Linguistic representation of Ukraine in English-language media discourse: a corpus-assisted approach

Solomiia Albota*

Lviv Polytechnic National University, Bandera Str., 12, Lviv, 79013, Ukraine

Abstract

This paper focuses on the representation of vision and perception of Ukrainian state within English-language media discourse, employing corpus linguistics methods and textual semantic analysis. The linguistic data has been allocated on the basis of *English Trends* target corpus, which is being kept up-to-date on progress tracing the latest news articles (7,754,337,212 tokens). Another, reference corpus – *English Web 2021 (enTenTen21)* – has been chosen with an intention to compare the perception of Ukrainian state abroad before and during the wartime. Using the *Sketch Engine* software package, a keyword analysis has been conducted. It has provided a continuum of Ukrainian world view from the perspective of English-language media perception incorporating lexical patterns. Statistical data has been assembled by discovering corpus keywords and their collocations concerning Ukraine. Their cases of usage within the linguistic context have been linguistically interpreted applying the textual semantic analysis. In order to reveal semantic prosody and statistics of the corpora linguistic data highlighting Ukraine in the news coverage, discourse themes have been allocated and linguistically interpreted using textual semantic analysis. Such textual software tools as *MonkeyLearn, Linguistic Inquiry and Word Count* and *Voyant* have been employed. A comparative analysis in terms of the linguistic data as well as statistical data has been performed. The author's views upon the key for the study corpus linguistic terms have been provided.

Keywords

corpus-assisted approach, corpus linguistics, textual semantic analysis, keywords analysis, collocations, English-language media discourse, Ukraine, semantic prosody, English-language corpora, linguistic analysis, statistical data, discourse theme

1. Introduction

The impact of news coverage on the common belief and perception has been widely recognized [1-3]. There is even a term – *vox populi* – the opinion of the majority of the people [4], which is vital when drawing portrait or image of the object under investigation. Ukraine has always attracted foreigners' attraction with their various attitudes and ways of perception, however, currently, Ukraine is covered in such an environment that no other

CLW-2024: Computational Linguistics Workshop at 8th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2024), April 12–13, 2024, Lviv, Ukraine

* Corresponding author.

solomiia.m.albota@lpnu.ua (S. Albota)

0000-0003-3548-1919 (S. Albota)

country would like to be in. Since russian-Ukrainian war news articles soared dramatically (what will be illustrated in sub-paragraph 4.2), media discourse has been involved in disseminating narratives about Ukrainian position imposed by martial law and military actions.

When there is an insight into media discourse, linguistic content and its social and linguistic context are considered. That is why a study of linguistic peculiarities of portraying Ukraine in the international arena is of importance given the full-scale military action, its coverage in English-language media discourse and general awareness. The purpose of the paper is to provide complex linguistic analysis of depicting Ukrainian position in the light of English-language media discourse incorporating corpus-based approach and textual semantic analysis. What concerns corpus linguistics, the focus corpus known for its genre of news articles *English Trends* and reference corpus *English Web 2021* are targeted to provide keywords analysis with a task to demonstrate collocational manner of the word *Ukraine* and its corpora statistical data like frequency values and keyness score to compare the vision of Ukrainian state abroad two years before and during the wartime encompassing a 5 year time span.

What is more, preliminary corpus-based findings concentrate on revealing discourse themes within the *English Trends* linguistic context in a form of concordance with reference to textual semantic analysis. By means of the latter there is a task to define either positive, negative or neutral semantic prosody of the discourse themes separately and after simple calculations – of one underlying theme to reflect a general view of Ukraine within English-language media discourse.

The paper fulfills its purpose by application of the following software tools: *Sketch Engine* – for the corpora management and sketching the linguistic data with its statistics; *Voyant* web-based tool – for the corpus visual representation and its statistics; *MonkeyLearn* analyzer and *Linguistic Inquiry and Word Count* digital tool– for proving semantic prosody of the discourse themes and demonstrating their statistical results.

2. Related Works

Media discourse encompasses information and clarification of disputes, effects of vox populi and discussion focusing on finding approaches to mutual agreement. In linguistics, media discourse is of importance in terms of: language variation and change (media discourse offers a rich source of linguistic data that reflects the diversity of language use in different contexts and communities; linguistic analysis of media language allows linguists to study language variation and change over time, including shifts in vocabulary, grammar, and discourse patterns); pragmatics and discourse analysis (the above mentioned discourse provides a fertile ground for studying pragmatics and discourse analysis; linguistic research of media texts helps uncover the pragmatic strategies employed by speakers and writers to achieve communicative goals, as well as the organization and structure of discourse in different media genres); language (the news genre media discourse is deeply intertwined with issues of power, ideology, and social identity; conducting lingual analysis allows researchers to examine how language is used to construct and reinforce power relations, shape public opinion, and perpetuate social inequalities within media contexts);

sociolinguistics (media discourse provides valuable insights into sociolinguistic phenomena, such as language variation and language attitudes; media language analysis can shed light on how linguistic features are associated with social factors such as gender, ethnicity, age, and social class, as well as how language ideologies are constructed and disseminated through media texts); cognitive linguistics (studying media discourse can also contribute to the understanding of cognitive processes involved in language comprehension and production; media texts research helps uncover how cognitive mechanisms such as inference, metaphor, and mental representation are activated and manipulated in the processing of media language); corpus linguistics (media corpora serve as valuable resources for corpus linguistics research; such investigation allows researchers to investigate language use patterns, frequency distributions, and collocational preferences across different media genres and discourse domains) [5].

Among functions news articles genre in terms of English-language media discourse can possess there are the following: informative (media news coverage provides information about current events, developments, and issues to the public, keeping them informed about what is happening locally, nationally, and globally), educational (media discourse helps educate the audience about various topics, including politics, science, culture, and economics, by presenting in-depth analysis, explanations, and background information), program-oriented (media news discourse enables the influence on public opinion and shapes the program set by highlighting certain issues or events over others, thereby directing attention to specific topics or concerns), convincing (news genre of the mentioned discourse often aims to persuade or influence the audience's opinions, attitudes, and behaviors through framing, tone, and selection of information), appealing (media news coverage can appeal to the public to take action or participate in social, political, or civic activities by raising awareness, fostering support for causes, or rallying people around certain ideas or movements), entertaining (while primarily focused on information dissemination, media news discourse also serves an entertainment function by engaging audiences through storytelling, human interest features, and engaging visuals), socializing (news article and reports help socialize individuals into societal norms, values, and expectations by providing insights into cultural practices, social dynamics, and community events), observing (media discourse monitors the actions of governments, institutions, and powerful entities, holding them accountable and ensuring transparency and accountability in society) [5].

By analyzing genuine texts, a corpus-based methodology has the potential to reveal aspects of language that were previously unknown offering a fresh perspective on the nature of language. Researchers can employ computer software to process and scrutinize language data, aiding in the identification of patterns in language usage [6, 7]. Scholars [8-10] characterize the term *corpus* as an enlarged dignified set of common texts that are formed intrinsically in a written or spoken way collected in a digital manner. This definition ensures the key features of the corpus indicating its importance. Intrinsic way of collecting texts means that such a collection is managed by the purpose of a scientist to compile the corpus. There is an option to compile a scientist's own corpus if there are not enough amount of certain texts required for the study. Another feature of the corpus – it has to enumerate texts that are actualized in the language so that they can be used as a real

language. Thus, the application of corpus-based approach tools has a great impact on the linguistic studies [11]. The past experience of emerging corpora proved that textual collection was compiled with different purposes within the teaching process such as dictionaries, parallel texts, etc. Afterwards, interest to the established corpus tools has risen, among which there are:

Corpus Query Tools: such tools enable scientists to search and retrieve linguistic data from corpora based on specific criteria. Examples illustrate: AntConc [11], Sketch Engine, Corpus Workbench (CWB), CQPweb, NoSketch Engine.

Concordance Tools: concordance instruments generate concordance lines, which display instances of a search term in context. Scientists can use concordances to analyze collocations, word usage patterns, and discourse structures. Exemplification involves: AntConc [11], MonoConc, WordSmith Tools.

Annotation Tools: notation instruments are used to manually or automatically annotate linguistic features in corpora, such as part-of-speech tagging, syntactic parsing, and named entity recognition. The following items included: TreeTagger, Stanford CoreNLP, NLTK (Natural Language Toolkit), Spacy.

Corpus Management Systems: these instruments facilitate the organization, storage, and retrieval of corpus data. They often include features for corpus indexing, metadata management, and corpus querying, for instance: Corpus Workbench (CWB), IMS Open Corpus Workbench (IMSCWB), Corpus Query Processor (CQP), Sketch Engine.

Statistical Analysis Tools: calculation research instruments allow researchers to conduct quantitative analyses of corpus data, such as frequency counts, collocation analysis, and measures of lexical diversity, for example: R, Python (with libraries such as NLTK, scikit-learn, and pandas) [12], SPSS (Statistical Package for the Social Sciences).

Vision Instruments: visualization tools help researchers visualize corpus data in various formats, such as graphs, charts, and heatmaps. Visualization aids in the exploration and interpretation of linguistic patterns and trends in corpora. These are illustrated: AntConc (for word frequency charts, concordance plots) [11], Voyant Tools, Wordle, Tableau.

Alignment Tools: coordination tools are used for aligning parallel or comparable corpora, which consist of translations or texts in different languages. These tools facilitate the study of translation, language variation, and cross-linguistic phenomena, exemplifying the following: GIZA++, hunalign, LF Aligner. These tools, among others, play a crucial role in corpus linguistics research, enabling researchers to explore linguistic patterns, conduct empirical analyses, and gain insights into language structure and use.

Semantics is characterized by designing of deeper semantic relationships among words occurring within collocations extracted by the corpus management software [13]. Semantic prosody, as the final corpus linguistic term employed in the paper, arose from the corpus linguistics. Semantic prosody, also known as discourse prosody, is used to indicate positive, negative or neutral associations of the words, collocations or the whole corpus statements [14]. Prosody was identified as hierarchic depending on the frequency of good, bad or neutral occurrences within the linguistic context, and compulsory possessing its explicitness – direct meaning of the word within the linguistic context. This term has been also characterized as an attitudinal preference [15] expressing its evaluative meaning. When there is a linguistic evaluating process, linguistic interpretation [16] attempting to

unite meaning with language is employed. Linguistic interpretation is inherent to the discourse analysis: understanding discourse is crucial in human communication, as it involves navigating social norms within a given situation to convey intentions effectively through language. Interpreting meaning in discourse goes beyond literal words as it is a dynamic exchange where linguists negotiate understanding within the context. Interpretation can be performed under either contextual or pretextual conditions but discourse analysis cannot [17].

