*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 11 Issue 2 February 2023

UN Sanctions on North Korea: A Big data Analysis

Namkil Kang

Far East University, South Korea

somerville@hanmail.net

DOI: 10.26821/IJSHRE.11.2.2023.110212

ABSTRACT

The ultimate goal of this paper is to analyze 40 articles of Google regarding UN sanctions on North Korea. This research was conducted by the software package NetMiner. A point to note is that one word has the highest frequency (1,243 tokens) and the highest proportion (0.521). A further point to note is that topic 12 was the most frequently used one, followed by topic 8, topic 11, and topic 15, in that order. It is worthwhile pointing out, on the other hand, that the word *sanction* was the most widely used one in 40 articles of Google, followed by the word *North Korea*, the word *missile*, the word *US*, and the word *UN*, in that order. When it comes to the distribution of brokerage, the liaison in 40 articles of Google was the most widely used one. Additionally, it is worthwhile noting that the Euclidean distance between the word *North* and *Pyongyang* is 2.449 (the lowest) and the Euclidean distance between *Security Council* and *North* is also 2.449 (the lowest). What this fact suggests is that they (*North* vs. *Pyongyang* and *Security Council* vs. *North*) have the highest similarity.

Keywords: UN sanctions, North Korea, big data, topic, keyword, brokerage, map

1. INTRODUCTION

The main purpose of this paper is to analyze 40 articles of Google regarding UN sanctions on North Korea. These articles were written from 29th, 11, 2019 to 21st, 2, 2023. Most importantly, this research was conducted by the software package NetMiner. First, we aim to explore the frequency and proportion of common nouns and proper nouns, which is mainly used for a keyword analysis. Second, we also aim to provide the frequency, proportion, and cumulative proportion of the total words used in 40 articles of Google. Third, we aim at inquiring into 15 topics consisting of 40 articles and main keywords consisting of 15 topics. Also, we aim to see how often each topic occurs in 40 articles of Google. Fourth, we aim to consider how often main keywords turn up in 40 articles of Google. Fifth, we aim to investigate the distribution of brokerage (the term of NetMiner) that means connecting clusters. More specifically, the distribution of the so-called coordinator, representative, gatekeeper, consultant, and liaison is dealt with. Also, the maps of brokerage including in-degree and out-degree are provided. Sixth, we attempt to inquire into structural equivalence (the term of NetMiner) that means that two nodes are equivalent if they are linked to identical neighbors. Especially, we look into the Euclidean distance among keywords. If two nodes have a low value, they have a high similarity and vice versa. Finally, a map capturing the Euclidean distance among nodes is provided.

2. Results

2.1. Word information

This section is focused on examining the frequency and proportion of common nouns and proper nouns. Also, this section centers on providing information on the frequency and proportion of the total words used in 40 articles of Google. Table 1 shows the frequency, proportion, and cumulative proportion of common nouns and proper nouns used in 40 articles of Google:

Table 1 Word information

Value	Frequency	Proportion	Cumulative Proportio

			n
Common Noun	1464	0.614	0.614
Proper Noun	920	0.386	1
Total	2384	1	

Perhaps it is worthwhile pointing out that 1,464 common nouns occurred in 40 articles of Google. Their proportion and cumulative proportion are 0.614. It is worth observing, on the other hand, that 920 proper nouns appeared in 40 articles. It is interesting to note that their proportion is 0.386. Finally, it should be pointed out that the total number of common nouns and proper nouns is 2,384.

Now attention is paid to word information:

Table 2 Word information

Value	Frequency	Proportion	Cumulative Proportion
1.0	1243	0.521	0.521
2.0	364	0.153	0.674
3.0	168	0.07	0.745
4.0	106	0.044	0.789
5.0	78	0.033	0.822
6.0	45	0.019	0.841
7.0	42	0.018	0.858
8.0	37	0.016	0.874
9.0	35	0.015	0.888
10.0	22	0.009	0.898
11.0	15	0.006	0.904
12.0	18	0.008	0.911
13.0	17	0.007	0.919
14.0	12	0.005	0.924
15.0	14	0.006	0.93
16.0	19	0.008	0.938
17.0	10	0.004	0.942
18.0	9	0.004	0.945
19.0	14	0.006	0.951
20.0	11	0.005	0.956
21.0	6	0.003	0.958
22.0	7	0.003	0.961
23.0	3	0.001	0.963
24.0	7	0.003	0.966
25.0	5	0.002	0.968
26.0	2	0.001	0.969
27.0	3	0.001	0.97
28.0	2	0.001	0.971
29.0	3	0.001	0.972
32.0	4	0.002	0.974
33.0	1	0	0.974
34.0	2	0.001	0.975
35.0	2	0.001	0.976
36.0	2	0.001	0.977
37.0	3	0.001	0.978
38.0	3	0.001	0.979
39.0	2	0.001	0.98

