

A Big Data Analysis of 378 KCI Articles on the Korean Wave *Hallyu*

Namkil Kang
Far East University
South Korea

Abstract

This paper aims to analyze the abstracts of 378 KCI (Korea Citation Index) articles on the Korean wave *Hallyu* published from 2000 to 2023. A point to note is seven nouns that are made up of a sentence obtained the highest frequency (490 tokens) among 4,400 nouns. A major point of this paper is that topic 8 was the most occurred one in 378 abstracts, followed by topic 10, topic 1 (topic 12), and topic 11, in descending order. A further point to note is that the word *Hallyu* is deemed to be the most significant in the word cloud of 378 abstracts. With respect to pivotal words, it is interesting to point out that the word *Hallyu* is the most central in 378 abstracts, followed by the word *study*, the word *culture*, the word *result*, and the word *Korea*, in that order. Finally, this paper argues that the word *Hallyu* is the highest in the degree centrality. Put differently, it has the most directly linked neighbors, thus being the highest. This in turn indicates that the word *Hallyu* is the most centric and pivotal in 378 abstracts.

Keywords: Hallyu, KCI, paper, big data, topic, keyword,

1. Introduction

The main goal of this paper is to analyze the abstracts of 378 KCI (Korea Citation Index) articles on the Korean wave *Hallyu* published from 2000 to 2023. We collected them in terms of the biblio data collector and analyzed them in terms of the software package NetMiner. First, we aim at going over the use and frequency of the so-called nouns that were used in 378 abstracts. Second, we attempt to consider 16 topics that showed up in 378 abstracts and their keywords. Also, we investigate how frequently 16 topics turned up in 378 abstracts. Additionally, we aim to probe into the map of 16 topics and their keywords. Third, we aim at looking into the word cloud of 378 abstracts in which representative keywords show up in different sizes. Fourth, we attempt to investigate what pivotal nouns that were used in 378 abstracts

are. Fifth, we aim to contemplate the so-called degree centrality and its map. The terminology degree centrality refers to the number of the neighbors of a word that is directly linked, hence implying that a core word is supposed to have a lot of its neighbors. This amounts to saying that the more a word has many direct neighbors, the more the word is core. In this paper, we attempt to find out words that have the higher degree centrality in the map.

2. Results

2.1. Nouns

In what follows, we aim to probe into the so-called use of names in 378 abstracts. Table 1 shows the classification and use of names that occurred in 378 abstracts:

Table 1 Use of nouns

Value	Frequency	Proportion	Cumulative Proportion
	3,954	0.899	0.899
Geographical Name	127	0.029	0.927
Organization Name	278	0.063	0.991
Personal Name	41	0.009	1

Total	4,400	1	
-------	-------	---	--

Perhaps it is worthwhile noting that the overall frequency of names is 4,400 tokens. More specifically, geographical names obtained 127 tokens, which account for 2.9%. When it comes to organization names, they obtained 278 tokens,

which account for 6.3%. Talking about personal names, they obtained 41 tokens, which account for 0.9%. Quite interestingly, the rest of nouns are 3,954 tokens, which account for 89.9%. Now take a look at Table 2:

Table 2 Use of nouns

Value	Frequency	Proportion	Cumulative Proportion
Common Noun	2,734	0.621	0.621
Proper Noun	1,666	0.379	1
Total	4,400	1	

It is worthwhile to consider common nouns. They showed up 2,734 times in 378 abstracts. That is, they obtained 2,734 tokens, which account for 62.1% among 4,400 nouns. When it comes to

proper nouns, they turned up 1,666 times in 378 abstracts. That is to say, they obtained 1,666 tokens, which account for 37.9%. The following table shows word length:

Table 3 Word length

Value	Frequency	Proportion	Cumulative Proportion
2.0	54	0.012	0.012
3.0	204	0.046	0.059
4.0	360	0.082	0.14
5.0	416	0.095	0.235
6.0	484	0.11	0.345
7.0	490	0.111	0.456
8.0	475	0.108	0.564
9.0	429	0.098	0.662
10.0	393	0.089	0.751
11.0	303	0.069	0.82
12.0	178	0.04	0.86
13.0	150	0.034	0.895
14.0	117	0.027	0.921
15.0	81	0.018	0.94
16.0	45	0.01	0.95
17.0	40	0.009	0.959
18.0	25	0.006	0.965
19.0	23	0.005	0.97

