

North Korea in Big Data

Namkil Kang

Far East University, South Korea

\*Corresponding Author

NAMKIL KANG

ABSTRACT

The ultimate goal of this paper is to provide an analysis of 481 YouTube videos in connection with North Korea. A point to note is that the six-word expression was the most widely used one in 481 YouTube videos. A further point to note is that the word North Korea was the most widely used one in 481 YouTube videos, followed by the word war, the word country, the word missile, the word person, the word Russia, the word US, and the word China, in that order. It is worth pointing out, on the other hand, that topic 7 was the most widely used one in 481 YouTube videos, followed by topic 4, topic 1 (topic 3), and topic 6, in descending order. With respect to the frequency of YouTube videos, it is interesting to note that the word Facebook occurred in 232 YouTube videos (the highest). This in turn indicates that it was the most widely used one in 481 YouTube videos. Finally, it is worthwhile noting that the words News, drill, missile, etc. are directly linked to the word North Korea. It is also worthwhile pointing out that the word weapon is linked to the word test and the latter is linked to the word missile again.

Keywords: North Korea, YouTube, big data, keyword, visualization

INTRODUCTION

The main purpose of this paper is to analyze 481 YouTube videos in connection with North Korea. First, we inquire into the length of words used in 481 YouTube videos, their frequency, their proportion, and their cumulative proportion. Second, we look into the frequency of the relevant words occurred in 481 YouTube videos. By doing so, one can see the frequency of main words related to North Korea. Third, we examine 8 topics that are formed by 5 keywords. 5 keywords consist of each topic, which constitutes each video. Fourth, we investigate the term degree, which refers to the frequency of YouTube videos. Finally, we provide the visualization of which words are closely related to North Korea. This visualization provides the links between North Korea and the relevant words. From this it seems clear that what YouTubers think about North Korea. This paper is organized as follows. In section 3.1, we argue that the six-word expression was the most widely used one in 481 YouTube videos. In section 3.2, we further argue that the word North Korea was the most widely used one in 481 YouTube videos, followed by the word war, the word country, and the word missile, in that order. In section 3.3, we show that topic 7 was the most widely used one in 481 YouTube videos, followed by topic 4, topic 1 (topic 3), and topic 6. In section 3.4, we further show that the word Facebook occurred in 232 YouTube videos (the highest). This in turn suggests that it was the most widely used one in 481 YouTube videos. In section 3.5, we show that the words News, drill, missile, etc. are directly linked to the word North Korea.

METHODS

The main goal of this paper is to provide an analysis of 481 YouTube videos collected on 14, November, 2022. We collected and analyzed 481 YouTube videos in terms of the YouTube data collector and Net Miner. The main purpose of this paper is to answer the following questions: Can we provide information on word length? Can we show the frequency of words that are closely related to North Korea? Can we show 5 keywords and 8 topics related to North Korea? Can we show the frequency of 481 YouTube videos in which the relevant words occur. Can we show the picture of which words are closely related to North Korea?

RESULTS

Word length

The goal of this section is to analyze the length of words used in 481 YouTube Videos. Table 1 shows the frequency of words, their length, their proportion, and their cumulative proportion:

Table 1: Word length

Value	Frequency	Proportion	Cumulative Proportion
-------	-----------	------------	-----------------------

2.0	784	0.02	0.02
3.0	2739	0.07	0.09
4.0	5074	0.13	0.22
5.0	5611	0.144	0.364
6.0	5684	0.146	0.51
7.0	5043	0.129	0.639
8.0	4007	0.103	0.742
9.0	3102	0.079	0.821
10.0	2126	0.054	0.876
11.0	1458	0.037	0.913
12.0	955	0.024	0.937
13.0	631	0.016	0.954
14.0	381	0.01	0.963
15.0	269	0.007	0.97
16.0	253	0.006	0.977
17.0	144	0.004	0.98
18.0	117	0.003	0.983
19.0	86	0.002	0.986
20.0	75	0.002	0.987
21.0	53	0.001	0.989
22.0	49	0.001	0.99
23.0	42	0.001	0.991
24.0	30	0.001	0.992
25.0	32	0.001	0.993
26.0	25	0.001	0.993
27.0	22	0.001	0.994
28.0	24	0.001	0.995
29.0	17	0	0.995
30.0	13	0	0.995
31.0	9	0	0.996
32.0	11	0	0.996
33.0	6	0	0.996
34.0	13	0	0.996
35.0	10	0	0.997
36.0	9	0	0.997
37.0	12	0	0.997
38.0	8	0	0.997
39.0	4	0	0.997
40.0	7	0	0.998
41.0	3	0	0.998

