International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 4 ǁ April 2023 ǁ PP. 565-577

English Grammar in Big Data

Namkil Kang
Far East University
South Korea

The ultimate goal of this article is to analyze 232 KCI (Korea Citation Index) articles concerning English grammar. This article shows that there was a publication of the most articles in December in 2011, 2013, and 2017. This article further shows that one word that occurred in 232 KCI articles has the highest frequency (1,085 tokens) and the highest proportion (0.517). Also, this article argues that the eight-word expression was the most preferred one for the authors of KCI articles, followed by the six-word expression, the seven-word expression, the five-word expression, and the ten-word expression, in that order. It is worthwhile noting, on the other hand, that in 232 KCI articles, topic 1 was the most widely used one, followed by topic 3, topic 7, and topic 2, in that order. With respect to the frequency of the relevant words occurred in 232 KCI articles, it is interesting to point out that the word English was the most occurred one, followed by the word grammar, the word study, the word student, the word learner, and the word result, in descending order. Finally, this article shows that the keywords sentence, language, use, Korean, analysis, form, grammar, and study are directly linked to English, whereas the keywords research, teaching, English, and student are directly linked to grammar.

Keywords: *English grammar, big data, keyword, topic, visualization, token*

Date of Submission: 08-04-2023 Date of acceptance: 22-04-2023

I. Introduction

The main purpose of this article is to analyze 232 KCI (Korea Citation Index) articles relevant to *English grammar*. We collected and analyzed big data (232 KCI articles published from 2002 to 2022) in terms of the biblio data collector and the software package NetMiner. First, we provide information on 232 KCI articles such as their frequency, their proportion, and their cumulative proportion. Second, we provide the frequency of nouns occurred in 232 KCI articles and their proportion. Third, we provide information on word length such as its frequency and its cumulative proportion. Fourth, we provide information on 7 topics and 5 keywords constituting them. Keywords consist of topics, which in turn constitute documents, namely KCI articles. Fifth, we provide information on degree (the term of NetMiner) which refers to the frequency of articles. Sixth, we aim to inquire into the frequency of main words occurred in 232 KCI articles. Finally, we provide the visualization of keywords relevant to *English grammar* through which we can see which words are closely related with the keyword *English grammar*.

II. Methods

The main goal of this article is to analyze 232 KCI articles published from 2002 to 2022 regarding *English grammar*. In this article, the biblio data collector and the software package NetMiner were used to collect and analyze 232 KCI articles. The main purpose of this article is to answer the following questions: Can we provide the frequency of 232 KCI articles published from 2002 to 2022? Can we provide the frequency of words including the proportion of the relevant nouns? Can we provide information on word length? Can we provide 7 topics and 5 keywords related to *English grammar*? Can we provide degree (the term of NetMiner) which refers to the frequency of articles? Finally, can we provide the visualization of which words are directly linked to the keyword *English grammar*?

III. Results

3.1. Information on 232 KCI articles

In what follows, we aim at providing information on the frequency of 232 KCI articles published from 2002 to 2022. Table 1 shows the frequency of 232 KCI articles, their proportion, and their cumulative proportion:

Table 1 Frequency of 232 KCI articles

Value	Frequency	Proportion	Cumulative Proportion
2002.08	2	0.009	0.009

2002.09	1	0.004	0.013
2002.12	1	0.004	0.017
2003.03	1	0.004	0.022
2003.12	2	0.009	0.03
2004.01	1	0.004	0.034
2004.03	2	0.009	0.043
2004.06	2	0.009	0.052
2004.12	2	0.009	0.06
2005.02	1	0.004	0.065
2005.06	1	0.004	0.069
2005.09	2	0.009	0.078
2006.02	3	0.013	0.091
2006.06	2	0.009	0.099
2006.08	1	0.004	0.103
2006.09	1	0.004	0.108
2006.12	1	0.004	0.112
2007.02	2	0.009	0.121
2007.06	4	0.017	0.138
2008.02	1	0.004	0.142
2008.04	1	0.004	0.147
2008.06	2	0.009	0.155
2008.08	1	0.004	0.159
2008.09	4	0.017	0.177
2008.11	1	0.004	0.181
2009.02	1	0.004	0.185
2009.03	1	0.004	0.19
2010.06	1	0.004	0.194
2010.08	2	0.009	0.203
2010.10	1	0.004	0.207
2010.11	1	0.004	0.211
2010.12	1	0.004	0.216
2011.02	2	0.009	0.224
2011.03	1	0.004	0.228
2011.08	1	0.004	0.233