Context plays a crucial role in understanding and linguistic interpretation of a meaning, as words are not perceived separately but are intricately linked to the surrounding context. The interpretation of one word relies on the relationship between text and context. Context is a dynamic concept, continuously evolving surroundings that facilitate interaction among communication participants and render the linguistic expression of their interaction [17]. Views on categorizing context differ among scholars [18, 19], with some proposing two categories, while others argue for three, four, or even six dimensions. Based on the various circumstances outlined in the preceding sources, there is an offer to differentiate between linguistic (it is crucial within this study), situational and cultural context (this can be partially interpreted in terms of textual semantic analysis in sub-paragraph 4.2). Unfortunately, there are no new sources to refer to contextual difference studies as each and every research goes back to 199*n*th definitions upon which it is reasonable to focus in linguistics [20].

Linguistic context [21] pertains to the context within a given discourse, encompassing the connections between words, phrases, sentences, and even paragraphs. For instance, consider the word warrior. Without the linguistic context, such as surrounding words and sentences, the precise meaning of the statement *They are warriors* cannot be determined. Linguistic context can be examined through such dimensions as deictic (during a language interaction, participants need awareness of their spatial and temporal surroundings; deitic expressions should be taken into consideration like *currently, afterwards*, spatial markers like *over here, over there,* and personal pronouns like *I, you*; these expressions are crucial when setting deitic roles stemming from the customary use of language where speakers communicate with others referring to themselves, specific venues or time span), co-text (focus on the preceding discourse coordinate – statements mentioned earlier; any statement following the first in a segment of discourse will have its interpretation heavily influenced by the preceding text, not only by phrases explicitly referring to it. According to scholars from past centuries, the interpretations of words within discourse are constrained by their co-text), and collocation (here is advocacy for acknowledging the significance of syntagmatic relations, such as those between *clap and hands*, *hiss and snake, blue and sky*, which refer to as collocation is taken into consideration; collocation isn't merely about the association of ideas. For example, *war is red*, war is indeed red as it is full of blood, we do not typically say *red war*, whereas the phrase *red rose* is commonly used).

Situational context [21], also known as the context of situation, pertains to the setting, timing, location, and the dynamic among participants during discourse. It involves considering the environment and the relationships among those involved. Traditionally, this theory is explored using the concept of register, which categorizes language use into three main components: field (the ongoing activity or purpose of communication; it

represents how language usage reflects the intentional role of the language user within the context of a text), tenor (relates to the social relationship manifested through discourse; it emphasizes that linguistic choices are influenced not only by the topic but also by the social dynamics between communicators), and mode (reflects the relationship between the language user and the medium of transmission; it distinguishes between communication channels that involve direct interaction and those that permit delayed interaction between participants). This approach aids in understanding how language interacts with its context.

Cultural context [21] encompasses the cultural norms, traditions, and historical background of language communities in which speakers are involved. Language is inherently social and intimately connected with the societal structure and values. Thus, it is inevitably influenced by various factors such as social roles (culture-specific functions that are established within a society and acknowledged by its members), status (denotes the relative social standing of individuals involved; each participant in a language event must be aware of, or make assumptions about, their status in relation to others, and often, status plays a significant role in determining who initiates the conversation), gender and age (frequently serve as determinants of, or interact with, social status; for instance, the terms of address used by one gender when speaking to an older individual may differ from those used by individuals of the same gender or age in similar situations), and other aspects of the cultural environment.

Having claimed media discourse to be the most studied and functional source of the linguistic findings, this paper will comprise three main linguistic directions: English-language media discourse, corpus-assisted approach and defining semantic prosody of the media discourse themes referring to Ukraine.

3. Methods and materials

There were different media discourse analyses concerning European political conflict, Israeli-Palestinian conflict, russian-Ukrainian war [22, 23] emphasizing the reflection of national image concerning conflict notion. All these studies mainly follow critical discourse analysis methodology, which is aimed more at defining social practice and social ideology [24-27]. In this paper, the attention is paid to the complex linguistic analysis which combines the two-stage methodology moving backward to detect the precedence of conflict and forward in order to outline possible threats to global peace.

The first methodological stage addresses the issues in terms of corpus-based approach applying *Sketch Engine* software corpus management tool. The reference to media discourse in this case is traced through target and reference corpora.

The target corpus under this investigation – English Trends (7,754,337,212 tokens and 6,627,303,274 words) – English-based corpus enumerating constantly updated news feeds (13 million words each day) The reference corpus – English Web 2021 (enTenTen21) – encompasses only linguistically valuable web content (61,585,997,113 tokens and 52,268,286,493 words), however, the texts were downloaded in October–December 2021 and January 2022 (https://www.sketchengine.eu/guide/). Corpora synchrony will be beneficial for distinguishing linguistic peculiarities of the *Ukraine* lexical patterns. Corpus-based approach to the study envisages the use of the following

methodology: application of *SketchEngine* software to conduct keywords analysis of the English trends corpus; using of *SketchEngine* software and *Voyant* textual software to discover collocations based on the words allocated by keywords analysis; textual semantic analysis of the corpora statements within the concordance on the basis of collocation occurrences and identification of semantic prosody of media discourse themes; linguistic analysis incorporating lexical patterns and their interpretation. It is also noteworthy that linguistic interpretation has been applied to the complex corpus-based approach to the study in order to linguistically comment each lexical occurrence within the Ukrainian context in chronological order.

Below, a list of corpus linguistic terms which are used in this paper and their explanation according to this research is provided.

Target, or also known as focus in *SketchEngine* software package guide (https://www.sketchengine.eu/guide/), corpus consists of a set of texts being investigated. Reference corpus involves being compared with those texts of target corpus. This is done for textual as well as statistical comparison (https://www.sketchengine.eu/guide/). With an intent to demonstrate perception of Ukraine in the international arena, media discourse was considered a source of linguistic data. Conducting keywords analysis is crucial for revealing collocations and lexical patterns of the word *Ukraine*.

Key or keyword has been studied as a notion across a culture, where its semantic development has been tracked over time [28]. That is why corpus keywords analysis is also known as corpus-based cultural keywords, which can be linguistically analyzed by searching collocations (collocates with nodes) – the frequency of words in proximity to keywords within the whole corpora [29, 30]. This corpus approach is called qualitative, and quantitative presupposes keyness simple math method according to which it is called corpus-comparative statistical keyword [28]. Cultural keyword absorbs all lexical words, while statistical keyword may cover also grammatical words. Objectivity established by engagement of corpus software between both corpus approaches is always relative to reference corpus [31].

Each collocation can be observed within a certain concordance – a set of lines illustrating the word detected on the basis of the corpus keywords analysis within a corpus, usually in the format of a KWIC, which stands for Key Word in Context and resembles the red text highlighted in a concordance with reference to the right and left lines (https://www.sketchengine.eu/guide/). These corpus concordance lines will be presented as corpus statements [32, 33] by which English-language media discourse utterances within the target corpus are meant. In the fourth paragraph they are usually allocated out of American, Australian, British, Canadian digital news articles and reports. Another term that helps to analyze the linguistic means within the corpus collocations is lexical patterns [34, 35]. The latter are considered within the linguistic context of English-language media discourse. Lexical patterns in this paper involve words that occur within the corpora statements with high frequency incorporating different forms of lexemes, and the meaning of some words referring to certain lexical pattern may alter. Utilizing computer software allows researchers to access various aspects of a corpus, including word lists, concordance lines, collocates, and lexical patterns. Via a corpus-based approach, scientists can impartially observe patterns in linguistic usage. Analyzing search items, lexical patterns,

and collocations from a corpus enables a qualitative examination, enhancing the reliability and comprehensiveness of findings. Incorporating corpus-based methods into discourse analysis can mitigate a scientist bias by leveraging the naturalness of language data and the objective assessment derived from extensive data analysis. Consequently, corpus-based approach is frequently integrated with discourse analysis to examine language use in the context of social events.

Corpus-assisted approach in linguistics has proved to be developing and combined along with sentiment analysis [36-39], which approaches our research to the second stage of the two-stage complex linguistic analysis, which concerns linguistic interpretation and textual semantic analysis both manually and using digital tools. By textual semantic analysis in this paper with a reference to corpus linguistics revealing the *Ukraine* word meaning in terms of English-language media discourse represented by target corpus linguistic context is considered.

Moreover, textual semantic analysis is applied to outline discourse themes highlighting main concerns of Ukraine from foreigners' perspective in a chronological order. English-language media discourse themes will be based on the linguistic analysis of the corpora collocations and their concordance statements and proved by digital software tools mentioned above. These themes represent particular linguistic context for defining semantic prosody of each corpus statement on the whole. In this study, semantic prosody refers to different ways of word or collocation associations which can be either positive, negative or neutral.

What concerns linguistic interpretation, it comprises both research stages: corpora-based approach requires interpretation of keywords and collocations as well as their statistical representation; defining semantic prosody of the discourse themes needs to be permeated with interpretation of corpus statements within the discourse linguistic context referring to the Ukrainian position over the last four years.

In this paper, a corpus-based approach is utilized to gather the wordlist data from news genre texts of the English-language media discourse using appropriate corpus software, extract concordance lines for revealing Ukrainian position attitude from the perspective of foreigners' perception, determine the frequency values of all statistical data defined, and analyze the linguistic context surrounding the word Ukraine in terms of discourse analysis.

4. Results

This paragraph entails the outcomes of two-stage complex linguistic analysis incorporating corpora-assisted approach to the English-language media discourse and identification of semantic prosody of the discourse themes via textual semantic analysis.

4.1. Applying corpora-based approach to the English-language media discourse vision of Ukraine

Ukrainian image in the international arena has occupied its niche long before such a harsh military conflict. Unfortunately, with sadness it is worth noting that for the last two years the interest in Ukraine was sparked by russian-Ukrainian war, which can be tracked within a linguistic research [40]. With an aim to linguistically highlight the importance of Ukrainian

state in the world under current martial law and trace the lexical patterns of Ukrainian world view from the perspective of English-language media discourse a range of corpus linguistic tools along with textual semantic analysis have been applied.

