitarise, depend, destroy, drag*) and the resulting dire situation, as indicated by lexemes with negative connotations (e.g. lemmas *dangerous, deplorable, desperate, despicable, disappointed, disgusting, madness*).

In order to verify the reliability of the inferences derived from the quantitative data, further evidence was sought by analyzing concordance lines, which provide contextualized use of lexical units. The analysis was conducted using the KWIK technique of the LanCSBox software. Examining the usage of keywords in context assists in validating initial hypotheses formulated on the basis of numeric findings. It is noteworthy, that decontextualized keywords may not be self-explanatory and could potentially lead the researcher to draw incorrect conclusions. Table 3 presents a subset of results obtained from the KWIK analysis.

**Table 3. Concordance lines sourced from EC**

Left context	Node	Right context
a negotiated end to their problems through	dialogue	On February 24, President Vladimir Putin stated
...for a political solution and diplomatic	dialogue	to resolve the crisis between Moscow and
commission's "attempts to engage in a constructive	dialogue	with Russian Federation...
and the best way forward was through	dialogue	and diplomacy. Speaking about the Ukraine war,

and called on Moscow to engage in	dialogue	to peacefully end the crisis. Nevertheless, Mr.
Ukraine to return to the path of	dialogue	and diplomacy. India has also sent humanitarian
of differences and disputes between countries through	dialogue	and consultation”, it made no explicit reference
wished the war could be ended through	dialogue	and diplomatic means as soon as possible.
Kremlin, the two leaders discussed “increasing economic	cooperation,	trade and military technical ties between China and
used in the war against Ukraine, and	cooperation	between Iran and Russia will continue without
...economic	cooperation	and regional issues, including a July deal
war in Ukraine. When the world needs	cooperation	and collaboration, rivalry and confrontation have increasingly
Beijing is increasingly worried about the war...	cooperation	in trade, agriculture, connectivity and other areas”.
is crucial for us to maintain close	cooperation	with our most important allies on issues
believes India was right to	abstain	from a United Nations General Assembly resolution
one of more than 30 countries to	abstain	from voting on UN resolutions condemning Russia
increasingly understood that African states preferred to	abstain	in votes related to Ukraine – in...
China among a handful of nations to	abstain	from voting on UN resolutions condemning Moscow’s
remained neutral and signalled its intention to	abstain	from Security Council votes on going to
Wang Yi has defended Beijing’s decision to	abstain	in a United Nations vote on the
condemning Russian actions in Ukraine, chosen to	abstain	on votes at the UN Security Council.
The world has now seen India	abstain	on United Nations(UN) Security Council, General

The obtained qualitative findings were consistent with the inferences drawn from the quantitative measurements. As expected, the EC emphasized the significance of dialogue and diplomacy in resolving the ongoing confrontation. Various Asian countries and leaders were mentioned advocating resolutions through dialogue, favoring diplomatic solutions over military actions. The themes of cooperation and collaboration, particularly in terms of economy, trade, and regional issues, were highlighted. Furthermore, a number of nations were identified for abstaining from voting on resolutions condemning Russia’s aggression in Ukraine. This indicates their stance of non-involvement. Overall, the analysis suggests a prevailing sentiment towards seeking peaceful resolutions and a reluctance to explicitly align themselves with either side in the conflict.

To conduct a more detailed analysis of the lexical features in the coverage of warfare, the N-gram tool of the Lancsbox was employed. As suggested by Brezina et al., this tool enables an in-depth examination of n-gram frequencies (i.e., multiword expressions) and produces a list of positive and negative n-grams by comparing two corpora. This procedure is valuable for describing linguistic characteristics, as n-grams reveal contiguous lexico-grammatical patterns [Brezina et al. 2012, 35–36].