2011.09	1	0.004	0.237
2011.12	5	0.022	0.259
2012.02	1	0.004	0.263
2012.03	2	0.009	0.272
2012.04	1	0.004	0.276
2012.06	2	0.009	0.284
2012.07	1	0.004	0.289
2012.08	3	0.013	0.302
2012.09	1	0.004	0.306
2012.11	1	0.004	0.31
2012.12	2	0.009	0.319
2013.02	1	0.004	0.323
2013.03	1	0.004	0.328
2013.05	1	0.004	0.332
2013.06	1	0.004	0.336
2013.07	1	0.004	0.341
2013.08	1	0.004	0.345
2013.09	4	0.017	0.362
2013.11	2	0.009	0.371
2013.12	5	0.022	0.392
2014.02	2	0.009	0.401
2014.03	2	0.009	0.409
2014.04	2	0.009	0.418
2014.05	1	0.004	0.422
2014.06	1	0.004	0.427
2014.08	2	0.009	0.435
2014.09	1	0.004	0.44
2014.11	2	0.009	0.448
2014.12	3	0.013	0.461
2015.01	1	0.004	0.466
2015.03	2	0.009	0.474
2015.05	1	0.004	0.478
2015.08	3	0.013	0.491
2015.09	1	0.004	0.496

2015.12	3	0.013	0.509
2016.01	2	0.009	0.517
2016.03	2	0.009	0.526
2016.04	2	0.009	0.534
2016.05	1	0.004	0.539
2016.06	2	0.009	0.547
2016.07	1	0.004	0.552
2016.08	3	0.013	0.565
2016.09	1	0.004	0.569
2016.11	1	0.004	0.573
2016.12	1	0.004	0.578
2017.02	4	0.017	0.595
2017.03	3	0.013	0.608
2017.04	3	0.013	0.621
2017.05	1	0.004	0.625
2017.06	2	0.009	0.634
2017.08	2	0.009	0.642
2017.09	3	0.013	0.655
2017.10	2	0.009	0.664
2017.11	4	0.017	0.681
2017.12	5	0.022	0.703
2018.02	3	0.013	0.716
2018.05	4	0.017	0.733
2018.06	1	0.004	0.737
2018.08	1	0.004	0.741
2018.09	3	0.013	0.754
2018.11	1	0.004	0.759
2018.12	3	0.013	0.772
2019.02	4	0.017	0.789
2019.03	2	0.009	0.797
2019.06	1	0.004	0.802
2019.07	3	0.013	0.815
2019.08	4	0.017	0.832
2019.11	1	0.004	0.836

2019.12	1	0.004	0.841
2020.03	2	0.009	0.849
2020.05	1	0.004	0.853
2020.08	3	0.013	0.866
2020.09	2	0.009	0.875
2020.10	1	0.004	0.879
2020.11	1	0.004	0.884
2020.12	1	0.004	0.888
2021.01	1	0.004	0.892
2021.02	1	0.004	0.897
2021.03	2	0.009	0.905
2021.04	1	0.004	0.909
2021.05	1	0.004	0.914
2021.06	1	0.004	0.918
2021.08	4	0.017	0.935
2021.09	1	0.004	0.94
2021.11	3	0.013	0.953
2021.12	3	0.013	0.966
2022.02	1	0.004	0.97
2022.03	1	0.004	0.974
2022.04	1	0.004	0.978
2022.06	1	0.004	0.983
2022.07	1	0.004	0.987
2022.08	1	0.004	0.991
2022.09	2	0.009	1
Total	232	1	

