Topic Modelling and Emotion Analysis of the Tweets of British and American Politicians on the Topic of War in Ukraine

Olena Karpina [a, *], Justin Chen [b]

[a] Lesya Ukrainka Volyn National University, Ukraine,
[b] Milton Academy, USA

Received August 26, 2022; Revised December 15, 2022; Accepted December 24, 2022

Abstract. This paper focuses on the content and emotive features of four politicians' posts that were published on their official Twitter accounts during the three-month period of the russian invasion of Ukraine. We selected two British politicians – Boris Johnson, the Prime Minister of the UK, and Yvette Cooper, the Labour MP and Shadow Home Secretary of the State for the Home Department – as well as two American politicians, President of the USA Joe Biden and Republican senator Marco Rubio. In the first phase, we constructed a dataset containing the tweets of the four politicians, which were selected with regard to the topic of war in Ukraine. To be considered approved, the tweets were supposed to contain such words as *Ukraine, russia, war, putin, invasion,* spotted in one context. In the second phase, we identified the most frequent lexical tokens used by the politicians to inform the world community about the war in Ukraine. For this purpose, we used Voyant Tools, a web-based application for text analysis. These tokens were divided into three groups according to the level of their frequency into most frequent, second most frequent and third most frequent lexical tokens. Additionally, we measured the distribution of the most frequent lexical tokens across the three-month time span to explore how their frequency fluctuated over the study period. In the third phase, we analysed the context of the identified lexical tokens, thereby outlining the subject of the tweets. To do this, we extracted collocations using the Natural Language Toolkit (NLTK) library. During the final phase of the research, we performed topic modelling using the Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model (GSDMM) and emotion analysis using the NRC Lexicon library.

Keywords: *lexical token, raw frequency, relative frequency, virtual discourse, topic modelling, emotion analysis, Twitter.*

Карпіна Олена, Чен Джастін. Тематичне моделювання й аналіз емоцій у твіт-повідомленнях британських та американських політиків на тему війни в Україні.

Анотація. Статтю присвячено дослідженню змістових та емотивних особливостей дописів чотирьох політиків, опублікованих в їхніх офіційних Твіттер акаунтах протягом трьох місяців російського вторгнення в Україну. Ми обрали двох британських політиків – прем'єр-міністра Великої Британії Бориса Джонсона та Іветт Купер, членкиню Палати громад від Лейбористської партії, тіньову міністерку внутрішніх справ Великої Британії – і двох американських політиків, – президента Джо Байдена та сенатора-республіканця Марко Рубіо. На першому етапі ми створили текстовий масив, що містить твіт-повідомлення

* Corresponding author. Olena Karpina, https://orcid.org/0000-0001-9520-074X, E-mail: karpina@vnu.edu.ua

East European Journal of Psycholinguistics, 9(2), 41–66. https://doi.org/10.29038/eejpl.2022.9.2.kar

чотирьох політиків, які було відібрано за тематикою війни в Україні. Для того, щоб вважатися схваленими, твіт-повідомлення повинні були містити такі слова, як Україна, росія, війна, путін, вторгнення, виявлені в одному контексті. На другому етапі ми визначили найчастотніші лексичні одиниці, що їх уживають політики для інформування світової спільноти про війну в Україні. Для цього ми скористалися програмою Voyant Tools, веб-додатком для текстового аналізу. Ці лексичні одиниці було розподілено на три групи за рівнем їхньої частотності: найчастотніші, другі за частотністю та треті за частотністю лексичні одиниці. Окрім того, ми виміряли розподіл найчастотніших лексичних одиниць у тримісячному часовому проміжку, щоб дослідити, як змінювалася їхня частотність протягом досліджуваного періоду. На третьому етапі ми проаналізували контекст ідентифікованих лексичних одиниць, які окреслюють тематику твіт-повідомлень у наборі даних. Для цього ми виявили колокації за допомогою бібліотеки NLTK. Під час останнього етапу дослідження ми виконали тематичне моделювання з використанням алгоритму вибірки Гіббса до багатоваріантного розподілу ймовірностей Діріхле та класифікацію емоцій за допомогою бібліотеки NRC Lexicon.

Ключові слова: *лексична одиниця, неопрацьована частотність, відносна частотність, віртуальний дискурс, тематичне моделювання, аналіз емоцій, Твіттер.*

Introduction

The discourse of social networks has repeatedly become an object of the research of scientists worldwide. Nerian (2018) investigated posts in social networks from the standpoint of linguistic pragmatics, interpreting them as a speech genre of Internet discourse that reveals the author's position in the public virtual space. Virtual communication in the social network Twitter (tweeting) was the subject of investigations by Goroshko (2011), who defines it as a genre of Internet communication and singles out its features: restricted length of messages, hashtags, and integration with other web services. Considering communicative strategies of social network users, Shvelidze (2021) outlines tweeting as a manifestation of solely English network discourse with strict genre characteristics. Poliakova (2021) focuses on the lexical aspect of political internet communication, assigning lexical means of tweeting to three vocabulary groups: neutral, politically marked, and emotionally expressive. The chapter in the work by the English researcher Crystal (2011), *Internet Linguistics: A Student Guide*, examines the Internet platform Twitter, where the author highlights the methodological, grammatical, structural-semantic, and pragmatic features of the study of tweets. Ukrainian linguist Nikolaieva (2019) studied vocabulary stratification in social media messages, characterizing social networks language as dynamic in terms of semantics and word formation.

The goal of our research was to perform lexical, semantic and emotional analysis of the tweets of English and American politicians in order to shape the general attitude of political elite of the USA and UK to the war in Ukraine, to disclose the topics of their major concern regarding the war, and to show the range of different emotions from one topic to another. The achievement of this goal became possible by securing the following objectives: to single out the most frequent lexical tokens in the datasets of every politician considered in the study; to define the collocations where these frequent lexical tokens occur; to perform the topic modelling based on the previously extracted collocations; to carry out the emotion analysis in terms of every topic. We chose two British politicians and two American politicians because both the U.S. and U.K. have had an impact on the Ukraine war without directly fighting in it. As a result of their involvement, we had a particular interest as to how these politicians described the war and their actions regarding it. We chose Boris Johnson, Prime Minister of the UK, and Yvette Cooper, the Labour MP and Shadow Home Secretary of the State for the Home Department, because of their important roles in the government and their membership in opposing political parties (Conservative Party and Labour Party, respectively). By examining two politicians from different political parties, we believed we would get two different opinions on the war and gain a more accurate understanding of the opinion of the country as a whole. We chose the U.S. President Joe Biden, a Democrat, and Senator Marco Rubio, a Republican, for the same reasons.

Method

Phase 1. Construction of the Dataset

At this phase, the construction of the dataset for further analysis is carried out. To achieve this objective, we skimmed the Twitter accounts of the four politicians and manually selected tweets about the war in Ukraine. We considered the posts relevant to the objectives of our study if they consisted of such linguistic tokens as *Ukraine, war, invasion, putin*[1]*, russia etc.*, occurring in one context. Initially, we created four separate word documents, each consisting of the validated tweets of one of the four politicians. As a result, we obtained a dataset with 20,084 total words.

Phase 2. Term Frequency and Distribution

To define the frequency of words, we made use of the online tool for text-mining *Voyant Tools*. For this purpose, we examined the dataset of each politician separately, singling out the 25 most frequent linguistic tokens of the tweets of each

[1] According to The Commission of Journalistic Ethics (CJE) the use of the stylistically coloured vocabulary, i.e proper names putin, russia etc. in lowercase, a tendency which appeared in journalistic texts after the full-scale invasion of Ukraine in February, 24, 2022, does not discriminate the ethnic group on the basis of national identity, being used with the reference to those people who identify themselves with the aggressor state supporting the policy of their leader and acting in accordance with it [www.imi.org.ua]. We share this tendency to lowercase all mentions of the country russia and its leader vladimir putin, considering it an act of consolidation of world antiwar forces.

politician. Then, the frequent words were divided into three groups. The first group contains the top five most frequent lexical tokens, used by every politician in the dataset. These tokens were compared to the five distinctive words in the dataset – high frequency words that were unique to a set of tweets of a particular politician compared to the sets of the other politicians in the whole dataset. We performed this process by uploading four datasets of tweets and automatically matching them against each other. The second and third group of lexical tokens, consisting of ten words each, come successively regarding the level of their frequency. We had to manually modify these lists, grouping the words with the same stem into one lexical token. Lexical tokens that represent such groupings are denoted with an asterisk (*) – for example, “military,” “militaristic,” and “militaries” would all be grouped under the lexical token “militar*.” We maintained this method of denoting words with the same stem throughout the paper.

