

## A Study on Changes in Korean Image of Foreign Tourists Using Big Data - Post COVID-19 -

Kyoung-Mi Yoo\*, Youn-Hee Choi\*\*, Gi-Hwan Ryu\*\*\*

\* Professor, Department of Tourism management, Institute of Information Technology,  
KwangWoon University

\*\* Ph. D student, Department of Immersive Content Convergence, Graduate School of  
KwangWoon University, Korea

\*\*\*Professor, Department of Tourism Industry, Graduate school of smart convergence,  
KwangWoon University

E-mail: {casino, gobrrr, allry}@kw.ac.kr

### Abstract

Currently, the Korean wave is not limited to popular culture, but has a significant impact not only on Korea's national image but also on the improvement of Korean companies' products and image of Korea. In this study, using Textom to confirm the change in foreign tourists' image of Korea, the data collection period was 1 year of 2020, when COVID 19 occurred, as a collection period for "Korea and foreigner" and related key words, each Hallyu content, and ranked in the top 80 keywords were derived. Centrality analysis for semantic network visualization was performed using UCINET6, and through CONCOR analysis, 7 groups 'K-Quarantine', 'K-Drama', 'K-Movie', 'K-Beauty', 'K-Shopping', It was clustered into 'K-Tech' and 'K-Pop'. As a result of the analysis, the image of Korea abroad generally recognized the Korean Wave as cultural content, but after the outbreak of COVID 19, it is judged that it has been recognized as a country with a successful case of K-Quarantine.

**Keywords:** Big data, Korean images, Semantic network analysis, Post COVID-19.

## 1. INTRODUCTION

The tourism industry, which was directly affected by COVID-19, which hit the world in 2020, contracted significantly, and exports of Hallyu cultural content products increased by 10.8%, despite a decrease of 87.3% in tourism exports[1]. Hallyu is not just a popular culture that spreads abroad, but has a significant impact on the image of Korea and the products of Korean companies[2]. In particular, Korea was recognized as a successful model country in the world for K- quarantine by sacrificing individual freedom in the initial response to the outbreak of COVID-19 and responding in an orderly and stable manner with the collectivism unique to Koreans[3]. As a result, the pride of being a Korean internally and externally, the changed national identity and prestige have been upgraded, and the national brand of Korea has risen significantly. Recently, studies on the positive effect of Hallyu cultural contents using big data on image recognition in Korea are being

actively conducted in various countries. Centered on Japan, the United States, and France, the higher the favorable feelings for Hallyu content, the more positive it will play in enhancing the national image of Korea[4]. In the Vietnamese market, the image of Korea is highly correlated with cosmetics, and it is analyzed that there is an economic effect that leads to an increase in exports[5]. In addition, as a result of analyzing the perception of Chinese tourists on Korean tourism image through Chinese media, there were a lot of positive reports about Korean tourism image, and it was written with a favorable attitude toward Korean tourism[6]. Therefore, in this study, using Textom, we want to find out what kind of changes have occurred in the image of Korea of foreign tourists after COVID-19. Textom is a solution program that can collect and refine data in a web environment using text mining technology, and even process matrix data generation [7].

## **2. RESEARCH METHOD**

### **2.1 Research project**

In order to understand the meaning of the network of core keywords, the following research tasks were established.

Research Project 1: To identify key keywords for the image of Korea through text mining of the collected data.

Research Project 2: To understand the semantic network between keywords related to Korean image.

Research Project 3: To understand the perception of foreign tourists about the change in Korea's image derived through the semantic network of key keywords.

### **2.2 Data collection**

The data collection period for this study was set from 1/1 to 12/31, 2020.

Data collection was conducted on blogs, cafes, news, web documents, and academic information provided by Naver, a social media, as well as news, web documents, Facebook and Twitter provided by Google. Textom 5.0, a big data solution service, was used to collect unstructured text data.