It is interesting to observe that in December in 2011, 2013, and 2017, 5 articles were published. Their figure is the highest and their proportion is 0.066. Note that as illustrated in Table 1, there was a publication of many articles in December. It is worthwhile noting, on the other hand, that in July (2007), September (2008, 2013), November (2017), May (2018), February (2019), and August (2019, 2021), 4 articles in connection with *English grammar* were published. It should also be pointed out that this figure is the second highest and that their proportion is 0.136. It must be noted, on the other hand, that in February (2006), August (2012), December (2014), August (2015), December (2015), August (2016), March (2017), April (2017), September (2017), February (2018), September (2018), December (2018), July (2019), August (2019), November (2021), and December (2021), 3 articles were published. Notice that their figure (3 articles) is the third highest and that their proportion is 0.208. It can thus be concluded that there was a publication of the most articles in December in 2011, 2013, and 2017.

3.2. Frequency of Nouns

This section centers on providing information on the frequency of nouns and their proportion. Table 2 shows the frequency of major nouns occurred in 232 KCI articles:

Table 2 Frequency of the relevant nouns

Value	Frequency	Proportion	Cumulative Proportion
1.0	1085	0.517	0.517
2.0	287	0.137	0.653
3.0	166	0.079	0.732
4.0	93	0.044	0.777
5.0	77	0.037	0.813
6.0	62	0.03	0.843
7.0	51	0.024	0.867
8.0	29	0.014	0.881
9.0	17	0.008	0.889
10.0	18	0.009	0.898
11.0	15	0.007	0.905
12.0	12	0.006	0.91
13.0	5	0.002	0.913
14.0	11	0.005	0.918
15.0	3	0.001	0.92
16.0	5	0.002	0.922
17.0	7	0.003	0.925
18.0	11	0.005	0.93
19.0	9	0.004	0.935
20.0	3	0.001	0.936
21.0	7	0.003	0.94
22.0	3	0.001	0.941
23.0	6	0.003	0.944
24.0	6	0.003	0.947
25.0	3	0.001	0.948
26.0	1	0	0.949
27.0	3	0.001	0.95
28.0	9	0.004	0.954
29.0	2	0.001	0.955
30.0	7	0.003	0.959
31.0	1	0	0.959
32.0	3	0.001	0.96
33.0	3	0.001	0.962

34.0	6	0.003	0.965
36.0	1	0	0.965
37.0	4	0.002	0.967
38.0	1	0	0.968
39.0	2	0.001	0.969
40.0	1	0	0.969
41.0	1	0	0.97
42.0	3	0.001	0.971
43.0	1	0	0.971
44.0	4	0.002	0.973
45.0	6	0.003	0.976
46.0	2	0.001	0.977
47.0	1	0	0.978
48.0	1	0	0.978
49.0	2	0.001	0.979
52.0	1	0	0.98
54.0	2	0.001	0.98
55.0	1	0	0.981
57.0	1	0	0.981
59.0	2	0.001	0.982
60.0	1	0	0.983
61.0	1	0	0.983
62.0	1	0	0.984
63.0	1	0	0.984
64.0	1	0	0.985
65.0	2	0.001	0.986
66.0	1	0	0.986
67.0	2	0.001	0.987
72.0	1	0	0.988
76.0	1	0	0.988
84.0	1	0	0.989
86.0	2	0.001	0.99
88.0	1	0	0.99
97.0	1	0	0.99

98.0	2	0.001	0.991
104.0	1	0	0.992
106.0	2	0.001	0.993
108.0	1	0	0.993
110.0	1	0	0.994
112.0	1	0	0.994
135.0	1	0	0.995
148.0	1	0	0.995
150.0	1	0	0.996
156.0	1	0	0.996
183.0	1	0	0.997
213.0	1	0	0.997
221.0	1	0	0.998
237.0	1	0	0.998
387.0	1	0	0.999
406.0	1	0	0.999
777.0	1	0	1
797.0	1	0	1
Total	2100	1	