Lexical computing tool – *SketchEngine* – a leading corpus management and corpus query instrument used by linguists, lexicographers, translators worldwide (https://www.sketchengine.eu/guide/). It is useful when conducting keywords analysis in order to detect the most frequent words that occur within the *English Trends* corpus and possess an ability to understand the main topic of the corpus. Keywords can be defined either qualitatively – when an author refers to them as the terms discovering their social and cultural importance or quantitively – in order to emphasize the importance of the word frequency in two corpora. In the paper, both quality and quantity approaches to keywords analysis will be used. Firstly, having set the criterion using *SketchEngine* for the corpus query within the news genre of the media discourse of the target *English Trends* corpus and reference *English Web* corpus, the following list of corpora keywords with their line numbers have been obtained (Fig.1). As the English Trends corpus is continuously updating and the author has earlier saved the screenshots like in Fig.1 and further within the paper, it would be wrong to use *SketchEngine* again with the same corpora queries due to failure in the interpretation and change in corpus keywords order provided. As the task of the paper is to track the position of word *Ukraine* within the English-language media discourse, using the linguistic qualitative corpus-based approach to the keywords analysis as well as linguistic interpretation, it is obvious that *Ukraine* as a word retains its "popularity" among media discourse news feeds. The lexical patterns of Ukrainian image from the perspective of English-language media discourse look the following way: *Ukraine* holds the sixth position, *Kyiv* – the seventh position, *Ukraine's* – the eighth position, *Zelensky* - the tenth position. The frequency values as well as keyness score have been calculated (Table 1).

Keywords provided in Table 1 have been arranged in descending order and taking into account statistical values, not the line number mentioned above. Given that the whole corpus has been under investigation, Ukrainian question has remains of great significance from July, 2020 up till now in 2024 (Fig.2, Fig.3).

Table 1

Frequency values and keyness score of the target and reference corpora (only keywords referring to Ukrainian current military state)

Keywords	Target corpus	Reference corpus	Keyness score
Ukraine	348.33	14.78	22.1
Ukrainian	150.27	9.05	15.1
putin	107.10	6.54	14.3
russia's	85.65	5.97	12.4
Ukraine's	50.65	1.37	21.8
Kyiv	49.27	1.30	21.9
vladimir	48.45	4.92	8.4
putin's	34.06	1.35	14.9

kremlin	26.75	1.95	9.4
Ukrainians	26.04	1.10	12.9
Zelensky	24.65	0.24	20.6
Volodymyr	18.88	0.29	15.4
Mariupol	14.75	0.11	14.2
Zelenskyy	13.06	0.02	13.8
Donetsk	12.60	0.71	8.0
Kherson	11.50	0.09	11.4
Kharkiv	11.08	0.32	9.1
Zaporizhzhia	7.25	0.03	8.0

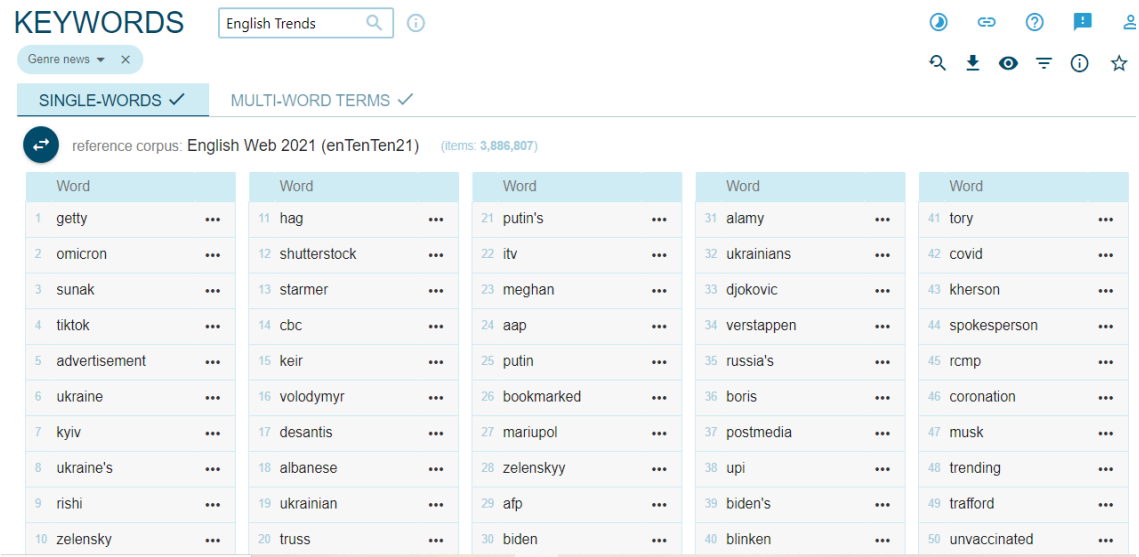

Figure 1: Corpora keywords

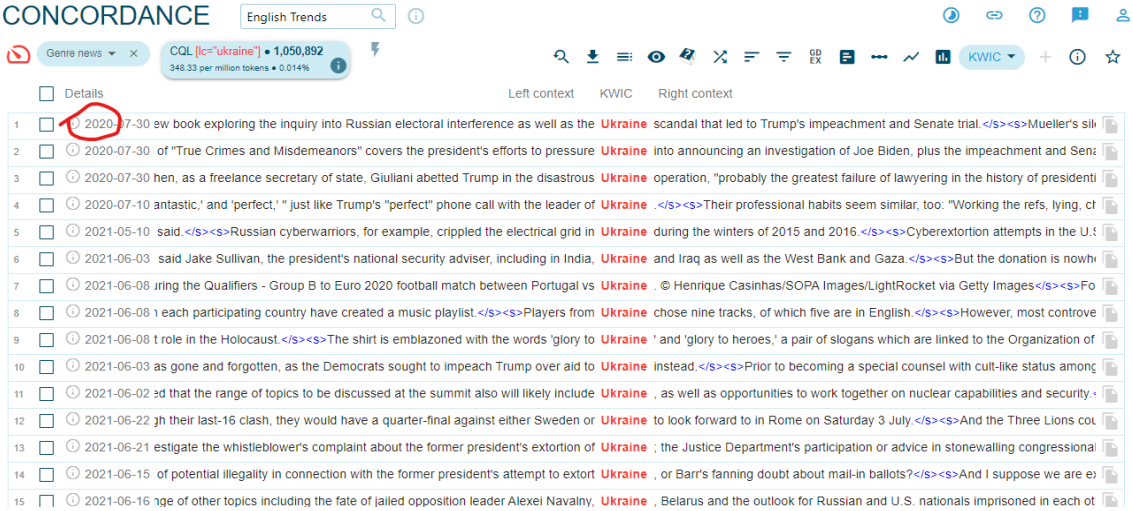

Figure 2: The *Ukraine* keyword's first occurrence

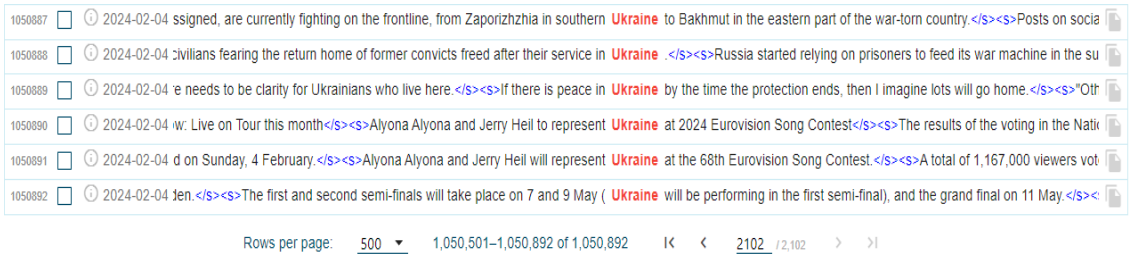

Figure 3: The *Ukraine* keyword's latest occurrence

Within the above-illustrated concordances there are statements referring to Ukraine covering not only current military state, that is why one of the research tasks is to discover media discourse themes and further define their semantic prosody.

The fact is that the frequency value of the word *Ukraine* within the target corpus (348.33) is higher than any other keyword in the corpora even outbidding *Covid, advertisement* and *getty* (Fig.4).

	Word	Focus	Reference	Score ?		Word	Focus	Reference	Score ?		Word	Focus	Reference	Score ?
1	getty	286.38	3.41	65.2	18	albanese	16.98	0.19	15.1	35	russia's	85.65	5.97	12.4
2	omicron	63.36	0.22	52.8	19	ukrainian	150.27	9.05	15.1	36	boris	61.69	4.21	12.0
3	sunak	53.01	0.22	44.1	20	truss	33.05	1.27	15.0	37	postmedia	12.05	0.09	12.0
4	tiktok	77.97	1.98	26.5	21	putin's	34.06	1.35	14.9	38	upi	18.48	0.62	12.0
5	advertisement	325.19	11.88	25.3	22	itv	47.77	2.29	14.8	39	biden's	44.20	2.79	11.9
6	ukraine	348.33	14.78	22.1	23	meghan	49.01	2.39	14.7	40	blinken	14.35	0.32	11.6
7	kyiv	49.27	1.30	21.9	24	aap	32.53	1.30	14.6	41	tory	49.47	3.37	11.6
8	ukraine's	50.65	1.37	21.8	25	putin	107.10	6.54	14.3	42	covid	213.00	17.53	11.5

Figure 4: The frequency value of the word *Ukraine* among others within the target corpus

So, keywords analysis provided a window into the Ukrainian state apprehension from the perspective of English-language media discourse: according to the frequency values *Ukraine* has remained the most debated word since 2020. Its lexical patterns (*Ukrainian, Ukraine's, Ukrainians*) indicate the importance and interest to our country. At the same time, even when lexical patterns of the presidents differ in values within the target corpus with the opposition prevailing (*putin, vladimir, putin's* = 189.61; *Zelensky, Volodymyr, Zelenskyy* =56.59), the keyness score concerning our president's lexical patterns increases (Zelensky, Volodymyr, Zelenskyy =49.8; *putin, vladimir, putin's* = 37.6). It is worth remarking that there is no lexeme like *russia* making Ukraine the most powerful keyword within the discourse. What concerns territories, Ukrainian lands dominate in total of almost all three-column frequency values (Table 1) and in quantity of territories mentioned (*Kyiv, Mariupol, Donetsk, Kherson, Kharkiv, Zaporizhzhia* = 106.45; 1.26; 16; *kremlin* = 26.75; 1.95; 9.4). What is meticulous about analysis in synchrony – second column represents the statistics of the reference corpus of 2021, and judging from its values there were low indicators of these lexical patterns, which means that only after 2022 a vivid debate about Ukrainian state took

place, unlike the second column of the opposition, whose territory aroused interest only before the russian-Ukrainian war.