### **2.3 Analysis method**

In the analysis method of this study, first, the collected atypical keywords were extracted using the Textome, and the importance was determined through the frequency of the keywords, and then text mining was performed. In addition, 80 core keywords were extracted by executing KrKwic by deleting irrelevant words such as 'nose', 'thing', 'dog', 'year', 'time', and 'middle' from the refined keywords. KrKwic finds and checks all the words used in the analyzed message and performs word frequency analysis to calculate how many times each word is used [8]. Second, using the UCINET 6 program, the Degree centrality, Colsonness centrality, and Betweenness centrality were analyzed as the top 80 key keywords, and the frequency and degree of correlation were visualized with the size of the node and the thickness of the line using NetDraw of UCINET 6. Finally, clusters of words with similarities were derived through CONCOR analysis, and the correlation between them was visualized.

### 3. RESEARCH RESULTS

#### 3.1 Keyword Frequency and TF-IDF Analysis

As a result of text mining the collected data, a total of 40,364 keywords were found, and the top 80 keywords were organized based on this.

As a result of keyword frequency analysis, the highest keyword was suggested as 'Quarantine'. This can be interpreted as the appearance of quarantine that did not exist before COVID-19 as a country recognized as a successful case of the K quarantine model around the world. In addition, the ranking of TF-IDF is very similar to the frequency ranking, indicating that key keywords are important in terms of weight. The results are shown in Table 1 below.

**Table 1. TF-IDF analysis of key keywords**

TF						TF-IDF					
R	Keyword	TF	R	Keyword	TF	R	Keyword	TF-IDF	R	Keyword	TF-IDF
1	Quarantine	10,867	41	Chart	1,483	1	Quarantine	22938.46	41	Content	5944.44
2	Beauty	10,231	42	Billboard	1,464	2	Beauty	22401.56	42	Fanaticism	5765.58
3	Drama	9,666	43	Netflix	1,392	3	Drama	20282.78	43	sales	5574.64
4	Shopping	7,750	44	Service	1,384	4	Shopping	17934.41	44	home shopping	5217.80
5	4 <sup>th</sup> Industry	7,605	45	Sales	1,375	5	4 <sup>th</sup> Industry	17377.75	45	interest	5070.01
6	Korea	7,375	46	Diffusion	1,355	6	Korea	14676.55	46	Netflix	5060.61
7	Food	5,583	47	Home shopping	1,350	7	Food	14182.80	47	Produce	5039.47
8	Actor	4,550	48	Increase	1,338	8	Actor	12008.85	48	Service	5034.14
9	Global	4,120	49	A.I	1,336	9	Global	10541.18	49	Diffusion	4949.35
10	Movie	3,846	50	Seoul	1,324	10	Movie	10534.42	50	Increase	4833.04
11	Pop	3,826	51	BTS	1,312	11	Pop	10008.24	51	Black Pink	4825.84
12	COVID-19	3,617	52	Attention	1,307	12	COVID-19	9695.87	52	BTS	4821.62
13	Domestic	3,579	53	Target	1,277	13	Online	9643.07	53	A.I	4812.32
14	Online	3,291	54	Change	1,260	14	Domestic	9382.93	54	Target	4773.32
15	USA	3,240	55	Wave	1,259	15	USA	9267.28	55	Future	4634.72
16	Export	2,856	56	Future	1,253	16	Export	9246.64	56	Seoul	4634.57
17	Safety	2,673	57	Produce	1,243	17	Safety	8728.20	57	Purchase	4628.00
18	Brand	2,662	58	Black Pink	1,202	18	Market	7188.64	58	Vaccine	4608.99
19	Market	2,451	59	Information	1,195	19	Bland	8041.83	59	Idol	4566.07
20	China	2,418	60	Purchase	1,184	20	China	7659.59	60	Fist	4529.46
21	Cosmetics	2,201	61	Idol	1,177	21	Cosmetics	7343.20	61	Change	4518.00
22	Hallyu	2,109	62	Vaccine	1,175	22	Market	7188.64	62	Mask	4514.56
23	Product	2,092	63	Digital	1,134	23	Video	6847.49	63	Information	4479.39
24	Video	2,002	64	Support	1,118	24	Hallyu	6846.97	64	Wave	4453.64
25	Technology	1,994	65	Investment	1,110	25	Japan	6706.98	65	Investment	4423.24
26	Government	1,945	66	Mask	1,063	26	Technology	6665.20	66	Digital	4352.36
27	Popularity	1,923	67	Kimchi	1,041	27	Product	6617.02	67	Event	4241.90
28	Japan	1,896	68	Event	1,027	28	Platform	6278.64	68	Support	4120.83
29	Overseas	1,891	69	Culture	1,013	29	Chart	6242.52	69	Kimchi	4080.75
30	Platform	1,832	70	Consumer	979	30	Goods	6219.04	70	Culture	3864.77

