

## THE UKRAINE WAR IN BBC NEWS: A BIG DATA ANALYSIS

**Namkil Kang**

Professor, College of Liberal Arts, Far East University, South Korea

### ABSTRACT

*This study aims to analyze 30 pieces of BBC news broadcasted in 2024 with respect to the Ukraine war. A point to note is that the word Ukraine was the most frequently occurred one, followed by the word Russia, the word war, and the name Putin, in that order. A further point to note is that the word NATO obtains 49 tokens and that it ranks eleventh. A major point of this paper is that the word Ukraine is the most pivotal one since it is biggest in the word cloud. In other words, this word was the most widely used and important one. Thus, this word is deemed to be significant in 30 pieces of BBC news. The word Russia is also pivotal and central since it is the second biggest in the word cloud. That is to say, it is one of the most centric words in 30 pieces of BBC news. It is worthwhile pointing out that the words Kremlin and Kharkiv show up in the center of the map, hence referring to the significant and centric ones in the network analysis. Finally, it must be stressed that the words aggression, alliance, allies, ammunition, blood, capitals, chance, chancellor, cities, dates, defense, diplomats, NATO, Ukraine, peace, Russia, etc. form a cohesive group (a cluster). The so-called cluster refers to forming a cohesive group.*

*Keywords: big data, network analysis, cluster analysis, frequency analysis, word cloud analysis*

### 1. INTRODUCTION

This research aims to provide an in-depth big data analysis of 30 pieces of BBC news broadcasted in 2024. In this paper, python was used as a research tool. First, we aim at providing the main contents of data, but we include only the first sentences of the relevant data. Second, we aim to abstract key words that occurred in 30 pieces of BBC news. Note that we abstracted only nouns after preprocessing the relevant sentences that occurred in 30 pieces of BBC news. Third, we aim to probe into the term frequency of the relevant nouns that showed up in 30 pieces of BBC news. The term frequency of the key nouns was abstracted in descending order. Fourth, we go over the word cloud of 30 pieces of BBC news. The so-called word cloud of python is produced in accordance with the term frequency of key nouns. Fifth, we contemplate the networks of 30 pieces of BBC news. A major point of the network analysis is that the central and pivotal words turn up in the center, whereas less central and pivotal words show up around the corner. The network analysis provides us with links among the relevant keywords. Sixth, we attempt to look into clusters in which similar key words form a cohesive group, depending on their similarities. Finally, we provide a cluster and the map of clusters.

### 2. RESULTS

#### 2.1 Contents of BBC News

This section is devoted to providing the first sentences of data (BBC news). Table 1 shows the main contents of BBC news:

**Table 1** Contents of 20 Pieces of BBC News

Number	Main Contents of BBC News
1	Two years on from Russia's full-scale invasion...
2	If you want to go to a concert in Kharkiv thes...
3	Russia is struggling to provide ammunition and...
4	Nato Secretary-General Jens Stoltenberg has wa...
5	Nearly two years since the Russian invasion of...
6	A new Russian school textbook has been...

7	The surprise announcement comes as a bill in C...
8	Three Ukraine-based Russian paramilitary group...
9	It was a dark night when the attack happened...
10	Nearly two years since the Russian invasion of...
11	Vladimir Putin has been in power since 2000...
12	The University of Hull has been supporting an ...
13	An animal lover is preparing to travel to Ukra...
14	Voronezh Governor Alexander Gusev said Maj Gen...
15	A Ukraine aid charity that was at threat of cl...
16	Moscow has launched a wide-ranging campaign te...
17	Three Ukraine-based Russian paramilitary group...
18	Apostolic Nuncio Visvaldas Kulbokas was told K...
19	Donald Trump will not fund Ukraine's fight aga...
20	It translates as "crooked horn", but President...

We included only the first sentences of 20 pieces of BBC news for the reason of space. We preprocessed the original text and abstracted only nouns. The so-called key nouns play a pivotal role in the network analysis, the frequency analysis, the cluster analysis, and the word cloud analysis.