Now, on the basis of keywords analysis with the help of Word Sketch corpus management tool a grammatical and collocational manner of the word *Ukraine* is being considered (Fig. 5).

Here is a list of collocates, by which a part of collocation which is dependent on the node. And the latter is a central word of a collocation. For instance, the first column on the left demonstrates the collocate *weak* to the node *Ukraine*. The first figure refers to the frequency value of its occurrence within the target corpus (2), and the second score indicated the strength of collocation (1.3). In comparison to other collocated in this Figure 5 it has been calculated that the highest frequency value of the collocation is that which constitutes the concordance statements of the year 2023 (Fig.6).

The strongest collocate is *moscow* in a relation to noun modified by *Ukraine* (10.3). However, when analyzed linguistically using textual semantic analysis and linguistic interpretation to the previous concordance statements of the collocation (Fig.6) – conflict issues with the opposition have been revealed, Fig.7 shows the collocate *moscow* in terms of the opponent's parade day.

The rest of the frequency values in their descending order have been underlined in red in Fig. 5 in order to form discourse themes uncovering the Ukrainian situation and its perception within the English-language media discourse. Also, having applied textual semantic analysis to 500 corpus concordance statements of the year 2024, a list of first ten frequent collocates to *Ukraine* have been explored. As this list in *SketchEngline* table does not change, here is its link (https://ske.li/collocates).

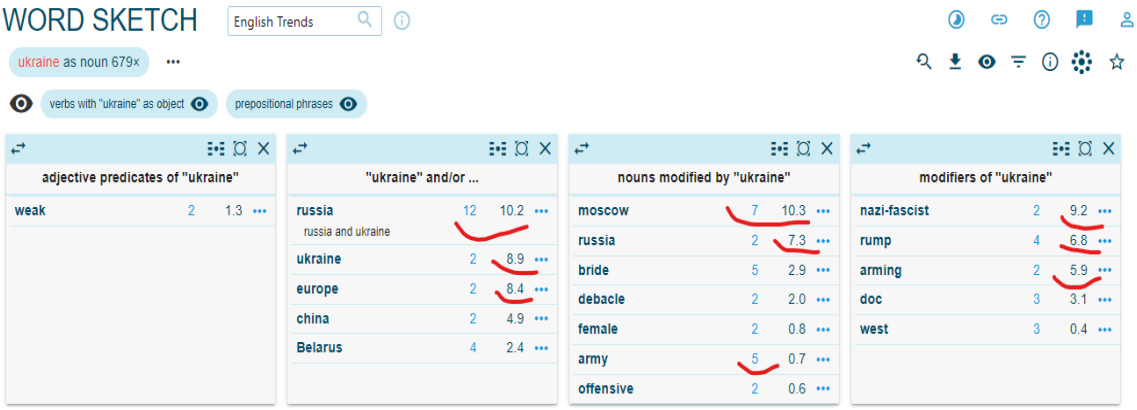

Figure 5: Grammatical and collocational manner of the word *Ukraine* using *SketchEngine* software

According to this link (https://ske.li/collocates) a column with Cooccurrences shows the total amount of the collocate occurrence in terms of concordance linguistic context parts either left or right, and the Candidates column indicates the amount of the collocate occurrences within the whole corpus. In comparison to the statistics of the corpus, English-language media discourse in 2024 demonstrated Ukraine by *war* and *invasion* the most frequent. With an aim to prove the perception of Ukrainian state by English-language media

discourse the Voyant software tool has been applied. A link to *Voyant* tool visualization of the word Ukraine within the English Trends corpus (https://voyant.lincsproject.ca/?corpus=867ea9347fb65a9dc3c847b7cca32004&visible=95&view=Cirrus) has been modelled. Another – summary link to the most frequent words in the corpus (https://voyant.lincsproject.ca/?corpus=867ea9347fb65a9dc3c847b7cca32004&view=Summary) has been provided.

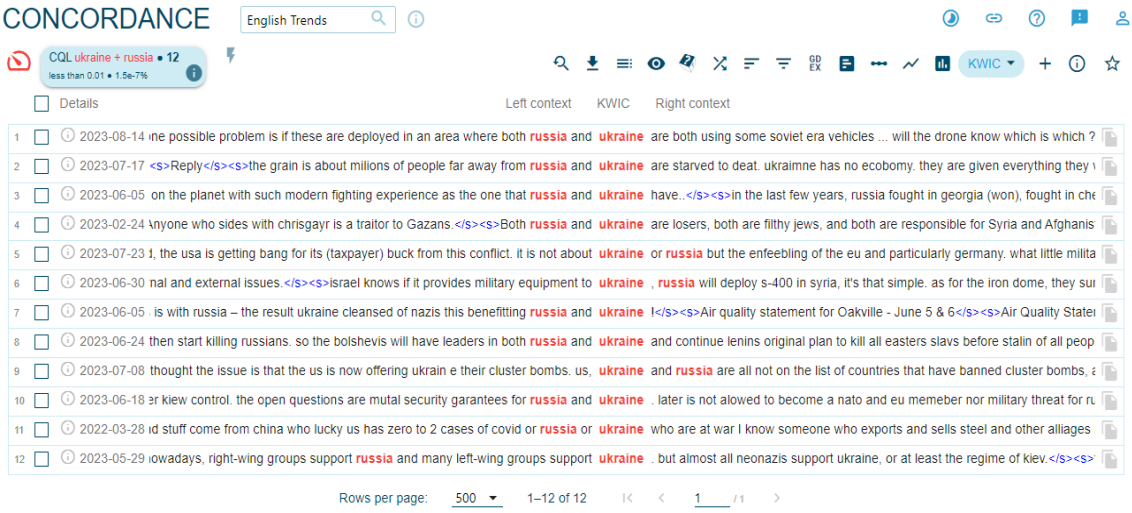

Figure 6: Concordance statements referring to the collocation *russian and Ukraine* (12) of the year 2023

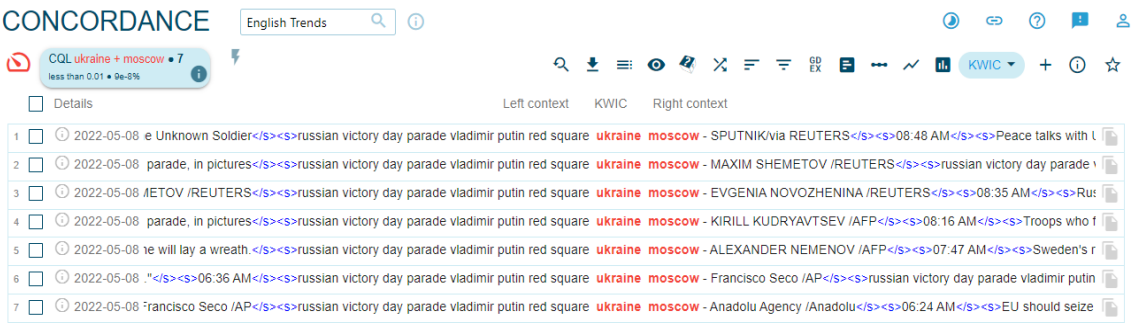

Figure 7: Concordance statements referring to the collocation *moscow and Ukraine* (12) of the year 2023

Regarding corpus-based approach to the linguistic representation of Ukraine it is noteworthy that *Voyant* web tool enables clear apprehension of the Ukrainian position within the English-language media discourse – russian-Ukrainian war in 2022 left an imprint on each and every news article in the world. When scaling the *Voyant* visual image in a downward position the first word after *Ukraine (12872) 2022 (8793) 02 (6288)* – *invasion* (1667) and the lexical patterns (*russia (4137), russian (3215)*). The first word of the latter lexical pattern was not inherent to the keywords in Table 1; however, it was noticed in visual representation of Ukrainian perception via Voyant tool

(https://voyant.lincsproject.ca/?corpus=867ea9347fb65a9dc3c847b7cca32004&visible=95&view=Cirrus) among the collocations. After further zooming such words as *president* (1488), *military* (1325), *Europe* (561), *security* (565), *sanctions* (733), *forces* (576), *border* (731), *troops* (1128), *support* (582), *people* (703), *crisis* (592), *conflict* (413), *government* (438), *allies* (384), *attack* (612), *minister* (480), *Crimea* (243), *aggression* (304), *weapons* (326), *threat* (253), *tensions* (476) have been selected along with their frequency values. The most frequent word - *president* (1488) – demonstrating interest among foreigners and inspiring trust among population.

4.2. Defining semantic prosody of the Ukrainian discourse themes through textual semantic analysis

The next linguistic analysis stage is to define the discourse themes within the English Trends corpus using textual semantic analysis and linguistic interpretation as well as contextual analysis of the corpus concordance statements which were briefly discovered in sub-paragraph 4.1. In other words, the generation of keywords analysis and linguistic analysis of collocates in terms of collocations helped to allocate the linguistic data giving prominence to target corpus in contrast with the reference corpus.

As for the division of corpus data into media discourse themes, time frames will be taken into consideration. In each discourse theme of the following year there is an example of corpus concordance statement either in a form of collocation or a word characterizing Ukraine and its state at certain period of time. By semantic textual analysis the filtering of the most appropriate corpus statements towards Ukrainian current state has been conducted. Sports reports or Ukrainian role in different contests have been neglected. In brackets, there is a reference whether the word or collocation is neutral, positive or negative with appropriate linguistic interpretative remarks, as well as corpus statement source. Each discourse theme is followed by Figure referring to the corpus semantic prosody of discourse themes linguistically and semantically analyzed using *Linguistic Inquiry and Word Count* tool as well as *MonkeyLearn* textual analytics tool. The latter is used to each and every corpus statement allocated for the linguistic and semantic analysis to prove the manual results of the linguistic interpretation (figure with results is given only after the discourse theme 1 of the tear 2020 not to overload the paper with images).

Discourse theme 1 of the year 2020: *the **Ukraine scandal** that led to Trump's impeachment* (*June, 2020, The Washington Post*), (negative semantics – willingness to involve Ukraine into disgrace, condemnation of Ukraine in the international arena); *president's efforts to **pressure Ukraine** into announcing an investigation of Joe Biden* (*June, 2020, The Washington Post*), (negative semantics – willingness to involve Ukraine into disgrace, coercion); *Giuliani abetted Trump in the **disastrous Ukraine operation*** (*June, 2020, The Washington Post*), (negative semantics – willingness to involve Ukraine into disgrace, incitement); *Trump's "perfect" phone call with the **leader of Ukraine*** (*June, 2020, The Washington Post*), (positive semantics – emphasis on tenacious human qualities, however, the statement tone with reference to the quoted lexeme *perfect* is tense).