31	Industry	1,826	71	Economy	948	31	Government	6194.94	71	Fan	3779.99
32	Enterprise	1,794	72	Singer	887	32	Billboard	6133.36	72	Zombie	3740.33
33	Broadcast	1,779	73	Fan	886	33	Popularity	6129.50	73	Consumer	3717.49
34	Contents	1,743	74	Parasite	873	34	Industry	6020.57	74	Economy	3677.68
35	Goods	1,737	75	Zombie	858	35	Overseas	6016.12	75	Mobile	3649.95
36	Recommend	1,734	76	Step	853	36	Broadcast	5984.58	76	Parasite	3640.33
37	Fanaticism	1,715	77	Youtube	845	37	Recommend	5984.53	77	Singer	3564.85
38	Group	1,637	78	Mobile	820	38	Group	5971.54	78	Step	3533.86
39	President	1,601	79	Entertainment	777	39	Enterprise	5967.81	79	Youtube	3465.64
40	Interest	1,511	80	Make up	763	40	President	5962.76	80	Make up	3358.08

### 3.2 Semantic Network Analysis

Centrality analysis was performed to check the frequency of the extracted 80 words and the connection relationship between each keyword. It was classified into degree centrality, closeness centrality, and betweenness centrality, and the results of the analysis are shown in Table 2 below. As a result of the degree centrality analysis, 'beauty', 'Korea', 'drama', 'quarantine', 'global', 'food' and 'shopping' appeared in the order. This can be interpreted as an increase in the number of mentions through social media due to the increased interest in K quarantine after COVID-19, if foreign tourists generally encountered Korean images such as beauty, drama, food, shopping, etc., which are Hallyu contents. As a result of closeness centrality analysis, the keywords that took first place with the highest value of 1.000 were 'drama', 'Korea', 'actor', 'global', 'pop', 'COVID19', 'domestic', 'online', 'USA', 'Safety', 'Market', 'Japan' and 'Overseas'. It can be judged that it has dominance among keywords, as it recorded a value of 0.900 or higher within the top 30 core keywords. Therefore, it can be said that it still plays a role as a strong keyword in the image of Korea regardless of before and after the outbreak of COVID-19. All of these words have high betweenness centrality, so when compared to other keywords, it can be seen that they are highly dependent on each other.[9] In other words, it can be interpreted that foreign tourists perceive Korea first, such as dramas, actors, K-pop, and COVID-19. Figure1 shows the visualization data of semantic network analysis to which the results of degree centrality, closeness centrality, and betweenness centrality are applied.

