



CARE FOR AND LIKE IN CORPORA AND CHATGPT

Namkil Kang

Far East University, South Korea

This paper aims at comparing *care for* and *like* in the British National Corpus, the Movie Corpus, and ChatGPT. When it comes to *care for* and *like* in the Movie Corpus, they have two things in common. That is to say, they had the highest use (744 tokens vs. 309,656 tokens) in the 2010s, while they had the lowest one (179 tokens vs. 22,703 tokens) in the 1930s. A point to note with respect to the BNC is that the phrase *care for people* is the most preferred (37 tokens) by the British, followed by *care for children*, *care for patients*, *care for others*, and *care for dementia*, in that order. A further point to note is that the phrase *like Mr* is the most preferred (172 tokens) by the British, followed by *like people*, *like hell*, *like water*, and *like children*, in descending order. This paper also shows that *care for* is 11.11% the same as *like* in the analysis of the BNC (the top 25). Finally, this paper argues that *care for* is 5.26% the same as *like* in the data from the ChatGPT (the top 20). It can thus be inferred that *care for* and *like* have a low degree of similarity.

Keywords: care for, like, token, British National Corpus, Movie corpus, ChatGPT

1. Introduction

This paper aims at comparing *care for* and *like* in two corpora and ChatGPT. We explore the similarity between *care for* and *like* in the Movie Corpus (MC), the British National Corpus (BNC), and the ChatGPT. First, we attempt to investigate the diachronic aspects of *care for* and *like* from the 1930s to the 2010s and compare them. We probe into the use of these two types for 90 years and contemplate the difference in their use between them during that period. Second, we aim to go over the collocations of *care for* and *like* in the BNC and look into the similarity between them. By examining the collocation of *care for* and that of *like*, we can see how similar they are in the BNC. We inquire into the collocations of these two types in descending order and examine which collocations are the preferable ones for the British. Third, we attempt to look into the collocations of *care for* and *like* from the AI-based ChatGPT that acquires a wealth of knowledge from its train data and webs. We try to observe how close *care for* and *like* are in the ChatGPT.

2. The Movie Corpus

This section is devoted to inquiring into the use of *care for* and that of *like* from the 1930s to the 2010s. Also, we look into the similarity between these two types in the Movie Corpus. Additionally, we investigate the use of *care for* and *like* in six countries' movies. Table 1 shows the use of *care for* and *like* for 90 years:

Table 1 Use of care for and like from the 1930s to the 2010s

	Care for	Like
ALL	3,429	848,199
1930s	179	22,703
1940s	282	37,266
1950s	291	39,878
1960s	250	32,570
1970s	231	37,040
1980s	272	53,843
1990s	457	96,688
2000s	723	218,555
2010s	744	309,656

It is interesting to observe that the overall use of *care for* was 3,429 tokens, while that of *like* was 848,199 tokens. Simply put, the use of *like* is 247 times higher than that of *care for*. This in turn implies that *like* (848, 199 tokens) was preferred over *care for* (3,429 tokens) by the movie writers of six countries.