As it can be seen from the corpus statements, 2020 was not marked greatly with the perception of Ukraine, only these four collocations represent negative semantic prosody

referring to the humiliation of Ukraine. To prove the semantic prosody with the above-mentioned digital tools, there is a controversy: the first corpus statement was both linguistically and statistically considered negative (Fig. 8), however, the second corpus statement was considered with positive semantics (Fig.9) but semantically there is an emphasis of Ukrainian negative engagement into the foreign political process, which allowed to linguistically interpret it with negative semantics. So, the *MonkeyLearn* tool analyzes the word pressure in collocation with efforts as a lexically neutral one, that is why semantics has been taken into consideration. The same situation is traced in Fig.10 – digital tool defined neutral semantic prosody of the first discourse theme. For statistical data neutral semantic prosody was considered but the trick with the semantics was worth noting.

Figure 8: Illustrating *MonkeyLearn* tool usage in order to prove negative semantic prosody of the discourse theme 1 of the year 2020

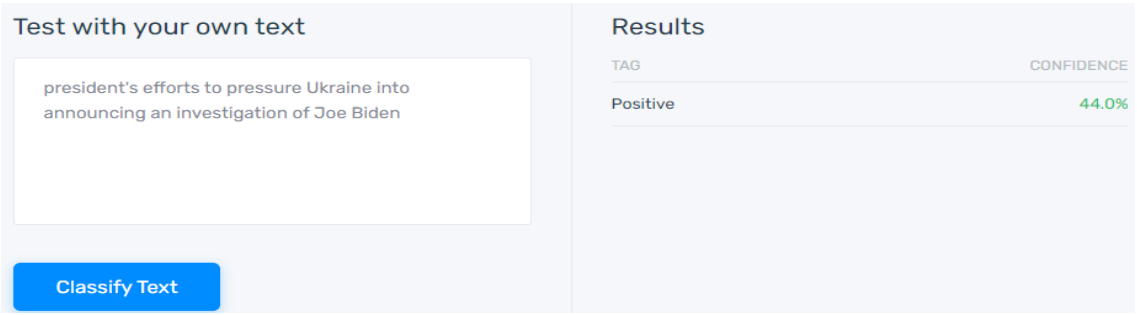

Figure 9: Controversy of revealing semantic prosody of the discourse theme 1 of the year 2020

Traditional LIWC Dimension	Your Text	Average for Social Media Language
I-words (I, me, my)	0.00	5.44
Positive Tone	1.75	5.93
Negative Tone	1.75	2.34

Figure 10: Detecting neutral semantic prosody of the discourse theme 1 of the year 2020 applying *Linguistic Inquiry and Word Count tool*

Discourse theme 2 of the year 2021 is represented by division into months:

June, 2021: ***crippled the electrical grid in Ukraine*** (*June, 2021, The Christian Science Monitor*), (negative semantics – attempt to hack Ukrainian electricity system); *former president's* ***extortion of Ukraine*** (*June, 2021, The Washington Post*), (negative semantics – attempt to blackmail Ukraine);

July, 2021: ***military threat to Ukraine*** (*July, 2021, CBN*), (negative semantics – opponent's willingness to escalate the conflict); ***tensions over Ukraine*** (*July, 2021, RAND Objective Analysis, Effective Solutions*), (negative semantics – intensity of growing conflict); ***situation*** in ***Ukraine escalated*** (*July, 2021, Der Spiegel*), (negative semantics – precedence of russian-Ukrainian war); ***threaten Ukraine*** (*July, 2021, The New York Times*), (negative semantics – opponent's willingness to escalate the conflict);

August, 2021: ***Russian aggression toward Ukraine*** (*August, 2021, CNBC*), (negative semantics – precedence of russian-Ukrainian war); *political* ***weapon against Ukraine****…****hurt Ukraine's economy*** (*August, 2021, CNBC*), (negative semantics – precedence of russian-Ukrainian war); *energy as a* ***weapon against Ukraine*** (*August, 2021, Google*), (negative semantics – precedence of russian-Ukrainian war); ***smuggling*** *across the territory's* ***border with Ukraine*** *(August, 2021, The Euronews*), (negative semantics – attempt to blackmail Ukraine);

September, 2021: *an* ***independent Ukraine*** *(September, 2021, The Euronews*), (positive semantics – willingness to self-defense of Ukraine); *the* ***anti-russia movement in Ukraine*** *(September, 2021, The Euronews)*, (negative semantics – intensity of growing conflict); *the* ***"historical unity" between Russia and Ukraine*** *(September, 2021,The Christian Science Monitor),* (negative semantics – obstinacy as a stumbling point); *if the* ***pipeline****…as a* ***weapon against Ukraine*** *(September, 2021, The Euronews)*, (negative semantics – precedence of russian-Ukrainian war); ***military action in Ukraine*** *(September, 2021, CBC*), (negative semantics – opponent's willingness to escalate the conflict);

October, 2021: ***tensions*** *between russia and Ukraine have been* ***soaring*** *(October, 2021, Express)*, (negative semantics – intensity of growing conflict); *rUSSIA will* ***attack Ukraine by air, land and sea*** *early next year (October,2021, Express),* (negative semantics – precedence of russian-Ukrainian war); ***fears*** *are* ***growing in Ukraine*** *that russia is mounting up to* ***launch a full-scale invasion of the country*** *(October,2021, Express)*, (negative semantics – intensity of growing conflict);

November, 2021: *russia will* ***enact a comprehensive invasion of Ukraine*** *by early 2022 (November, 2021, Express)*, (negative semantics – precedence of russian-Ukrainian war); *to* ***stoke unrest in Ukraine*** *(November, 2021, Express),* (negative semantics – intensity of growing conflict); *to* ***resolve tensions*** *between Ukraine and its neighbor russia* (*November, 2021, ABCnews*), (positive semantics – a call for peace); *to* ***invade Ukraine*** *that borders Belarus to the south* (*November, 2021, Independent*), (negative semantics – precedence of russian-Ukrainian war); *to* ***conduct aggressive actions against Ukraine*** *(November, 2021, The Standard)*, (negative semantics – precedence of russian-Ukrainian war);

December, 2021: *urged the Kremlin to "****de escalate" the crisis over Ukraine*** *(December, 2021, SKY),* (positive semantics – a call for peace); *A* ***russian invasion of Ukraine*** *would set off a major national security crisis for Europe (December, 2021, The New York Times)*, (negative semantics – suffer from worldwide consequences); ***significant aggressive moves***

against Ukraine *(December, 2021, The New York Times),* (negative semantics – intensity of growing conflict); ***invasion fears grow*** *(2021, 9News),* (negative semantics – intensity of growing conflict); an ***imminent invasion of Ukraine*** *(December, 2021, Express)*, (negative semantics – intensity of growing conflict); to ***invade Ukraine*** *(December, 2021, Express),* (negative semantics – precedence of russian-Ukrainian war); *deter Putin from* ***launching an invasion into Ukraine*** *(December, 2021, CNN)*, (positive semantics – a call for peace); *Western leaders urge Russia to* ***cool Ukraine Tensions*** *(December, 2021, Independent),* (positive semantics – a call for peace); *deterring Russian* ***invasion of Ukraine*** *(December, 2021, FoxNews),* (positive semantics – a call for peace); *We're told* ***russia will invade Ukraine****, China will attack Taiwan (December, 2021, The Sun)*, (negative semantics – inevitable prediction).

In the previous discourse theme, the words in the statements were not omitted as there were only four of them. Here (1379 corpus concordance statements) and in the following themes a shortage of statements due to their size and intention to demonstrate a variety of collocations was considered. This theme is characterized by frequency of such lexical patterns as *threat, tension, escalation, aggression, weapon, military, action, attack, fears, invasion, crisis.* They all have a negative association when talking about precedence to russian-Ukrainian war. By applying textual semantic analysis to the above highlighted corpus collocations individual cases of positive semantics occurred though it did not lead to peace. A variety of synonyms and synonymic collocations to the war precedence like *unrest, conflict, aggression, escalation, aggressive action, military action, invasion, tension, military threat, military fear* was considered, which only intensifies the calamity. The most frequent lexical pattern of *invasion, invade* signals about negative semantic prosody of the discourse theme of 2021 (Fig.14).

Traditional LIWC Dimension	Your Text	Average for Social Media Language
I-words (I, me, my)	0.00	5.44
Positive Tone	2.21	5.93
Negative Tone	2.34	2.34

Figure 11: Proving negative semantic prosody of the discourse theme 2 of the year 2021 applying *Linguistic Inquiry and Word Count tool*

Discourse theme 3 of the year 2022 is demonstrated by division into months:

January, 2022: ***russia against Ukraine*** *as the sort of thing that would be considered an act of war in the U.S. (January, 2022, The Washington Post),* (negative semantics – hostility narrative between neighboring countries); *if russia further* ***invade Ukraine*** *(January, 2022, Voyager),* (negative semantics – conditions under which war can start); *allies will 'respond decisively' if Russia* ***invades Ukraine*** *(January, 2022, The Washington Post),* (negative semantics – conditions under which war can start); *100,000 troops near the border, made* ***further moves against Ukraine*** *(January, 2022, The Telegraph),* (negative semantics –

readiness to combat against our country); *intent on* ***invading Ukraine*** *(January, 2022, The Telegraph),* (negative semantics – readiness to combat against our country); *the* ***Ukraine crisis*** *would lead to a "complete rupture of relations." (January, 2022, The Telegraph),* (negative semantics – conditions under which war can start); *massing tens of thousands of* ***troops along its border with Ukraine*** *(January, 2022, The Guardian),* (negative semantics – readiness to combat against our country); *If Russia* ***invades Ukraine*** *and* ***incorporates Ukrainian territory*** *into the russian state (January, 2022, CBSnews),* (negative semantics – readiness to combat against our country); *The russians are coming as close as they can to* ***surrounding eastern Ukraine*** *(January, 2022, CBSnews),* (negative semantics – readiness to combat against our country); *he may well* ***invade Ukraine*** *if he does not get what he wants (January, 2022, CBSnews),* (negative semantics – conditions under which war can start); *putin is* ***putting pressure on Ukraine*** *and the West to act more in his interests (January, 2022, CBSnews),* (negative semantics – readiness to combat against our country); *russian* ***aggression against Ukraine*** *will be met with "serious consequences," (January, 2022, CBSnews),* (negative semantics – readiness to combat against our country); *any new russian* ***incursion into Ukraine*** *would be met with harsh, intensified and extraordinary economic sancti (January, 2022, CBSnews),* (negative semantics – readiness to combat against our country);