Table 4. Key n-grams

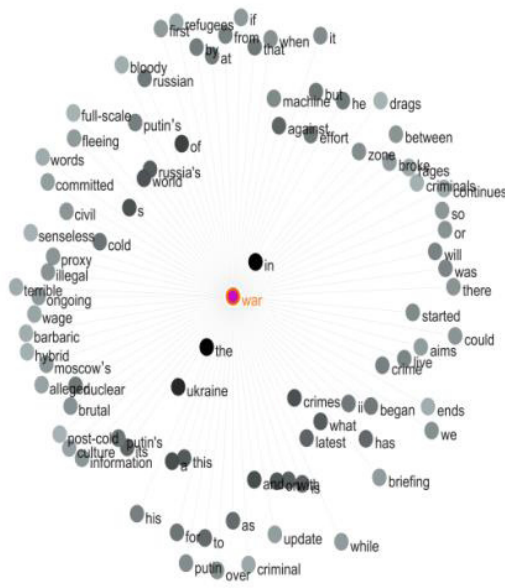
Key n-grams + (EC)	Key n-grams – (WC)
ukraine war	military aid
new delhi	russian attacks
russia's invasion	the alliance
ukraine crisis	an existential
ukraine conflict	nuclear annihilation
volodymyr zelenskyy	red army
the russia	russian men
the covid-19	strikes in
united nations	putin's invasion
russian president	helping to
russia's war	occupied areas
ongoing war	destruction of
war-torn ukraine	donetsk region
to condemn	the killing
indian students	symbol of
business news	crimes in
invasion of	of attrition
stranded in	mass grave
dialogue and	biden putin
war-hit ukraine	killed thousands
food security	prisoner swap

Based on the data presented in Table 4, both corpora contain lexical items that reflect condemnation of military aggression. In the EC, these include n-grams such as *Russia's invasion*, *Russia's war*, *to condemn*, *invasion of*, *war-torn Ukraine*, *war-hit Ukraine*. In the WC, the corresponding n-grams are *Russian attacks*, *nuclear annihilation*, *Putin's invasion*. Additionally, several n-grams highlight the consequences of war, such as *an existential*, *occupied areas*, *destruction of*, *the killing*, *crimes in*, *of attrition*, *mass grave*, *killed thousands*. It is noteworthy that lexeme *kill* appears twice in the list of most frequent n-grams. Based on the lexical data provided, it can be observed that the theme of war proceedings and its consequences is more prominent lexically in the Western news coverage compared to its Eastern counterpart.

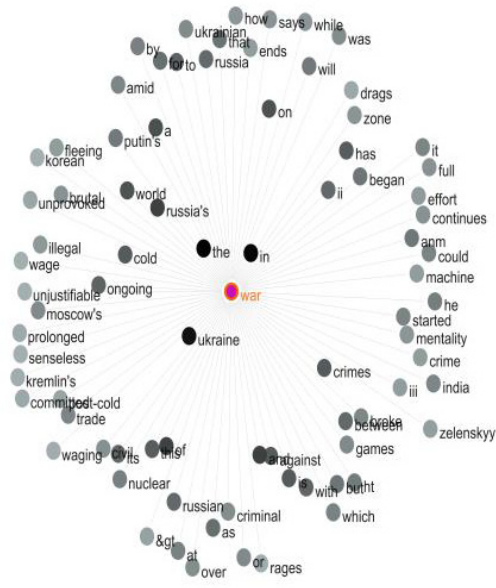
In contrast, the Eastern news release depicts the war using n-grams such as *Ukraine war*, *Ukraine crisis*, *Ukraine conflict*, and *Russia's war*. The expression of attitudes towards the war is somewhat ambiguous in the EC, often portraying the war as Ukraine-induced conflict or crisis. Interestingly, n-grams such as *Volodymyr Zelenskyy*, *Russian President* are salient in the EC, while in the WC, n-grams like *Putin's invasion* and *Biden Putin* appear on the list. Once again, the high salience of n-grams like *business news*, *dialogue*, and *food security* in the EC highlights the importance placed on economic stability and cooperation in the Eastern news coverage. These themes, however, are not prominently represented by the n-grams in the WC, which could indicate that they are of lesser concern in the news reports on the war.

Finally, the GraphColl tool was utilized to examine the collocates of the lexeme *war* and provide a summary of the main themes present in the target corpora. As noted by Kutter and Kantner, analyzing collocates may signal the attitudes conveyed by these collocates [Kutter, Kantner 2012]. The outcomes of the collocation analysis are depicted in Graphs 1 and 2, which visualize the collocates. Additionally, Tables 5 and 6 present statistical values that elucidate collocation patterns.





