

On Consist of and Made up of: A Corpora-Based Analysis*

Namkil Kang
Far East University, South Korea

Abstract—The ultimate goal of this article is to show that *consist of* and *made up of* are somewhat different in their use. It is worth pointing out that the type *made up of* (1,129 tokens) may be preferable to the type *consist of* (993 tokens) in the UK. A further point to note is that *consist of* is 57.14% the same as *made up of*. Talking about *consist of* and *made up of* in the Hansard Corpus (20, April 2022. Online <https://english-corpora.org/hansard/>), the former is preferable to the latter in the British parliament. Quite interestingly, *consist of* was the most preferable one (942 tokens) for British politicians in 1960, but it was the least preferred one (49 tokens) in 1800. On the other hand, *made up of* reached a peak (1,020 tokens) in the British parliament in 1990. More importantly, the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>) shows that *consist of people* is the most frequently used one in the UK, whereas *made up of people* and *made up of representatives* are the most widely used ones. Finally, it is worthwhile noting that 6.66% of forty five nouns are the collocations of both *consist of* and *made up of*.

Index Terms—British National Corpus, Hansard Corpus, consist of, made up of

I. INTRODUCTION

As Murphy (2016, 2019) points out, *consist of* and *made up of* are used synonymously. As argued by Kang (2022), the British National Corpus (20, April 2022. Online <https://corpus.byu.edu/bnc>) is a database that is heavily used in English learning. The main purpose of this paper is to compare *consist of* with *made up of* in the British National Corpus (20, April 2022. Online <https://corpus.byu.edu/bnc>) and the Hansard Corpus (20, April 2022. Online <https://english-corpora.org/hansard/>). First, we consider the similarity between *consist of* and *made up of* by examining seven genres. Also, we measure the distance between *consist of* and *made up of* in seven genres. By doing so, we can see how similar *consist of* and *made up of* are in seven genres. Second, we consider which type is the preferred one in the British parliament. Also, we consider when *consist of* and *made up of* were the most frequently used in the British parliament. Finally, we consider the collocations of *consist of* and *made up of* in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). By using the software package NetMiner, we can see how similar *consist of* and *made up of* are. This paper is organized as follows. In section 2, we argue that *made up of* may be favored over *consist of* in their use in the UK. We further argue that *consist of* is 57.14% the same as *made up of*. We also contend that *consist of* and *made up of* show the highest similarity in the non-academic genre. In section 3, we maintain that *consist of* was preferable to *made up of* in the British parliament. More interestingly, *consist of* and *made up of* reached a peak in 1960 and 1990, respectively. In section 4, we argue that *consist of people* is the most frequently used one (4 tokens) in the UK, whereas *made up of people* and *made up of representatives* are the most widely used ones (19 tokens vs. 19 tokens). Finally, we show that 6.66% of forty five nouns are the collocations of both *consist of* and *made up of*.

II. CONSIST OF AND MADE UP OF IN THE BNC

In what follows, we consider the similarity between *consist of* and *made up of* by examining seven genres. Table 1 shows the frequency of *consist of* and *made up of* in seven genres:

TABLE 1
FREQUENCY OF CONSIST OF AND MADE UP OF IN SEVEN GENRES (BNC)

GENRE	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-ACAD	ACADMIC	MISC
Consist of	993	23	22	67	25	230	318	308
Made up of	1,129	53	49	102	89	273	272	291

An immediate question is “Which type is the preferred one in the UK?” Table 1 clearly shows that *made up of* is preferable to *consist of*. To be more specific, the frequency of *consist of* is 993 tokens, whereas that of *made up of* is 1,129 tokens. This amounts to saying that the British prefer using *made up of* rather than using *consist of*.