20.0	14	0.003	0.973
21.0	9	0.002	0.975
22.0	16	0.004	0.979
23.0	13	0.003	0.982
24.0	10	0.002	0.984
25.0	9	0.002	0.986
26.0	11	0.003	0.988
27.0	8	0.002	0.99
28.0	7	0.002	0.992
29.0	6	0.001	0.993
30.0	1	0	0.993
31.0	1	0	0.994
32.0	7	0.002	0.995
33.0	2	0	0.996
34.0	3	0.001	0.996
35.0	3	0.001	0.997
36.0	2	0	0.998
37.0	1	0	0.998
38.0	1	0	0.998
39.0	1	0	0.998
40.0	1	0	0.998
41.0	1	0	0.999
43.0	1	0	0.999
47.0	1	0	0.999
57.0	1	0	0.999
58.0	1	0	1
64.0	1	0	1
78.0	1	0	1
Total	4,400	1	

It is worth noticing that the number of seven nouns that consist of a sentence is 490 and that they account for 11.1% (the highest). Quite interestingly, the number of six nouns that are made up of a sentence is 484 and they account for 11% (the second highest). It is also interesting to consider the number of eight nouns that constitute a sentence. They obtained 475 tokens, which account for 10.8% (the third highest). It is

very interesting that the number of nine nouns that consist of a sentence is 429 and they account for 9.8%. It is vital that seven nouns that are made up of a sentence were the most widely used, followed by six nouns, eight nouns, and nine nouns, in that order. We thus conclude that seven nouns that are made up of a sentence obtained the highest frequency (490 tokens) among 4,400 nouns.

2.2. 16 topics and their keywords

This section is focused on inquiring into 16 topics that occurred in 378 abstracts and their keywords.

The following table shows 16 topics that showed up in 378 abstracts and their keywords:

Table 4 16 topics and their keywords

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
Topic-1	paper	film	role	South Korea	Koreans
Topic-2	culture	Hallyu	identity	literature	pop
Topic-3	Hallyu	Korean	study	language	method
Topic-4	medium	term	music	article	globalization
Topic-5	Hallyu	research	characteristic	value	consumption
Topic-6	Hallyu	Korea	dramas	student	drama
Topic-7	Hallyu	research	study	phenomenon	fan
Topic-8	image	effect	intention	Korea	attitude
Topic-9	Wave	wave	country	product	culture
Topic-10	Hallyu	tourism	study	tourist	analysis
Topic-11	brand	consumer	product	star	effect
Topic-12	China	Hallyu	industry	policy	program
Topic-13	content	study	development	dance	direction
Topic-14	group	fashion	analysis	factor	difference
Topic-15	Japan	Hallyu	image	hallyu	relationship
Topic-16	hallyu	France	Hallyu	la	el

It is worth mentioning that the keywords *culture*, *Hallyu*, *identity*, *literature*, and *pop* consist of topic 2. This in turn implies that those five keywords are deemed to be much used in 378 abstracts. Most importantly, the word *Hallyu* is referred to as the symbol of the total Korean culture, hence the second keyword. More interestingly, topic 4 includes the keywords *medium*, *term*, *music*, *article*, and *globalization*. The third keyword *music*

can be seen as the third most widely used in topic 4. It must be stressed that the keywords *Hallyu*, *tourism*, *study*, *tourist*, and *analysis* are made up of topic 10. It is particularly noteworthy that the keywords *group*, *fashion*, *analysis*, *factor*, and *difference* constitute topic 14. The second keyword *fashion* also has something to do with *Hallyu*, hence the second keyword.