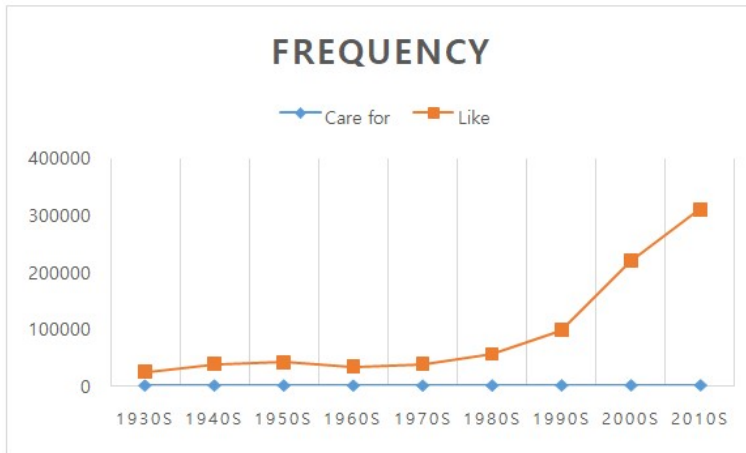
As illustrated in Table 1, the use of *care for* in the 1930s was 179 tokens, which was the lowest from the 1930 to the 2010s. This in turn indicates that *care for* during this period was the least preferred. Its use slightly increased to 103 tokens in the 1940s. There was a gradual rise (291 tokens) in the use of *care for* in the 1950s. That is to say, there was an increase of 9 tokens in the 1950s. Note, however, that there was a sudden drop (250 tokens) in the use of *care for* in the 1960s. More specifically, there was a fall of 41 tokens in the 1960s. Quite interestingly, there was a steady decrease (231 tokens) in the use of *care for* from the 1960s to the 1970s. In the 1970s, there was a decline of 19 tokens. More interestingly, there was a gradual rise (272 tokens) in the use of *care for* in the 1980s. The same can be said of the 1990s. The use of *care for* steadily increased to 457 tokens in the 1990s. Most interestingly, the use of *care for* dramatically increased to 723 tokens in the 2000s. Finally, *care for* reached a peak (744 tokens) in the 2010s. This in turn suggests that during that period, *care for* was the most preferred one (744 tokens) for the movie writers of six countries.

Now attention is paid to *like*. It should be noted that *like* had the lowest use (22,703 tokens) in the 1930s. This in turn indicates that during this period, it was the least preferred. Notice, however, that there was a gradual rise (37,266 tokens) in the use of *like* in the 1940s. To be more specific, there was an increase of 14,563 tokens in the 1940s. Similarly, the use of *like* steadily increased to 39,878 tokens in the 1950s. It must be stressed, however, that there was a sudden decline (32,570 tokens) in the use of *like* in the 1960s. However, the use of *like* gradually increased to 37,040 tokens in the 1970s. Exactly the same can be said of the 1980s and the 1990s. The use of *like* steadily increased to 53,843 tokens and 96,688 tokens, respectively. Most importantly, there was a dramatic increase (218,555 tokens) in the use of *like* in the 2000s. More specifically, there was a sharp rise of 121,867 tokens in the 2000s. Finally, *like* reached a peak (309,656 tokens) in the 2010s, which in turn suggests that it was the most preferred (309,656 tokens) by the movie writers of six countries. We thus conclude

that *like* had the highest use (309,656 tokens) in the 2010s, while it had the lowest one (22,703 tokens) in the 1930s.

Now attention is paid to Figure 1:

Figure 1 Diachronic aspects of care for and like



Quite interestingly, *care for* and *like* have two things in common. That is to say, they had the highest use (744 tokens vs. 309,656 tokens) in the 2010s, while they had the lowest one (179 tokens vs. 22,703 tokens) in the 1930s. However, *care for* does not pattern with *like*. When it comes to the former, it shows an upward trend in its use except for the 1960s and the 1970s. Talking about the latter, it also shows an upward trend in its use, but in the 1960s, it shows a downward trend, hence leading to the difference between them. Another difference between them is their average frequency. The average frequency of *care for* is 381 tokens, while that of *like* is 94,244 tokens. This in turn indicates that the movie writers of six countries preferred using *like* to using *care for* in their movies. One more difference is that there was a dramatic increase (723 tokens) in the use of *care for* in the 2000s, while there was a dramatic one (218,555 tokens and 309,656 tokens) in the use of *like* in the 2000s and the 2010s, respectively. We thus conclude that *care for* and *like* are partly similar and partly different in their use.

Now let us have a look at Table 2:

Table 2 Use of care for and like in the movies of six countries

Country	US/CA	UK/IE	AU/NZ	Misc
Care for	2,492	661	46	230
Like	677,188	108,717	13,405	48,889

Quite interestingly, *care for* was the most widely used (2,492 tokens) in American and Canadian movies. However, it was less frequently used (661 tokens) in British and Irish movies. The same can be said about *like*. It ranks first (677,188 tokens) in American and Canadian movies. It ranks second (108,717 tokens) in British and Irish ones. We thus conclude that *care for* and *like* were the most widely used (2,492 tokens vs. 677,188 tokens) in American and Canadian movies, while they were less widely used (661 tokens vs. 108,717 tokens) in British and Irish ones.