Graph 1. Collocates of lemma war in WC



Graph 2. Collocates of lemma war in EC

Table 5. Collocates of the node “war” in the WC

Position	Collocate	Statistical value	Frequency (collocate)	Frequency (corpus)
L	ukraine	23.102177785897524	1335	12671
R	crimes	22.0494867813496	320	362
L	world	19.37008032441109	297	1854
L	russia’s	19.23159703187148	208	701
L	cold	18.452899126711827	94	111
R	effort	17.31769603244759	81	156
R	crime	16.646515191220473	59	96
L	putin’s	16.396893845624206	93	447
L	nuclear	15.349660467699778	80	588
L	civil	14.68368443202131	28	40
R	started	14.673037740612518	42	136
L	proxy	14.42957171003764	29	53
L	brutal	14.216580423796588	34	99
L	illegal	14.124209536753465	31	80
L	ongoing	13.345192372778138	25	72
L	moscow’s	13.32464912833821	27	92
L	russian	13.046258163779056	79	2795
L	wage	12.883108042594772	16	26
L	fleeing	12.407698541562178	24	122
L	alleged	12.392883046921712	18	52
R	rages	12.361756337133954	14	25
R	criminals	12.261402898506658	11	13
L	post-cold	12.227403950510583	10	10
L	committed	12.116644591092838	100	21
L	senseless	11.964369544676783	10	12
L	barbaric	11.946117840120564	12	21
L	hybrid	11.791917615205444	11	18
R	drags	11.227403950510613	10	20

L	terrible	11.193013818829455	13	45
R	criminal	11.134295590279523	15	72
L	bloody	11.090507749455721	12	38
L	full-scale	11.08990042676063	10	22
L	putin	10.823505287942458	38	1452
L	refugees	10.051521095429964	17	222

Table 6. Collocations of the node “war” in the EC

Position	Collocate	Statistical value	Frequency (collocate)	Frequency (corpus)
L	ukraine	25.19300680326859	2053	11725
L	russia’s	21.37055249301071	426	1482
L	cold	20.473213419580603	177	198
R	crimes	20.402130154822906	177	208
L	ongoing	18.533727787372943	132	315
R	games	15.50908883248764	36	52
L	putin’s	15.46166824345105	63	288
R	started	15.08936343007364	44	127
R	criminal	14.841772640439554	32	58
L	russian	14.701963182838748	107	2389
R	mentality	14.596658982154109	24	29
L	civil	14.545775881748149	29	53
L	nuclear	13.90000286532326	50	425
L	brutal	13.876952754713475	27	68
R	russia	13.848986994484017	103	3849
L	trade	13.577497697633861	42	315
L	post-cold	13.51353335688237	15	15
L	moscow’s	13.29969973319298	35	221
L	illegal	13.220067496351449	22	58
R	ends	13.1581338677703	19	39
R	crime	13.121607732875747	19	40
R	drags	12.713611382237081	13	17
R	continues	12.670606845621347	28	175
L	unjustifiable	12.206108095014816	10	11
L	fleeing	11.967802132110036	19	89
L	wage	11.758645854718903	10	15
L	unprovoked	11.567229130325332	14	47
L	prolonged	11.547751315755184	12	30
R	ukrainian	11.510930795502542	34	700
L	senseless	11.343609477207073	10	20
L	waging	11.171158669503617	11	30
R	rages	11.123852592864562	11	31
L	kremlin’s	10.685730239808064	11	42
L	committed	10.553144574169703	13	76
R	zelenskyy	10.447977198889589	18	217

The findings reveal several similarities in the lexical patterns of both corpora. The most statistically significant collocates in the WC include *Ukraine*, *crimes*, *world*, *Russia’s*, *cold*, while in the EC, they include *Ukraine*, *Russia’s*, *cold*, *crimes*, *ongoing*. The presence of collocates such as *Ukraine*, *crimes*, *Russia’s* and *cold* indicates a high level of convergence between the target corpora. Both corpora include a number of collocates revealing negative discourse prosodies, with a more pronounced attitude observed in the WC.