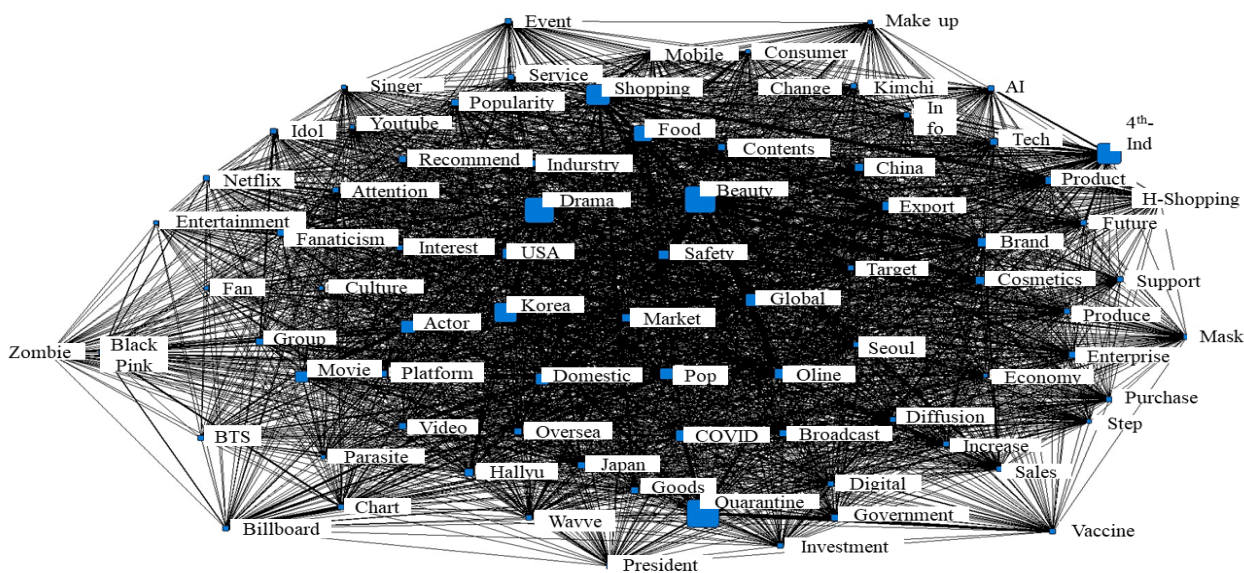


Figure 1. Visualization result of semantic network analysis

Table 2. Analysis of centrality of top 30 keywords

Rank	Keywords	N	Degree centrality	Rank	Closeness centrality	Rank	Betweenness centrality	Rank
1	Quarantine	10,867	.085	4	.963	22	.092	36
2	Beauty	10,231	.150	1	.988	14	.100	17
3	Drama	9,666	.105	3	1.000	1	.118	1
4	Shopping	7,750	.072	7	.963	22	.083	46
5	4 <sup>th</sup> Industry	7,605	.044	16	.878	30	.034	72
6	Korea	7,375	.107	2	1.000	1	.118	1
7	Food	5,583	.073	6	.988	14	.100	17
8	Actor	4,550	.044	16	1.000	1	.118	1
9	Global	4,120	.082	5	1.000	1	.118	1
10	Movie	3,846	.035	21	.975	19	.105	16
11	Pop	3,826	.048	13	1.000	1	.118	1
12	COVID-19	3,617	.047	14	1.000	1	.118	1
13	Domestic	3,579	.063	8	1.000	1	.118	1
14	Online	3,291	.060	9	1.000	1	.118	1
15	USA	3,240	.050	12	1.000	1	.118	1
16	Export	2,856	.052	10	.975	19	.084	44
17	Safety	2,673	.039	20	1.000	1	.118	1
18	Brand	2,662	.052	10	.963	22	.080	51
19	Market	2,451	.046	15	1.000	1	.118	1
20	China	2,418	.043	18	.975	19	.084	43
21	Cosmetics	2,201	.042	19	.963	22	.088	40
22	Hallyu	2,109	.033	24	.988	14	.112	15
23	Product	2,092	.031	27	.929	29	.054	65
24	Video	2,002	.024	33	.988	14	.114	14
25	Technology	1,994	.019	35	.940	27	.058	64
26	Government	1,945	.024	33	.940	27	.092	35
27	Popularity	1,923	.035	21	.963	22	.091	36
28	Japan	1,896	.029	30	1.000	1	.118	1
29	Overseas	1,891	.034	23	1.000	1	.118	1
30	Platform	1,832	.031	27	.988	14	.102	16