## 2.2. Abstracting Nouns

The goal of this section is to preprocess the original text. We eliminated verbs, adjectives, adverbs, prepositions, etc. and abstracted only nouns for four analyses (the network analysis, the word cloud analysis, the frequency analysis, and the cluster analysis):

**Table 2** Preprocess of the Original Text

<b>Abstracting Nouns</b>
'years', 'Russia', 's', 'invasion', 'capitals', 'backing', 'Kyiv', 'January', 'European', 'Union', 'package', 'grants', 'loans', 'Ukraine', 'government', 'services', 'EU', 'target', 'shells', 'beginning', 'month', 'EU', 'diplomats', 'plans', 'top', 'Peace', 'Facility', 'weapons', 'Kyiv', 'Nato', 'year', 'members', 'alliance', 'target', 'output', 'defence', 'EPA', 'Medics', 'soldierEPA', 'politicians', 'support', 'Ukraine', 'ground', 'blood', 'weapons', 'balance', 'Europe', 'Ukraine', 'aggression', 'stocks', 'ammunition', 'weapons', 'Europe', 'Ukraine', 'UK', 'Foreign', 'Secretary', 'David', 'Cameron', 'House', 'Lords', 'week', 'weapons', 'systems', 'expense', 'expiry', 'dates', 'allies', 'countries', 'Eastern', 'Europe', 'legacy', 'ammunition', 'suitable', 'weapons', 'stocks', 'war', 'leaders', 'weapons', 'Kyiv', 'Ukraine', 'talks', 'Fiona', 'Hill', 'future', 'security', 'row', 'Germany', 'Taurus', 'range', 'miles', 'UK', 'Storm', 'Shadows', 'Ukraine', 'Many', 'allies', 'Taurus', 'Ukraine', 'chance', 'lines', 'German', 'Chancellor', 'Olaf', 'Scholz', 'cities', 'escalation', 'concert', 'Kharkiv', 'days', 'Ukraine', 'city', 'kilometres', 'border', 'mass', 'gatherings', 'start', 'Russia', 's', 'invasion', 'events', 'place', 'years', 'silence', 'Kharkiv', 'National', 'Opera', 'Ballet', 'sound', 'life', 'Kharkiv', 'life', 'theatre', 'director', 'Ihor', 'Touluzov', 'Demand', 'kind', 'event', 'Singing', 'streets', 'bunker', 'theatre', 'auditorium', 'flights', 'stairs', 'dress', 'circle', 'chandeliers', 'champagne', 'lot', 'sound', 'music', 'stage', 'spotlights', 'rows', 'seats', 'company', 'singers', 'dancers', 'musicians', 'audience', 'hall', 'feeling', 'stage', 'lots', 'people', 'Natalia', 'Babarok', 'artists', 'Natalia', 'Babarok', 'audiences', 'artists', 'stage', 'handful', 'times', 'February', 'fire', 'curtain',