Additionally, we differentiated between raw frequency – the actual number of occurrences of a term in a document – and relative frequency – the ratio of the frequency of a specific term to the frequency of all words in the given corpus. Raw frequency was calculated to classify the most frequent terms into three groups. In our opinion, the frequency of terms measured in the number of occurrences provides a more comprehensive view of the lexical use of the politicians. Relative frequency was necessary to outline the distribution of frequent terms in time, taking into consideration the fact that the number of tweets about the war posted every month was different within the period considered: for example, the tweets collected in May were considerably smaller in number than those collected in February-March. Consequently, showing the number of occurrences would be insufficient to measure the importance of a term in datasets different in size. To measure the peculiarities of the distribution of frequent terms, we divided each document containing the tweets of one particular politician into three parts, each of them referring to different months of the war (we combined February and March into one month, as the war had started at the end of February). Therefore, the first part consists of the tweets posted in February-March, the second part represents the posts from April, and the last one refers to May. As a result, the values of raw frequency differed from the values of relative frequency (see the tables in Results section).

In our Python-based collocation analysis, we first preprocessed the raw tweets to gain more significant collocations (eliminated meaningless collocations with numbers, emojis, or tokens like “the” or “a”). Using Python’s regular expression library and “demoji” library, we removed emojis, URLs, non-letter characters, and extra whitespace. We also used the Word Net Lemmatizer and stopwords from Natural Language Toolkit (NLTK) to lemmatize the text and remove stopwords. Additionally, we manually stemmed all the word variations of Ukraine, Britain, and russia into *Ukrain, russia,* and *brit*, respectively. This way, we could take into account all the word-forming variants of the countries (for example, Britain, Britains, British) and get a more accurate count of the collocations that include those tokens. Furthermore, to account for spelling variations of *Zelenskyy* and any references to his Twitter account (@ZelenskyyUa), we converted all the variations of the name (*Zelenskiy*, *Zelensky*, etc.) into *Zelensky*. We then used the NLTK collocations library to collect the most frequent bigrams, trigrams, and quadgrams that appeared in the tweets of each politician (Bird, 2006).

Phase 3. Topic Modelling

To conduct topic modelling with unsupervised learning, we used the Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model (GSDMM) as proposed by Yin & Wang (2014). We first preprocessed each tweet as outlined in phase 2. The GSDMM algorithm – which Weisser et al. (2022) found performs better on short texts like tweets than its more popular counterpart, LDA (Latent Dirichlet Allocation) – groups the tweets of each politician into unnamed "clusters," or topics, based on text similarity. Using the results of the unsupervised clustering, we examined the five most common tokens of each cluster to manually assign each cluster a topic.

Phase 4. Emotion Analysis

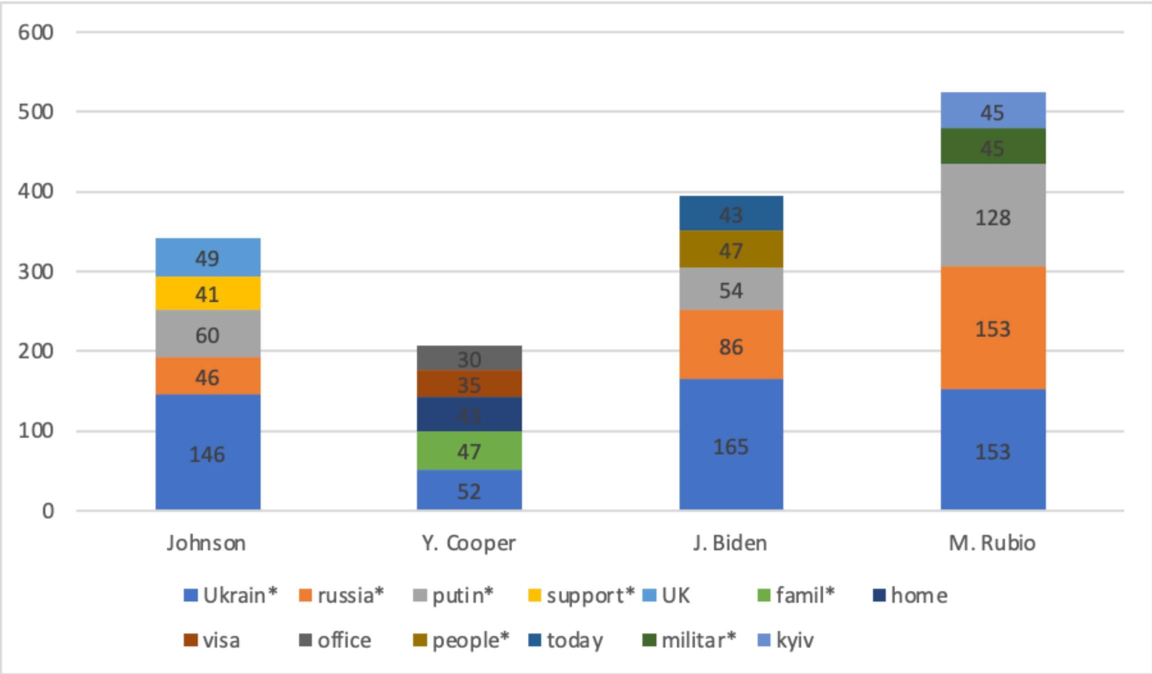
Within each topic, we converted the tweets into lists of tokens and then evaluated the emotions of the tweets using the NRC Lexicon, created by Mohammad & Turney (2013); it contains approximately 27,000 words and recognizes the following ten emotions: fear, anger, anticipation, trust, surprise, positive, negative, sadness, disgust, and joy. Each word in the lexicon has certain emotional scores assigned to it, and so the scores for each word add up to a final emotion score for each emotion in that tweet. We used the total emotion scores of all the tweets in a topic to determine the emotions the politician felt towards that topic.

Results

Frequency and Term Distribution

Considering the frequency of linguistic tokens used by each politician, we obtained different results. The only token found in the top five most frequent terms in the datasets of the four politicians is Ukrain*. Some other terms were shared by two or three politicians. For example, the tokens *putin** and russia* appeared among the top five frequent terms in the tweets of three out of the four politicians. We also observed the terms that were restricted to the frequent lexical use of only one politician (see UK*, home*, kyiv*, militar*). Having combined the top five most frequent lexical tokens of each politician, we obtained thirteen terms total, taking into consideration that some terms overlapped in the datasets. The overall view that summarises these frequencies can be seen in Fig. 1.

Figure 1
Top Frequent Terms



The detailed results on the frequency and term distribution are specified with the reference to each politician.

Boris Johnson

The dataset with the tweets of Boris Johnson contains 4,308 total words and 1,112 unique word forms, the frequent words being as follows:

Group 1. Most Frequent Terms: ukrain* (146); UK (49); putin* (60); russia* (46) support* (41);

Group 2. Second Most Frequent Terms: president (28); zelenskyyua (26); people* (25); speak* (spoke*) (23); invasion (invading) (22); militar* (21); countr* (20); help (20) stand* (21); economic (20);

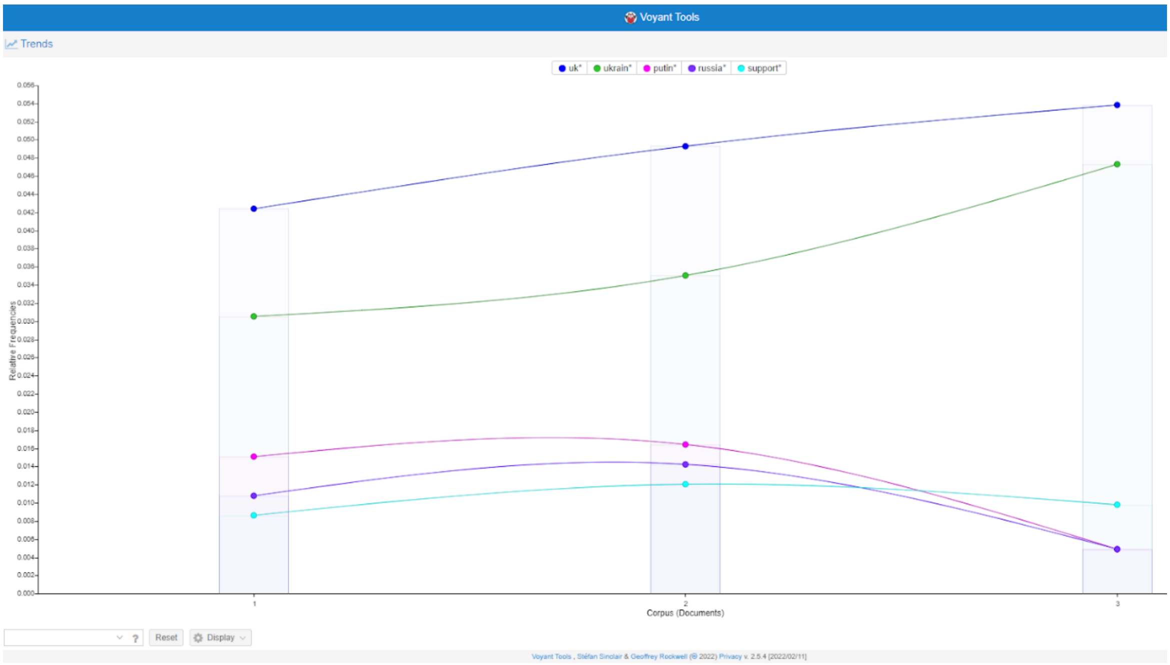
Group 3. Third Most Frequent Terms: today (19); continue* (18); free* (18) sanction* (17); NATO (14); aid (13); ensure (13); step* (13); barbaric (12); partners (12).

