

A Comparative Analysis of Sorry for and Sorry about in Two Corpora



Namkil Kang

Far East University South Korea

ABSTRACT: The ultimate goal of this paper is to provide a comparative analysis of *Sorry for* and *Sorry about* in the COCA and BNC. When it comes to the COCA, it is interesting to point out that the types *Sorry for* and *Sorry about* show the same pattern in the TV/movie and fiction genres and the academic genre, whereas they show a different pattern in the blog, web, magazine, newspaper genres and the spoken genre. With respect to the BNC, it is significant to note that *Sorry for* and *Sorry about* show the same pattern in the fiction genre and the spoken genre, whereas they show a different pattern in the newspaper and magazine genres and the misc, non-academic and academic genres. It can thus be inferred that these two types are synonymously used, but they are slightly different from each other in their use. A major point to note that in the COCA, *Sorry for* is the furthest type from *Sorry about* in the TV/movie genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. A further point to note is that in the BNC, *Sorry for* is the furthest type from *Sorry about* in the spoken genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. The COCA clearly shows that *Sorry for people* is the most preferable one for Americans, followed by *Sorry for Mr.*, *Sorry for me*, and *Sorry for Mrs.*, in that order. The COCA shows, on the other hand, that *Sorry about dinner* is the most preferable one for Americans, followed by *Sorry about lunch*, *Sorry about Mr.*, *Sorry about agent*, and *Sorry about that*, in descending order. The COCA clearly indicates that the expression *Sorry for being* is the most preferable one among Americans, followed by *Sorry for interrupting*, *Sorry for having*, *Sorry for getting*, and *Sorry for making*, in descending order. The COCA also shows that the expression *Sorry about being* is the most preferable one for Americans, followed by *Sorry about getting*, *Sorry about calling*, and *Sorry about leaving (Sorry about having)*, in that order. It is worth noting that only two nouns (*Mr.* and *things*) can be the collocations of both *Sorry about* and *Sorry for*. It is also worth pointing out that eleven of thirteen seven gerunds are linked to both *Sorry about* and *Sorry for*. This in turn suggests that *Sorry about* and *Sorry for* are interchangeably used, but they are somewhat different from each other in their use.

KEYWORDS: corpus, COCA, BNC, type, token, sorry for, sorry about,

1. INTRODUCTION

The main goal of this paper is to provide a comparative analysis of *Sorry for* and *Sorry about* in the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). As Murphy (2016, 2019) points out, *Sorry for* and *Sorry about* are synonymously used:

- (1) I am sorry for shouting at you yesterday. (=sorry about shouting at you yesterday).
 - (2) Sorry for the delay. (=Sorry about the delay)
- (Murphy 2019)

We provide a comparative analysis of *Sorry for* and *Sorry about* in the eight genres of the COCA and the seven genres of the BNC. Also, we observe similarities between *Sorry for* and *Sorry about* in terms of the Euclidean distance in the COCA and BNC. It shows how much they are related to each other. Additionally, we provide a collocation analysis in the COCA and make a difference between *Sorry for* and *Sorry about* in terms of the software NetMiner. This paper argues that the types *Sorry for* and *Sorry about* are interchangeably used, but they are slightly different from each other in their use. The organization of this paper is as follows. In section 2.1, we argue that the types *Sorry for* and *Sorry about* show the same pattern in the TV/movie and fiction genres and the academic genre, whereas they show a different pattern in the blog, web, magazine, newspaper genres and the spoken genre. In section 2.2, we show that *Sorry for* is the furthest type from *Sorry about* in the TV/movie genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. In section 3.1, we contend that *Sorry for* and *Sorry about* show the same pattern in the fiction genre and the spoken genre, whereas they show a different pattern in the newspaper and magazine genres and the misc, non-academic and academic genres. This in turn indicates that these two types are interchangeably used, but they are somewhat different