Interestingly, *consist of* ranks first (318 tokens) in the academic genre, whereas *made up of* ranks first (291 tokens) in the miscellaneous genre. Quite interestingly, there is no similarity between *consist of* and *made up of* in rank-one. With

* This work was supported by the 2022 Far East University Research Grant (FEU2022S02).

respect to the academic genre, it is worth noting that the frequency of *consist of* (318 tokens) is somewhat higher than that of *made up of* (272 tokens). This in turn implies that the British prefer using *consist of* rather than using *made up of* in the academic field. With respect to the miscellaneous genre, on the other hand, it is interesting to note that *consist of* (308 tokens) is preferable to *made up of* (291 tokens).

Noteworthy is that *consist of* ranks second (308 tokens) in the miscellaneous genre, whereas *made up of* ranks second (273 tokens) in the non-academic genre. Again, there is no similarity between *consist of* and *made up of* in rank-two. More interestingly, in the non-academic genre, the frequency of *made up of* (273 tokens) is much higher than that of *consist of* (230 tokens). This amounts to saying that the British prefer *made up of* to *consist of* in the non-academic genre.

It is worth pointing out that *consist of* ranks third (230 tokens) in the non-academic genre, whereas *made up of* ranks third (272 tokens) in the academic genre. Quite interestingly, *consist of* does not show the same property as *made up of* in rank-three, hence showing no similarity.

It is worthwhile mentioning that the types *consist of* and *made up of* are the fourth most frequently used ones (67 tokens vs. 102 tokens) in the magazine genre. Interestingly, there is a similarity between *consist of* and *made up of* in the magazine genre.

It is worthwhile noting that *consist of* and *made up of* are the fifth most frequently used ones (25 tokens vs. 89 tokens) in the newspaper genre. Quite interestingly, there is a close similarity between *consist of* and *made up of* in the newspaper genre. It should be pointed out, however, that the frequency of *made up of* (89 tokens) is three times higher than that of *consist of* (25 tokens) in the newspaper genre. We take this as implying that British journalists prefer to use *made up of* rather than use *consist of* in their newspapers.

It is interesting to note that *consist of* (23 tokens) and *made up of* (53 tokens) are the sixth most frequently used ones in the spoken genre. Again, there is a similarity between *consist of* and *made up of* in the spoken genre, hence the same pattern in rank-six. Interestingly, the frequency of *made up of* (53 tokens) is higher than that of *consist of* (23 tokens) in the spoken genre. This might imply that the British prefer to use *made up of* rather than use *consist of* in the spoken genre.

Finally, it is worthwhile pointing out that *consist of* and *made up of* are the seventh most frequently used ones (22 tokens vs. 49 tokens) in the fiction genre. More interestingly, there is a similarity between *consist of* and *made up of* in the fiction genre, thus showing a high similarity in rank-seven. It must be noted, however, that *made up of* is favored over *consist of* in the fiction genre. We infer this from the fact that the frequency of *made up of* (49 tokens) is even higher than that of *consist of* (22 tokens). To sum up, *consist of* and *made up of* show the same pattern in four genres, but they show a different pattern in three genres. From all of this, we may conclude that *consist of* is 57.14% the same as *made up of*.

Now let us consider the use of *consist of* and *made up of* in seven genres:

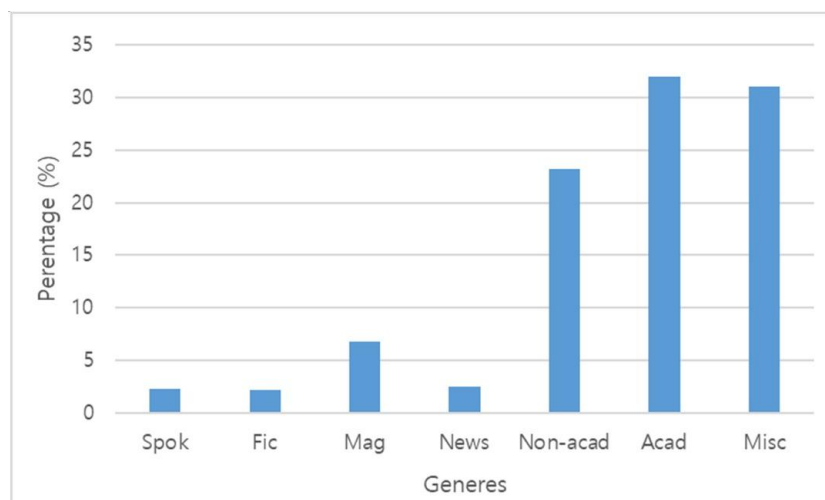


Figure 1 Percentage of *consist of* in Seven Genres

As can be seen from Figure 1, the academic field is the most influenced by *consist of*, followed in descending order by the miscellaneous genre, and the non-academic genre.

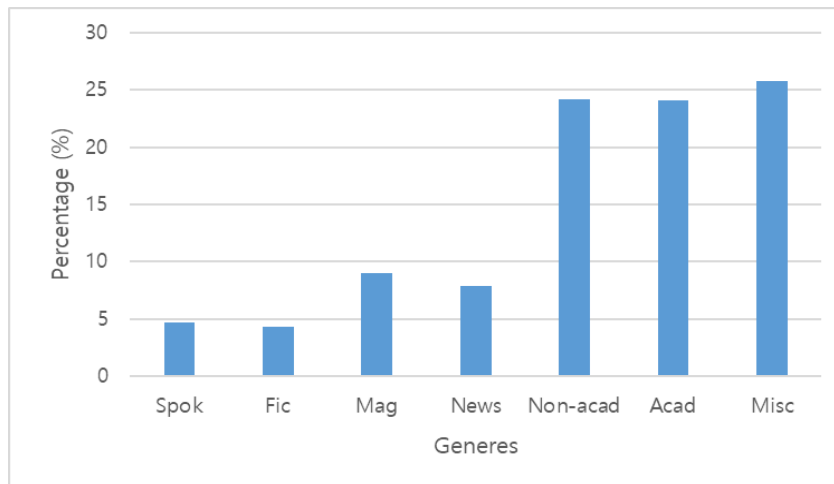


Figure 2 Percentage of made up of in Seven Genres

As shown in Figure 2, the miscellaneous genre is the most influenced by *made up of*, followed by the non-academic genre, and the academic genre.

Now attention is paid to the distance between *consist of* and *made up of* in seven genres. Here the Euclidean distance provides an indication of how similar *consist of* and *made up of* are in seven genres. We define the Euclidean distance in the following way:

(1) The Euclidean distance:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

The following table shows the degree of the similarity between *consist of* and *made up of* in seven genres:

TABLE 2
EUCLIDEAN DISTANCE BETWEEN CONSIST OF AND MADE UP OF IN THE BNC

Genre	Spok	Fic	Mag	News	Non-acad	Acad	Misc
Percentage of consist of	2.31	2.21	6.74	2.51	23.16	32.02	31.01
Percentage of made up of	4.69	4.34	9.03	7.88	24.18	24.09	25.77
Euclidean distance	2.38	2.13	2.29	5.37	1.02	7.93	5.24

More interestingly, *consist of* is the furthest from *made up of* in the academic genre. The Euclidean distance between *consist of* and *made up of* in the academic genre is 7.93. Thus, the types *consist of* and *made up of* show a low similarity. Quite interestingly, in the non-academic field, *consist of* is the nearest to *made up of*. Their distance is 1.02, which is the lowest. Thus, the types *consist of* and *made up of* show a high similarity. We thus conclude that in the non-academic genre, *consist of* is the nearest to *made up of*.