Distinctive words: UK (49), zelenskyyua (26), regime (9), defensive (9), putin's (8).

Table 1
Boris Johnson Top Five Term Distribution

Term	**February-March**		**April**		**May**	
	raw fr.	**rel. fr.**	**raw fr.**	**rel. fr.**	**raw fr.**	**rel. fr.**
Ukrain*	85	.0305	32	.0350	29	.0473
putin*	42	.0150	15	.0140	3	.0065
UK	32	.0114	13	.0142	4	.0065
russia*	30	.0107	13	.0142	3	.0065
support*	24	.0086	11	.0120	6	.0097

Figure 2
Boris Johnson Relative Term Frequency Distributed in Time



The chart represents the change in frequencies of the top five most frequent terms. The terms *UK* and *Ukrain** display a fairly constant increase in their usage. Conversely, such terms as *putin** and *russia**, which were used with nearly equal frequency during the periods of February-March and April, started to lose their popularity and showed a significant decline in May. Interestingly, the fact that the frequency of the term *support** remained stable during the whole time span suggests that the ideas of assistance and guidance delivered in the time of need to those who suffer from russian aggression were among top priorities for the British Prime Minister.

Yvette Cooper

The dataset made up with the tweets of Yvette Cooper is twice as small as the dataset of Boris Johnson, comprising 2,608 total words and 745 unique word forms.

Frequent lexical tokens:

Group 1. Most Frequent Terms ukrain* (52); famil* (47); home (43)*; visa (35); office (33).

Group 2. Second Most Frequent Terms people (25); help* (21); UK (19); govt (government) (18); delay* (17); need* (17); week* (17); shame* (16); wait* (16); sanctuary (12).

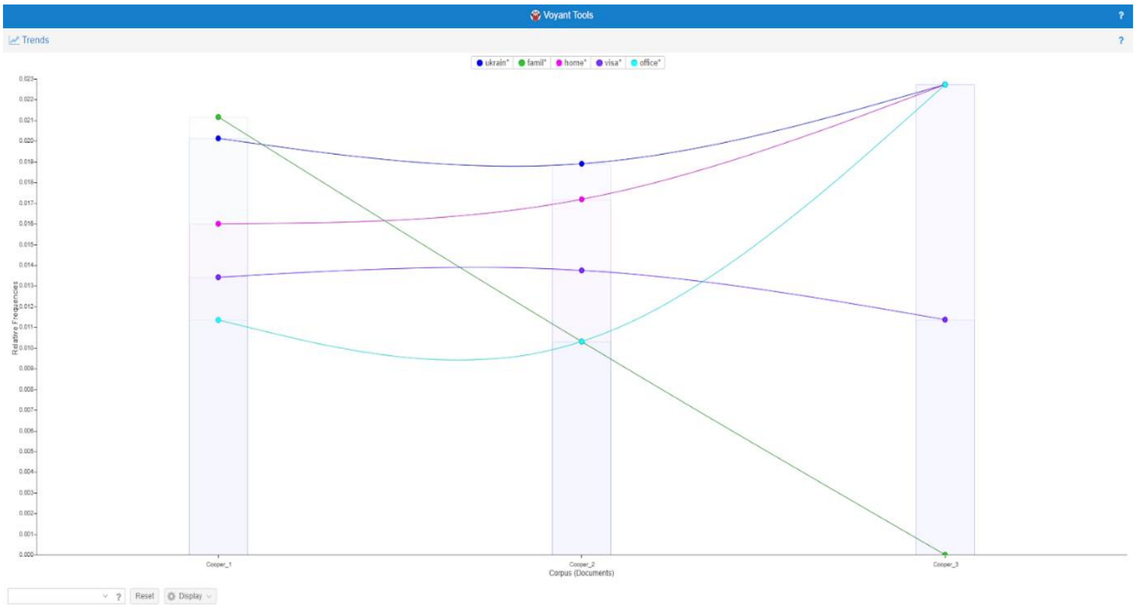
Group 3. Third Most Frequent Terms scheme (12); war (12); minister* (11); desperate (10); thousands (10); Brit* (10); let* (10); long (9); Priti (Patel) (11); refugee* (9).

Distinctive words: office (30), delays (16), shameful (13), waiting (12), sanctuary (12).

Table 2
Yvette Cooper Top Five Term Distribution

Term	**February-March**		**April**		**May**	
	raw fr.	**rel. fr.**	**raw fr.**	**rel. fr.**	**raw fr.**	**rel. fr.**
Ukrain*	39	.0201	11	.0289	2	.0227
famil*	41	.0211	6	.0103	-	-
home*	31	.0159	10	.0171	2	.0227
visa*	26	.0134	8	.0137	1	.0113
office*	22	.0113	6	.0103	2	.0227

Figure 3
Yvette Cooper Relative Term Frequency Distributed in Time



The terms *Ukrain**, *home** and *office* tend to increase in number, showing a steady rise. The term *visa* was fairly stable over the whole period of research. It is noteworthy that the term *famil**, which was the most frequent term during the February-May period, demonstrated a dramatic decline in April and completely disappeared from the lexicon of Yvette Cooper in May.

Joe Biden

The dataset of the current U.S. President is just slightly larger than the dataset of the head of the British Parliament, consisting of 5,362 total words and 1,210 unique word forms.

Frequent lexical tokens:

Group 1. Most Frequent Terms ukrain* (165); russia* (86); putin* (54); people* (47); today (43);

Group 2. Second Most Frequent Terms support* (37); united (35); war (34); assistance (33); states (26); allies (23); aggression (21); humanitarian (21); continue (20); world (19);

Group 3. Third Most Frequent Terms defend (17); partners (17); president (17); fight (16); security (16); economic (14); minister (13); weapons (13); met (12); additional (11); country (11).

Distinctive words: united (35), states (26), aggression (21), costs (10), unjustified (8)

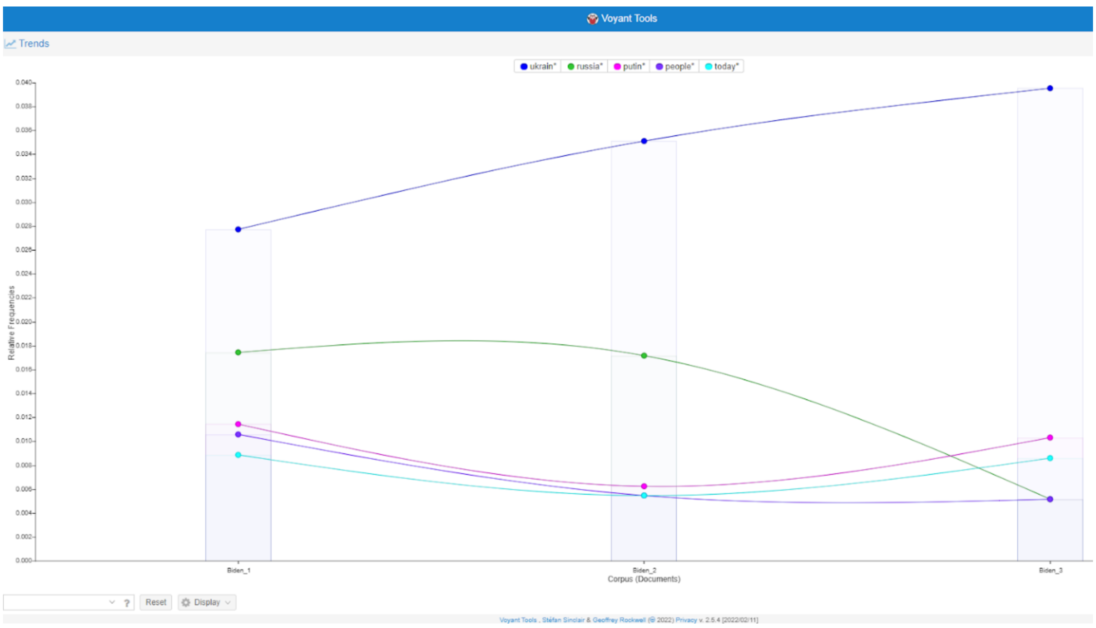
Table 3

Joe Biden Top Five Term Distribution

Term	**February-March**		**April**		**May**	
	raw fr.	**rel. fr.**	**raw fr.**	**rel. fr.**	**raw fr.**	**rel. fr.**
Ukrain*	97	.0277	45	.0351	2	.0395
russia*	61	.0174	22	.0171	3	.0051
putin*	40	.0114	8	.0062	6	.0103
people*	37	.0105	7	.0054	3	.0051
today*	31	.0088	7	.0054	5	.0085