February, 2022: ***tensions between russia and Ukraine*** *reached new heights (February, 2022, Express),* (negative semantics – a signal for military affair); *described an* ***ongoing "tug of war" in Ukraine*** *between russia and the West (February, 2022, Independent,* (negative semantics – readiness to strike war against our country); *has ordered russian troops to* ***invade Kremlin-backed areas of Ukraine*** *on what he calls a "peacekeeping mission." (February, 2022, Buzzfeednews),* (negative semantics – war started); *as putin* ***massed forces around Ukraine*** *(February, 2022, Buzzfeednews),* (negative semantics – war started); *russia's* ***illegal invasion of Ukraine*** *(February, 2022, Inews),* (negative semantics – was as a genocide started); *"feel[s] like a* ***war's begun" between Ukraine and russia*** *(February, 2022, FoxNews),* (negative semantics – war started); *the* ***sheer scale of the militarisation of the border in Ukraine*** *is something that we would have thought a relic of the past. (February, 2022, Independent),* (negative semantics – war started);

March, 2022: *show* ***support for Ukraine*** *(March, 2022, Newslanes),* (positive semantics – in search for humanitarian aid); *russia should* ***stop its bombardment of Ukraine*** *(March, 2022, CTVnews),* (positive semantics – attempts of foreign partners to call for peace); *"all foreigners* ***willing to defend Ukraine*** *and world order" (March, 2022, The Conversation),* (positive semantics – attempts of foreign partners to call for peace);

April, 2022: *the Antonov 'Mriya' was* ***destroyed in Ukraine*** *(April, 2022, Daily Telegraph),* (negative semantics – evidence of war ruins); *Almost* ***300,000 refugees from Ukraine*** *have been registered in Germany as of Friday (April, 2022, Daily Telegraph),* (negative semantics – harsh war consequences, positive semantics – support of foreign partners); *show their* ***solidarity with Ukraine*** *(April, 2022, Daily Telegraph),* (positive semantics – support of foreign partners); ***losses*** *related to the* ***war in Ukraine*** *(April, 2022, The New York Times),* (negative semantics – harsh war consequences);

May, 2022: ***'nightmare situation'*** *during visit to* ***Ukraine refugee centre*** *(May, 2022, The Standard), (*negative semantics – harsh war consequences, positive semantics – support of foreign partners);

June, 2022: *announced another $800million of* ***additional weapons aid to Ukraine*** *(June, 2022, Mail Online),* (positive semantics – in search for humanitarian aid);

July, 2022: ***volunteers*** *are crucial in* ***helping Ukraine to win the war*** *(July, 2022, Inews),* (positive semantics – manifestation of assistance);

August, 2022: *deliver* ***humanitarian aid to Ukraine*** *(August, 2022, The Baltic Times,* (positive semantics – in search for humanitarian aid);

September, 2022: *vladimir putin (right) wants to* ***end the war in Ukraine*** *'as soon as possible (September, 2022, Mail Online),* (positive semantics of waiting for peace, vain hopes though);

October, 2022: ***'massive' attack on military targets and energy infrastructure across Ukrain****e using high-precision weapons (October, 2022, Mail Online),* (negative semantics – harsh war consequences);

November, 2022: *increase* ***solidarity*** *and maximise* ***support*** *for* ***Ukraine*** *(November, 2022, The Telegraph),* (positive semantics – support of foreign partners);

December, 2022: *those* ***fleeing the conflict in Ukraine*** *could come to the UK for up to three years (December, 2022, The Standard),* (negative semantics – harsh war consequences, aid of foreign partners – positive semantics).

The year of 2022 was illustrated with 45 564 corpus statements highlighting the tragedy of Ukraine, and semantic prosody indicates tears, deaths, victims, ruins – total Ukrainian genocide. There are lexical patterns of *invade, invasion, invaded, will invade, has invaded* signaling the semantic divergence between the previous discourse theme and this one: 2021 was considered a precedence to possible combat actions where *invasion* was only a hypothesis, however, in January of 2022 there was a real threat of an attack by changing direction sharply into real war. Here, semantic prosody is more negative than in both previous discourse themes (Fig.12). Positive semantics of help, shelter and support of foreign partners was revealed.

Traditional LIWC Dimension	Your Text	Average for Social Media Language
I-words (I, me, my)	0.14	5.44
Positive Tone	1.78	5.93
Negative Tone	4.38	2.34

Figure 12: Proving negative semantic prosody of the discourse theme 3 of the year 2022 applying Linguistic Inquiry and Word Count tool

Discourse theme 4 of the year 2023 is illustrated by division into months:

January, 2023: *significant majorities came together in* ***support of Ukraine*** *despite the best efforts of some (January, 2023, CNN),* (positive semantics – support of foreign partners); *the*

ongoing war in Ukraine *captured the headlines (January, 2023, The Star Phoenix),* (negative semantics – war growing into a bitter one);

February, 2023: *enabled and supported this* ***further invasion of Ukraine*** *(February, 2023, CBC),* (negative semantics – war growing into a bitter one); *the need to continue to* ***support and defend Ukraine*** *(February, 2023, Independent),* (positive semantics – support of foreign partners); *shifting towards* ***providing F-16s to Ukraine*** *(February, 2023, Inews),* (positive semantics – support of foreign partners); *As the* ***war rages on in Ukraine*** *(February, 2023, The Star)*, (negative semantics – war growing into a bitter one);

March, 2023: *ending with* ***no consensus on the Ukraine*** *(March, 2023, CTVnews),* (negative semantics – war growing into a bitter one); *war. monitoring railroad routes used for the* ***transport of weapons into Ukraine*** *(March, 2023, Express),* (positive semantics – support of foreign partners);

April, 2023: *Volodymyr Zelenskyy* ***boosts Ukraine ties*** *with Poland (April, 2023, Euronews),* (positive semantics – support of foreign partners); *US has* ***supplied so much to Ukraine*** *(April, 2023, Express),* (positive semantics – support of foreign partners); ***on-going invasion of Ukraine****(April, 2023, Mail Online),* (negative semantics – war growing into a bitter one);

May, 2023: *the* ***longest and bloodiest battle of the war in Ukraine*** *(May, 2023, Independent),* (negative semantics – war growing into a bitter one);

June, 2023: *but for lasting peace in Europe,* ***Ukraine must be liberated*** *when they ascend to full membership (June, 2023, Express),* (positive semantics – support of foreign partners); *actively considering* ***sending cluster munitions to Ukraine*** *to help Kyiv's counteroffensive punch through Russia's defenses (June, 2023, Politico),* (positive semantics – support of foreign partners);

July, 2023: *the commanders, hailed as* ***heroes in Ukraine***, ***led*** *last year's* ***defence*** *(July, 2023, The Guardian),* (positive semantics – feeling of being proud for Ukraine); ***cluster munitions*** *will be a* ***'game changer' in Ukraine*** *(July, 2023, CNN),* (positive semantics – support of foreign partners);

August, 2023: *saying that* ***supporting Ukraine*** *is not charity, but an* ***investment****. (August, 2023, Politico),* (positive semantics – support of foreign partners, however, with controversy); *the union needs to* ***convince Ukraine to hold negotiations*** *with russia, (August, 2023, CNN),* (positive semantics – a call for peace); ***strikes against Ukrainian ports*** *and threatened to* ***hand Ukraine "an ecological catastrophe"****(August, 2023, The Guardian),* (negative semantics – war growing into a bitter one);

September, 2023: *the numbers of* ***mines in the battlefield Ukraine is encountering*** *are at a historic high (September, 2023, Independent),* (negative semantics – war growing into a bitter one);

October, 2023: *He said that* ***Ukraine would prevail against*** *russia (October, 2023, The New York Times),* (positive semantics – winning prediction); ***help her community in Ukraine preserve and revive*** *its cultural identity (October, 2023, Global Issues),* (positive semantics – revitalization of Ukrainian culture); *Both are* ***helped by a charity, Save Ukraine****, which aims to reunite families (October, 2023, The Guardian),* (positive semantics – support of foreign partners);

November, 2023: *U.N. resolutions calling for russia to* ***withdraw from Ukraine*** *(November, 2023, Time),* (positive semantics – a call for peace); *they hope to* ***tie support for Ukraine*** *and Israel together. (November, 2023, The Independent),* (positive semantics – support of foreign partners);

December, 2023: ***soaring deaths in his Ukraine war*** *(December, 2023, The Independent),* (negative semantics – war growing into a bitter one); *"russia must lose and* ***Ukraine must win*** *(December, 2023, Euronews),* (positive semantics – winning prediction); ***attacks continued across Ukraine*** *(December, 2023, Metro),* (negative semantics – war growing into a bitter one).

This discourse theme is characterized by decreasing the amount of corpus statements – 16857, the textual semantic analysis of which allowed to dwell upon the positive semantic prosody (Fig.13) mainly represented by such lexical patterns as *support, defend, provide, transport, boost, supply, liberate, send, prevail, help, win.* They all were characterized by changing their forms but only in present times, which can be linguistically interpreted as a willingness to overcome last war torments and hope for support and inevitable victory. The usage of collocation *ongoing war* intensifies the tedious process of military support. Calls for peace and predictions concerning Ukrainian victory never stop notwithstanding the fact that amount of deaths and ruins do not stop either.