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

from each other in their use. In section 3.2, we argue that *Sorry for* is the furthest type from *Sorry about* in the spoken genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. This in turn suggests that these two types show deep similarities in the academic genre. In section 4, we contend that *Sorry for people* is the most preferable one for Americans, followed by *Sorry for Mr.*, *Sorry for me*, *Sorry for Mrs.*, *Sorry for kids* (*Sorry for swearing*), *Sorry for guys*, and *Sorry for rambling*, in descending order. We also contend that *Sorry about dinner* is the most preferable one for Americans, followed by *Sorry about lunch*, *Sorry about Mr.*, *Sorry about agent*, *Sorry about that* (*Sorry about things*), and *Sorry about breakfast*, in that order. We show that only two nouns (*Mr.* and *things*) can be the collocations of both *Sorry about* and *Sorry for*. This in turn implies that these two types are slightly different from each other in their use. In section 5, we maintain that the expression *Sorry for being* is the most preferable one among Americans, followed by *Sorry for interrupting*, *Sorry for having*, *Sorry for getting*, *Sorry for making* (*Sorry for taking*), *Sorry for calling*, and *Sorry for saying*, in descending order. We also maintain that the expression *Sorry about being* is the most preferable one for Americans, followed by *Sorry about getting*, *Sorry about calling*, *Sorry about leaving* (*Sorry about having*), *Sorry about coming*, and *Sorry about posting*, in that order. Finally, we show that eleven of thirteen seven gerunds are linked to both *Sorry about* and *Sorry for*. This in turn implies that *Sorry about* and *Sorry for* are interchangeably used, but they are somewhat different from each other in their use.

2. A GENRE ANALYSIS OF SORRY FOR AND SORRY ABOUT IN THE COCA

2.1. The Genre Frequency of Sorry for and Sorry about in the COCA

In what follows, we aim to provide a comparative analysis of *Sorry for* and *Sorry about* in the COCA. Table 1 shows the genre frequency of *Sorry for* and *Sorry about* in the COCA:

Table 1. Genre Frequency of Sorry for and Sorry about in the COCA

Genre	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
<i>Sorry for</i>	14,641	2,491	1,924	4,755	1,523	2,705	539	608	96
<i>Sorry about</i>	10,172	504	360	7,297	543	1,306	78	74	10

An important question is “Which type is preferred by Americans?” Table 1 clearly indicates that *Sorry for* is frequently used by Americans. The overall frequency of *Sorry for* is 14,641 tokens, whereas that of *Sorry about* is 10,172 tokens. This in turn shows that the type *Sorry for* is preferred over the type *Sorry about* by Americans. It seems thus reasonable to hypothesize that Americans prefer using *Sorry for* to using *Sorry about*.

An immediate question is “In which genre(s) are the types *Sorry for* and *Sorry about* the most commonly used?” Table 1 clearly shows that *Sorry for* is the most widely used in the TV/movie genre. Likewise, *Sorry about* is the most frequently used in the TV/movie genre. Thus, it is reasonable to assume that Americans prefer *Sorry for* and *Sorry about* the most in the TV/movie genre. It should be noted, however, that *Sorry about* is favored over *Sorry for* in the TV/movie genre. Most importantly, the types *Sorry for* and *Sorry about* show the same characteristic with respect to their ranking in the TV/movie genre of the COCA. It can be inferred from this that the types *Sorry for* and *Sorry about* reveal very close similarities.

It is noteworthy that the types *Sorry for* and *Sorry about* are the second most preferred ones in the fiction genre. This in turn implies that Americans are keen on using the types *Sorry for* and *Sorry about* in their novels. That the types *Sorry for* and *Sorry about* rank second in the fiction genre suggests that they share the same property. It should be pointed out, however, that *Sorry for* is preferable to *Sorry about* in the fiction genre. This indicates that Americans prefer using *Sorry for* to using *Sorry about* in their novels.

It is interesting to point out that the type *Sorry for* is the third most preferred one in the blog genre, whereas the type *Sorry about* is the third most preferred one in the spoken genre. This in turn implies that the types *Sorry for* and *Sorry about* are somewhat different from each other in their use. It is interesting to note that *Sorry for* is preferable to *Sorry about* in the spoken genre. This in turn suggests that Americans prefer using *Sorry for* to using *Sorry about* in daily conversation. It is worth pointing out, on the other hand, that *Sorry for* is preferred over *Sorry about* by American bloggers.

It is worthwhile noting that the type *Sorry for* is the fourth most preferred by Americans in the web genre, whereas the type *Sorry about* is the fourth most preferred by them in the blog genre. This indicates that the type *Sorry for* does not show the same property as the type *Sorry about* in their ranking. It seems thus reasonable to assume that they do not reveal close similarities in their frequency. It is interesting to point out that the type *Sorry for* is favored over the type *Sorry about* in the web genre. The frequency of the type *Sorry for* is five times higher than that of the type *Sorry about* in the web genre.