'props', 'bits', 'scenery', 'stage', 'throne', 'bicycle', 'racks', 'costumes', 'weeks', 'scale', 'invasion', 'troops', 'missile', 'Chunks', 'stone', 'side', 'building', 'windows', 'roof', 'times', 'staff', 'flames', 'risk', 'life', 'remains', 'troops', 'city', 'rocket', 'Belgorod', 'Russia', 'seconds', 's', 'lot', 'people', 'place', 'director', 'Just', 'month', 'barrage', 'missiles', 'areas', 'city', 'people', 'Seven', 'civilians', 'February', 'oil', 'depot', 'drone', 'attack', 'fire', 'street', 'children', 'Artists', 'action', 'theatre', 'February', 'Volodymyr', 'Kozlov', 'Thousands', 'Kharkiv', 'residents', 'metro', 'explosions', 'Russia', 'ammunition', 'weapons', 'war', 'Ukraine', 'officials', 'challenges', 'equipment', 'material', 'official', 'concerns', 'provision', 'weapons', 'Ukraine', 'war', 'year', 'supply', 'ammunition', 'arms', 'looks', 'factor', 'Russia', 'ammunition', 'production', 'capabilities', 'needs', 'Ukraine', 'conflict', 'official', 'Moscow', 'supply', 'sources', 'ammunition', 'weapons', 'term', 'solution', 'impact', 'sanctions', 'cause', 'Sanctions', 'hard', 'delays', 'costs', 'inability', 'access', 'components', 'Russia', 'production', 'systems', 'repairs', 'systems', 'term', 'consequences', 'quality', 'weapons', 'Russia', 'advances', 'town', 'Avdiivka', 'hand', 'battlefield', 'officials', 'Moscow', 's', 'ability', 'successes', 'cost', 'terms', 'casualties', 'Russia', 'shells', 'rate', 'Ukraine', 'times', 'estimates', 'Russia', 'prisoners', 'war', 'Avdiivka', 'Dozens', 'Russian', 'troops', 'die', 'Ukraine', 'air', 'strike', 'Zelensky', 'warns', 'deficit', 'weapons', 'claims', 'supply', 'pessimism', 'quarters', 'Ukraine', 's', 'prospects', 'failure', 'year', 'counter', 'comments', 'officials', 'attempt', 'fact', 'Moscow', 'problems', 'predictions', 'Russia', 'missiles', 'weapons', 'official', 'Russia', 'sources', 'arms', 'drones', 'missiles', 'Iran', 'ammunition', 'stocks', 'North', 'Korea', 'quality', 'negotiations', 'visit', 'Russia', 's', 'Defence', 'Minister', 'Sergei', 'Shoigu', 'North', 'Korea', 'official', 'Russia', 'equipment', 'defence', 'industry', 'countries', 'India', 'weaponry', 'forces', 'air', 'force', 'year', 'claims', 'weapons', 'anniversary', 'Russia', 'invasion', 'Ukraine', 'February', 'assessment', 'officials', 'Russia', 'goals', 'Ukraine', 'Russia', 'plan', 'term', 'manpower', 'resources', 'difference', 'neighbour', 'Russia', 'ammunition', 'Ukraine', 'Russia', 'limits', 'supply', 'possibility', 'Ukraine', 'allies', 'support', 'supply', 'arms', 'looks', 'factor', 'war', 'outcome', 'US', 'package', 'Ukraine', 'Congress', 'concern', 'Russia', 'supply', 'Ukraine', 'support', 'Nato', 'Secretary', 'Jens', 'Stoltenberg', 'US', 'failure', 'assistance', 'Ukraine', 'impact', 'battlefield', 'Congress', 'aid', 'package', 'Republicans', 'House', 'Representatives', 'aid', 'package', 'Kyiv', 'Russia', 'attacks', 'cities', 'months', 'fighting', 'forces', 'ruins', 'Avdiivka', 'frostbite', 'toll', 'troops', 'man', 'fist', 'split', 'peach', 'rock', 'army', 'surgeon', 'fingers', 'case', 'frostbite', 'morning', 'makeshift', 'cottage', 'ruins', 'front', 'line', 'town', 'Avdiivka', 'Il', 'hands', 'surgeon', 'enough', 'weapons', 'Vadym', 'Injured', 'Ukrainian', 'soldier', 'Russia'

We included the part of the preprocessed data for the reason of space. As can be seen from Table 2, we eliminated adjectives, adverbs, verbs, prepositions, etc. from the original text. Simply put, we abstracted only nouns from the original text. This process is for the frequency analysis, the network analysis, the word cloud analysis, and the cluster analysis, namely the big data analysis.