3. The Collocations of care for and like in the BNC

The goal of this section is to go over the collocations of *care for* and *like* in the BNC. Also, we aim to compare the collocation of *care for* and that of *like* in the BNC. Table 3 shows the collocation of *care for* in the BNC:

Table 3: Collocation of care for in the BNC

Number	Collocation of care for	Frequency
1	care for people	37
2	care for children	32
3	care for patients	19
4	care for others	15
5	care for dementia	6
6	care for women	6
7	care for relatives	5
8	care for babies	5
9	care for parents	4
10	care for trees	3
11	care for books	3
12	care for animals	3
13	care for adult	2
14	care for coffee	2
15	care for clients	2
16	care for families	2
17	care for life	2
18	care for detail	2
19	care for Londoners	2
20	care for members	2
21	care for mothers	2
22	care for nature	2
23	care for baby	2
24	care for politics	2
25	care for sufferers	2

It is interesting to note that *care for people* had the highest use (37 tokens) in the BNC. This in turn implies that the phrase *care for people* is the most preferable one (37 tokens) for the British. It is interesting to observe the phrase *care for children*. It ranks second (32 tokens) in the BNC. This in turn means that the phrase *care for children* is the second most widely used (32 tokens) in the UK. Also, it must be stressed that the collocation *patients* is the third most frequently used (19 tokens) in the BNC. It should also be noted that the phrase *care for others* ranks fourth (15 tokens) in the BNC. It is also worth noticing that the phrase *care for dementia* is the fifth most widely used (6 tokens). From all of this, it is evident that the phrase *care for people* is the most preferred (37 tokens) by the British, followed by *care for children*, *care for*

patients, care for others, and care for dementia, in that order. Additionally, it is worth observing that *care for coffee* ranks twelfth (2 tokens) in the BNC. It is also appropriate to mention that *care for Londoners* ranks twelfth (2 tokens) in the BNC. We thus conclude that the phrase *care for people* is the most preferable one (37 tokens) among the British.

Now have a look at the collocation of *like* in the BNC:

Table 4 Collocation of like in the BNC

Number	Collocation of like	Frequency
1	like Mr	172
2	like people	165
3	like hell	150
4	like water	106
5	like children	98
6	like Mrs	75
7	like men	72
8	like home	68
9	like sir	65
10	like others	61
11	like things	59
12	like women	59
13	like death	51
14	like dogs	51
15	like school	49
16	like animals	47
17	like blood	44
18	like ice	44
19	like fish	43
20	like shit	43
21	like manner	42
22	like food	41
23	like thunder	41
24	like flies	40
25	like coffee	37

It is worth mentioning that *like Mr* had the highest use (172 tokens) in the BNC. This in turn indicates that the phrase *like Mr* is the most preferable one (172 tokens) among the British. Note that the phrase *like Mr* is followed by the phrase *like people*. More specifically, the latter is the second most widely used (165 tokens) in the BNC. It is interesting to observe that the phrase *like hell* ranks third (150 tokens) in the BNC. It must also be noted that the phrase *like water* is the fourth most frequently used (106 tokens) in the BNC. What is interesting about the phrase *like children* is that it ranks fifth (98 tokens) in the BNC. It can thus be inferred that the phrase *like Mr* is the most preferred (172 tokens) by the British, followed by *like people*, *like*

hell, like water, and like children, in descending order. Notice that *care for people* ranks first (37 tokens) in the BNC, while *like people* ranks second (165 tokens). Quite interestingly, *care for coffee* ranks thirteenth (2 tokens) in the BNC, while *like coffee* ranks twenty fifth (37 tokens). More interestingly, *care for others* ranks fourth (15 tokens) in the BNC, while *like others* ranks tenth (61 tokens). Most importantly, 5 of 45 nouns are the collocations of both *care for* and *like*. To be more specific, the nouns *people, children, others, women, and coffee* are the collocations of both *care for* and *like*. This amounts to saying that *care for* is 11.11% the same as *like* in the top 25. It seems thus appropriate to conclude that *care for* and *like* have a low similarity in the top 25.