In the WC, the immediate left-hand collocates of the node *war* include terms like *cold*, *Putin's*, *nuclear*, *proxy*, *brutal*, *illegal*, *wage*, *fleeing*, *senseless*, *barbaric*, *hybrid*, *terrible*, *refugees*, and *post-cold*. Among the right-hand collocates are lemmas like *crimes*, *rages*, *criminals*, *drags*. In the EC, the left-hand collocates of *war* include terms such as *cold*, *ongoing*, *Putin's*, *Russian*, *civil*, *nuclear*, *brutal*, *trade*, *post-cold*, *Moscow's*, *illegal*, *unjustifiable*, *unprovoked*, *prolonged*, and *senseless*. Notably, the current war is identified as *Russia's*, *Putin's*, *Moscow's* and *Kremlin's* in the EC. The use of possessive case demonstrates a more determined stance taken by the Eastern region towards the issue of war, compared to the previous findings obtained through keyword analysis. Thus, both target corpora include lexical items that point to the condemnation of war, its illegality, crimes committed, and the fact that it was unprovoked and unjustifiable.

Dissimilarities can be detected in the usage of the lemma *Putin*, which is among the immediate collocates of the node *war* in the WC, whereas the lemma *Zelenskyy* does not feature in this collocation network. In contrast, both *Putin* and *Zelenskyy* are collocates of war in the EC, with *Putin* being closer to the node, indicating a stronger association between the node *war* and its collocate *Putin* in the EC. Furthermore, lemma *Putin's* is a left-hand collocate, occupying the subject and attribute position in the sentence structure, whereas lemma *Zelenskyy* appears as a right-hand collocate, occupying the object position in the sentence structure. This suggests that the Ukrainian president is presented not as the main doer of the action as opposed to lemma *Putin*, but as a recipient or the one being acted upon. Similar to the previous findings of this study, the word *trade* is a collocate of the node *war* in the EC, indicating a particular concern for economic issues that is absent in the list of collocates of *war* in the WC.

The synthesis of the collected data revealed common themes that were emphasized in the target corpora (Table 7). Due to the fact that many narratives intertwine, the division of linguistic data into distinct thematic groups is not straightforward.

**Table 7. Common themes and lexical choices in the Target Corpora**

Themes EC	Lexical choices	Themes WC	Lexical choices
Concerned parties	India, Beijing, Delhi, China, Taiwan, Ukraine, Russia, Zelenskyy, Putin	Agents	Zelenskyy, Putin, Moscow, Kremlin, China
Stance	Russia's invasion, Ukraine crisis, Ukraine conflict, Russia's war, abstain	Warfare	retreat, soldiers, victory, battle, kill, occupy, military aid, nuclear, prisoner swap
Condemnation/ Attitudes towards war	war-torn Ukraine, condemn, war-hit Ukraine, crimes, illegal, unjustifiable, unprovoked, senseless	Condemnation	Russian attacks, Putin's invasion, crimes
Economic stability	trade, food, sanction, security, supply, economic, economy	Descriptors of war and its consequences	hospital, chilling, horrors, brutal, illegal, ongoing, fleeing, senseless, barbaric, terrible, bloody, destruction, attrition, mass grave
Cooperation and dialogue	dialogue, cooperation, bilateral, relations	Victims	family, refugees, killed thousands
Covid-19	Covid-19	–	–

The tabulated data reveal both points of convergence and divergence in the war discourses of the contrasted corpora. Despite both corpora exhibiting negative discourses regarding the ongoing military intervention, dissimilarities were observed in several aspects. While in the EC, the themes of economic stability, cooperation and dialogue, and

Covid-19 pandemic were profusely discussed, the WC placed greater emphasis on issues associated with war itself, including warfare, detailed description of war and its consequences, and the plight of its victims.

### *Discussion*

To comparatively analyze the lexical patterning of war coverage in the Eastern and Western news releases, various techniques such as keyword analysis, n-gram analysis, and collocation analysis were applied. The findings revealed both similarities and some points of divergence in terms of lexical choices and prevalent discourses.