It is worthwhile saying that one word that appeared in 232 KCI articles has the highest frequency (1,085 tokens) and the highest proportion (0.517). More specifically, its proportion and its cumulative proportion are 0.517, respectively. It is worth mentioning, on the other hand, that the frequency of two words that occurred in 232 KCI articles is 287 tokens. This figure is the second highest and their proportion and their cumulative proportion are 0.137 and 0.653, respectively. Quite interestingly, the frequency of three words that appeared in 232 KCI articles is 166 tokens (the third highest). Their proportion is 0.079 and their cumulative proportion is 0.732. It must also be stressed that there are four words whose frequency is 93 tokens (the fourth highest). Their proportion and their cumulative proportion are 0.044 and 0.777, respectively. Finally, it is interesting to note that there are five words whose frequency is 77 tokens. As indicated in Table2, this figure is the fifth highest and their proportion and their cumulative proportion are 0.037 and 0.813, respectively. We thus conclude that one word that occurred in 232 KCI articles has the highest frequency (1,085 tokens) and the highest proportion (0.517).

3.3. Word length

In the following, we aim at providing information on word length, its frequency, and its proportion. Table 3 shows the frequency of word length, its proportion, and its cumulative proportion:

Table 3 Word length

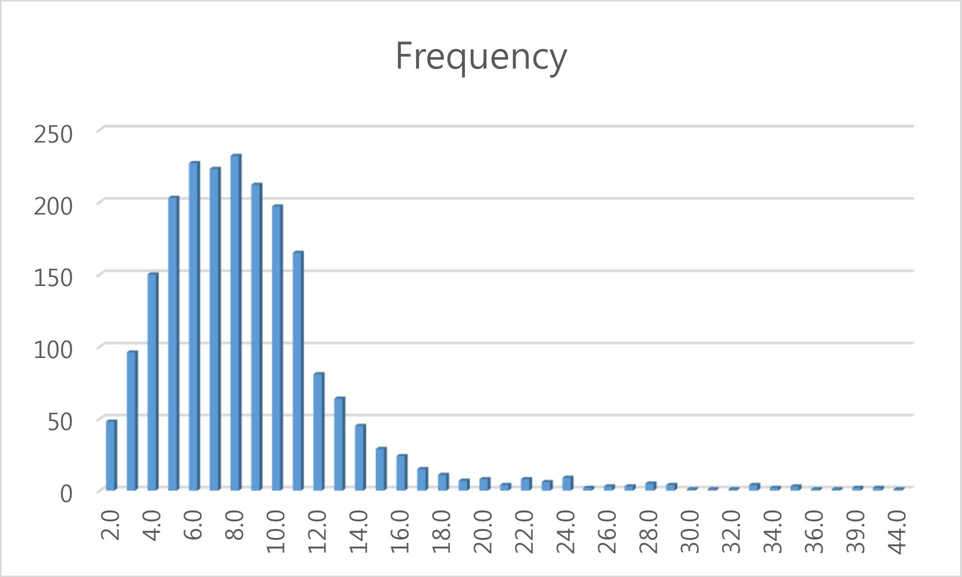
Value	Frequency	Proportion	Cumulative Proportion
2.0	48	0.023	0.023
3.0	96	0.046	0.069
4.0	150	0.071	0.14
5.0	203	0.097	0.237