Now we turn to the frequency of 16 topics:

Table 5 Frequency of 16 topics

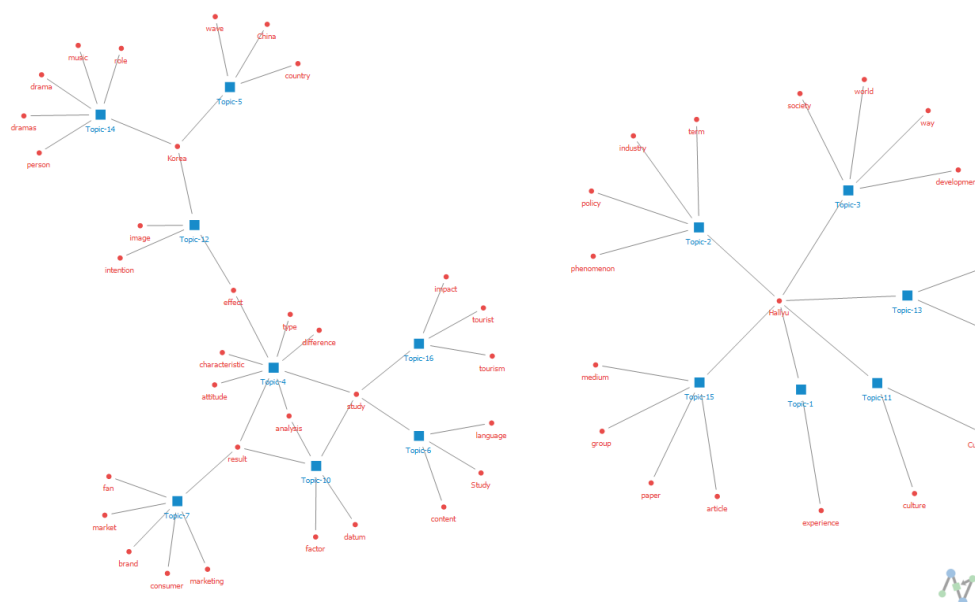
	Frequency of 16 topics
Topic-1	28
Topic-2	29
Topic-3	17
Topic-4	22

Topic-5	23
Topic-6	22
Topic-7	21
Topic-8	48
Topic-9	23
Topic-10	33
Topic-11	24
Topic-12	28
Topic-13	17
Topic-14	14
Topic-15	22
Topic-16	7

It is worth observing that topic 8 showed up 48 times in 378 abstracts (the highest). As exemplified in Table 4, the keywords *image*, *effect*, *intention*, *Korea*, and *attitude* consist of topic 8. That the frequency of topic 8 is the highest indicates that those five keywords were the most widely used in 378 abstracts. It is interesting to point out that topic 8 is followed by topic 10. More specifically, topic 10 occurred 33 times in 378 abstracts (the second highest). As observed earlier, topic 10 includes the keywords *Hallyu*, *tourism*, *study*, *tourist*, and *analysis*. Those keywords count as the second most widely used in 378 abstracts. Note that topic 10 is followed by topic 1 and topic 12. As

indicated in Table 4, the keywords *paper*, *film*, *role*, *South Korea*, and *Koreans* constitute topic 1. On the other hand, the keywords *China*, *Hallyu*, *industry*, *policy*, and *program* are made up of topic 12. More interestingly, topic 1 and topic 12 showed up 28 times in 378 abstracts. Finally, it is significant to note that topic 11 appeared 24 times in 378 abstracts. As illustrated in Table 4, topic 11 includes the keywords *brand*, *consumer*, *product*, *star*, and *effect*. It therefore seems appropriate to mention that topic 8 was the most occurred one in 378 abstracts, followed by topic 10, topic 1 (topic 12), and topic 11, in descending order. Now let us turn to the map of 16 topics:

Figure 1 Topics



It is important to mention that topic 14, topic 5, and topic 12 are linked to the keyword *Korea*. That is to say, three topics have the keyword *Korea* in common. It should be pointed out that topic 4, topic 16, topic 6, and topic 10 have a commonality. That is, those four topics have the keyword *study* in common. On the other hand, topic 7, topic 4, and topic 10 are linked to the keyword *result*. This in turn indicates that the keyword *result* is supposed to be a central word in 378 abstracts. It is also appropriate to mention that topic 2, topic

15, topic 1, topic 11, topic 13, and topic 3 also have a commonality. More specifically, those six topics have the keyword *Hallyu* in common. This seems to suggest that the keyword *Hallyu* counts as the most significant and pivotal in 378 abstracts.