42.0	7	0	0.998
43.0	5	0	0.998
44.0	3	0	0.998
46.0	3	0	0.998
47.0	2	0	0.998
48.0	4	0	0.998
49.0	2	0	0.998
50.0	3	0	0.998
51.0	2	0	0.998
52.0	1	0	0.999
53.0	1	0	0.999
54.0	1	0	0.999
55.0	1	0	0.999
57.0	2	0	0.999
58.0	1	0	0.999
60.0	2	0	0.999
61.0	2	0	0.999
62.0	1	0	0.999
63.0	1	0	0.999
64.0	1	0	0.999
65.0	3	0	0.999
66.0	1	0	0.999
68.0	1	0	0.999
69.0	2	0	0.999
70.0	3	0	0.999
71.0	1	0	0.999
72.0	3	0	0.999
78.0	1	0	0.999
79.0	1	0	0.999
80.0	1	0	0.999
83.0	2	0	0.999
85.0	1	0	0.999
90.0	1	0	0.999
93.0	1	0	0.999
94.0	1	0	0.999
95.0	1	0	0.999
96.0	1	0	0.999
98.0	1	0	1
99.0	1	0	1
101.0	2	0	1

105.0	2	0	1
107.0	1	0	1
109.0	1	0	1
111.0	1	0	1
115.0	1	0	1
116.0	1	0	1
117.0	1	0	1
131.0	1	0	1
138.0	1	0	1
141.0	1	0	1
145.0	1	0	1
161.0	1	0	1
193.0	1	0	1
245.0	1	0	1
272.0	1	0	1
Total	39028	1	

It is significant to note that the six-word expression has the highest frequency and the highest proportion. More specifically, its frequency is 5,684 tokens, its proportion is 0.146, and its cumulative proportion is 0.51. It is interesting to point out that the frequency of the five-word expression is the second highest. Its frequency is 5,611 tokens, its proportion is 0.144, and its cumulative proportion is 0.364. As can be seen from Table 1, the frequency of the four-word expression is the third highest (5,074 tokens). It is noteworthy that the frequency of the seven-word expression ranks fourth. To be more specific, its frequency is 5,043 tokens and its proportion is 0.129. Quite interestingly, the eight-word expression ranks fifth (4,007 tokens). Its proportion is 0.103 and its cumulative proportion is 0.742. From all of this, it is clear that the six-word expression was the most preferred one for YouTubers, followed by the five-word expression, the four-word expression, the seven-word expression, and the eight-word expression, in that order. We thus conclude that the six-word expression was the most widely used one in 481 YouTube videos.

### The Frequency of words related to North Korea

The goal of this section is to examine the frequency of words related to *North Korea*. Table 2 shows the frequency of major words which are closely related to *North Korea*:

**Table 2:** Frequency of major words occurred in 481 YouTube videos

Number	Words	Part of Speech(POS)	Frequency	Word Length
1	North Korea	Proper Noun	3456	11
2	war	Common Noun	3152	3
3	country	Common Noun	2574	7
4	missile	Common Noun	2549	7
5	person	Common Noun	2373	6
6	Russia	Proper Noun	2256	6
7	world	Proper Noun	2250	5
8	US	Proper Noun	1952	2
9	China	Proper Noun	1926	5
10	time	Common Noun	1782	4
11	ang	Common Noun	1729	3
12	Kim	Proper Noun	1646	3
13	news	Common Noun	1418	4
14	South Korea	Proper Noun	1369	11
15	Ukraine	Proper Noun	1246	7
16	year	Common Noun	1209	4

17	Korea	Proper Noun	1194	5
18	korea	Common Noun	1188	5
19	USA	Proper Noun	1165	3
20	America	Proper Noun	1136	7
21	Japan	Proper Noun	1123	5
22	man	Common Noun	1092	3
23	NK	Proper Noun	1047	2
24	weapon	Common Noun	1046	6
25	thing	Common Noun	1017	5
26	day	Common Noun	944	3
27	video	Common Noun	872	5
28	mo	Common Noun	783	2
29	North	Proper Noun	751	5
30	News	Proper Noun	747	4
31	life	Common Noun	744	4
32	way	Common Noun	744	3
33	God	Proper Noun	734	3
34	north	Proper Noun	676	5
35	nuke	Common Noun	653	4
36	India	Proper Noun	632	5
37	Trump	Proper Noun	629	5
38	power	Common Noun	623	5
39	lang	Common Noun	622	4
40	mga	Common Noun	618	3
41	one	Common Noun	614	3
42	fact	Common Noun	602	4
43	peace	Common Noun	600	5
44	nothing	Common Noun	594	7
45	nation	Common Noun	569	6
46	leader	Common Noun	562	6
47	rocket	Common Noun	560	6
48	lol	Common Noun	548	3
49	guy	Common Noun	547	3
50	money	Common Noun	542	5
51	china	Common Noun	534	5
52	Ang	Proper Noun	533	3
53	everyone	Common Noun	530	8
54	War	Proper Noun	518	3
55	something	Common Noun	508	9
56	thanks	Common Noun	497	6
57	ng	Common Noun	492	2
58	ko	Common Noun	491	2
59	sa	Common Noun	486	2
60	test	Common Noun	463	4
61	channel	Common Noun	461	7
62	problem	Common Noun	450	7
63	end	Common Noun	440	3
64	government	Common Noun	437	10
65	jet	Common Noun	431	3
66	anything	Common Noun	428	8
67	force	Common Noun	423	5
68	russia	Proper Noun	421	6