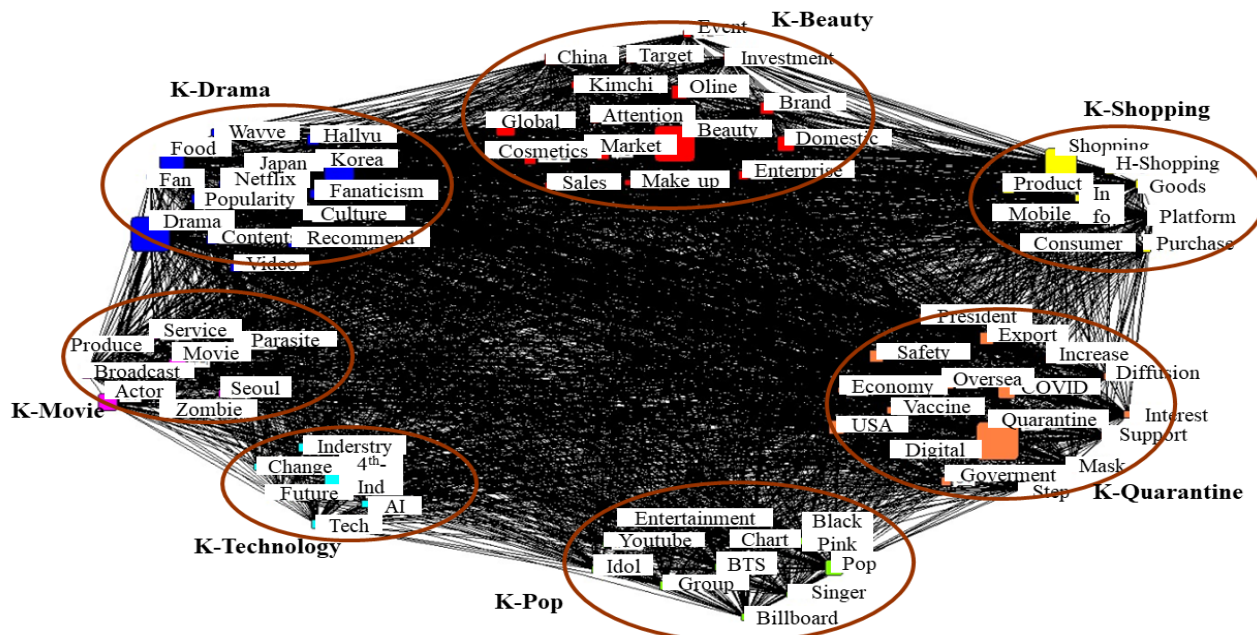
### 3.3 CONCOR analysis

To analyze the correlation between keywords, the CONCOR analysis, a methodology that groups keywords occupying a similar position among the top 80 keywords and analyzes their relationship patterns, was used. As a result of CONCOR analysis, they were grouped into 7 groups as shown in <Table 3>. The first cluster was named K quarantine, 17 related keywords, and the second cluster was named K drama, with 15 related keywords. Here, 'Food' was expected to be classified as one central cluster because it ranked 7th in the frequency ranking, but as a result of the analysis, 'Food' was included in the K-drama. It can be speculated that the reason that 'food' and 'kimchi' were included in the K-drama is because foreign tourists are exposed to more Korean food and kimchi through dramas than expected. The third was named K-Movie, and the related keywords were 8, the fourth was K-Beauty, 15 related keywords, and the fifth was K-Shipping, and 9 related keywords were found. The sixth was K-technology, and the related keywords were 6, and the seventh was K-pop, and there were 10 related keywords. From this, it can be seen that foreign tourists' image of Korea is

greatly affected by the Korean Wave.

**Table 3. Result of CONCOR analysis**

Group	Keywords	Number
K-Quarantine	Quarantine, Vaccine, President, Support, Mask, Interest, Stage, Increase, USA, COVID-19, Overseas, Safety, Digital, Economy, Export, Diffusion, Government	17
K-Drama	Drama, Hallyu, Fan, Recommend, Video, Netflix, Food, Kimchi, Contents, Culture, Fanaticism, Popularity, Korea, Wavve, Japan	15
K-Movie	Service, Zombie, Movie, Broadcast, Seoul, Actor, Parasite, Produce	8
K-Beauty	Beauty, Global, Sales, Domestic, Brand, Cosmetics, Attention, Market, Target, Enterprise, Online, Event, China, Makeup, Investment	15
K-Shopping	Shopping, Home hopping, Consumer, Platform, Mobile, product, Information, Purchase, Goods	9
K-Technology	4th Industry, Future, AI, Technology, Change, Industry	6
K-Pop	Pop, Entertainment, BTS, Blackpink, Billboard, Idol, Singer, Chart, Group, Youtube	10