It is significant to note that the type *Sorry for* is the fifth most preferred one in the spoken genre, whereas the type *Sorry about* is the fifth most preferred one in the web genre. This in turn shows that the type *Sorry for* does not show the same property as the type *Sorry about*. Simply put, they do not reveal close similarities in their use. It is important to note that the frequency of *Sorry for* is two times higher than that of *Sorry about* in the spoken genre. This in turn suggests that *Sorry for* is favored over *Sorry about* in

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

America.

It is interesting to note that the type *Sorry for* is the sixth most preferred one in the newspaper genre, whereas the type *Sorry about* is the sixth most preferred one in the magazine genre. Again, the types *Sorry for* and *Sorry about* show a different pattern in the newspaper and magazine genres. It is worth pointing out that the frequency of *Sorry for* is eight times higher than that of *Sorry about* in the newspaper genre. This may imply that American journalists prefer using *Sorry for* to using *Sorry about* in their newspapers. It should be pointed out, on the other hand, that the type *Sorry for* is favored over *Sorry about* in the magazine genre. More specifically, the frequency of *Sorry for* is 539 tokens, whereas that of *Sorry about* is 78 tokens. The frequency of *Sorry for* is six times higher than that of *Sorry about* in the magazine genre. Thus, the type *Sorry for* is preferred over the type *Sorry about* by American journalists.

Noteworthy is that the type *Sorry for* is the seventh most preferred one in the magazine genre, whereas the type *Sorry about* is the seventh most preferred one in the newspaper genre. Again, the types *Sorry for* and *Sorry about* show a different property in the magazine and newspaper genres. It is worthwhile pointing out that the types *Sorry for* and *Sorry about* have the lowest frequency in the academic genre. We thus conclude that the types *Sorry for* and *Sorry about* show the same pattern in the TV/movie and fiction genres and the academic genre, whereas they show a different pattern in the blog, web, magazine, newspaper genres and the spoken genre. This in turn suggests that the types *Sorry for* and *Sorry about* are synonymously used, but they are slightly different from each other in their use.

Now let us turn our attention to the use of *Sorry for* and *Sorry about* in the eight genres of the COCA:

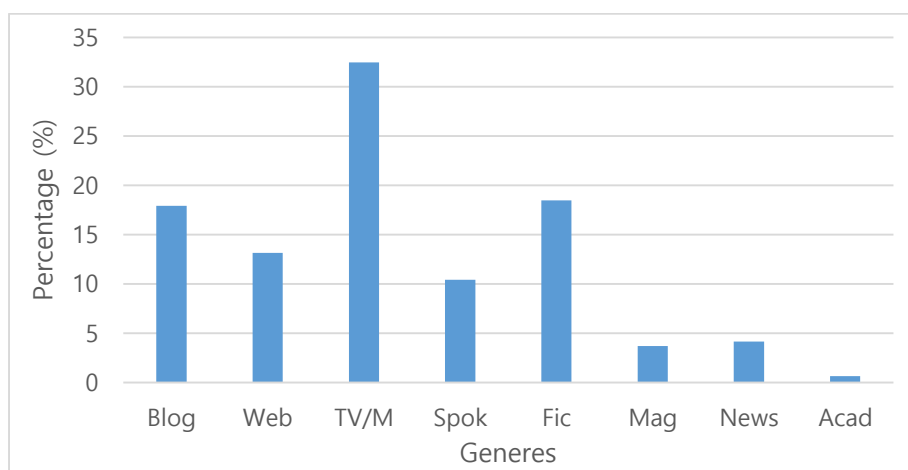


Figure 1. Percentage of the use of *Sorry for* in the eight genres in the COCA

With respect to the use of the type *Sorry for*, it is interesting to note that the TV/movie genre is the most influenced by it, followed by the fiction genre, the blog genre, the web genre, the spoken genre, the newspaper genre, the magazine genre, and the academic genre, in descending order.