40.0	1	0	0.98
41.0	1	0	0.981
42.0	1	0	0.981
43.0	2	0.001	0.982
44.0	1	0	0.982
46.0	1	0	0.983
47.0	1	0	0.983
48.0	1	0	0.984
50.0	2	0.001	0.984
51.0	1	0	0.985
52.0	1	0	0.985
55.0	1	0	0.986
56.0	1	0	0.986
57.0	2	0.001	0.987
58.0	1	0	0.987
59.0	1	0	0.988
60.0	1	0	0.988
63.0	1	0	0.989
65.0	1	0	0.989
68.0	1	0	0.99
70.0	1	0	0.99
71.0	1	0	0.99
74.0	2	0.001	0.991
79.0	1	0	0.992
82.0	1	0	0.992
84.0	1	0	0.992
86.0	2	0.001	0.993
88.0	1	0	0.994
95.0	1	0	0.994
102.0	1	0	0.995
107.0	1	0	0.995
109.0	1	0	0.995
110.0	1	0	0.996
119.0	1	0	0.996
127.0	1	0	0.997
131.0	1	0	0.997
132.0	1	0	0.997
155.0	1	0	0.998
167.0	1	0	0.998
175.0	1	0	0.999
196.0	1	0	0.999
543.0	1	0	1
667.0	1	0	1
Total	2384	1	

It is worth pointing out that one word has the highest frequency (1,243 tokens) and the highest proportion (0.521). It is interesting to observe two words. It is important to note that one word is followed by two words in their frequency. More specifically, there occur two words in 40 articles of Google and their frequency is 364 tokens (the second highest). Their proportion and cumulative proportion are 0.153 and 0.674, respectively. It is interesting to observe that there occur three words whose frequency is 168 tokens (the third highest). Finally, it should be noted, on the other hand, that four words appear in 40 articles and that their frequency is 106 tokens (the fourth highest). Additionally, their proportion and cumulative proportion are 0.044 and 0.789, respectively.

2.2. Topic information

The goal of this section is to examine 15 topics and their five keywords. Table 3 shows 15 topics and 5 keywords occurred in 40 articles of Google:

Table 3 Topic info

	1st Key word	**2nd Key word**	**3rd Key word**	**4th Keyword**	**5th Key word**
Topic-1	export	ban	vessel	import	product
Topic-2	sanction	North Korea	UN Security Council	year	resolution
Topic-3	state	North Korea	member	government	security
Topic-4	sanction	activity	year	Pyongyang	measure
Topic-5	Security Council	weapon	official	situation	aid
Topic-6	North Korea	UN	country	access	restriction
Topic-7	sanction	regime	Security Council	UN	action
Topic-8	US	South Korea	Japan	United States	policy
Topic-9	North Korea	sanction	Sanctions	talk	Guidance
Topic-10	right	sanction	report	measure	issue
Topic-11	China	Russia	sanction	UN Security Council	Ukraine
Topic-12	missile	test	resolution	response	launch
Topic-13	DPRK	regulation	good	activity	service
Topic-14	North	program	sanction	weapon	Koreas
Topic-15	person	entity	asset	restriction	EU

It is worth noting that topic 2 includes the keywords *sanction*, *North Korea*, *UN Security Council*, *year*, and *restriction*. As illustrated in Table 3, the word *sanction* is the first keyword, thereby indicating that it was the most occurred one in topic 2. It is important to mention, on the other hand, that the keywords *North Korea*, *UN*, *country*, *access*, and *restriction* consist of topic 6. It is also worthwhile noting that topic 12 contains *missile*, *test*, *resolution*, *response*, and *launch*. As alluded to in Table 3, the first keyword is the word *missile*, which in turn suggests that it was the most widely used one in topic 12.