Figure 4

Joe Biden Relative Term Frequency Distributed in Time



We may observe that the term Ukrain*, starting from the highest value among all the other frequent terms, has shown a more significant increase in May. Conversely, the term russia*, which was used with the same relative frequency in the February-March and April periods, showed a sharp decline in May. The terms *putin** and *today** have shown a similar tendency to slightly decrease in April and then gradually rise to the same value in May. The term *people*, which was only fourth place among the terms of particular importance in February-March, was lowered to fifth place, showing a steady decline in May.

Marco Rubio

The U.S. Senator for Florida tends to be the most eloquent expressing his attitude to the war in Ukraine – his dataset is the largest, consisting of 7,757 total words and 1,899 unique word forms. However, we did not manage to define any tweets referring to the topic of the war in May. Consequently, the term distribution was measured in terms of two periods: February-March and April.

Frequent lexical tokens:

Group 1. Most Frequent Terms ukrain* (153); russia* (153); putin* (128); militar* (45); kyiv (45);

Group 2. Second Most Frequent Terms invad*(invasion) (37); force* (31); nato (20); oil (18); cit*(25); day* (24); new (17); plan* (20); govt (government) (24); suppl* (20);

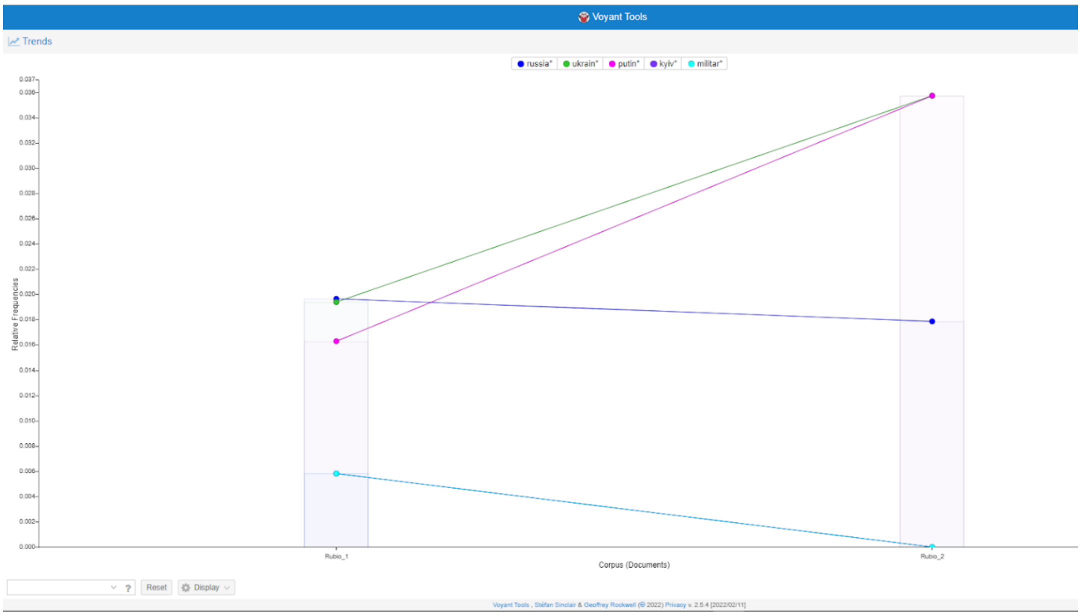
Group 3. Third Most Frequent Terms power (17); control*(16); weapon* (16); never (18); nuclear (15); hours (14); long* (19); people (14); biden (13); cut (13).

Distinctive words: putin (126), russia (117), strikes (12), costly (12), puppet (10).

Table 4
Marco Rubio Top Five Term Distribution

Term	**February-March**		**April**	
	raw fr.	**rel. fr.**	**raw fr.**	**rel. fr.**
Ukrain*	150	.0193	2	.0357
russia*	152	.0196	1	.0178
putin*	126	.0162	2	.0357
militar*	45	.0058	-	-
Kyiv	45	.0058	-	-

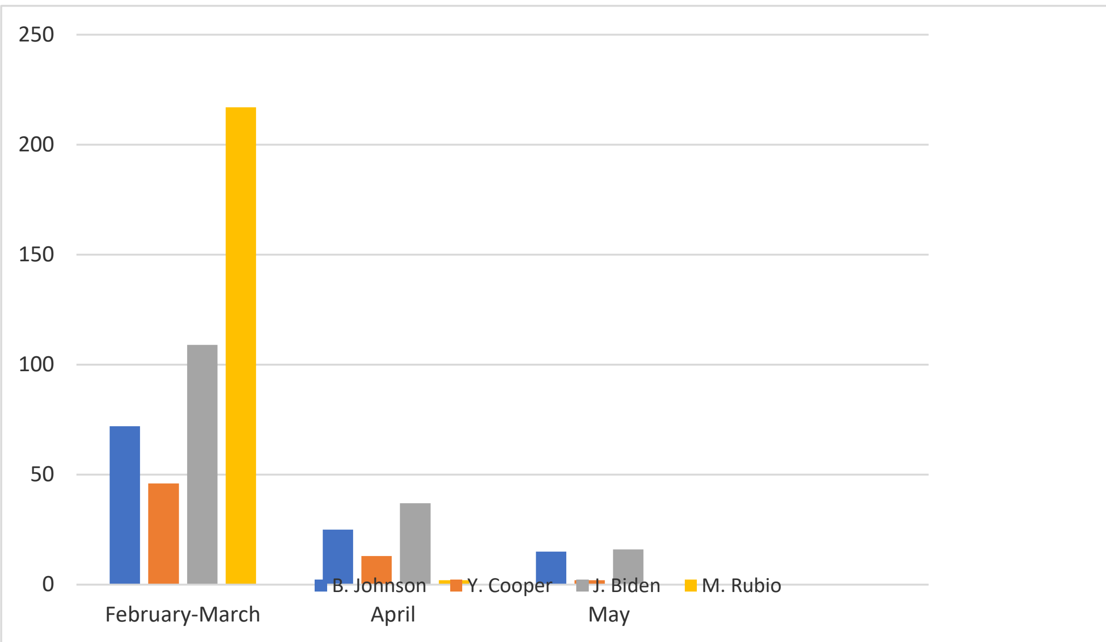
Figure 5
Marco Rubio Relative Term Frequency Distributed in Time



The terms ukrain* and putin* showed a tendency to increase in their frequency values, while the term russia* was evenly distributed between the two periods. The terms *militar**and *Kyiv* did not display much changes at all, as their usage was restricted to the first period only.

The difference between the values of raw and relative frequency indicated in the tables can be explained by the density of tweets posted during the period of research (see Fig. 6).

Figure 6
The Density of Tweets in the Period of February-May



If, at the beginning of the war, the number of tweets posted by the politicians varied from 46 to 217, in May we may observe a considerable decline in the frequency of the posts, which lowered to 2-16 posts.