III. CONSIST OF AND MADE UP OF IN THE HANSARD CORPUS

In this section, we aim to consider the diachronic aspects of *consist of* and *made up of* from 1800 to 2000. Table 3 shows the use of *consist of* and *made up of* in the British parliament (1800-2000):

TABLE 3
FREQUENCY OF CONSIST OF AND MADE UP OF IN THE HANSARD CORPUS

Period	Consist of	Made up of
1800	49	9
1810	63	16
1820	74	25
1830	256	67
1840	206	52
1850	507	64
1860	397	94
1870	524	94
1880	598	176
1890	534	142
1900	629	212
1910	795	227
1920	604	266
1930	801	291
1940	820	348
1950	825	356
1960	942	525
1970	919	669
1980	918	851
1990	837	1,020
2000	453	591

An important question is “Which type was the preferable one for British politicians?” Table 3 clearly indicates that *consist of* was the preferred one in the British parliament from 1800 to 2000. That is to say, the frequency of *consist of* is 11,751 tokens, but that of *made up of* is 6,095 tokens. This shows that *consist of* was preferable to *made up of* in the British parliament. Note that as observed earlier, the British prefer to use *made up of* rather than use *consist of*. The following graph shows the diachronic use of *consist of* and *made up of* from 1800 to 2000:

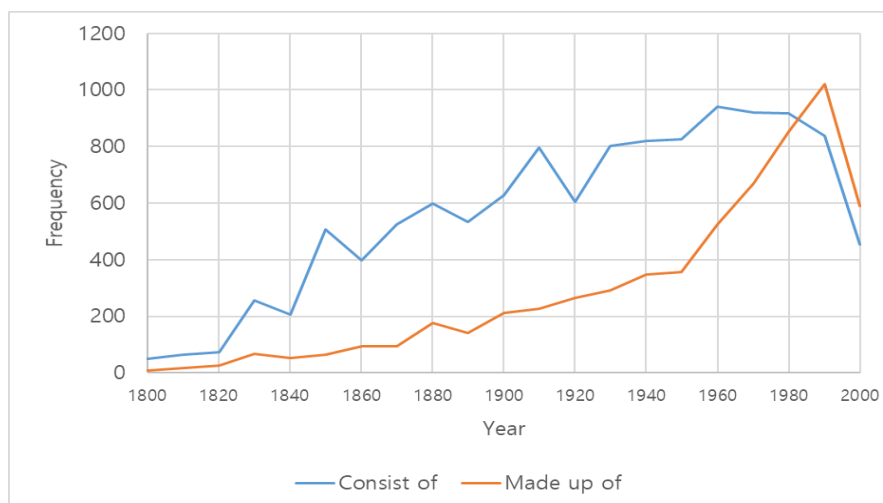


Figure 3 Use of consist of and made up of in the British Parliament From 1800 to 2000

It is interesting to note that the frequency of *consist of* increased from 1800 to 1850. Note that there was a decline (a decline of 110 tokens) in 1860. With respect to the frequency of *consist of*, it is interesting to point out that there was a gradual increase (an increase of 201 tokens) from 1860 to 1880. Notice, however, that there was a sudden fall (a decline of 64 tokens) in the frequency of *consist of* in 1890. Interestingly, the frequency of *consist of* increased steadily (a rise of 408 tokens) from 1890 to 1960 except 1920. However, the figure of *consist of* decreased suddenly (a decline of 489 tokens) from 1960 to 2000.

Interestingly, the figure of *made up of* increased gradually (a rise of 58 tokens) from 1800 to 1830. Note, however, that the figure of *made up of* decreased suddenly (a decrease of 15 tokens) in 1840. After this period, there was a gradual rise (an increase of 124 tokens) in the frequency of *made up of* from 1840 to 1880. With respect to the frequency of *made up of*, it is worth noting that there was a sudden decrease (a decline of 34 tokens) in 1890. Quite interestingly, the figure of *made up of* increased dramatically (a rise of 878 tokens) from 1890 to 1990. Notice that there was a decrease of 429 tokens in 2000. Most interestingly, *made up of* had the highest frequency (1,020 tokens) in 1990. Most importantly, *consist of* was always preferable to *made up of* from 1800 to 1980. Note, however, that *made up of* was favored over *consist of* in the British parliament in 1990 and 2000.