**Figure 2. Visualization result of CONCOR analysis**

#### 4. CONCLUSION

In this study, various keywords were collected by using Textome in order to confirm the change of foreign tourists' image of Korea. As a result of the collected data, a total of 40,364 keywords were found, and the top 80 keywords with high TF and TF-IDF values were summarized through text mining, and degree centrality, closeness centrality, and betweenness centrality were analyzed for semantic network visualization. In addition, as a result of conducting CONCOR analysis to visualize the network connection and see how key

keywords are clustered, they were classified into 7 groups. Before COVID-19, the image of Korea perceived abroad was generally recognized as Hallyu contents such as 'K-pop', 'drama', 'actor', 'beauty', and 'shopping'. As can be seen, it can be judged that the country has been recognized as a successful case of K quarantine in the world. The data collection of this study was collected in the form of unstructured text data provided by Textom, but characteristics such as recognition level by generation and country were not reflected. Therefore, if big data research that reflects these characteristics is carried out in future research, it is thought that more meaningful research results can be drawn in developing tourism trends and tourism types.

## REFERENCES

- (1) B. J.G Jeon, S.N Kim, “2020 Study on ripple effects of Hallyu” , KOFICE, pp. 13, 2020.
- (2) Tserendulam Dorjmaa, Taek-Soo Shin(2019), “The Impact of Hallyu (Korean Wave) Property on the Attitude of Korean Image and Product Perception in Mongolian Market : Focusing on consumers in Mongolia and Ulaanbaatar”, *Korean Journal of Business Administration*, Vol.32(12), 2019.12, 2131-2156(26 pages), doi : 10.18032/kaaba.2019.32.12.2131.
- (3) Ji-Won Han, Yung-Wook Kim(2021), “A Content-Analyzing Study on Korean Newspapers’ Quoting Foreign News about South Korea’s Response to COVID-19: Focusing on the National Image and National Identity”, *Korean Journal of Communication & Information*, Vol.105, pp. 98-133 (36 pages), doi : 10.46407/kjci.2021.02.105.98
- (4) Hyo-Jin Moon, Sung-Hyun Park(2012), A Study the Relation between Popular Factors and Likability of Hallyu and the National Image, *Korean Academic Society for Public Relations*, Vol.16, Issue 4, p. 247-280, doi: 10.15814/jpr.2012.16.4.247
- (5) Je-Hong Lee(2015), “A Study on Korea Country Image and Cosmetics Brand Image in Vietnam Market by the Korean Wave”, *Korea Association for International Commerce and Information*, 2015, Vol.17, no.3, pp. 73-91 (19 pages), doi : 10.15798/kaici.17.3.201509.73.
- (6) Chen Yu-xuan, Min Bo-young(2019), “A Study on the Image of Korea Tourism through Big Data Analysis : focusing on Chinese Press Reports”, *Tourism Studies*, Vol.43(7), 173, pp.85-102, doi :10.17086/JTS.2019.43.7.85.102.
- (7) Big Data Analysis Solution, TEXTOM Manual, v5.0.
- (8) Han-Woo Park, Ldydesdorff(2004), “Understanding the KrKwic: A computer program for the analysis of Korean text: For news on local innovations provided by Daum.net”, *Journal of The Korean Data Analysis Society*, Vol.6(5), p1377-1387.
- (9) Hye-Jin Lee, Young-Ok Kang (2020), “Understanding Tourist’s Region of Attraction and Image of City through Social Network Data Analysis”, *Journal of the Korean Urban Geographical Society*, Vol.23(1), p.101-114, doi : 10.21189/JKUGS.23.1.8.