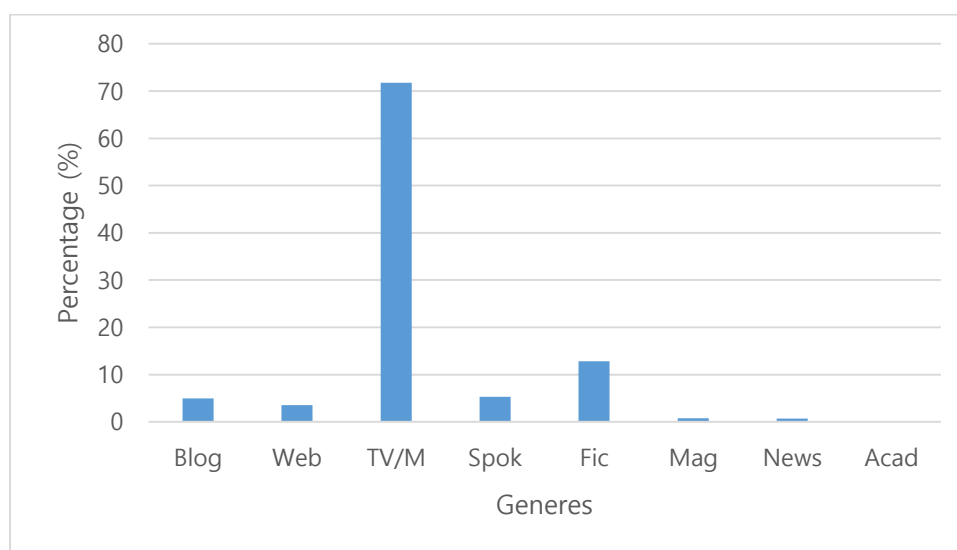


Figure 2. Percentage of the use of *Sorry about* in the eight genres in the COCA

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

It is interesting to point out that the TV/movie genre is the most influenced by the type *Sorry about*, followed by the fiction genre, the spoken genre, the blog genre, the web genre, the magazine genre, the newspaper genre, and the academic genre, in that order.

2.2. The Euclidean Distance in the Eight Genres of the COCA

Now attention is paid to the distance between the type *Sorry for* and the type *Sorry about* in the COCA. The Euclidean distance provides the index about how much *Sorry for* and *Sorry about* are related to each other. It shows the degree of similarities between *Sorry for* and *Sorry about*. We adopt the Euclidean distance as follows:

(2) The Euclidean distance

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Note that the more the figure of the Euclidean distance is low, the more the types *Sorry for* and *Sorry about* indicate close similarities:

Table 2. Euclidean distance between two types in the eight genres of the COCA

	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Percentage of Sorry for	17.91	13.14	32.47	10.4	18.47	3.68	4.15	0.6
Percentage of Sorry about	4.95	3.53	71.73	5.33	12.83	0.76	0.72	0.09
Euclidean distance	12.96	9.61	39.26	5.07	5.64	2.92	3.43	0.51

It is worth noting that *Sorry for* is the furthest type from *Sorry about* in the TV/movie genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. Thus, it is reasonable to assume that the types *Sorry for* and *Sorry about* show deep similarities in the academic genre. Similarly, they reveal very close similarities in the magazine genre. In the magazine genre, the figure of the Euclidean distance is 2.92, which is the second lowest in the eight genres of the COCA. On the other hand, the types *Sorry for* and *Sorry about* show close similarities in the newspaper genre. In the newspaper genre, the figure of the Euclidean distance is 3.43, which is the third lowest in the eight genres of the COCA. We thus conclude that *Sorry for* is the nearest type to *Sorry about* in the academic genre. This in turn implies that the types *Sorry for* and *Sorry about* are interchangeably used, but they are slightly different from each other in their use.

3. A GENRE ANALYSIS OF SORRY FOR AND SORRY ABOUT IN THE BNC

3.1. The Genre Frequency of Sorry for and Sorry about in the BNC

In the following, we aim to provide a comparative analysis of the types *Sorry for* and *Sorry about* in the seven genres of the BNC. Table 3 indicates the genre frequency of *Sorry for* and *Sorry about* in the seven genres:

Table 3. Genre Frequency of Sorry for and Sorry about in the BNC

Genre	All	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-ACAD	ACAD	MISC
Sorry for	920	150	474	43	123	41	9	80
Sorry about	419	162	212	4	4	2	2	31

An important question is “Which type is preferred by the British?” Table 3 clearly indicates that the type *Sorry for* is the preferable one for the British. As shown in Table 3, the overall frequency of the type *Sorry for* is 920 tokens, whereas that of the type *Sorry about* is 419 tokens. This in turn shows that the type *Sorry for* is preferred by the British. It seems thus reasonable to hypothesize that the British prefer using *Sorry for* to using *Sorry about*. It is significant to note that the type *Sorry for* is preferred over the type *Sorry about* by Americans and the British, thus showing that Americans and the British reveal the same pattern with respect to the use of these two types.