One major area of convergence was the unanimous condemnation of the military aggression. Both target corpora explicitly identified the aggression as Russian and described it as an invasion. However, there were notable differences between the two corpora regarding the framing of the conflict. In the EC, n-grams such as *Ukraine crisis* and *Ukraine conflict* were frequently mentioned, while these were absent in the WC. This suggests that certain media sources in the EC were more reluctant to directly attribute the cause of the war to Russia, instead presenting it as a conflict with internal factors within Ukraine. By doing so, they place an emphasis on the internal nature of the affairs, not immediately connected to Russia's aggression. This echoes the findings by Kryzhanivska, who noted indirect references used by some media sources [Kryzhanivska 2022]. The researcher noted that by employing terms like *civil war* to describe the military intervention, media sources inadvertently or intentionally reflected the Russian rhetoric of portraying the conflict as an internal matter. Also, as suggested by Lukin, avoiding or replacing of the term "war" can be a deliberate strategy to divert public attention from the harshness and destruction it entails. Linguistic dissociation is attained through the use of more objective and unemotional lexis. In this way, the public remains oblivious to the realities of war [Lukin 2019].

Furthermore, both news outlets exhibited negative discourse prosodies when referring to the ongoing war, utilizing lexemes with negative evaluative connotations. However, the WC demonstrated a richer lexical description of war, its consequences, and the destitute conditions of the affected side compared to the EC. The principal divergence between the two corpora was observed in the prevailing themes of the conflict discourse. The EC focused on the internal affairs of the respective countries even within the context of war, as evidenced by the choice of proper names such as *India*, *Beijing*, *Delhi*, and *China*. On the other hand, the WC more frequently featured lexemes like *Putin*, *Moscow*, *Kremlin*, *Russia*, and *China*, among others. This finding replicates the results of a study conducted by Lőrincz, which observed the abundant use of lexemes with negative evaluative connotations in describing the attitude towards war [Lőrincz 2022]. It also resonates with the findings of Kryzhanivska, who identified significant differences in the lexical preferences of Russian, Ukrainian, and English media sources [Kryzhanivska 2022]. Specifically, the Russian source employed neutral language to present its official stance on war issues, strategically framing the current war as unavoidable and justifiable. The Western news coverage exhibited a higher level of focus on the warfare itself, as evidenced by the greater lexical density observed in the keyword and n-gram analyses. In contrast, the EC frequently addressed the Covid-19 pandemic, economic stability, business and trade, as well as food supply. These themes were significantly less prominent in the WC and were not included in the analyzed data.

The overall data analysis indicates that the EC emphasized the themes of dialogue, cooperation, and a certain reluctance among some media outlets to unequivocally mention Russia as the victimizer and Ukraine as the victimized. This indicates a certain degree of detachment from the ongoing military intervention and a prioritization of the region's internal affairs and interests. This finding is partially consistent with previous research that depicted China as actively seeking peace and dialogue with international community. The

news reports often portrayed the Chinese government as attempting to maintain a neutral role, while the consequences of the war were seldom discussed [Lin 2023]. In contrast, the US news release in the above research exhibited a preoccupation with economic issues, which contradicts the findings of this study.

Furthermore, the analysis of the data reveals that the target corpora consistently portray Russia and its leader as the aggressor, employing the possessive case to emphasize their role (e.g., *Putin's invasion*, *Putin's war*, *Russia's invasion*, *Russia's war*). Interestingly, 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. This syntactic pattern suggests that lemma *Putin* is often associated with the agent or instigator of the action. This finding aligns with previous research that depicted Russia as the main cause of the war and a perceived threat to the region [Salihoglu and Karatepe 2023]. It also corresponds to the portrayal of Western support for Ukraine and expressions of sympathy toward its citizens [Kryzhanivska 2022].

The comparative analysis of dominant discourses in the Eastern and Western news outlets provides a synthesis of the findings obtained through various research tools employed in this study. The examination of themes and their lexical patterns reveals some shared tendencies in the news coverage. On one hand, there is a consensus in portraying Russia as the victimizer and Ukraine as the victimized. However, a certain level of ambiguity is observed in the Eastern news coverage, which refers to the ongoing war using n-grams such as *Ukraine conflict* and *Ukraine crisis*.

Regarding the parties involved in the conflict, the Eastern news release tends to focus more on regional issues, while the Western news release places a stronger emphasis on the immediate participants of the military intervention. Both news releases express condemnation of the war, but the Western news coverage specifically delves into the procedures associated with the war and extensively utilizes lexemes with negative evaluative connotations to reflect the region's attitude towards the unfolding events.