Collocation analysis

Table 5
Most Frequent Boris Johnson Collocations

Bigram Count	Trigram Count	Quadgram Count
('president', 'zelensky'): 21	('spoke', 'president', 'zelensky'): 9	('continue', 'step', 'militar', 'economic'): 2
('support', 'ukrain'): 20	('putin', 'must', 'fail'): 6	('evening', 'spoke', 'president', 'zelensky'): 2
('ukrain', 'people'): 13	('ensure', 'putin', 'fails'): 4	('militar', 'economic', 'diplomatic', 'support'): 2
('invasion', 'ukrain'): 9	('economic', 'support', 'ukrain'): 3	('people', 'ukrain', 'slava', 'ukrain'): 2
('putin', 'regime'): 9	('putin', 'barbaric', 'invasion'): 3	('ukrain', 'putin', 'must', 'fail'): 2

Boris Johnson tends to use the token *ukrain** and its word-forming variants (*Ukrainian*, *Ukrainians*) in bigrams with *support* (20), *people* (13), and *invasion* (9).

Yvette Cooper

Table 6
Most Frequent Yvette Cooper Collocations

Bigram Count	Trigram Count	Quadgram Count
('home', 'office'): 30	('people', 'fleeing', 'war'): 4	('people', 'fleeing', 'war', 'europe'): 3
('ukrain', 'famil'): 10	('elderly', 'parents', 'ukrain'): 3	('admit', 'security', 'checks', 'done'): 2
('priti', 'patel'): 8	('fleeing', 'war', 'europe'): 3	('famil', 'arrive', 'uk', 'without'): 2

Bigram Count	Trigram Count	Quadgram Count
('turned', 'away'): 8	('home', 'office', 'still'): 3	('existing', 'community', 'sponsorship', 'scheme'): 2
('home', 'secretary'): 7	('admit', 'security', 'checks'): 2	('ministers', 'officials', 'admit', 'security'): 2

In Yvette Cooper's tweets, the token *ukrain** appeared most frequently with *famil** (10) in bigrams and the tokens *elderly* and *parents* (3) in trigrams. Neither of the tokens *putin* nor *russia* appeared in the most frequent collocations.

Joe Biden

Table 7
Most Frequent Joe Biden Collocations

Bigram Count	Trigram Count	Quadgram Count
('support', 'ukrain'): 27	('putin', 'price', 'hike'): 5	('unprovoked', 'unjustified', 'attack', 'ukrain'): 3
('united', 'states'), 26	('hold', 'russia', 'accountable'): 4	('putin', 'war', 'choice', 'ukrain'): 3
('ukrain', 'people'): 23	('putin', 'war', 'choice'): 4	('support', 'ukrain', 'face', 'russia'): 3
('allies', 'partners'), 16	('ukrain', 'people', 'defend'): 4	('ukrain', 'people', 'defend', 'countr'): 3
('assistance', 'ukrain'): 11	('support', 'ukrain', 'people'): 4	('united', 'states', 'allies', 'partners'): 3

Joe Biden tended to use the token *ukrain** in bigrams with *support* (27), *people* (23), and *assistance* (11). Additionally, the tokens *ukrain** and *people* appeared in trigrams with *defend* (4) and *support* (4).

Marco Rubio

Table 8
Most Frequent Marco Rubio Collocations

Bigram Count	Trigram Count	Quadgram Count
('war', 'crimes'): 9	('costly', 'militar', 'victory'): 6	('either', 'costly', 'militar', 'victory'): 4

Bigram Count	Trigram Count	Quadgram Count
('costly', 'militar'): 7	('matter', 'many', 'cities'): 4	('costly', 'militar', 'victory', 'occupation'): 3
('ukrain', 'militar'): 7	('nuclear', 'power', 'plant'): 4	('largest', 'nuclear', 'power', 'plant'): 2
('cease', 'fire'): 6	('install', 'puppet', 'govt'): 3	('quality', 'life', 'rapidly', 'deteriorating'): 2
('russia', 'invaders'): 6	('people', 'never', 'accept'): 3	('nuclear', 'power', 'plant', 'ukrain'): 2

In Marco Rubio's top 5 most frequent bigrams, trigrams, and quadgrams, the token *ukrain** collocates with only one token, militar* (7). Similarly, in the most frequent collocations, the token *russia** appears only once – as a bigram with *invaders* (6).

Topic Modelling and Emotion Analysis

After using the GSDMM algorithm for unsupervised clustering, we assigned each cluster of tweets a topic based on the 5 most frequent tokens in the cluster. We then used the NRC Lexicon to calculate the emotion scores for each topic.

Boris Johnson

Table 9
Topics of Boris Johnson's Tweets

Topic	Putin	Ukraine Support	All Tweets
Most Frequent Tokens	('ukrain', 109) ('putin', 51) ('russia', 43) ('uk', 40) ('zelensky', 27)	('ukrain', 37) ('uk', 10) ('freedom', 9) ('putin', 8) ('support', 7)	ukrain* (146) putin* (60) uk (49) russia* (46) support* (41)
NLTK Emotion Score	positive: 234 negative: 188 trust: 152 fear: 149 anger: 111 anticipation: 68 sadness: 59 disgust: 53 joy: 49 surprise: 29	positive: 79 anticipation: 60 trust: 51 negative: 42 surprise: 37 joy: 31 fear: 30 anger: 30 disgust: 16 sadness: 15	positive: 319 negative: 228 trust: 216 fear: 182 anger: 143 anticipation: 97 joy: 78 sadness: 74 disgust: 65 surprise: 40
Number of Tweets	86	29	116

The vast majority of Boris Johnson's 116 tweets about the war in Ukraine were either about putin or about support for Ukraine. When talking about putin, Johnson kept his tweets positive, with high scores for positive emotion (234) and trust (152). However, he also expressed high amounts of negative emotion (188) and fear (149) in his tweets about putin. In his tweets that showed support for Ukraine, Johnson had the highest emotion score for positive emotion (79), followed by anticipation (60) and trust (51). For all of his tweets on the war in Ukraine, Boris Johnson expressed positivity (319) the most.

Yvette Cooper

Table 10

Topics of Yvette Cooper's Tweets

Topic	Home Office	UK Help	All Tweets
Most Frequent Tokens	('home', 27) ('ukrain', 24) ('still', 22) ('office', 20) ('families', 19)	('ukrain', 26) ('uk', 16) ('family', 14) ('people', 13) ('home', 13)	ukrain* (52) famil* (47) home (43)* visa (35) office (30)
NLTK Emotion Score	negative: 80 anticipation: 44 positive: 42 fear: 38 sadness: 34 trust: 21 anger: 20 joy: 17 surprise: 15 disgust: 14	positive: 56 negative: 54 trust: 35 fear: 35 anticipation: 29 sadness: 24 joy: 22 anger: 15 disgust: 11 surprise: 8	negative: 143 positive: 106 anticipation: 82 fear: 77 trust: 61 sadness: 60 joy: 42 anger: 36 disgust: 26 surprise: 26
Number of Tweets	34	23	61

Yvette Cooper's tweets on the war focused entirely on the UK government, specifically about the Home Office of the UK and the help the UK could provide. She rendered her concern about the complicated procedure of entering the UK and the incompetence of the government to take over the responsibilities. When talking about the Home Office and its actions regarding Ukraine, Yvette Cooper expressed high negative emotion (80) and anticipation (44). In her tweets about the UK's support for Ukraine, Yvette Cooper showed nearly equal positive (56) and negative (54) emotion scores, as well as the same emotion scores for trust (35) and fear (35). For all of her tweets on the war in Ukraine, Yvette Cooper showed the highest score for negative emotion (143).

Joe Biden

Table 11
Topics of Joe Biden's Tweets

Topic	Ukraine Support	Putin	Russian Sanctions	All Tweets
Most Frequent Tokens	('ukrain', 142) ('russia', 53) ('people', 41) ('support', 32) ('assistance', 32)	('putin', 20) ('ukrain', 18) ('prices', 10) ('war', 8) ('russia', 8)	('russia', 25) ('today', 9) ('economy', 7) ('sanctions', 5) ('ruble', 4)	ukrain* (165) russia* (86) putin* (54) people* (47) today (43)
NLTK Emotion Score	positive: 330 trust: 225 negative: 150 fear: 149 anticipation: 127 anger: 85 joy: 73 surprise: 37 sadness: 20 disgust: 20	negative: 54 positive: 43 fear: 34 trust: 31 anger: 25 joy: 21 anticipation: 16 sadness: 15 disgust: 10 surprise: 10	positive: 37 negative: 33 fear: 29 trust: 26 anger: 16 anticipation: 12 joy: 9 sadness: 7 disgust: 6 surprise: 3	positive: 424 trust: 291 negative: 240 fear: 216 anticipation: 155 anger: 127 joy: 106 surprise: 50 sadness: 42 disgust: 36
Number of Tweets	106	23	18	151

Joe Biden's tweets on America's support for Ukraine mostly expressed positive emotion (330) and trust (225). When talking about putin, Joe Biden had the highest score for negative emotion (54). In his tweets about the economic sanctions on russia, Joe Biden expressed positive emotion (37) the most, followed by negative emotion (33) and fear (29). When tweeting about the war in Ukraine as a whole (all tweets), Joe Biden clearly expressed positive emotion (424) the most; his emotion score for positivity vastly exceeded those for all the other emotions. The next highest emotion scores were for trust (291) and negative emotion (240).