An important question is “In which genre are both *Sorry for* and *Sorry about* the most widely used?” Table 3 clearly shows that they are the most widely used in the fiction genre, hence the same pattern. This in turn suggests that British writers are keen on using *Sorry for* and *Sorry about* in their novels. It must be noted, however, that the type *Sorry for* is favored over the type *Sorry about* in the fiction genre. The frequency of *Sorry for* is two times higher than that of *Sorry about*.

It is important to note that the types *Sorry for* and *Sorry about* are the second most preferred ones in the spoken genre. Again, the type *Sorry for* reveals the same property as the type *Sorry about* in the spoken genre of the BNC. That the types *Sorry for* and *Sorry about* rank second in the spoken genre suggests that the British prefer using these two types in daily conversation. It should be noted,

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

however, that *Sorry about* is preferred over *Sorry for* by the British in daily conversation.

It is worth pointing out that the type *Sorry for* is the third most preferred one in the newspaper genre, whereas the type *Sorry about* is the third most preferred one in the misc genre. As indicated in Table 3, the types *Sorry for* and *Sorry about* do not show the same characteristic in their use. It is worthwhile noting that the type *Sorry for* is favored over the type *Sorry about* in the misc genre. On the other hand, in the newspaper genre, the type *Sorry for* is preferable to the type *Sorry about*. This may imply that British journalists prefer using *Sorry for* to using *Sorry about*. More interestingly, the frequency of *Sorry for* is thirty times higher than that of *Sorry about*.

It is noteworthy that the type *Sorry for* is the fourth most preferred one in the misc genre, whereas the type *Sorry about* is the fourth most preferred one in the newspaper and magazine genres. Again, these two types do not show close similarities since they belong to different genres. It must be pointed out, however, that the type *Sorry for* is preferable to the type *Sorry about* in the magazine genre. This in turn shows that British journalists prefer using *Sorry for* to using *Sorry about* in their magazines.

It is interesting to note that the type *Sorry for* is the fifth most preferred one in the magazine genre, whereas the type *Sorry about* is the fifth most preferred one in the academic and non-academic genres. Again, these two types do not show close similarities in the three genres of the BNC. It should be pointed out that *Sorry for* ranks sixth and seventh in the non-academic and academic genres, respectively, but *Sorry about* ranks fifth in the non-academic and academic genres. Again, these two types do not show close similarities in their use. We thus conclude that these two types show the same pattern in the fiction genre and the spoken genre, whereas they show a different pattern in the newspaper and magazine genres and the misc, non-academic and academic genres. From this, it is clear that these two types are interchangeably used, but they are somewhat different from each other in their use.

Now attention is paid to the use of *Sorry for* and *Sorry about* in the seven genres of the BNC:

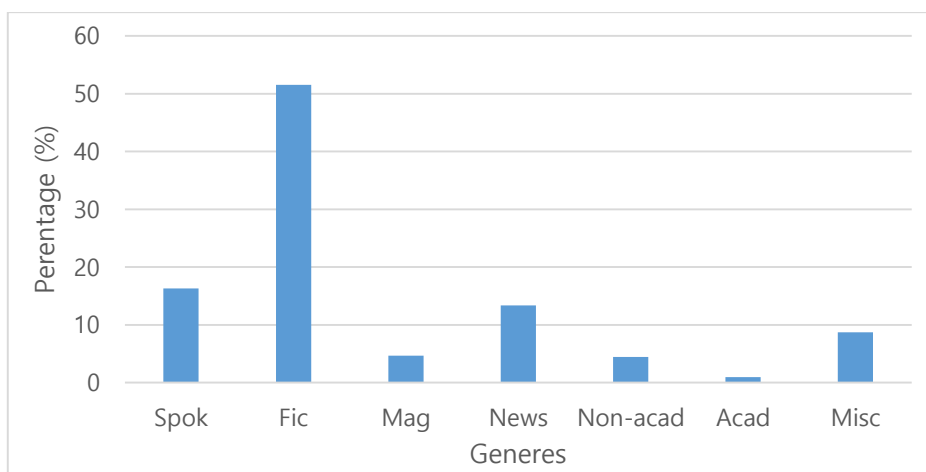


Figure 3. Percentage of the use of Sorry for in the seven genres of the BNC

It is worth noting that the fiction genre is the most influenced by the type *Sorry for*, followed by the spoken genre, the newspaper genre, the misc genre, the magazine genre, the non-academic genre, and the academic genre, in that order.