Lastly, the Eastern news coverage frequently addressed the themes of economic stability, international collaboration, and the impact of the Covid-19 pandemic, which were not prominently featured in its Western counterpart. This resonates with previous research conducted by Sun, who found that Chinese media “concentrate more on the conflict's political and economic consequences on the international order” rather than draw attention to the war per se. The country's neutrality was explained by the reluctance “to damage the important strategic partnership with Russia” [Sun 2022, 55].

Overall, it is important to acknowledge that the presence of disparities in news coverage was anticipated, given the artificial division between the Eastern and Western worlds, which allows for a diverse range of voices and perspectives to emerge.

### ***Limitations***

It is important to acknowledge that the comparison between the Eastern and Western corpora was based on English language media sources, which may have different focuses and therefore influence linguistic choices. Western newspapers primarily target a domestic audience, shaping their political agenda based on internal priorities. On the other hand, English language news releases in the Eastern corpus are more geared towards an international readership. Consequently, the agenda reflected in English articles may differ from that of media published in the respective languages of the Eastern region.

This study does not claim generalizability of its findings due to the size and scope of the target corpora. Furthermore, it is essential to recognize that researcher bias may have influenced the interpretation of the findings. Although care was taken to mitigate it, it is nearly inevitable in discourse analysis. However, the use of corpus linguistics methodology in exploring lexical preferences helped to address the potential impact of personal

bias and minimize subjectivity. Despite these limitations, the study yielded many interesting insights into the lexical choices and prevalent discourses in the analyzed textual data.

### Conclusions

This study has undertaken to examine comparatively the lexical features present in the prevalent war discourses in the Western and Eastern media. By examining the similarities and differences in news releases and the conveyed attitudes, the study offers a better understanding of the underlying ideologies that can potentially shape the outcomes of events. Language has been recognized to carry ideological implications, thus analyzing the language used to depict the ongoing war sheds light on hidden meanings and strategies employed to sway public perceptions. To serve the objectives of this study, two specialized corpora consisting of newspaper articles from prominent sources were compiled and analyzed. The study was based on the premise that lexical choices could signal the two regions' underlying ideologies with respect to the issue of war disseminated and instilled through news release.

Both similarities and differences were detected in the coverage of the ongoing war within the target corpora, highlighting general trends in the perceptions of recent events in the Eastern and Western regions. Thus, all media sources were unanimous in condemning the military aggression, identifying the aggressor side and expressing sympathy towards Ukraine as the victimized party. However, there were notable differences in the extent of focus on the warfare, and on the internal affairs between the EC and the WC. Both corpora attributed the main responsibility for the war to Russia, demonstrating a shared perspective. Nevertheless, the Eastern news release occasionally exhibited ambivalence in their attitudes by using n-grams such as *Ukraine crisis* and *Ukraine conflict* to refer to the ongoing war. In contrast, the WC provided extensive coverage of the aggressors and the proceedings associated with the war. The description of the war was particularly detailed and expressive in the WC, utilizing emotionally-charged language, while the EC demonstrated a more reserved approach. Interestingly, the EC displayed a greater emphasis on economic stability and the impact of Covid-19, even within the context of the war. Additionally, the EC highlighted the importance of cooperation and dialogue in the face of the conflict, expressing a reluctance to become directly involved (e.g., with the lemma *abstain*).

To gain deeper insights into the linguistic features of prominent discourses across different news outlets, it would be valuable to analyze corpora comprising more recent reports with a broader geographical coverage. Furthermore, conducting a comparative analysis of reports from diverse news outlets, including independent web-based resources, would be highly informative. This would provide a more comprehensive picture of the linguistic choices and discourses employed by media regarding the studied topic.

As a final note, we must never disregard words, since they are propaganda's weapons always preceding wars, much like guns and soldiers are of military forces. In today's era of information technologies, words spread instantaneously reaching millions in a matter of seconds. This necessitates critical analysis of the ideas encoded through linguistic means. By engaging in such analysis, we can gain a deeper understanding of the ideologies, motivations and perspectives that can potentially influence the course and outcomes of events.