69	lot	Common Noun	415	3
70	NATO	Proper Noun	395	4
71	point	Common Noun	394	5
72	comment	Common Noun	393	7
73	South	Proper Noun	386	5
74	drill	Common Noun	382	5
75	army	Common Noun	381	4
76	Iran	Proper Noun	373	4
77	Taiwan	Proper Noun	371	6
78	friend	Common Noun	349	6
79	Hindi	Proper Noun	348	5
80	Putin	Proper Noun	347	5
81	part	Common Noun	343	4
82	ally	Common Noun	341	4
83	po	Common Noun	341	2
84	someone	Common Noun	341	7
85	use	Common Noun	334	3
86	side	Common Noun	332	4
87	water	Common Noun	332	5
88	air	Common Noun	331	3
89	para	Common Noun	331	4
90	Biden	Proper Noun	329	5
91	Jesus	Proper Noun	327	5
92	everything	Common Noun	327	10
93	history	Common Noun	327	7
94	kim	Common Noun	324	3
95	mag	Common Noun	324	3
96	reason	Common Noun	323	6
97	threat	Common Noun	320	6
98	today	Common Noun	320	5
99	attention	Common Noun	314	9

It is worth noting that the word *North Korea* was the most frequently used one in 481 YouTube videos. More specifically, it has the highest frequency (3,456 tokens) and the highest proportion. It is worthwhile pointing out that the word *war* is the second most frequently used one (3,152 tokens). It should be pointed out, on the other hand, that the word *country* was the third highest (2,574 tokens). Additionally, it must be mentioned that the word *missile* ranks fourth (2,549 tokens). Quite interestingly, the word *person* is the fifth highest (2,373 tokens). It should be noted, on the other hand, that the word *Russia* ranks sixth (2,256 tokens). It is worthwhile noting that the word *world* is the seven highest (2,250 tokens). More interestingly, the words *US* and *China* rank eighth and ninth, respectively. From all of this, it is evident that the word *North Korea* was the most widely used one in 481 YouTube videos, followed by the word *war*, the word *country*, the word *missile*, the word *person*, the word *Russia*, the word *US*, and the word *China*, in that order. In addition, it is worth observing that the name *Biden* ranks ninetieth (329 tokens). Also, it is worth mentioning that the name *Kim* was the ninetieth highest (324 tokens). We thus conclude that the word *North Korea* was the most frequently used one (3,456 tokens) in 481 YouTube videos.

### Topic Information

In this section, we aim to consider 8 topics and 5 keywords related to *North Korea*. Table 3 shows 8 topics and 5 keywords that constitute each topic:

**Table 3:** Topic information

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
<b>Topic-1</b>	News	youtube	Hindi	Channel	TV
<b>Topic-2</b>	NA	VIRAL	KOREA	UKRAINE	ORAS

<b>Topic-3</b>	news	News	tv9hindi	Bharatvarsh	Follow
<b>Topic-4</b>	Russia	today	Subscribe	news	missile
<b>Topic-5</b>	use	TV	video	content	Copyright
<b>Topic-6</b>	Korea	fact	north	news	Zee
<b>Topic-7</b>	missile	South Korea	North	test	News
<b>Topic-8</b>	com	BTS	CBS News	shorts	Korea

It is important to note that topic 4 includes 5 keywords such as *Russia, today, Subscribe, news, and missile*. The 1<sup>st</sup> keyword among 5 keywords is *Russia*, which in turn indicates that it was the most frequently used one among 5 keywords. It is interesting to note, on the other hand, that topic 6 is constituted by 5 keywords such as *Korea, fact, north, news, and Zee*. The 1<sup>st</sup> keyword is *Korea*, which we take as indicating that it was the most widely used one among five keywords. It is interesting to point out that 5 keywords such as *missile, South Korea, North, test, and News* consist of topic 7. It seems reasonable to assume that the keyword *missile* was much used in 481 YouTube videos in connection with *North Korea*. Finally, it should be pointed out that 5 keywords such as *com, BTS, CBS News, shorts, and Korea* constitute topic 8. In this topic, the 2<sup>nd</sup> keyword is *BTS*, which we think of as a symbol of South Korea.