Traditional LIWC Dimension	Your Text	Average for Social Media Language
I-words (I, me, my)	0.00	5.44
Positive Tone	3.56	5.93
Negative Tone	3.01	2.34

Figure 13: Proving positive semantic prosody of the discourse theme 4 of the year 2023 applying *Linguistic Inquiry and Word Count tool*

Discourse theme 5 of the year 2024: *relentlessly reporting on president* ***vladimir putin's atrocities in Ukraine*** *and the escalating repression in russia (January, 2024, Politico),* (negative semantics – war growing into a bitter one); *Canada will be making a $650 million 'multi-year commitment' for* ***further Ukraine Aid*** *(January, 2024, CTVnews),* (positive semantics – support of foreign partners); *There was talk that under no circumstances* ***would Ukraine be given combat aircraft*** *(January, 2024, The Baltic Times),* (negative semantics – delaying help distracts from victory); *have* ***promised to donate a significant number of fighter jets to Ukraine*** *(January, 2024, The Baltic Times),* (positive semantics – support of foreign partners); *russia started its* ***current war with Ukraine*** *(January, 2024, Foxnews),* (negative semantics – war growing into a bitter one); ***Ukraine*** *and russia* ***announce*** *largest* ***prisoner swap*** *since start of war (January, 2024, The Guardian),* (positive semantics in terms of war achievements); *he seeks* ***glory by conquering Ukraine*** *(January, 2024, Politico),* (negative semantics – superiority syndrome); *the russian war of* ***aggression against Ukraine*** *is also a "war against culture". (January, 2024, Deutschland),* (negative semantics – hindering Ukrainian culture revitalization); *"The* ***war in Ukraine*** *is not over - quite the*

contrary, unfortunately (January, 2024, Deutschland), (negative semantics – war growing into a bitter one); ***Ukraine*** *was now* ***suffering from*** *a* ***deficit of air defence missiles*** *(January, 2024, Mirror),* (negative semantics – a lack of armament); *an attempt to* ***disencourage*** *the west* ***from supplying Ukraine with weapons and resources*** *(January, 2024, Mirror),* (negative semantics – political and manipulative intervene tactics); *the biggest issue in her mailbox –* ***beating the war in Ukraine*** *or immigration (January, 2024, The Guardian),* (negative semantics – political and manipulative intervene tactics); *the* ***Ukraine peace talks*** *aimed to finalise principles "for a lasting and just peace (January, 2024, The Guardian),* (positive semantics – a call for peace); *Denmark will allocate a* ***new aid package to Ukraine*** *in the amount of more than 21 million dollars for the restoration (January, 2024, The Guardian),* (positive semantics – support of foreign partners); *Sweden jointly announced a new package of* ***military assistance for Ukraine*** *worth about 240 million euros, which included infantry fighting vehicles (January, 2024, The Guardian),* (positive semantics – support of foreign partners); ***cruelty of war*** *in so many parts of the world,* ***especially in Ukraine*** *(January, 2024, OSVnews),* (negative semantics – accent on war in Ukraine); *the situation generally is "****not looking good for Ukraine****", as it deals with shortages of ammunition, low morale among its troop (January, 2024, BBC),* (negative semantics – war growing into a bitter one); ***hamper Western support for Ukraine*** *and also increase the risk of a trade war (January, 2024, Politico),* (negative semantics – political and manipulative intervene tactics); *russia's* ***continued aggression towards Ukraine*** *(January, 2024, The Week),* (negative semantics – war growing into a bitter one); *EU leaders* ***exhort allies to do more for Ukraine*** *as ministers debate ways to fill its ammunition gap (January, 2024, The Seattle Times),* (positive semantics – a call for support of Ukraine).

As the year has just started, obviously, there will be still a room for further research but in comparison to the theme of the whole year 2021, there are 1352 corpus statements for 1 month – January – where Ukrainian position and perception may be linguistically and semantically discovered. The most debated topic remains support for Ukraine and now *continued aggression* was added to the synonyms to *ongoing war* and *current war*. Unlike the previous themes, last discourse theme of the year 2024 uncovered political and manipulative tactics of withholding military aid for Ukraine in order to lose this battle. Here, it is similar to the previous theme in terms of a call for peace and more assistance. Considering the military situation, semantic prosody is essentially negative (Fig.14) involving attacks, a lack of necessary armament and financial support.

Traditional LIWC Dimension	Your Text	Average for Social Media Language
I-words (I, me, my)	0.00	5.44
Positive Tone	2.08	5.93
Negative Tone	5.34	2.34

Figure 14: Proving negative semantic prosody of the discourse theme 5 of the year 2024 applying Linguistic Inquiry and Word Count tool

For comparing both semantic prosody of the discourse themes and its statistical data, Table 2 is provided.

Table 2
Frequency values and keyness score of the target and reference corpora (only keywords referring to Ukrainian current military state)

Discourse themes	Semantic prosody	Positive statistics	Negative statistics
1 of the year 2020	neutral*	1.75	1.75
2 of the year 2021	Negative	2.21	2.34
3 of the year 2022	negative	1.78	4.38
4 of the year 2023	positive	3.56	3.01
5 of the year 2024	negative	2.08	5.34
Total of all 5 themes	negative	11.38	16.82

neutral* - Iinguistically and semantically it was considered negative, with digital tools – neutral

Each discourse themes are permeated with Ukraine. Two years before war negative semantic prosody was only slowly growing either to recall the eternal conflict or to overcome the opponent. The year of the war 2022 and two years afterwards are characterized with bitter military actions and strive for sovereignty. To briefly summarize the complex linguistic analysis and statistically compare the semantic prosody of the discourse themes, it is worth noting that Ukraine within the English-language media discourse has been greatly debated – it was covered in the news more as the victim of the aggression with all war precedence and current consequences. Imposition of martial law on Ukraine has led to its perception as of military state, and total negative semantic prosody proved it (16.82%). However, the key principle of Ukrainian position is to survive and prosper after the victory. This positivity (11.38%) has been achieved by those heroes who defend our land, foreign partners' support and world justice.

5. Discussions

Proliferation of information technologies and spread of communicative means has led to distribution of large arrays of data which produces a strong impact on the range of utterances related to a person in particular situations. Since Ukraine has become the key indicator of concern worldwide, there is a wide range of studies referring to the Ukrainian position in the international arena [41, 42, 5]. The most cases of research are based on the news media coverage of Ukraine, however, discovering either European views only or political sphere and national security issues. The main function of media discourse in this paper is to discover what it reveals about the Ukrainian community but from the perspective of English-speaking foreigners, which was possible to trace in *English Trends*

corpus. Unlike the previous studies based on content analysis of European journals and magazines mentioned above, this research is crucial in shaping the social perception of Ukraine in the international arena extending to American, Canadian and Australian news articles and reports. However, this target corpus unlike those from the previous studies has one drawback in terms of illustrating the linguistic data for the second time – it is constantly updating and statements within the corpus are increasing in chronological order thus making it complicated to provide link to the data analyzed.

Not so many studies have been recently conducted incorporating semantic (discourse) prosody [43, 44], which emphasizes the significance to mark it in this study revitalizing corpus linguistics key terms by linguistically interpreting and adjusting them to this paper. Here, the approach to study semantic prosody presupposed allocation of discourse themes which fostered the process of linguistic interpretation of the Ukrainian position, whereas the previous studies were not aimed at singling out common themes for certain periods of time, only a heap of persistent data. Also, in this paper there is an attempt to maximally structure both linguistic and statistical data revealed.

There are two ways of applying corpus-based approach tools: the first way is the so-called management corpus guide enabling insights into key terms and sequence of corpora usage, which can be seen in [44]; the second way allows a scientist to experience his/her own corpus managing process navigating through the digital environment either setting a query and obtaining results in charts, diagrams, tables and pictures or conducting linguistic analysis through analysis of corpus concordance linguistic context, which can be traced in [43]. This paper is aimed at implementing both ways of applying corpus-based approach as all sequent research steps have been followed by the *Sketch Engine* tool guide and personal research corpus query settings. It is noteworthy that a choice of studying representation Ukraine by linguistic means has been substantiated not only by the increased attention to Ukraine in linguistic studies but also by applying corpora keywords analysis according to which Ukraine has proved to be the most frequent word within the corpora news articles genre of the English-language media.

All in all, such author's vision to linguistic analysis of media discourse plays a central role in advancing our understanding of language structure, use, and function within contemporary society, offering valuable insights into the complex interplay between language, communication and linguistic context. That is why, linguistic analysis in terms of media discourse is essential when investigating the object of the research. Analyzing the utterances-statements within the media discourse can help detect conflict precedence, its manifestations influencing the formation of linguistic and social vision of Ukraine.

6. Conclusions

To conclude, Ukraine was linguistically represented within the English-language media discourse as a state which is struggling and does not surrender. Corpus-assisted approach to the research allowed to identify the word *Ukraine* and its lexical patterns (*Ukrainian, Ukrainians, Ukrainian's*) among the rest of the corpora keywords (in terms of target corpus *English Trends* as of genre of news articles and reference corpus *English Web 2021*) comprising keywords analysis along with allocating collocations of the word *Ukraine*.

Statistical data representation of frequency values towards keywords and collocations as well as keywords keyness score was provided and linguistically interpreted.

Theoretical insights into media discourse, corpus linguistics, linguistic interpretation and context along with author's own interpretation of the linguistic key terms were targeted at enhancing the study's comprehension. Corpus-based discoveries helped to advance in disclosing discourse themes with reference to textual semantic analysis, they were divided according to their chronology within the target corpus: *discourse theme of the year 2020, discourse theme of the year 2021, discourse theme of the year 2022, discourse theme of the year 2023, discourse theme of the year 2024* with their appropriate division into months. After a comprehensive linguistic analysis it can be stated that first two discourse themes have a topic in common – precedence to russian-Ukrainian war with semantics of hope and foreign calls not to intervene in Ukraine; the year of 2022 – the year of imposing martial law and war; and next two discourse themes – growing military concerns till now. With the emergence of political topics regarding Ukraine over the years such collocations concerning perception of Ukraine were chronologically subjected to the following lexical interchange: *no intention to invade, a call for deescalate, a call for not to attack* (2020), *cyber-attacks, growing conflict*, *feud with Ukraine* (2021), *a move on Ukraine, incursion into Ukraine, intervention, assault on Ukraine, bombardments* (2022). For the last two years lexical patterns underwent a move from *current conflict, ongoing war* to *continued aggression*. So, both lexically and semantically military "abscess" in Ukraine was considered. Negative semantic prosody permeates the representation of Ukraine in news coverage, mainly in American, British, Australian and Canadian reports, however, by 5.44% lower than negative prosody, positivity was expressed by a great amount of direct speech in utterances predicting success of Ukraine and inspiring it with hope for a bright future along with the support of partners.

The complex linguistic analysis was accompanied by the appropriate software package tools. *SketchEngine* was used for introducing corpus-assisted approach to the study and depicting the linguistic data statistically. For linguistic data mining, corpus representation, data analytics, word frequency distribution and visual representation of the object under investigation *Voyant* digital tool was applied. Such sentiment analyzers as *MonkeyLearn* and *Linguistic Inquiry and Word Count* demonstrated their functionality in verification process of semantic prosody of the discourse themes of the study and statistical results.

It would be valuable to linguistically analyze linguistic corpora of various genres and conduct comparative analysis of representation of Ukraine both in international and domestic arena.