Now let us consider how often each topic occurred in 40 articles of Google:

Table 4 Topic frequency

	# of sentences
Topic-1	78
Topic-2	31
Topic-3	92
Topic-4	46
Topic-5	42
Topic-6	41
Topic-7	75
Topic-8	141
Topic-9	48
Topic-10	94
Topic-11	120
Topic-12	164
Topic-13	82
Topic-14	87
Topic-15	111

It is interesting to consider topic 12. It is probably worthwhile pointing out that topic 12 occurred 164 times in 40 articles of Google (the highest). As observed earlier, the keywords *missile*, *test*, *resolution*, *response*, and *launch* constitute topic 12. It is particularly noteworthy that topic 12 is followed by topic 8. To be more specific, topic 8 appeared 141 times in 40 articles (the second highest). As indicated in Table 3, the five keywords *US*, *South Korea*, *Japan*, *United States*, and *policy* consist of topic 8. It is interesting to note that topic 11 appeared 120 times (the third highest). It must be stressed, on the other hand, that topic 15 occurred 111 times in 40 articles of Google. It seems thus reasonable to conclude that topic 12 was the most frequently used one, followed by topic 8, topic 11, and topic 15, in that order.

2.3. Frequency of words

In what follows, we aim at considering how often major words occurred in 40 articles of Google. Our list was cut off in the top 30:

Table 5 Major words and their frequency

Number	Word	Frequency
1	sanction	557
2	North Korea	453
3	missile	164
4	US	148
5	UN	137
6	resolution	116
7	China	115
8	test	113
9	program	109
10	country	101
11	weapon	94
12	North	93
13	DPRK	89
14	Pyongyang	85
15	person	83
16	Security Council	81
17	activity	80
18	Russia	76
19	South Korea	70
20	right	67
21	measure	66
22	regime	65
23	year	62
24	UN Security Council	61
25	entity	60
26	Koreas	54
27	Japan	54
28	response	53
29	export	53
30	launch	51

It is interesting to note that the word *sanction* occurred 557 times in 40 articles of Google (the highest frequency). This in turn implies that it was the most widely used one. It is interesting to observe the word *North Korea*. It appeared 453 times in 40 articles (the second highest). We take this as suggesting that the word *North Korea* was the second most occurred one. It is worth observing, on the other hand, that the word *missile* occurred 164 times in 40 articles of Google (the third highest). It must also be emphasized that the word *US* appeared 148 times (the fourth highest). It seems thus reasonable to assume that the word *US* was the fourth most widely used one in 40 articles. Perhaps it is worthwhile saying that the word *UN* occurred 137 times in 40 articles of Google. It can thus be inferred that the word *sanction* was the most frequently used one in 40 articles of Google, followed by the word *North Korea*, the word *missile*, the word *US*, and the word *UN*, in that order. We thus conclude that the word *sanction* was the most widely used one in 40 articles of Google.

2.4. Brokerage

In the following, we aim at looking into the distribution of brokerage (the term of NetMiner). The term brokerage refers to the connection of clusters. It can be sub-classified into the so-called coordinator, representative, gatekeeper, consultant, and liaison:

Table 6 Distribution of Brokerage (Out-degree)

MEASURES	VALUE					
	COORDINATOR	GATEKEEPER	REPRESENT	ITINERANT	LIAISON	TOTAL
MEAN	0.048	0.376	0.457	2.484	27.689	31.054
STD.DEV.	0.229	1.759	1.974	47.055	454.71	501.679
MIN.	0	0	0	0	0	0
MAX.	3	24	29	1,795	17,299	19,094