### REFERENCES

- Ali M. A. and Omar A. (2016), "Discourse and Manipulation in the Representation of the Russian Military Intervention in the Syrian Civil War", *International Journal of Linguistics*, Vol. 8, No. 3, pp. 129–140. DOI: <https://doi.org/10.5296/ijl.v8i3.9498>
- Alzahrani S. et al. (2018), "Framing shifts of the Ukraine conflict in pro-Russian news media", in H. Bisgin, R. Thomson, A. Hyder, C. Dancy (eds), *Social, Cultural, and Behavioral Modeling*,

11<sup>th</sup> International Conference, SBP-BRiMS 2018, Proceedings, pp. 303–314. (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 10899 LNCS), Springer Verlag. DOI: [https://doi.org/10.1007/978-3-319-93372-6\\_34](https://doi.org/10.1007/978-3-319-93372-6_34)

Bednarek M. (2008), “Semantic preference and semantic prosody re-examined”, *Corpus Linguistics and Linguistic Theory*, Vol. 4, No. 2, pp. 119–139. DOI: <https://doi.org/10.1515/CLLT.2008.006>

Blommaert J. and Bulcaen C. (2000), “Critical discourse analysis”, *Annual Review of Anthropology*, Vol. 29, pp. 447–466. DOI: <https://doi.org/10.1146/annurev.anthro.29.1.447>

Brezina V., McEnery T. and Wattam S. (2015), “Collocations in context: a new perspective on collocation networks”, *International Journal of Corpus Linguistics*, Vol. 20, Issue 2, pp. 139–173. DOI: <https://doi.org/10.1075/ijcl.20.2.01bre>

Brezina V., Weill-Tessier P. and McEnery A. (2021), #LancsBox v. 6.x. [software package].

Gabrielatos C. (2018), “Keyness analysis: Nature, metrics and techniques”, in C. Taylor & A. Marchi (eds), *Corpus Approaches to Discourse: A Critical Review*, Routledge, pp. 225–258.

Ilin I. and Nihatova O. (2023), “The ‘Brotherly People’ Metaphor and the Russian-Ukrainian Irredentist War: A Corpus-Based Study”, *Czech Journal of International Relations*, Vol. 58, No. 2, pp. 7–41. DOI: <https://doi.org/10.32422/mv-cjir.504>

Kryzhanivska A. (2022), “The Beginnings: Russian-Ukrainian War in European, Ukrainian, and Russian Media in 2014–15”, *Balkanica et Slavica*, Vol. 2, No. 1, pp. 9–26. DOI: <http://doi.org/10.30687/BES/2785-3187/2022/01/001>

Kutter A. and Kantner C. (2012), “Corpus-Based Content Analysis: A Method for Investigating News Coverage on War and Intervention”, *International Relations Online Working Paper*, 2012/01, Stuttgart University, Stuttgart, available at: [https://www.sowi.uni-stuttgart.de/dokumente/forschung/irowp/IROWP\\_Series\\_2012\\_1\\_Kutter\\_Kantner\\_Corpus-Based\\_Content\\_Analysis.pdf](https://www.sowi.uni-stuttgart.de/dokumente/forschung/irowp/IROWP_Series_2012_1_Kutter_Kantner_Corpus-Based_Content_Analysis.pdf) (accessed April 10, 2023).

Lin J. (2022), “Russian-Ukrainian War: A comparative analysis of framing in Los Angeles Times and China Daily”, in *Radbound University*, June 13, available at: <https://theses.uibn.ru.nl/handle/123456789/12838?locale-attribute=en> (accessed April 10, 2023).

Lőrincz M. (2022), “A corpus analysis of the representation of war in Ukraine in the international press”, *The Journal of V. N. Karazin Kharkiv National University. Series: Foreign Philology. Methods of Foreign Language Teaching*, Issue 96, pp. 61–68.

Lukin A. (2019), *War and its Ideologies: A Social-Semiotic Theory and Description*, Springer, Singapore. DOI: <https://doi.org/10.1007/978-981-13-0996-0>

McLuhan M. (2005), *The medium is the message*, Gingko Press.

Mercier A. (2005), “War and media: Constancy and convulsion”, *International Review of the Red Cross*, Vol. 87, pp. 649–659.