6.0	227	0.108	0.345
7.0	223	0.106	0.451
8.0	232	0.11	0.561
9.0	212	0.101	0.662
10.0	197	0.094	0.756
11.0	165	0.079	0.835
12.0	81	0.039	0.873
13.0	64	0.03	0.904
14.0	45	0.021	0.925
15.0	29	0.014	0.939
16.0	24	0.011	0.95
17.0	15	0.007	0.958
18.0	11	0.005	0.963
19.0	7	0.003	0.966
20.0	8	0.004	0.97
21.0	4	0.002	0.972
22.0	8	0.004	0.976
23.0	6	0.003	0.979
24.0	9	0.004	0.983
25.0	2	0.001	0.984
26.0	3	0.001	0.985
27.0	3	0.001	0.987
28.0	5	0.002	0.989
29.0	4	0.002	0.991
30.0	1	0	0.991
31.0	1	0	0.992
32.0	1	0	0.992
33.0	4	0.002	0.994
34.0	2	0.001	0.995
35.0	3	0.001	0.997
36.0	1	0	0.997
38.0	1	0	0.998
39.0	2	0.001	0.999
43.0	2	0.001	1

44.0	1	0	1
Total	2100	1	

It is significant to note that the eight-word expression was the most widely used one in 232 KCI articles. More specifically, it has the highest frequency (232 tokens) and the highest proportion (0.11). It is also worth observing that the six-word expression was the second highest (227 tokens) in 232 KCI articles. Its proportion and its cumulative proportion are 0.108 and 0.345, respectively. Quite interestingly, the seven-word expression was the third highest. Its frequency is 223 tokens and its proportion and its cumulative proportion are 0.106 and 0.451, respectively. It must be pointed out, on the other hand, that the five-word expression was the fourth highest (203 tokens). Its proportion is 0.097 and its cumulative proportion is 0.237. Additionally, it is interesting to point out that the ten-word expression was the fifth highest (197 tokens). It seems thus reasonable to assume that the eight-word expression was the most preferred one for the authors of KCI articles, followed by the six-word expression, the seven-word expression, the five-word expression, and the ten-word expression, in that order. Figure 1 clearly shows that the eight-word expression was the most widely used one and followed by the six-word expression:

Figure 1 Frequency of word length



3.4. Topics and keywords

In what follows, we aim to provide 7 topics and 5 keywords constituting them. Table 4 shows 7 topics and 5 keywords. Note that 5 keywords are made up of each topic, which in turn constitutes articles:

Table 4 Topic Info

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
Topic-1	grammar	English	student	study	language
Topic-2	English	learner	grammar	study	test
Topic-3	grammar	student	group	English	study
Topic-4	grammar	English	item	analysis	model
Topic-5	English	Korean	Grammar	structure	study
Topic-6	construction	English	verb	clause	sentence
Topic-7	English	grammar	textbook	Grammar	book

It is particularly noteworthy that the keywords *grammar*, *English*, *student*, *study*, and *language* constitute topic 1. It should be noted, on the other hand, that topic 2 is formed by the keywords *English*, *learner*, *grammar*, *study*, and *test*. In topic 2, the 1st keyword is *English*, which in turn implies that it was the most preferred one among five keywords. More interestingly, topic 5 is constituted by the keywords *English*, *Korean*, *Grammar*, *structure*, and *study*. Again, the 1st keyword is *English*, which we think of as the most widely used among topic 5. It is worthwhile noting, on the other hand, that topic 7 is constituted by the keywords *English*, *grammar*, *textbook*, *Grammar*, and *book*. As can be seen from Table 4, *grammar* was the most widely used one as the 1st keyword, whereas *English* was the most frequently used one as the 2nd keyword.

Now attention is paid to the use of each topic:

Table 5 Frequency of each topic

	# of documents
Topic-1	70
Topic-2	24
Topic-3	57
Topic-4	14
Topic-5	22
Topic-6	17
Topic-7	28

It is important to mention that topic 1 appeared in 70 articles (the highest). As observed in Table 4, topic 1 is constituted by the keywords *grammar*, *English*, *student*, *study*, and *language*. It is worthwhile pointing out, on the other hand, that topic 3 occurred in 57 articles (the second highest). The keywords *grammar*, *student*, *group*, *English*, and *study* consist of topic 3. It should also be pointed out that the keywords *English*, *grammar*, *textbook*, *Grammar*, and *book* are made up of topic 7 and that it appeared in 28 articles. Finally, topic 2 occurred in 24 articles (the fourth highest). As observed earlier, the keywords *English*, *learner*, *grammar*, *study*, and *test* constitute topic 2. From all of this, it seems evident that topic 1 was the most preferred by the authors of KCI articles, followed by topic 3, topic 7, and topic 2, in descending order.