IV. COLLOCATIONS OF CONSIST OF AND MADE UP OF IN THE BNC

In the following, we aim to provide a collocation analysis of *consist of* and *made up of* in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). This list (Table 4) was cut off in the top 24:

TABLE 4
COLLOCATION OF CONSIST OF

Number	Collocation	Frequency
1	Consist of people	4
2	Consist of fish	2
3	Consist of cash	2
4	Consist of money	2
5	Consist of molecules	2
6	Consist of modules	2
7	Consist of mixtures	2
8	Consist of hundreds	2
9	Consist of interviews	2
10	Consist of periods	2
11	Consist of persons	2
12	Consist of representatives	2
13	Consist of sheets	2
14	Consist of volunteers	2
15	Consist of water	2
16	Consist of words	2
17	Consist of obsidian	1
18	Consist of objects	1
19	Consist of oats	1
20	Consist of noun	1
21	Consist of non-lawyers	1
22	Consist of neutrons	1
23	Consist of music	1
24	Consist of mortgage	1

Now which collocation is the most preferred by the British? Table 4 indicates that the expression *consist of people* is the most commonly used one (4 tokens). This in turn indicates that *consist of people* is the most preferable one (4 tokens) for the British. It is interesting to note, on the other hand, that *consist of mixtures* ranks second (2 tokens) in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). It is also interesting to point out that *consist of representatives* ranks second (2 tokens) in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). Finally, it is worthwhile mentioning that *consist of water* ranks second (2 tokens) in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). It can thus be concluded that the expression *consist of people* is the most preferred one (4 tokens).

Now consider Table 5:

TABLE 5
COLLOCATION OF MADE UP OF

Number	Collocation of made up of	Frequency
1	Made up of people	19
2	Made up of representatives	19
3	Made up of members	10
4	Made up of parts	7
5	Made up of individuals	6
6	Made up of players	4
7	Made up of staff	4
8	Made up of street	4
9	Made up of thousands	4
10	Made up of units	4
11	Made up of waves	4
12	Made up of hundreds	3
13	Made up of electrons	3
14	Made up of couples	3
15	Made up of cells	3
16	Made up of amino	3
17	Made up of quarks	3
18	Made up of women	3
19	Made up of components	2
20	Made up of attitudes	2
21	Made up of atoms	2
22	Made up of employees	2
23	Made up of elements	2
24	Made up of experts	2

Table 5 clearly indicates that *made up of people* and *made up of representatives* are the most widely used ones (19

tokens vs. 19 tokens) in the UK. This in turn suggests that *made up of people* and *made up of representatives* are the most preferred ones (19 tokens) in the UK. As illustrated in Table 5, *made up of people* and *made up of representatives* are the most preferred (19 tokens vs. 19 tokens) by the British, followed by *made up of members*, *made up of parts*, *made up of individuals*, and *made up of players*, in that order. Quite interestingly, *consist of people* and *made up of people* rank first (4 tokens vs. 19 tokens) in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). On the other hand, *consist of representatives* and *made up of representatives* rank second (2 tokens vs. 19 tokens) in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). It is worth observing that *made up of members* ranks third (10 tokens) in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>). We thus conclude that *made up of people* and *made up of representatives* are the most preferred (19 tokens vs. 19 tokens) by the British.

Now let us consider the visualization of the collocations of *consist of* and *made up of* in the top 24:

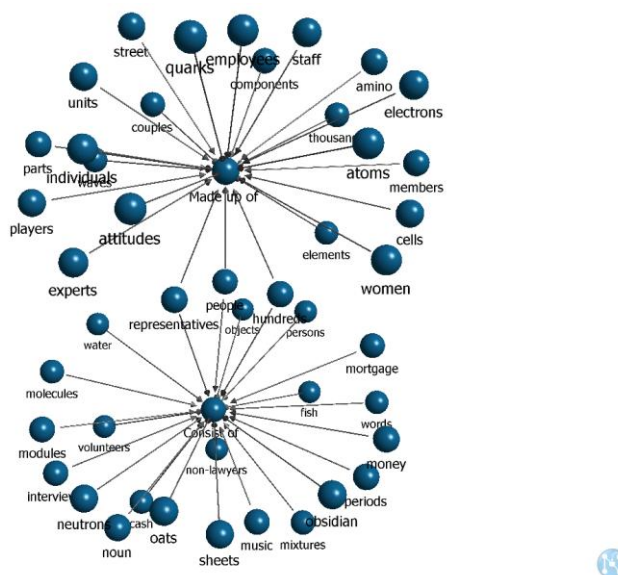


Figure 4 Visualization of the Collocations of *consist of* and *made up of* in the BNC

As exemplified in Figure 4, twenty four nouns are linked to *consist of* and *made up of*, respectively. Most importantly, three nouns are linked to both *consist of* and *made up of*, which indicates that they are the collocations of both *consist of* and *made up of*. The three nouns linked to *consist of* and *made up of* are *people*, *representatives*, and *hundreds*. Most importantly, 6.66% of forty five nouns are the collocations of *consist of* and *made up of*. It can thus be concluded that the types *consist of* and *made up of* are somewhat different in their collocations.

V. CONCLUSION

To sum up, we have compared *consist of* with *made up of* in the British National Corpus (20, April 2022. Online <https://corpus.byu.edu/bnc>) and the Hansard Corpus (20, April 2022. Online <https://english-corpora.org/hansard/>). In section 2, we have argued that the British prefer using *made up of* rather than using *consist of*. We have further argued that *consist of* is 57.14% the same as *made up of*. Also, we have shown that *consist of* is the nearest to *made up of* in the non-academic genre. In section 3, we have maintained that *consist of* was preferable to *made up of* in the British parliament. Quite interestingly, *consist of* was the most preferable one (942 tokens) for British politicians in 1960, whereas it was the least preferred one (49 tokens) in 1800. On the other hand, *made up of* reached a peak (1,020 tokens) in the British parliament in 1990. In section 4, we have contended that *consist of people* is the most frequently used one (4 tokens) in the UK, whereas *made up of people* and *made up of representatives* are the most widely used ones (19 tokens vs. 19 tokens). Finally, we have shown that 6.66% of forty five nouns are the collocations of both *consist of* and *made up of*. This amounts saying that *consist of* and *made up of* may be low similarity synonyms in the BNC (20, April 2022. Online <https://corpus.byu.edu/bnc>).

REFERENCES

- [1] *British National Corpus (BNC)*. 20, April 2022. Online <https://corpus.byu.edu/bnc>
- [2] *Hansard Corpus (HC)*. 20, April 2022. Online <https://english-corpora.org/hansard/>
- [3] Kang, N. (2022). A Comparative Analysis of Impressed by and Impressed with in Two Corpora. *Theory and Practice in Language Studies*, 11(11): 819-827.
- [4] Murphy, R. (2016). *Grammar in Use*. Cambridge University Press. Singapore.
- [5] Murphy, R. (2019). *English Grammar in Use*. Cambridge University Press. New York.



Namkil Kang was born in Chungbuk, South Korea on 27, August, 1964. He received a BA degree (English Language and Literature) at Chungbuk National University in South Korea in 1988. He received an MA degree (English Linguistics) at Sogang University in South Korea in 1990. He received an MPhil degree (Linguistics) at the University of Oxford in 1998. Finally, he received a Ph.D. degree (Linguistics) at the University of Oxford in 2000.

He has been with Far East University in South Korea since 2002. He is a professor at Far East University in South Korea. He is working on syntax, corpus linguistics, and English education at this moment. He has published about 100 articles and three books. He is a recent ex-chairman of the Joongwon Linguistic Society of South Korea and a chairman of the Jungang English Language and Literature Association of South Korea.