References

[1] A. Schulz, W. Wirth, P. Müller, We Are the People and You Are Fake News: A Social Identity Approach to Populist Citizens' False Consensus and Hostile Media Perceptions, Communication Research 47(2) (2020) 201-226. doi:10.1177/0093650218794854

[2] Sh. Gearhart, A. Moe, B. Zhang, Hostile media bias on social media: Testing the effect of user comments on perceptions of news bias and credibility, Human Behavior and Emerging Technologies 2(2) (2020) 140-148. doi: 10.1002/hbe2.185

[3] J. Li, M.-H. Su, Real Talk About Fake News: Identity Language and Disconnected Networks of the US Public's "Fake News" Discourse on Twitter, Social Media + Society 6(2) (2020). doi:10.1177/2056305120916841
[4] N. Mede, M. Schäfer, J. Metag, Cognitio populi – Vox populi: Implications of science-related populism for communication behavior, Communications (2023). doi: https://doi.org/10.1515/commun-2022-0059
[5] O. H. Tkachenko, O. P. Sydorenko, I. O. Melnikova, N. R. Myronets, A. H. Smus, The media image of Ukraine: An European view, Linguistics and Culture Review, 5(S4) (2021) 203-217. doi: 10.21744/lingcure.v5nS4.1574
[6] Zdebskyi, V. Lytvyn, Y. Burov, Z. Rybchak, P. Kravets, O. Lozynska, R. Holoshchuk, S. Kubinska, A. Dmytriv, Intelligent system for semantically similar sentences identification and generation basedon machine learning methods, in: Proceedings of the 4th. International conference on computational linguistics and intelligent systems, 2020, pp. 317-346.
[7] V. Vysotska, A. Sachenko, V. Lytvyn, Y. Burov, Information resources processing using linguistic analysis of textual content, in: Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2017, pp. 573–578.
[8] I. Pilán, P. Lison, L. Øvrelid, A. Papadopoulou, D. Sánchez, M. Batet, The Text Anonymization Benchmark (TAB): A Dedicated Corpus and Evaluation Framework for Text Anonymization, Computational Linguistics 48 (4) (2022) 1053–1101 doi: 10.1162/coli_a_00458
[9] Y. Hathaway, They Made us into a Race. We Made Ourselves into a People: A Corpus Study of Contemporary Black American Group Identity in the Non-Fictional Writings of Ta-Nehisi Coates, Corpus Pragmatics 5 (2021) 313–334. doi: 10.1007/s41701-021-00101-8
[10] M. Leedham, T. Lillis, A. Twiner, Exploring the core 'preoccupation' of social work writing: A corpus-assisted discourse study, Journal of Corpora and Discourse Studies 3 (2022) 1-26. doi: 10.18573/jcads.26
[11] O. Levchenko, N. Kunanets, A. Hadzalo, The Application of AntConc Concordancer in Linguistic Researches, in: Proceedings of the IEEE 13th International Scientific and Technical Conference on Computer Sciencesand Information Technologies, 2018, pp. 144-147.
[12] I. Khomytska, I. Bazylevych, V. Teslyuk, I. Karamysheva, The chi-square test and data clustering combined for author identification, in: Proceedings of IEEE XVIIIth. Scientific and Technical Conference, 2023.
[13] N. Cheilytko, R. von Waldenfels, Exploring Word Sense Distribution in Ukrainian with a Semantic Vector Space Model, in: Proceedings of the 2nd Ukrainian Natural Language Processing Workshop (UNLP), 2023, pp. 73–78.
[14] D. Hauser, N. Schwarz, Semantic Prosody: How Neutral Words with Collocational Positivity/Negativity Color Evaluative Judgments, Current Directions in Psychological Science 32(2) (2023) 98-104. doi: 10.1177/09637214221127978
[15] W. Liu, A Corpus-Based Study of Semantic Prosody and Semantic Preference of "Seem", Open Journal of Social Sciences 8 (2020) 282-291. doi: 10.4236/jss.2020.85020.
[16] B. Waldon, Linguistic interpretation as inference under argument system uncertainty: the case of epistemic must, in: Proceedings of the Probability and Meaning Conference (PaM 2020), 2020, pp. 34–40.

[17] D. Mazzarelia, "I didn't mean to suggest anything like that!": Deniability and Context Reconstruction, Mind and Language 38 (1) (2023) 218-236. doi: 10.1111/mila.12377
[18] E. Dahlborg, Å. Boman, H. Eriksson, E. Tengelin, Encircling discourses – A guide to critical discourse analysis in caring science, Scandinavian Journal of Caring Sciences 38 (1) (2024) 177-184. doi: 10.1111/scs.13194
[19] Dilafruz, Xabibiloevna "Discourse and Discourse Analyses in Linguistics." Best Journal of Innovation in Science, Research and Development (2023): 31–36.
[20] G. Finch, The Linguistic Context, in: How to Study Linguistics, Palgrave Study Guides: Literature, Palgrave, London, 2023, pp. 12-45. doi: 10.1007/978-0-230-80213-1_2
[21] L. Song, The Role of Context in Discourse Analysis, Journal of Language Teaching and Research 6(1) 2010 876-879. doi: 10.4304/jltr.1.6.876-879
[22] A. Mosheer Amer, Israeli Onslaught on Gaza: A Critical Discourse Analysis, International Journal of English Linguistics 12(5) (2022). doi:10.5539/ijel.v12n5p1
[23] S. Bhowmik, J. Fisher, Framing the Israel-Palestine conflict 2021: Investigation of CNN's coverage from a peace journalism perspective, Media, Culture & Society 45(5) (2023) 1019-1035. doi: 10.1177/01634437231154766
[24] Marta, Strukowska. "Front-Page Coverage of the First-Day Russian Invasion on Ukraine in 2022. The Ethnographic Discourse Analysis Based on the Polish Press Reportage." Academic Journal of Modern Philology 18 (2022): 157-175.
[25] A. Alyahya, Critical Discourse Analysis of Newspaper Articles Declaring the Outbreak of War in Ukraine: The Washington Post vs. The Moscow Times, International Journal of Society, Culture & Language (Themed Issue on Language, Discourse, and Society) 11(2) (2023) 47-59. doi: 10.22034/ijscl.2023.1982871.2969
[26] L. Pavlichenko, Polarization in Media Political Discourse on the War in Ukraine: Critical Discourse Analysis, Visnyk Universytetu imeni Alfreda Nobelia. Seriya: Filologichni Nauki, 24 (2) (2022) 241-223. doi: 10.32342/2523-4463-2022-2-24-18
[27] C. Ononiwu, Ideology and cognitive stereotypes in media representation of the Russia–Ukraine conflict, Media, War & Conflict (2023). doi: 10.1177/17506352231201743
[28] K. O'Halloran, How to use corpus linguistics in the study of media discourse, in: A. O'Keeffe, M. McCarthy (Ed.), The Routledge Handbook of Corpus Linguistics, 2010, pp.563–577.
[29] D. Gablasova, V. Brezina, T. McEnery, Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence, A Journal of Research in Language Studies 67(S1) (2017) 155-179. doi: 10.1111/lang.12225
[30] P. Rebuschat, D. Meurers, T. McEnery, Language Learning Research at the Intersection of Experimental, Computational, and Corpus-Based Approaches, A Journal of Research in Language Studies 67(S1) (2017) 6-13. doi: 10.1111/lang.12243
[31] J. Gao, Q. Pham, Ch. Polio, The role of theory in quantitative and qualitative second language learning research: A corpus-based analysis, Research Methods in Applied Linguistics, 1(2) (2022). doi: 10.1016/j.rmal.2022.100006
[32] Das, S., Ghosh, S., Kolya, Unparalleled sarcasm: a framework of parallel deep LSTMs with cross activation functions towards detection and generation of sarcastic statements. Lang Resources & Evaluation 57 (2023). doi: 10.1007/s10579-022-09622-3
[33] A. Izes, Fact Versus Opinion in US Defamation Law: A Corpus and Appraisal Analysis of Speaker Stance Toward Reputational Harm, International Journal for the Semiotics of Law 36 (2023) 1185–1216 doi: 10.1007/s11196-023-09981-2

[34] S. Sonbul, El-D. DAS, K. Conklin, G. Carrol, “Bread and butter” or “butter and bread”? Nonnatives’ processing of novel lexical patterns in context, Studies in Second Language Acquisition 45(2) (2023) 370-392. doi:10.1017/S0272263122000237
[35] Runze, Xu, Raksangob, Wijitsopon. “Corpus Linguistics and Cinematic Discourse: Lexical Bundles in Mainstream Film Scripts.” LEARN Journal: Language Education and Acquisition Research Network 16(1) (2023): 545–574.
[36] C. Pérez, The representation of migration in parliamentary settings: critical cross-linguistics corpus-assisted discourse analyses, Humanities and Social Sciences Communications 965 (10) (2023) doi: 10.1057/s41599-023-02496-y
[37] M. Naldi, S. Petroni, A Testset-Based Method to Analyse the Negation-Detection Performance of Lexicon-Based Sentiment Analysis Tools, Computers 12(1) (2023) 18. doi: 10.3390/computers12010018
[38] Y. Altameemi, M. Altamimi, Thematic Analysis: A Corpus-Based Method for Understanding Themes/Topics of a Corpus through a Classification Process Using Long Short-Term Memory (LSTM), Applied Sciences 13(5) (2023) 3308 doi: 10.3390/app13053308Y
[39] James, Pustejovsky, Sabine, Bergler, Peter, Anick. “Lexical Semantic Techniques for Corpus Analysis.” Computational Linguistics 1993: 331-358.
[40] S. Albota, Linguistic and semantic representation of the WAR concept by phraseological units, in: Proceedings of the 2nd International workshop on social communication and information activity in digital humanities, 2023, pp. 355–364.
[41] Hassan, Majhool Al-Hasnawi. “The Russian-Ukrainian war and its implications for international security: a geopolitical vision.” Texas Journal of Multidisciplinary Studies 9 (2022): 177–192.
[42] G. Tonlorenzi. The Ukrainian Russian Conflict Between the International Liberal System and International Realist System, in: C. Peña, J. Ramírez (Eds.), Threats to Peace and International Security: Asia versus West. Advanced Sciences and Technologies for Security Applications, Springer, Cham, 2023, pp. 62–68. doi: 10.1007/978-3-031-28336-9_4
[43] B. Malory Polarized Discourses of Abortion in English: A Corpus-based Study of Semantic Prosody and Discursive Salience, Applied Linguistics. doi: 10.1093/applin/amad042
[44] J. Sarhad, R. Mahmood, A Corpus-based Study of Semantic Prosody across a Native Corpus, Journal of Garmian University 10(3) (2023) 902-909. doi: 10.24271/garmian.2023.10378