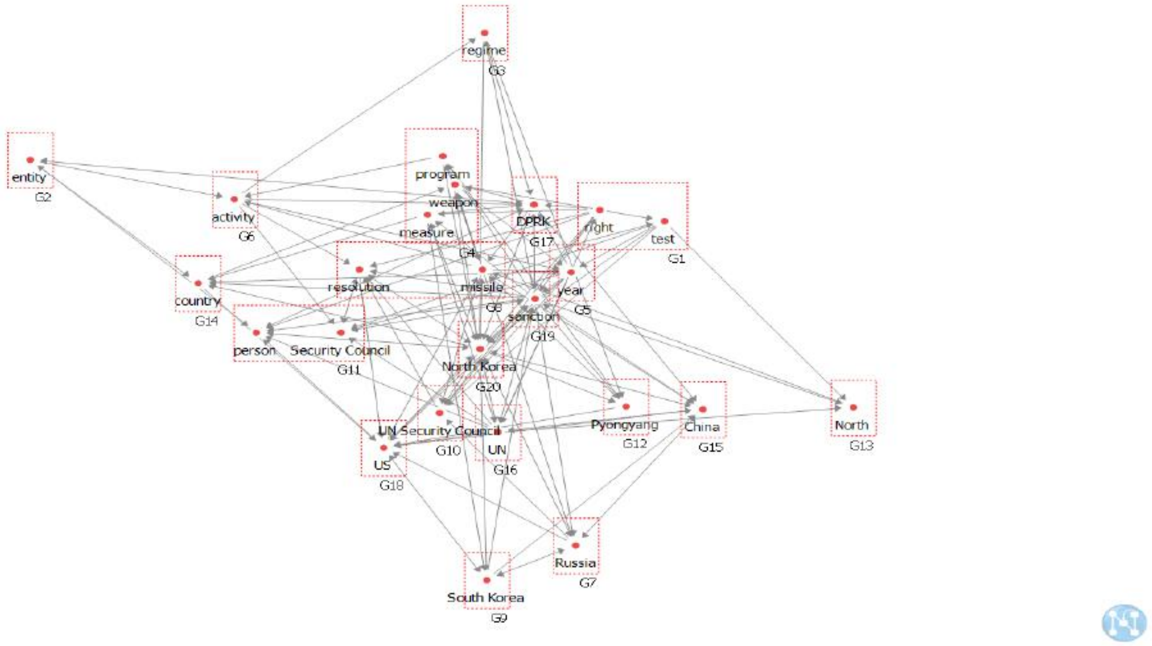
To begin with, let us define the term coordinator. The term coordinator refers to the role of mediating between two nodes that belong to a group. On the other hand, the term representative indicates that it receives a link from a node within the same group and gives it to another node that belongs to a different group. The term gatekeeper indicates that it receives a link from a node that belongs to a different group and gives it to another node in the same group. The term itinerant indicates that it receives a link from a node in a different group and gives it to another node in a different group. Finally, the term liaison refers to the role of mediating between two nodes in different groups. Quite interestingly, the max value of a coordinator is 6, whereas that of a gatekeeper is 21. Here the max value refers to the total number of brokerage. On the other hand, the max value of a represent is 31, whereas that of an itinerant is 1,691. Finally, the max value of a liaison is 17,403, which is the highest. We thus conclude that the total number of brokerage is 19,094 and the proportion of the liaison is the highest.

Now attention is paid to brokerage (Out-degree). Out-degree brokerage refers to the connector of links go ing outside:

It is interesting to point out that the total number of brokerage (Out-degree) is 19,094. When it comes to the coordinator, its max value is 3. On the other hand, the max value of a gatekeeper is 24. This in turn indicates that the number of the gatekeeper is eight times higher than that of the coordinator. Talking about the represent, its number is slightly higher than the number of the gatekeeper. It is interesting to consider the itinerant. There is a dramatic increase in its number. The max value of the itinerant is 1,795, which ranks second. Finally, it is important to mention that the max value of the liaison is 17,229 (the highest). As observed earlier, the liaison undertakes the task of mediating between two nodes in different groups. We thus conclude that the liaison in 40 articles was the most widely used one.

Now let us turn to the map of brokerage occurred in 40 articles of Google. We got rid of many nodes for the reason of space and included core nodes only:

Figure 1 Map of brokerage (In-degree)



As exemplified in Figure 1, we included 25 nodes only. Note, however, that there are too many links among nodes, so it is very difficult for us to capture all brokerage. Yet, NetMiner provides information related to Figure 1:

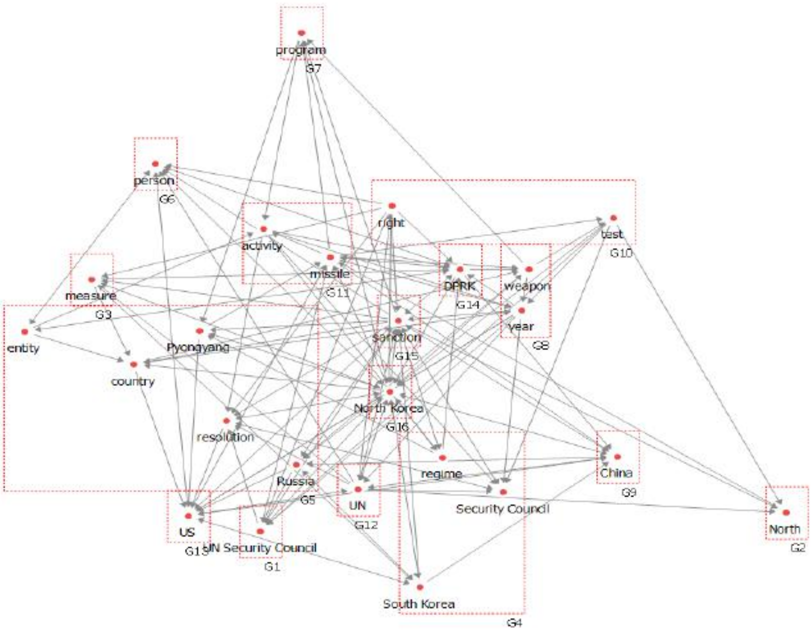
Table 7 Distribution of brokerage

MEASURES	VALUE					
	COORDINATOR	GATEKEEPER	REPRESENT	ITINERANT	LIAISON	TOTAL
MEAN	0	0.12	0.08	0.44	20.72	21.36
STD.DEV.	0	0.588	0.392	1.023	31.679	32.478
MIN.	0	0	0	0	2	2
MAX.	0	3	2	5	145	150

It is interesting to consider the coordinator. As shown in Table 7, there is no coordinator in Figure 1. On the other hand, there are three gatekeepers in Figure 1. In Figure 1, the number of the represent is 2, whereas that of the itinerant is 5. Quite interestingly, the number of the liaison is 145.

Now let us consider the map of brokerage (Out-degree):

Figure 2 Map of brokerage (Out-degree)



Again, we included 25 nodes only for the reason of space. As indicated in Table 8, there is no coordinator in Figure 2. However, there are 6 gatekeepers in Figure 2. Additionally, there are 7 represents in Figure 2, whereas there are 8 itinerants. Finally, the number of the liaison is 142, which is the highest. For the map of similar expressions, see Kang (2022a, 2022b, 2022c, 2022d).

Table 8 Distribution of brokerage (Out-degree)

MEASURES	VALUE					
	COORDINATOR	GATEKEEPER	REPRESENT	ITINERANT	LIAISON	TOTAL
MEAN	0	0.4	0.56	0.68	19.72	21.36
STD.DEV.	0	1.265	1.627	1.667	30.986	32.478
MIN.	0	0	0	0	2	2
MAX.	0	6	7	8	142	150

2.5. Structural equivalence

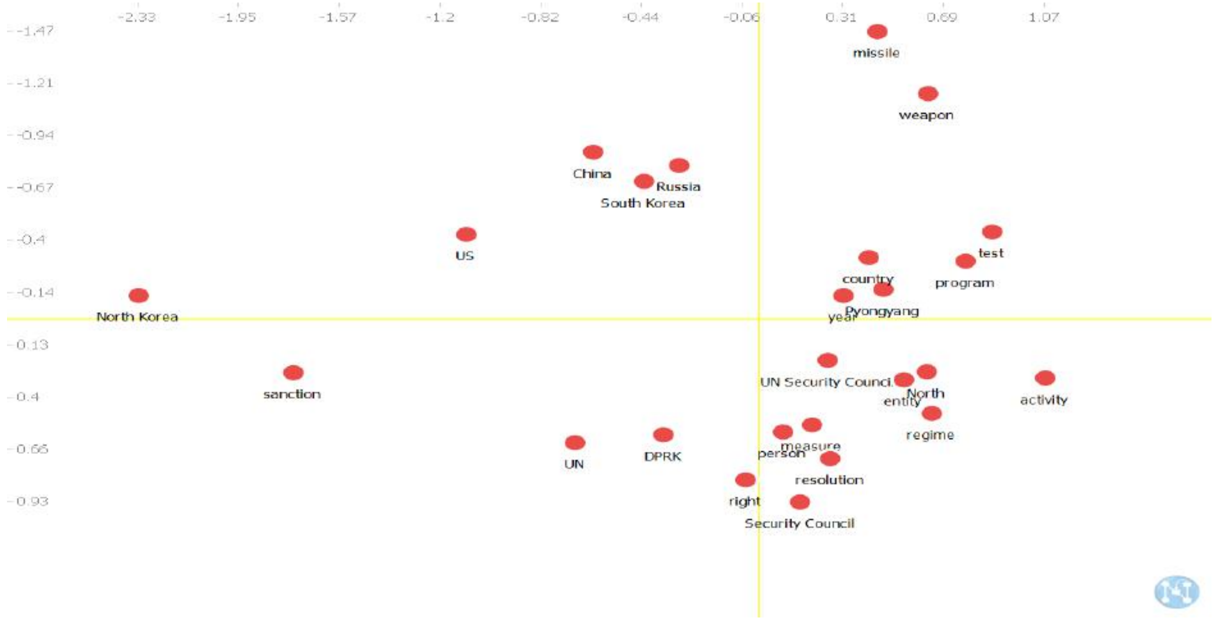
In what follows, we aim to investigate structural equivalence (the term of NetMiner) that indicates that two nodes are equivalent if they are linked to identical neighbors. Especially, we inquire into the Euclidean distance among major keywords:

Table 9 Euclidean distance

	China	DPRK	North	North Korea	Pyongyang	Russia	Security Council
China	0	4.359	3.742	4.583	3.464	2.828	4
DPRK	4.359	0	3.873	4.472	3.873	4.359	3.873
North	3.742	3.873	0	5	2.449	3.464	2.449
North Korea	4.583	4.472	5	0	5	4.583	4.796
Pyongyang	3.464	3.873	2.449	5	0	3.162	3.464
Russia	2.828	4.359	3.464	4.583	3.162	0	4
Security Council	4	3.873	2.449	4.796	3.464	4	0

It is important to note that the word *North* is the nearest to the word *Pyongyang*. More specifically, the Euclidean distance between the word *North* and *Pyongyang* is 2.449, which is the lowest. This in turn indicates that they have a high similarity. Nodes that have structural equivalence share a similarity with each other. Exactly the same can be said of the words *Security Council* and *North*. The Euclidean distance between them is also 2.449, which is the lowest. This in turn suggests that they exhibit a high degree of similarity. Note that the Euclidean distance between *China* and *China* itself is 0, but it is only one node. The Euclidean distance works for two nodes. It must also be stressed that the Euclidean distance between *China* and *Russia* is 2.828, which is the second lowest. This in turn means that two countries share a similarity with each other.

Now attention is paid to a map that shows the Euclidean distance among nodes:



Notice that *China* is near to *Russia*, but *China* is far from *North Korea*. Quite interestingly, the word *missile* is near to the word *weapon*, but they are far from the word *Security Council*. Note that the word *North Korea* is near to the word *sanction*, but they are far from the word *missile*. To sum up, nodes that have a high degree of similarity gather together, but nodes that have a low degree of similarity are far from one another.

3. Conclusion

To sum up, we have analyzed 40 articles of Google regarding UN sanctions on North Korea.

These data were analyzed by the software package NetMiner. In section 2.1, we have shown that one word has the highest frequency (1,243 tokens) and the highest proportion (0.521). In section 2.2, we have also shown that topic 12 was the most frequently used one, followed by topic 8, topic 11, and topic 15, in that order. In section 2.3, we have argued that the word *sanction* was the most frequently used one in 40 articles of Google, followed by the word *North Korea*, the word *missile*, the word *US*, and the word *UN*, in that order. In section 2.4, we have investigated the distribution of brokerage. Quite interestingly, the liaison in 40 articles of Google was the most widely used one. In section 2.5, we have contended that the Euclidean distance between the word *North* and *Pyongyang* is 2.449 (the lowest) and the Euclidean distance between *Security Council* and *North* is also 2.449 (the lowest). What this fact suggests is that they (*North* vs. *Pyongyang* and *Security Council* vs. *North*) have the highest similarity.

References

[1]. Kang, N. (2022a). A Comparative Analysis of Search for and Look for in Four Corpora. *Advances in Social Sciences Research Journal* 9 (3): 168-178.
[2]. Kang, N. (2022b). A Comparative Analysis of Impressed by and Impressed with in Two Corpora. *Theory and Practice in Language Studies* 12 (5): 819-827.
[3]. Kang, N. (2022c). On Speak to and Talk to: A Corpora-based Analysis. *Theory and Practice in Language Studies* 12 (7):1262-1270.
[4]. Kang, N. (2022d). On Speak with and Talk with: A Corpora-based Analysis. *International Journal of Social Science and Human Research* 5 (8): 3354-3360.