Marco Rubio

Table 12
Topics of Marco Rubio's Tweets

Topic	Ukraine	Putin	Ukraine Nuclear Plants	All Tweets
Most Frequent Tokens	('ukrain', 91), ('russia', 79), ('putin', 62), ('kyiv', 43), ('military', 31)	('putin', 65) ('russia', 60) ('ukrain', 45) ('would', 19) ('nato', 18)	('ukrain', 17) ('russia', 14) ('plant', 8) ('nuclear', 7) ('fire', 7)	ukrain* (153) russia* (153) putin* (128) militar* (45) kyiv (45)
NLTK Emotion Score	negative: 189 positive: 182 fear: 168 anticipation: 108 trust: 94 anger: 84 sadness: 76 joy: 45 disgust: 41 surprise: 28	negative: 207 fear: 155 positive: 129 anger: 95 trust: 72 sadness: 67 anticipation: 66 disgust: 46 surprise: 36 joy: 32	negative: 46 fear: 32 positive: 25 anger: 22 trust: 17 sadness: 16 anticipation: 10 surprise: 9 disgust: 9 joy: 3	negative: 442 fear: 355 positive: 336 anger: 201 anticipation: 184 trust: 183 sadness: 159 disgust: 96 joy: 80 surprise: 73
Number of Tweets	111	89	20	220

Most of Marco Rubio's tweets were either about Ukraine, putin, or Ukrainian nuclear power plants. On the topic of Ukraine, Marco Rubio had similar scores for negative emotion (189) and positive emotion (182). The third and fourth highest emotion scores were for fear (168) and anticipation (108). Marco Rubio's tweets on putin had the highest score for negative emotion (207), followed by fear (155) and positive emotion (129). When talking about the Ukrainian nuclear power plants, Marco Rubio had emotion scores that followed the same pattern as his tweets on putin: the three highest scores were for negative emotion (46), fear (32), and positive emotion (25). In fact, the whole dataset of his tweets on the war in Ukraine followed this pattern; negative emotion (442) was followed by fear (355) and positive emotion (336).

Discussion

The most frequent terms in the Boris Johnson dataset are Ukrain*, russia*, putin*, UK*, and support*. The gradual decrease of the use of the terms putin* and russia* is regarded as a shift in the focus of attention from the atrocities committed by russian soldiers to the ways of support for Ukraine, delivered by the UK. The politician notes the active participation of his country in providing assistance to Ukraine and expresses confidence in further support. This is evidenced by tokens such as *support*, *aid*, and *stand*, used in the same context and evenly distributed in the tweets during all the periods under consideration. For example:

> *The **UK stands** with **Ukraine** – we will send further defensive **aid** and they have our full backing in the negotiations.*

A similar tendency was observed in the tweets of Joe Biden, where Ukrain*, the token with the highest frequency value in the whole dataset, displayed a sharp increase in May, compared to its use in February-March. In the first period (February-March), the US president speaks about the support of Ukrainians, reports on the acts of assistance, and admires the brave resistance of Ukrainian people. During the second period (April), the token Ukrain* was found in the tweets highlighting the importance of international unity and cooperation; for example, the tweets mentioned the battle for Kyiv and announced new programs for Ukrainian refugees. The tweets from the third period (May) are concentrated on the continuing efforts to support Ukraine, providing military, economic and financial support. For example:

> *Today, the United States is announcing that we intend to provide an additional **$500 million** in direct economic assistance to the Ukrainian government. This brings our total economic support for Ukraine to **$1 billion** in the past two months.*

The frequency of terms used by Yvette Cooper shows her preoccupation with the idea of facilitating the process of entering the UK for Ukrainians, elaborating the procedure with less restrictions**.** She appeals to the Home Office, which is directly responsible for visas and immigration. This focus explains the high values of frequency of terms such as home* (43), office (30), and visa (35), which were evenly distributed during the whole period of research. For example:

> ***Home Office** is still causing long delays for Ukrainian refugees. Thousands of desperate **families** who have applied are still waiting weeks & hearing nothing. Why is the **Home** Secretary so incapable of getting a grip on this?*

The term famil* (47), which was the most frequent in the first period, has reached the lowest value among the top five frequent terms in the second period and completely disappeared in the third one. Having examined a wider context of this term, we discovered that initially it was primarily used with the regard to the socalled Family Scheme, allowing the relatives from the UK and Ukraine to reunite. The decrease of the term's usage in the following periods may be explained by the fact that due to the efforts of Yvette Copper bringing the problem of issuing visas to fleeing Ukrainians to public notice, the problem was partly solved, and the term family* started appearing with the reference to British families ready to welcome Ukrainian refugees according to the Sponsorship Scheme.

The most frequent terms in Marco Rubio dataset, Ukrain* and putin*, demonstrated the tendency to increase in number, reaching the same values in April. These two tokens were primarily spotted in the same context, where the Senator presents his reflections on war questions, makes predictions about the next steps that are likely to be taken by putin in the nearest future, and updates the followers on the current situation in Ukraine. For example:

> ***#Putin*** *would not hesitate to stage or carry out a biological weapon false flag in* ***#Ukraine*** *and this is the kind of messaging you would see as a prelude to him doing that.*

A considerable number of his predictions mention Kyiv, the capital of Ukraine, as an important strategical object, which explains the equal distribution of this term. The examples of the use of this term are shown in the tweets below.

> *As I have been saying now for 10 days the #Putin plan is no longer to take over most of #Ukraine The plan is to annex coastal south, lay siege to* ***#Kyiv*** *& 5 cities in North, degrade Ukraines military & factories & then offer cease fire on terms he will claim are a strategic victory*

> *#Putin still wants to capture #Kyiv & install a puppet govt But when he realizes that's not feasible he will: 1. Focus on destroying as much of @DefenceU as possible 2. Then offer cease fire that imposes neutrality on #Ukraine & recognizes #Crimea & #Donbas as part of #Russia*

While examining the connection between the frequency of terms and collocations, we noticed that not all terms with high frequency values were productive in forming collocations. For example, the constituents of the most productive bigram *president Zelensky* in Boris Johnson dataset; *Unites States* in Joe Biden dataset belong only to the second level of frequency, while the constituents of the most frequent bigram *war crimes* in Marco Rubio dataset do not belong to frequent terms at all.

A considerable number of frequent collocations are formed with the help of distinctive words selected from the datasets in view of their uniqueness compared to the rest of the datasets in the whole dataset, which do not necessarily overlap with frequent terms. For example, both terms from the collocation *United States* were defined as distinctive words of Joe Biden, at the same time the terms were referred to the second level of frequency. Conversely, the term *costly*, defined as a distinctive word of Marco Rubio, though not found among the most frequent terms, turned out to demonstrate a high productivity in making collocations (20 collocations), unlike the term *Kyiv* which belongs to top five frequent terms of Marco Rubio, but was not found among frequent collocations. The term which turned out to be the most productive in making collocations is *Ukrain**.

We discovered some similarities between the use of lexical tokens in the datasets of the American politicians: the lists of their top five most frequent terms overlap in three tokens: ukraine*, russia* and putin*. Lexical preferences of British politicians displayed contrastive results: the top five lists of Boris Johnson and Yvette Cooper overlap only in one term – Ukrain*.

The collocations of Boris Johnson's tweets reveal his extremely disapproving attitude towards putin's actions. The token *putin* appeared 9 times as a bigram with *regime*, a term with a negative connotation that Boris Johnson used to draw attention to putin's distasteful dictatorial powers (Table 5). Furthermore, Boris Johnson's 2nd most and 4th most frequent trigram was *putin must fail* and *ensure putin fails*, clear indicators that Boris Johnson thinks of putin as the main adversary in the war (Table 5). Boris Johnson's description of putin's military choices as *barbaric* in the 5th most frequent trigram, *putin barbaric invasion*, showcases how immoral he thinks putin's actions are (Table 5).