Now attention is paid to the frequency of YouTube videos:

**Table 4:** Frequency of YouTube videos

	# of documents
<b>Topic-1</b>	66
<b>Topic-2</b>	19
<b>Topic-3</b>	66
<b>Topic-4</b>	79
<b>Topic-5</b>	38
<b>Topic-6</b>	47
<b>Topic-7</b>	122
<b>Topic-8</b>	44

It is worthwhile mentioning that topic 7 occurred in 122 YouTube videos. As observed earlier, 5 keywords such as *missile, South Korea, North, test, and News* constitute topic 7. This in turn suggests that these keywords were much used in 481 YouTube videos. It is interesting to note, on the other hand, that topic 4 appeared in 79 YouTube videos. As can be seen from Table 3, topic 4 is constituted by 5 keywords such as *Russia, today, Subscribe, news, and missile*. Quite interestingly, topic 1 and topic 3 occurred in 66 YouTube videos. As illustrated in Table 3, topic 1 includes 5 keywords such as *News, YouTube, Hindi, Channel, and TV*, whereas topic 3 includes 5 keywords such as *use, TV, video, content, and Copyright*. It must be noted, on the other hand, that topic 6 appeared in 47 YouTube videos. From all of this, it is clear that topic 7 was the most widely used one in 481 YouTube videos, followed by topic 4, topic 1 (topic 3), and topic 6, in descending order. We thus conclude that topic 7 that occurred in 122 YouTube videos was the most frequently used one in 481 YouTube videos.

### Words and degree

The goal of this section is to examine the frequency of YouTube videos in which many words related to *North Korea* were used. The list was cut off in the top 100. Table 5 shows the frequency of YouTube videos:

Number	Words	Degree
1	Facebook	232
2	news	220
3	facebook	208
4	com	198
5	missile	197
6	News	196

7	South Korea	195
8	Subscribe	193
9	channel	170
10	North Korea	143
11	Twitter	128
12	video	121
13	youtube	118
14	Instagram	107
15	Follow	103
16	YouTube	97
17	update	91
18	Channel	87
19	North	87
20	TV	85
21	Watch	83
22	NorthKorea	81
23	US	80
24	story	80
25	Japan	74
26	southkorea	74
27	instagram	73
28	Korea	72
29	test	72
30	day	71
31	person	71
32	country	68
33	information	68
34	Live	66
35	India	64
36	America	62
37	launch	61
38	Business	60
39	world	59
40	Russia	56
41	business	54
42	Kim, Jongun	54
43	Missile	53
44	politics	53
45	time	53
46	drill	52
47	Connect	50
48	Fair	50
49	App	49
50	website	49
51	issue	48
52	Bollywood	47
53	Political	47
54	Videos	47
55	hour	47
56	military	47
57	scholarship	47
58	Hindi	46



59	World	46
60	South	45
61	google	45
62	border	44
63	use	44
64	KOREA	43
65	water	43
66	Pyongyang	42
67	air	42
68	purpose	42
69	sea	42
70	source	42
71	today	42
72	Seoul	41
73	South	41
74	medium	41
75	visit	41
76	weapon	41
77	China	40
78	NORTH	40
79	app	40
80	tension	40
81	year	40
82	Apps	39
83	Disclaimer	39
84	Google	39
85	Visit	39
86	war	39
87	Copyright	38
88	Download	38
89	Wednesday	38
90	reporting	38
91	tip	38
92	Ukraine	37
93	research	37
94	Missiles	36
95	allowance	36
96	criticism	36
97	right	36
98	Website	35
99	United States	34
100	coast	34

It is significant to note that the word *Facebook* occurred in 232 YouTube videos (the highest). This in turn implies that it was the most widely used one in 481 YouTube videos. It is interesting to point out, on the other hand, that the word *news* appeared in 220 YouTube videos (the second highest). It is also interesting to note that the word *missile* occurred in 197 YouTube videos (the fifth highest). Quite interestingly, the word *South Korea* appeared in 195 YouTube videos. It can thus be inferred that the word *Facebook* was the most widely used one in 481 YouTube videos, followed by the word *news*, the word *missile*, and the word *South Korea*, in that order. More importantly, the word *US* occurred in 80 YouTube videos (it ranks twenty third). On the other hand, the word *Pyongyang* appeared in 42 YouTube videos (it ranks sixty sixth). Finally, it is worth observing that the word *Seoul* ranks seventy second. More specifically, it occurred in 41 YouTube videos. We thus conclude that the word *Facebook* that occurred in 232 YouTube videos was the most widely used one.