4. The Collocations of care for and like in the ChatGPT

In what follows, we probe into the collocations of *care for* and *like* in the ChatGPT. Also, we attempt to contemplate the similarity between *care for* and *like* in the ChatGPT. Table 5 shows the collocation of *care for* in the ChatGPT:

Table 5 Collocation of care for in the ChatGPT

Number	Collocation of care for
1	children
2	plants
3	elderly
4	environment
5	patients
6	people
7	family
8	pets
9	health
10	safety
11	nature
12	others
13	needy
14	orphans
15	wildlife
16	community
17	well-being
18	humanity
19	the sick
20	individuals

It is interesting to note that the first collocation that the ChatGPT recommended is the noun *children*. This in turn indicates that there is a high probability that the phrase *care for children* can be used. Perhaps it is worthwhile noting that the second recommendation of the ChatGPT

is the noun *animals*, as illustrated in Table 5. This in turn means that train data include the phrase *care for animals*. Again, there is a high probability that this phrase can be used. Quite interestingly, the third collocation that the ChatGPT recommended is the noun *plants*. This in turn implies that the ChatGPT predicts that the phrase *care for plants* can also be much used. This prediction stems from its train data and webs. It is also worth noticing that the fourth recommendation of the ChatGPT is the collocation *elderly*. Another recommendation of the ChatGPT is the noun *environment*. This amounts to saying that the phrase *care for environment* also occurs in train data. It should also be noted that the phrase *care for pets* was recommended by the ChatGPT. We thus conclude that the first recommendation of the ChatGPT is the phrase *care for children*.

Now let us take a look at Table 6:

Table 6 Collocation of like in the ChatGPT

Number	Collocation of care for
1	food
2	music
3	movies
4	books
5	animals
6	cars
7	sports
8	travel
9	art
10	photography
11	fashion
12	technology
13	games
14	dancing
15	nature
16	desserts
17	science
18	hobbies
19	people
20	flowers

It is particularly noteworthy that the first collocation that the ChatGPT recommended is the noun *food*. This in turn suggests that there is a high probability that the phrase *like food* can be used. It should also be pointed out that the second recommendation of the ChatGPT is the collocation *music*. It is thus safe to assume that the phrase *like music* also appears in train data.

It would be unfair not to note that the third collocation that the ChatGPT recommended is the noun *movies*. More interestingly, the phrase *like technology* was also recommended by the ChatGPT. Most importantly, only 2 of 38 nouns are the collocations of both *care for* and *like* in the top 20. This in turn indicates that *care for* is 5.26% the same as *like* in the analysis of the top 20. From this it seems clear that *care for* and *like* have a low similarity in the top 20. We thus conclude that *care for* and *like* are synonyms, but they have a low degree of similarity in the top 20. For the analysis of synonyms and big data, see Kang (2022a, 2022b, 2022c, 2022d, 2023a, 2023b).

5. Conclusion

To sum up, we have compared *care for* and *like* in two corpora and ChatGPT. In section 2, we have argued that *care for* and *like* have two things in common. That is to say, they had the highest use (744 tokens vs. 309,656 tokens) in the 2010s, while they had the lowest one (179 tokens vs. 22,703 tokens) in the 1930s. In section 3, we have argued that the phrase *care for people* is the most preferred (37 tokens) by the British, followed by *care for children*, *care for patients*, *care for others*, and *care for dementia*, in that order. We have also maintained that the phrase *like Mr* is the most preferred (172 tokens) by the British, followed by *like people*, *like hell*, *like water*, and *like children*, in descending order. We have also shown that *care for* is 11.11% the same as *like* in the analysis of the BNC (the top 25). Finally, we have argued that *care for* is 5.26% the same as *like* in the data from the ChatGPT (the top 20). It can thus be inferred that *care for* and *like* have a low degree of similarity.

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.