3.5. Frequency of articles

In the following, we aim to provide degree (the term of NetMiner) which refers to the frequency of articles. Table 5 shows the frequency of main words occurred in 232 KCI articles:

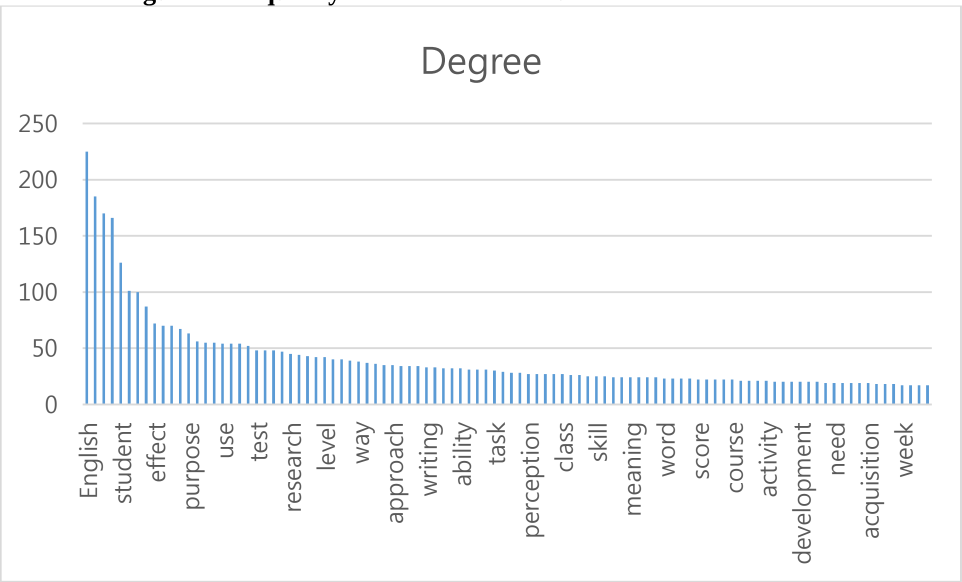
Table 5 Frequency of articles

Number	**Word**	**Degree**
1	English	225
2	grammar	185
3	study	170
4	Grammar	166
5	student	126
6	learner	101
7	result	100
8	language	87
9	effect	72
10	learning	70
11	analysis	70
12	paper	67
13	purpose	63
14	finding	56
15	teaching	55
16	group	55
17	use	54
18	school	54
19	Korean	54
20	instruction	52
21	test	48
22	sentence	48

23	knowledge	48
24	form	47
25	research	45
26	questionnaire	44
27	implication	43
28	method	42
29	level	42
30	datum	40

It is significant to note that the word *English* appeared in 225 articles (the highest). This in turn implies that it was the most preferable one for the authors of KCI articles. It is worthwhile pointing out, on the other hand, that the word *grammar* occurred in 185 articles (the second highest). Quite interestingly, the word *study* appeared in 170 articles (the third highest). It should also be mentioned that the word *Grammar* ranks fourth. To be more specific, it occurred in 166 articles. Additionally, noteworthy is that the word *student* appeared in 126 articles (it ranks fifth). It is also interesting to observe that the word *learner* appeared in 101 articles. From all of this, it seems clear that the word *English* was the most preferable one among the authors of KCI articles, followed by the word *grammar*, the word *study*, the word *Grammar*, the word *student*, and the word *learner*, in that order. It should be noted, on the other hand, that the word *teaching* ranks fifteenth (It occurred in 55 articles). More interestingly, the word *questionnaire* appeared in 44 articles (It ranks twentieth). It can thus be concluded that the word *English* was the most preferred one for the authors of KCI articles and followed by the word *grammar*. Figure 2 shows the frequency of main words occurred in 232 KCI articles:

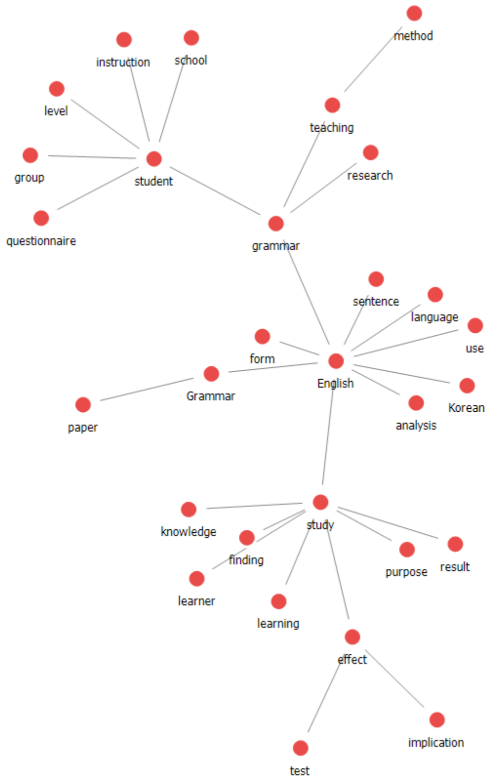
Figure 2 Frequency of main words occurred in 232 KCI articles



3.6 The visualization of words related with English grammar

The goal of this section is to provide the visualization of main words that are closely related with *English grammar*. As exemplified in Figure 3, this visualization shows the links between *English grammar* and main words. Notice that these words frequently occurred in 232 KCI articles:

Figure 3 Visualization of words related with English grammar



It is worthwhile noting that the words *sentence*, *language*, *use*, *Korean*, *analysis*, *form*, *grammar*, and *study* are directly linked to *English*. It is interesting to observe, on the other hand, that the words *research*, *teaching*, *English*, and *student* are directly linked to *grammar*. Quite interestingly, the words *knowledge*, *finding*, *learner*, *learning*, *effect*, *purpose*, and *result* are linked to the word *study*. For the visualization of synonyms and keywords, see Kang (2022a, 2022b, 2022c, 2022d, 2023a, 2023b). To sum up, this visualization provides the links between *English grammar* and the relevant keywords.

IV. Conclusion

To sum up, we have analyzed 232 KCI articles related to *English grammar*. In section 3.1, we have shown that there was a publication of the most articles in December in 2011, 2013, and 2017. In section 3.2, we have argued that one word that appeared in 232 KCI articles has the highest frequency (1,085 tokens) and the highest proportion (0.517). In section 3.3, we have maintained that the eight-word expression was the most preferred one for the authors of KCI articles, followed by the six-word expression, the seven-word expression, the five-word expression, and the ten-word expression, in that order. In section 3.4, we have contended that topic 1 was the most preferred by the authors of KCI articles, followed by topic 3, topic 7, and topic 2. In section 3.5, we have shown that the word *English* was the most preferable one among the authors of KCI articles, followed by the word *grammar*, the word *study*, the word *Grammar*, the word *student*, and the word *learner*, in that order. In section 3.6, we have provided the relevant links showing that the words *sentence*, *language*, *use*, *Korean*, *analysis*, *form*, *grammar*, and *study* are directly linked to *English*, whereas the words *research*, *teaching*, *English*, and *student* are directly linked to *grammar*.

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.

[5]. Kang, N. (2023a). K-Pop in BBC News: A Big Data Analysis. *Advances in Social Sciences Research Journal*, *10*(2), 156-169.

[6]. Kang, N. (2023b). K-Dramas in Google: A NetMiner Analysis. *Transaction on Engineering and Computing Sciences*, *11*(1), 193-216.