Additionally, Boris Johnson makes sure to point his criticisms mostly at putin, not at russia as a whole country. The token *russia* didn't show up in the top 5 most frequent bigrams, trigrams, or quadgrams, and although it is the 4th most frequent token in Boris Johnson's tweets, this ranking is lower than that of Marco Rubio's or Joe Biden's tweets (1st and 2nd most frequent token, respectively), suggesting that Boris Johnson does not blame russia for the war, but putin (Table 5). Indeed, Boris Johnson's tweet on March 13th, 2022, supports this idea:

> *To the people of Ukraine: Slava Ukraini. To the people of Russia: I do not believe this war is in your name. This crisis, this tragedy, can and must come to an end. Because the world needs a free and sovereign Ukraine.*

However, the token *putin* is sometimes used synonymously with the token *russia** in the context of the far-reaching isolation of this country from the rest of the world. Boris Johnson arouses the topic of elimination of russian banks from SWIFT, reducing dependence on russian oil, etc. An example of this usage is shown by Boris Johnson's tweet on April 9th:

> *The UK will send more defensive weapons to Ukraine and will work with G7 partners*
> ***to target every pillar of the Russian economy** to ensure putin fails.*

The emotion analysis of Boris Johnson's tweets demonstrates his consistently positive outlook on the war. In all his tweets and all his topics, positive emotion had the highest emotion score. He likely kept his attitude positive to prove to the public that his actions were helping the Ukrainian cause. In order to gain public approval, Boris Johnson would want to make sure that he described his actions in the best possible light, demonstrating that his way was working and that he was doing everything he could to support Ukraine.

With trigrams such as *putin must fail* and *putin barbaric invasion*, we would expect that the "putin" topic would have negative emotions as the highest emotion scores. And yet, two of the three highest emotion scores are positive (positive emotion and trust), with positive emotion having the highest score (Table 9). One explanation for this ranking is that Boris Johnson wants his tweets to show positive support for Ukraine (as evidenced by his high scores of positive emotion), and so he would offset and exceed any negative mentions of putin with positive remarks about Ukraine in order to make the main topic of his tweet about supporting Ukraine. For example, on March 15th, Boris Johnson tweeted:

Putin's barbaric actions murdering Brent Renaud and other innocent civilians are testing not just Ukraine but all of humanity.
Speaking to President @ZelenskyyUa I assured him that we will continue to do all that we can to bring an end to this disastrous conflict.

Despite mentioning "Putin's barbaric actions" at the beginning, Boris Johnson ends the tweet with claims of fervent support, counteracting the negative emotions surrounding putin with positive ones about supporting Ukraine. As a result, this tweet, which was classified under the "putin" topic, had a positive emotion score of 5 but a negative emotion score of only 3.

The ranking of anticipation as the 2nd highest emotion score in the "Ukraine Support" topic could be attributed to Boris Johnson's promises of success or intended outcomes of the UK's actions (Table 9). Consider the following tweet on March 14th:

I am hugely grateful to our NHS staff, partners and Polish friends for their support in bringing Ukrainian children who need lifesaving medical care to the UK.
We will do all we can to support them whilst they continue their critical cancer treatment.

While he does not explicitly promise anything, Boris Johnson pledges that the UK will do all they can to support the NHS and Ukraine. Due to this vow, the tweet's second highest emotion score was for anticipation (2).

Yvette Cooper's collocations indicate that she frequently mentions the Home Office of the UK and Priti Patel, the current Home Secretary (Table 6). Given Yvette Cooper's position of Shadow Home Secretary, her narrow focus on her department seems appropriate. Furthermore, Yvette Cooper seems to mention the conflict in Ukraine only in regard to visas, refugees, and other national security or immigration matters; *famil**, *home*, *visa*, and *office* were four of her five most frequent tokens, and the token *ukrain** collocated most frequently with *famil** and *elderly parents* (Table 6). Additionally, unlike the other politicians studied in this paper, Yvette Cooper's five most frequent tokens do not include *russia* or *putin* (Table 10). These two tokens also do not make up any of the five most frequent bigrams, trigrams, or quadgrams, further showcasing that Yvette Cooper focuses mostly on how the war in Ukraine affects her department, and not on the conflict itself (Table 6).

She tends to be critical towards her country's overall involvement, as negative emotion had the highest emotion score for all of her tweets (Table 10). Yvette

Cooper's negativity likely stems from her membership in the Labour Party, the opposing party to the majority party in the UK government, the Conservatives. As a Labour Party MP (Member of Parliament), Yvette Cooper likely disagrees with many of the decisions that the Conservative majority made, resulting in tweets with strong negative emotions. Unlike Boris Johnson or Joe Biden, she will not be held responsible for her country's actions in the war, and she is also not the one making all her country's decisions, so she does not feel the need to shed every action in a positive light like a President or Prime Minister would.

Yvette Cooper's overall negativity transfers over to the Home Office, as the highest emotion score by far for that topic is also negative emotion (Table 10). Specifically, Yvette Cooper often targets Priti Patel, the Home Secretary, in a critical manner. Priti's membership in the Conservative Party likely also plays a role in Yvette Cooper's negative tone and disapproval.

However, when talking only about the UK's aid for Ukraine, Yvette Cooper seems to be more conflicted, with nearly equal scores of positive and negative emotions, followed by the exact same emotion scores for trust and fear. Her more positive attitude towards the UK's support for Ukraine emphasizes how much more critical she is towards the Home Office's actions. Her negative attitude towards the Home Office and not towards the rest of the UK indicates that she finds many more issues with the UK's immigration and visa policies regarding Ukraine, but that she is more neutral regarding other areas of the UK's involvement in the war.

Joe Biden's collocations demonstrate his continuous support to the Ukraine cause. Joe Biden uses a variety of words to express this support, the most frequent tokens paired with *ukrain** being *support*, *allies*, *partners*, *assistance*, and *defend* (Table 7).

Although Joe Biden frequently describes the war in Ukraine as putin's "war of choice" (based on the trigram *putin war choice*), he doesn't make as clear of a distinction between putin and russia as Boris Johnson does, as evidenced by Joe Biden's 2nd most frequent trigram *hold russia accountable* (Table 7). Joe Biden makes clear that he knows the war is putin's choice and not russia's, but he must hold all of russia accountable in order to stop putin. Like Boris Johnson, Joe Biden also emphasizes the illegitimacy and immorality of russia's attack on Ukraine, describing the invasion as "unprovoked" and "unjustified" in his most frequent quadgram *unprovoked unjustified attack ukrain* (Table 7).

The range of topics identified in the dataset includes the topic of how the global economy is affected by aggressive russian policy. Joe Biden regularly mentions the "putin price hike" in his tweets – the term is his most frequent trigram, used to define the increase of gas prices. He frequently refers to the inflation increase in America, which he blames on putin's invasion of Ukraine, as the "putin price hike" likely in order to deflect blame from himself. Therefore, he uses the trigram *putin price hike* even more than the trigram *support ukrain people*, perhaps because the resulting inflation increase is more relevant to the American people than the war itself (Table 7).

For all of Joe Biden's tweets, positive emotion had the highest emotion score, followed by trust; like Boris Johnson, Joe Biden likely keeps his tone positive to gain public approval by assuring Americans that his actions are working and supporting Ukraine well. For similar reasons, the highest emotion scores for the "Ukraine Support" topic are also positive emotion and trust. His tweets likely have high scores for trust because he wants to demonstrate to the public that he trusts and believes in both the Ukraine cause and his methods so that Americans will also share his beliefs. An example of this trust is shown in a March 27th tweet:

> *Rather than breaking Ukrainian resolve, Russia's brutal tactics have only strengthened it. Rather than driving NATO apart, the West is now stronger and more united than it has ever been.*

Joe Biden stated to the public his firm belief that russia's tactics aren't working and that the West is getting stronger and stronger. The tweet showed his trust in the methods and righteousness of their cause, and so trust (2) had the highest emotion score for this tweet.

Topic modelling of Joe Biden's tweets revealed that he has a considerable amount of tweets on russian sanctions, further showcasing that although he understands that putin is at fault for the war, he must hurt russia as well in order to stop putin. The highest emotion score for the "russian Sanctions" topic is positive emotion, a surprising fact given that economic sanctions usually have a negative connotation. However, the high positivity may be because the enacting of sanctions are positive for the U.S. and Ukraine, as the intended outcome is to bring an end to the war. An example that supports this idea is the following tweet on March 1st:

> *I just spoke with President Zelenskyy to discuss our continued support for Ukraine — including security assistance and humanitarian aid — as it defends itself against Russian aggression. We will hold Russia accountable, and our sanctions are already having a devastating impact.*

In this tweet, although Joe Biden mentions that the sanctions are "having a devastating impact," they are put in place to aid and support Ukraine. Therefore, positive emotion (6) had a higher emotion score for this tweet than negative emotion (3).