### 2.3 The Frequency Analysis

This section focuses on providing the term frequency analysis of the relevant nouns. As exemplified in Table 3, the term frequency of the relevant nouns occurred in descending order:

**Table 3** The Term Frequency

Number	Word	Frequency
1	Ukraine	220
2	Russia	209
3	war	126
4	Putin	114
5	president	60
6	years	60
7	Vladimir	58
8	people	54

9	forces	53
10	year	51
11	NATO	49
12	defense	48
13	invasion	47
14	Moscow	44
15	time	42
16	children	38
17	troops	35
18	US	34
19	weapons	34
20	city	33
21	minister	32
22	February	31
23	Volodymyr	31
24	world	29
25	kremlin	28
26	Kyiv	28
27	Avdiivka	27
28	man	27
29	Europe	26
30	drones	25
31	Russian	25
32	region	24
33	Russians	24
34	aid	23
35	attack	23
36	group	23
37	security	23
38	term	23
39	ammunition	22
40	part	22
41	state	22
42	life	21
43	army	20
44	exercise	20
45	home	20

It must be stressed that the noun *Ukraine* obtains the highest term frequency. More specifically, the word obtains 220 tokens. This in turn indicates that this noun was the most widely used in 30 pieces of BBC news. It is worth observing is that the word *Ukraine* is followed by the word *Russia*. To be more specific, the word *Russia* obtains 209 tokens (the second highest). We take it as indicating that it was the second most frequently used in 30 pieces of BBC news. Particularly noteworthy is that the word *war* obtains 126 tokens (the third highest), thereby counting as the third most important in 30 pieces of BBC news. It is worth noticing that the word *war* is followed by the name *Putin*. The latter obtains 114 tokens and it ranks fourth in 30 pieces of BBC news. This in turn implies that the name *Putin* was the fourth most frequently occurred one in 30 pieces of BBC news. It therefore seems reasonable to contend that the word *Ukraine* was the most frequently occurred one, followed by the word

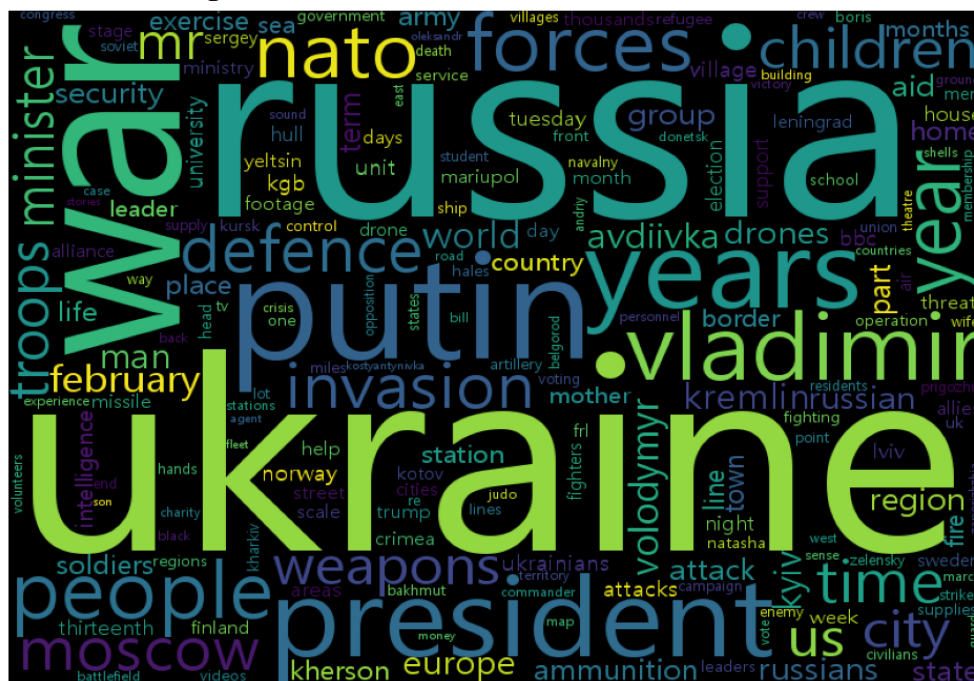
## International Journal of Applied Engineering & Technology

*Russia*, the word *war*, and the name *Putin*, in that order. It is worth noting that the word *NATO* obtains 49 tokens and that it ranks eleventh. More importantly, the word *troops* obtains 35 tokens and it ranks seventeenth. It should be pointed out that the word *Moscow* obtains 44 tokens and that it ranks fourteenth. We thus conclude that the noun *Ukraine* was the most widely occurred one in 30 pieces of BBC news.