2.3. The word cloud of 378 abstracts

In what follows, we aim at providing the word cloud of 378 abstracts and see what is going on in the picture. Figure 2 shows the degree of the centrality and importance of the relevant nouns that showed up in 378 abstracts:

9	Korean	122
10	analysis	119
11	country	113
12	effect	103
13	image	92
14	product	91
15	Study	91
16	China	85
17	industry	81
18	factor	77
19	purpose	75
20	development	73
21	paper	72
22	phenomenon	71
23	relationship	70
24	order	67
25	influence	67
26	person	64
27	world	63
28	medium	63
29	market	58
30	way	56
31	survey	55
32	addition	55

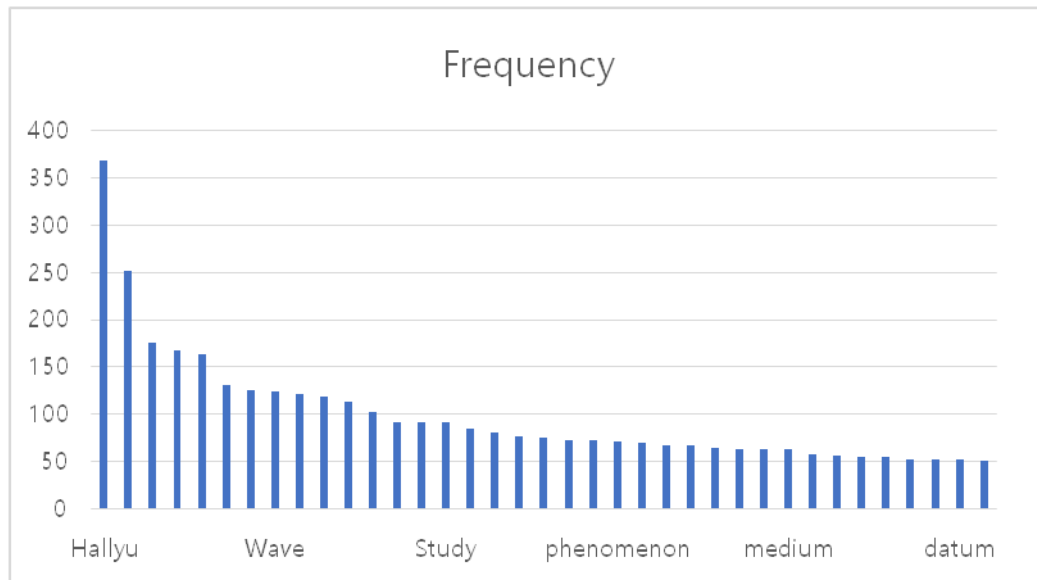
33	intention	52
----	-----------	----

Particularly noteworthy is that the word *Hallyu* obtained the highest frequency (368 tokens) in 378 abstracts. More specifically, it turned up 368 times in 378 abstracts. We take that as suggesting that this word appears to be the most centric in 378 abstracts. It is worthwhile to consider the word *study*. This word occurred 252 times in 378 abstracts. This is considered to be the second most centric. What is interesting is that the word *study* is followed by the word *culture*. To be more

specific, the latter showed up 176 times in 378 abstracts, thus being regarded as the third most

occurred one in 378 abstracts. It is worth pointing out that the word *result* appeared 168 times. This amounts to saying that this word is one of the most key words in 378 abstracts. It would be unfair not to contend that the word *Korea* showed up 164 times in 378 abstracts. The following graph shows the use and frequency of key words that occurred in 378 abstracts. We take this as confirming evidence that the word *Hallyu* is the most significant in 378 abstracts, followed by the word *study*, the word *culture*, the word *result*, and the word *Korea*, in that order. We thus conclude that the word *Hallyu* is the most frequently used in 378 abstracts:

Figure 3 Use and frequency of key nouns

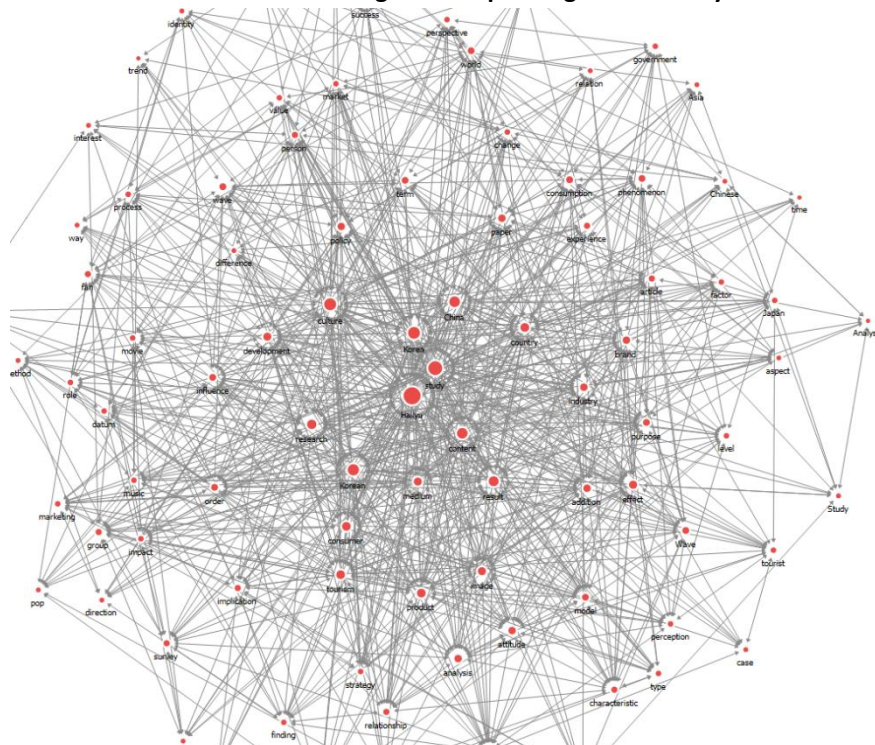


2.5. Degree centrality

This section focuses on defining degree centrality and providing its map. The terminology degree centrality indicates the number of neighbors linked

directly. Simply put, if a word has many neighbors that are directly linked, then the word is a core one. We try to find out some core words that have many neighbors that are directly linked in this map:

Figure 4 Map of degree centrality



Most importantly, the word *Hallyu* is the highest in the degree centrality. Put differently, it has the

most directly linked neighbors, thus being the highest. This in turn means that the word *Hallyu* is the most centric and significant in 378 abstracts. It

is interesting to mention the word *study*. It is the second highest in the degree centrality. That is to say, it is the second most centric one, hence one of some core words. It must be noted that the word *culture* is the third highest in the degree centrality, as illustrated in Figure 4. We can take this as confirming evidence that this word is one of

central words in 378 abstracts. It is worthwhile noting that the word *Korea* is the fourth highest in the degree centrality. This implies that this counts as one of some core words that occurred in 378 abstracts. We thus conclude that the word *Hallyu* is regarded as the most centric in the degree centrality. For the analysis of big data and corpus, see Kang (2023a, 2023b, 2023c, 2023d, 2023e, 2023f).

3. Conclusion

To sum up, we have analyzed the abstracts of 378 KCI (Korea Citation Index) articles on the Korean wave *Hallyu* published from 2000 to 2023. In section 2.1, we have shown that seven nouns that are made up of a sentence obtained the highest frequency (490 tokens) among 4,400 nouns. In section 2.2, we have maintained that topic 8 was the most occurred one in 378 abstracts, followed by topic 10, topic 1 (topic 12), and topic 11, in descending order. In section 2.3, we have contended that the word *Hallyu* is deemed to be the most significant in the word cloud of 378 abstracts. In section 2.4, we have argued that the word *Hallyu* is the most centric in 378 abstracts, followed by the word *study*, the word *culture*, the word *result*, and the word *Korea*, in that order. Finally, we have further argued that the word

Hallyu is the highest in the degree centrality. Simply put, it has the most directly linked neighbors, thus being the highest. This in turn means that the word *Hallyu* is the most centric and significant in 378 abstracts.

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.
- [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.