The collocations of Marco Rubio's tweets reveal that, like Boris Johnson and Joe Biden, Marco Rubio also emphasizes the immorality of putin's invasion of Ukraine. Marco Rubio's most frequent bigram was *war crimes* in reference to many of the russian military's actions. Furthermore, the only token that *russia* collocated with in his most frequent collocations was *invaders* (Table 8). Marco Rubio also frequently mentions the idea that russia will at best have a "costly military victory" (his most frequent trigram), as the war will take an immense toll on russia even if they win (Table 8).

Emotion analysis shows that negative emotion and fear had the highest emotion scores in all of Marco Rubio's tweets. As a Republican, the opposing party to Joe

Biden's Democrats, Marco Rubio would have a more critical view on America's and Joe Biden's actions regarding the war in Ukraine. Although both political parties support Ukraine, Marco Rubio and other Republicans have views different from Joe Biden's on how America should act on the war. This fact is likely also why negative emotion has the highest emotion score for Marco Rubio's "Ukraine" topic: not because he doesn't support Ukraine, but because he doesn't support some of America's actions involving the war. Additionally, like Yvette Cooper, Marco Rubio is not the one directly making the decisions concerning the war, so he would not feel pressured to praise all his country's actions like Joe Biden and Boris Johnson would.

Topic modelling demonstrates that Marco Rubio often discusses Ukraine's nuclear power plants when talking about the war, unlike the other politicians studied in this paper. Marco Rubio worries that if russian soldiers damage a nuclear power plant, a devastating radiation leak could occur, further damaging Ukraine. He makes sure to address this possibility and to draw attention to the nuclear power plants to emphasize their importance and the danger they possess.

Conclusions

The frequency of terms, collocations, topic modelling, and emotion analysis performed in this research revealed the attitudes of politicians regarding different aspects of the war in Ukraine. Topic modelling indicated that when the politicians tweeted about the war, their tweets tended to fall under one of two topics: putin's actions or Ukrainian support. One exception to this trend was Yvette Cooper, who focused mostly on her own department within the UK government, the Home Office. Additionally, Joe Biden and Marco Rubio showed interest in the topics of russian sanctions and Ukrainian nuclear plants, respectively. Emotion analysis demonstrated that Joe Biden and Boris Johnson, the leaders and figureheads of their respective nations, often expressed positive emotions such as trust and positivity in their tweets, especially the ones supporting Ukraine, in order to shed their actions regarding the war in the best possible light. Even their tweets regarding vladimir putin contained high amounts of positive emotion. Meanwhile, Yvette Cooper and Marco Rubio, two politicians of lower rank than Boris Johnson and Joe Biden, tweeted much more critically about their country's actions in the war, expressing negative emotions such as fear and negativity. Furthermore, Yvette Cooper and Marco Rubio are members of the political parties that are opposite to those of Boris Johnson and Joe Biden, and so their different political views likely resulted in more critical opinions of their country's response.

References

Горошко Е. И., Полякова Т. Л. (2011) Лингвистические особенности англоязычного твиттера. Учені *записки Таврійського національного університету імені Вернадського. Сер. Філологія. Соціальні комунікації, Т. 24*(63). № 2(1), 53–58. URL: http://repository.kpi.kharkov.ua/handle/KhPI-Press/49133

Нерян С. О. (2018) Допис у соцмережі як мовленнєвий жанр інтернет- комунікації. *Науковий вісник Херсонського державного університету, Сер. Лінгвістика, 33*(1), 66-70. URL: https://journals.indexcopernicus.com/api/file/viewByFileId/708692.pdf

Ніколаєва, Т. М. (2019). Лексико-семантичні аспекти мови соціальних мереж. *Закарпатські філологічні студії, 9*(2), 96-101. URL: https://dspace.uzhnu.edu.ua/jspui/handle/lib/33170

Полякова Т. Л. (2021). Лексичні засоби в жанрі твітинг в англомовній політичній інтернет-комунікації. *Закарпатські філологічні студії, 14*(1), 177-181. https://doi.org/10.32782/tps2663-4880/2020.14-1.32

Швелідзе Л. Д. (2021) Мовні засоби реалізації комунікативних стратегій у дискурсі соціальних мереж (на матеріалі української та англійської мов) (дис. … канд. філол. наук). Донецький національний університет імені Василя Стуса, Вінниця. URL: https://abstracts.donnu.edu.ua/article/view/9878

Bird, S. (2006). NLTK: the natural language toolkit. Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, (69-72). https://doi.org/10.3115/1225403.1225421

Crystal, D. (2011). A Microexample: Twitter. In *Internet Linguistics: A Student Guide*. (pp. 36-56). London and New York: Routledge. Taylor & Francis Group.

Mohammad, S. M., & Turney, P. D. (2013). NRC Emotion Lexicon. National Research Council, Canada. https://doi.org/10.4224/21270984

Weisser, C., Gerloff, C., Thielmann, A., Python, A., Reuter, A., Kneib, T., & Säfken, B. (2022). Pseudo-document simulation for comparing LDA, GSDMM and GPM topic models on short and sparse text using Twitter data. *Computational Statistics, 1-28*. http://dx.doi.org/10.1007/s00180-022-01246-z

Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge discovery and data mining, (233-242). https://doi.org/10.1145/2623330.2623715

References (translated and transliterated)

Bird, S. (2006). NLTK: the natural language toolkit. Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, (69-72). https://doi.org/10.3115/1225403.1225421

Crystal, D. (2011). A Microexample: Twitter. In Internet Linguistics: A Student Guide. (pp. 36-56). London and New York: Routledge. Taylor & Francis Group.

Goroshko, E. I., Poliakova, T. L. (2011). Lingvisticheskiie osobennosti angloiazychnogo tvittera. [Linguistic features of English-language Twitter]. *Ucheni Zapysky Tavriiskoho Natsionalnoho Universytetu Imeni Vernadskoho. Filolohiia. Sotsialni Komunikatsii Series, 24*(63), No 2–1, 53–58. Retrieved from http://repository.kpi.kharkov.ua/handle/KhPI-Press/49133

Mohammad, S. M., & Turney, P. D. (2013). NRC Emotion Lexicon. National Research Council, Canada. https://doi.org/10.4224/21270984

Nerian, S. O. (2018). Dopys u sotsmerezhi yak movlennievyi zhanr internet-komunikatsii. [A post in a social network as a speech genre of Internet communication]. *Naukovyi Visnyk Khersonskoho Derzhavnoho Universytetu. Linhvistyka Series*, *33*(1), 66-70. Retrieved from https://journals.indexcopernicus.com/api/file/viewByFileId/708692.pdf

Nikolaieva, T. M. (2019) Leksyko-semantychni aspekty movy sotsialnykh merezh [Lexico-semantic aspects of the social networks language]. *Zakarpatski Filolohichni Studii*, *9*(2), 96-101. Retrieved from https://dspace.uzhnu.edu.ua/jspui/handle/lib/33170

Poliakova, T. L. (2021). Leksychni zasoby v zhanri tvitynh v anhlomovnii politychnii internet-komunikatsii [Lexical means in the genre Tweeting in English political Internet communication]. *Zakarpatski Filolohichni Studii*, *14*(1), 177-181. https://doi.org/10.32782/tps2663-4880/2020.14-1.32

Shvelidze, L. D. (2021) Movni zasoby realizatsii komunikatyvnykh stratehii u dyskursi sotsialnykh merezh (na materiali ukrainskoi ta anhliyskoi mov) [Linguistic Means of Implementation of Communicative Strategies in the Social Media Discourse (Based on the Ukrainian and English Languages)]. Unpublished doctoral dissertation. Extended abstract. Vasyl' Stus Donetsk National University, Vinnytsya. Retrieved from https://abstracts.donnu.edu.ua/article/view/9878

Weisser, C., Gerloff, C., Thielmann, A., Python, A., Reuter, A., Kneib, T., & Säfken, B. (2022). Pseudo-document simulation for comparing LDA, GSDMM and GPM topic models on short and sparse text using Twitter data. *Computational Statistics,* 1-28. http://dx.doi.org/10.1007/s00180-022-01246-z

Yin, J., & Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge discovery and data mining, (233-242). https://doi.org/10.1145/2623330.2623715

Source

CJE gives recommendations for the use of words “orcs,” “ruscists,” and “putin” in the media. Retrieved from https://imi.org.ua/en/news/cje-gives-recommendations-for-the-use-of-words-orcs-ruscists-and-putin-in-the-media-i45817 (date of access: 7.12.2022)