### 2.4 The Word Cloud of 30 Pieces of BBC News

In this section, we explore the so-called word cloud of 30 pieces of BBC news. The word cloud is the reflection of the central and pivotal words in documents. In this section, we provide the word cloud of the pivotal and central words, as shown in Figure 1. It is important to mention that the word *Ukraine* is the most pivotal one since it is biggest in the word cloud. In other words, this word was the most widely used and important one. Thus, this word is deemed to be significant in 30 pieces of BBC news. It is also important to note that the word *Russia* is also pivotal and central since it is the second biggest. That is to say, it is one of the most centric words in 30 pieces of BBC news. We take it as meaning that this word is the second most significant one in 30 pieces of BBC news. It should be noted that the word *Russia* is followed by the word *war*. The latter is the third biggest, hence referring to a key noun. It is particularly noteworthy that the name *Putin* is the fourth biggest. Thus, it counts as one of the pivotal and centric words in 30 pieces of BBC news. We thus conclude that the noun *Ukraine* is deemed to be the most significant in 30 pieces of BBC news.

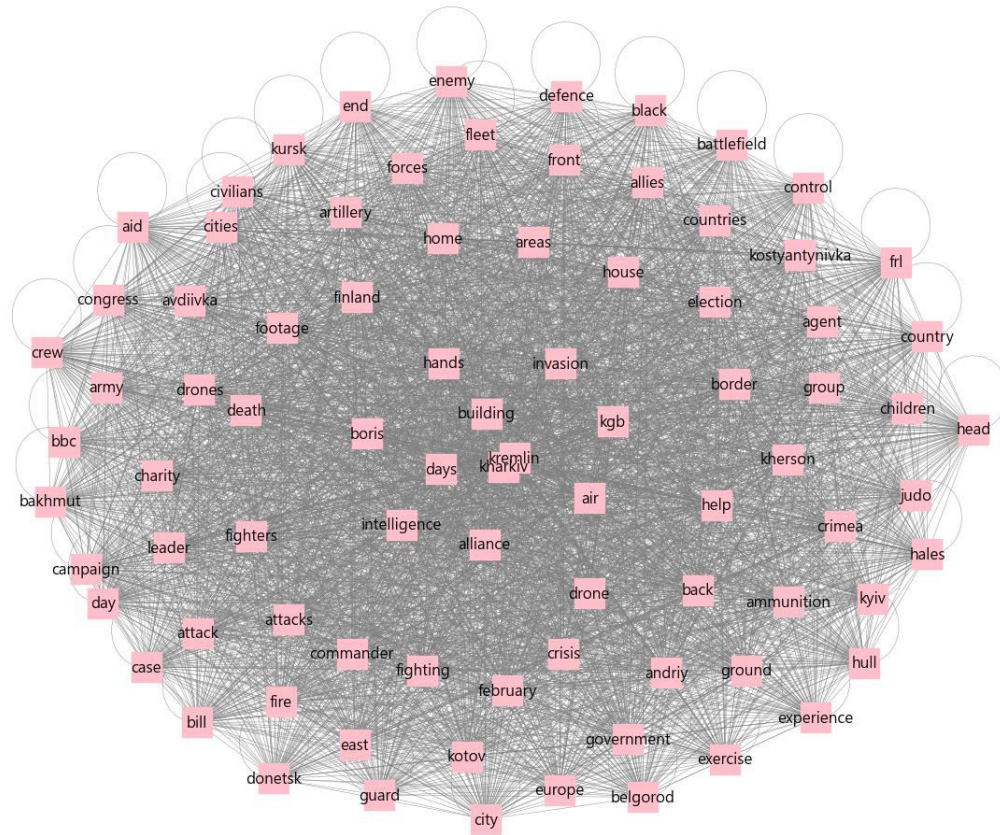
Figure 1 The Word Cloud of 30 Pieces of BBC News