Orpin D. (2005), “Corpus Linguistics and CDA: examining the ideology of sleaze”, *International Journal of Corpus Linguistics*, Vol. 10, Issue 1, pp. 37–61. DOI: <https://doi.org/10.1075/ijcl.10.1.03orp>

O’Halloran K. (2010), “How to use corpus linguistics in the study of media discourse”, in O’Keeffe A. and M. McCarthy (eds), *The Routledge Handbook of Corpus Linguistics*, Routledge Handbooks in Applied Linguistics, Routledge, Abingdon, pp. 563–576.

Rayson P. (2019), “Corpus analysis of key words”, in Chappelle Carol A. (ed.), *The Concise Encyclopedia of Applied Linguistics*, Wiley, pp. 320–326.

Salihoglu S. and Karatepe C. (2023), “Investigation of online news about Ukraine-Russia war: A corpus-based critical discourse analysis”, *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, Issue 33, 1171–1186. DOI: <https://doi.org/10.29000/rumelide.1279172>

Shevchenko I., Morozova I. and Shevchenko V. (2022), “Nominatsiyi ahresiyi Rosiyi proty Ukrayiny v anhlomovnykh media: kohnityvno-prahmatychnyy analiz”, *The Journal of V. N. Karazin Kharkiv National University. Series: Foreign Philology. Methods of Foreign Language Teaching*, Issue 95, pp. 70–77. (In Ukrainian). DOI: [10.26565/2227-8877-2022-95-09](https://doi.org/10.26565/2227-8877-2022-95-09)

Sun W. (2022), “Division of Chinese opinion on Russia-Ukraine war: Analyzing key actors’ strategic framing on Weibo”, *European Journal of Law and Political Science*, Vol. 1, No. 5, pp. 48–56. DOI: <http://dx.doi.org/10.24018/ejpolitics.2022.1.5.45>

*М. І. Леєрінц*

**Компаративний корпусний аналіз висвітлення  
російсько-української війни  
в англомовній східній і західній пресі**

У статті представлено результати компаративного аналізу мовних особливостей дискурсу російсько-української війни у східних і західних ЗМІ, виконаного із застосуванням методів корпусної лінгвістики. Вивчення спільних і відмінних рис у мовних засобах висвітлення війни дає змогу розкрити сутність прихованих ідеологій, що поширюються через ЗМІ, які опосередковано впливають на хід і кінцевий результат подій. Базу дослідження становлять два спеціалізованих корпуси, згенерованих з релізу східної (475506 слів) і західної (531973 слова) преси. Текстові дані було проаналізовано із застосуванням методів аналізу ключових слів, аналізу n-грам і колокацій, з подальшою контекстною інтерпретацією мовних одиниць.

На підставі аналізу виявлено як спільні риси, так і розбіжності в мовних засобах конструювання ключових нарративів. Негативна дискурсивна просодія простежується у виборі мовних засобів в обох аналізованих корпусах. В обох інформаційних джерелах наявне засудження воєнної інтервенції, водночас зафіксовано розбіжності у ступені акцентування позиції щодо війни та перебігу воєнних дій. У східних джерелах спостерігається неоднозначність позиції стосовно зазначеного конфлікту з тенденцією до його ідентифікації з внутрішньо зумовленими чинниками щодо України. Натомість західні джерела зосереджують значну увагу на висвітленні розгортання воєнних подій і змалюванні її наслідків, використовуючи емоційно-забарвлені лексичні засоби. Крім того, східні ЗМІ в контексті війни надають першочерговості економічній стабільності в регіоні, подоланню наслідків пандемії COVID-19 із наголошенням потреби діалогу та міжнародної співпраці. З'ясовано, що східні джерела акцентують політику нейтралітету, якої дотримуються офіційні представники регіону, на відміну від західних ЗМІ, які виявляють непохитність позиції в засудженні агресивних дій з боку Росії та безумовній підтримці української сторони. Отже, критичний аналіз меседжів, закодованих у мовних засобах, що поширюються медіа, набуває великого значення в контексті війни, оскільки обґрунтованість суджень у воєнному дискурсі є першоосновою об'єктивності серед читацької аудиторії.

**Ключові слова:** англомовні медіа; війна; дискурс; Захід; корпусна лінгвістика; лексичні особливості; Схід; Україна

*Стаття надійшла до редакції 16.07.2023*