### 2.5. The Network Analysis

In what follows, we probe into the networks of the relevant words that occurred in 30 pieces of BBC news. The several pivotal and significant keywords turn up in the center, whereas less pivotal and significant words show up at the corner in the map. It is interesting to observe that the words *Kremlin* and *Kharkiv* occur in the center of the map. These words indicate two important places. The word *Kharkiv* refers to one of cities in Ukraine. It is worth noting that these two places were the most significantly used in the map. Note, however, that the words *enemy*, *Europe*, *campaign*, *army*, *Congress*, *drones*, *aid*, etc. turn up at the corner of the map. They indicate less significant ones in the network analysis. More importantly, the words *invasion*, *KGB*, *alliance*, etc. show up around the center of the map, thus referring to key words. We thus conclude that the words *Kremlin* and *Kharkiv* are deemed to be significant and centric ones in the network analysis.

Figure 2 The Map of Networks



2.6 The Cluster Analysis

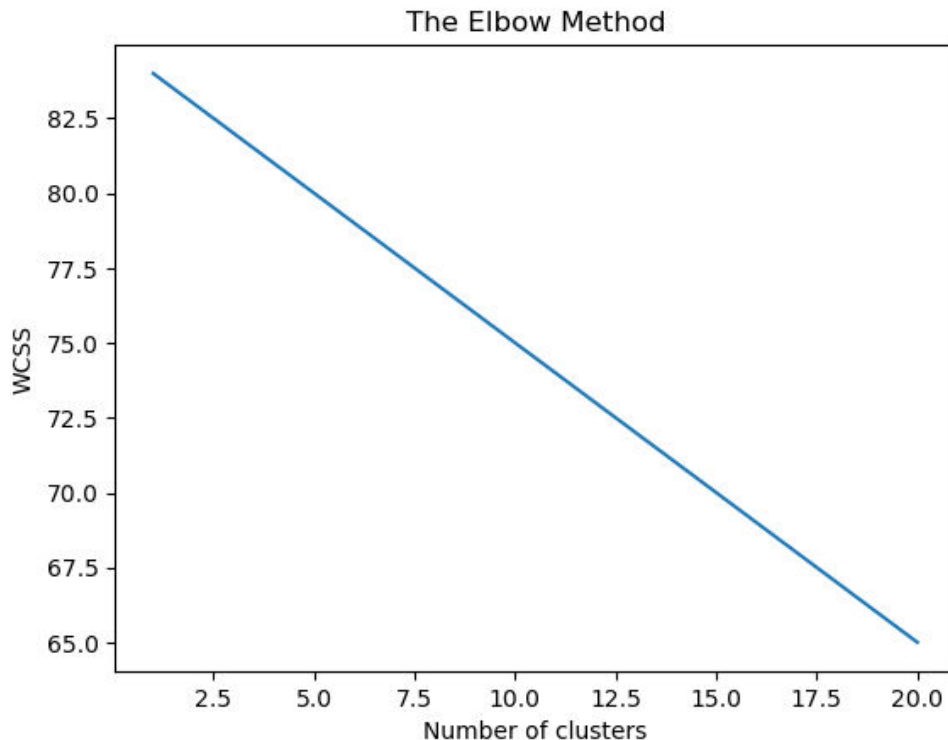
In the following, we aim at going over the so-called cluster. The term cluster refers to forming a cohesive group by grouping the related words. In this section, we abstracted a group (cluster 1) from 30 pieces of BBC news. Table 4 shows a cluster in which the relevant words form a cohesive group:

Table 4 the Cluster of 30 Pieces of BBC News

The Cluster of Words
Aggression, alliance, allies, ammunition, blood, Cameron, capitals, chance, chancellor, cities, dates, David, defense, diplomats, eastern, Europe, European, expense, expiry facility, foreign, future, German, Germany, government, grants, invasion, January, leaders, legacy, loans, month, NATO, output, package, peace, plans, politicians, range, Russia, secretary, security, services, shadows, shells, stocks, storm, support, systems, talks, target, top, UK, Ukraine, cast, union, war, week, years

As exemplified in Table 4, the words *aggression, alliance, allies, ammunition, blood, capitals, chance, chancellor, cities, dates, defense, diplomats, NATO, Ukraine, peace, Russia*, etc. form a cohesive group by linking the relevant words. Python produced many clusters, but we included cluster 1 since it contained many words. The following graph shows the distance among words. The so-called distance is calculated with the Euclidean distance. Note that the more there are more clusters, the closer the Euclidean distance among words is:

**Figure 3** the Distance among Words



The term WCSS refers to the Euclidean distance. As can be seen from the graph, the more there are more clusters, the closer the Euclidean distance among words is. It seems thus reasonable to contend that as illustrated in Table 4, those words form a cohesive group since the Euclidean distance among them is closer than the other words that occurred in 30 pieces of BBC news. As exemplified in Figure 3, when the number of clusters is 20, the Euclidean distance among words is the lowest. Note, however, that when the number of the cluster is about 1, the Euclidean distance is the highest. For the big data analysis, see Kang (2023a, 2023b, 2023c, 2023d, 2023e, 2023f, 2024a).

### 3. CONCLUSION

To sum up, we have analyzed 30 pieces of BBC news broadcasted in 2024 with respect to the Ukraine war. In section 2.3, we have contended that the word *Ukraine* was the most frequently occurred one, followed by the word *Russia*, the word *war*, and the name *Putin*, in that order. We have also contended that the word *NATO* obtains 49 tokens and that it ranks eleventh. In section 2.4, we have maintained that the word *Ukraine* is the most pivotal one since it is biggest in the word cloud. In other words, this word was the most widely used and important one. Thus, this word is deemed to be significant in 30 pieces of BBC news. The word *Russia* is also pivotal and central since it is the second biggest in the word cloud. That is to say, it is one of the most centric words in 30 pieces of BBC news. In section 2.5, we have argued that the words *Kremlin* and *Kharkiv* are deemed to be significant and centric ones in the network analysis. In section 2.6, we have shown that the words *aggression*, *alliance*, *allies*, *ammunition*, *blood*, *capitals*, *chance*, *chancellor*, *cities*, *dates*, *defense*, *diplomats*, *NATO*, *Ukraine*, *peace*, *Russia*, etc. form a cohesive group by linking the relevant words.

### REFERENCES

- [1] Kang, N. (2023a). K-Pop in BBC News: A Big Data Analysis. *Advances in Social Sciences Research Journal* 10(2), 156-169.
- [2] Kang, N. (2023b). K-Dramas in Google: A NetMiner Analysis. *Transaction on Engineering and Computing Sciences* 11(1), 193-216.

---

*International Journal of Applied Engineering & Technology*

---

- [3] Kang, N. (2023c). A Comparative Analysis of Tolerate and Put up with in the COCA. *Semiconductor and optoelectronics* 42(1): 1468-1476.
- [4] Kang, N. (2023d). Sure of and Sure about in Corpora and ChatGPT. *Journal of Harbin Engineering University* 44(7): 1347-1351.
- [5] Kang, N. (2023e). Turn out adj and Turn out to be adj in the Now Corpus and ChatGPT. *Journal of Harbin Engineering University* 44(8): 825-831.
- [6] Kang, N. (2023f). Care for and Like in Corpora and ChatGPT. *Semiconductor and optoelectronics* 42(2): 188-198.
- [7] Kang, N. (2024a). A Big Data Analysis of a Hot Political Issue. *Studies in Linguistics* 70: 149-165