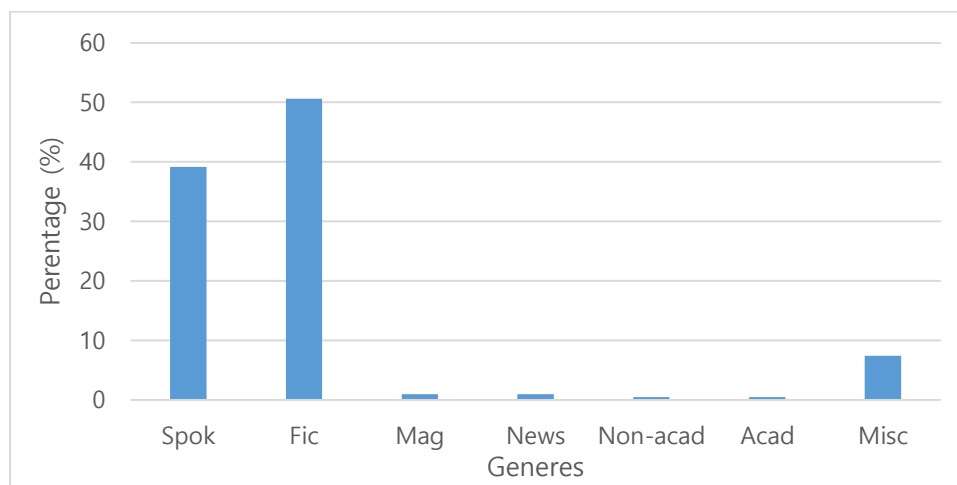


Figure 4. Percentage of the use of Sorry about in the seven genres of the BNC

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

It is worthwhile pointing out that the fiction genre is the most influenced by the type *Sorry about*, followed by the spoken genre, the misc genre, the magazine genre (the newspaper genre), and the non-academic genre (the academic genre), in descending order.

3.2. The Euclidean Distance in the BNC

Now let us turn our attention to the Euclidean distance:

Table 4. Euclidean distance in the Seven Genres of the BNC

	SPOK	FIC	MAG	NEWS	NON-ACAD	ACAD	MISC
Percentage of Sorry for	16.30	51.52	4.67	13.36	4.45	0.97	8.69
Percentage of Sorry about	39.14	50.59	0.95	0.95	0.47	0.47	7.39
Euclidean distance	22.84	0.93	3.72	12.41	3.98	0.5	1.3

Note that the more the figure of the Euclidean distance is low, the more *Sorry for* and *Sorry about* show close similarities. It is important to note that *Sorry for* is the furthest type from *Sorry about* in the spoken genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. This in turn indicates that these two types show deep similarities in the academic genre. In the case of the fiction genre, the figure of the Euclidean distance is 0.93, which is the second lowest. This indicates that these two types show very close similarities. It must be noted, however, that the figure of the Euclidean distance in the other genres is not low, compared with the academic genre and the fiction genre. This seems to suggest that these two types do not reveal very close similarities in the other genres. We thus conclude that the types *Sorry for* and *Sorry about* are interchangeably used, but they are slight different from each other in their use.

4. THE COLLOCATION OF SORRY FOR AND SORRY ABOUT IN THE COCA

In this section, we aim to provide a collocation analysis of the types *Sorry for* and *Sorry about* in the COCA. Table 5 shows the collocation of the type *Sorry for* in the COCA:

Table 5. Collocation of Sorry for in the COCA

Number	Collocation of Sorry for	Frequency
1	Sorry for people	62
2	Sorry for Mr	12
3	Sorry for me	9
4	Sorry for Mrs	8
5	Sorry for kids	7
6	Sorry for swearing	7
7	Sorry for guys	6
8	Sorry for rambling	5
9	Sorry for things	5
10	Sorry for women	5
11	Sorry for men	4
12	Sorry for Miss	4
13	Sorry for Tilikum	4
14	Sorry for governor	3
15	Sorry for Dr	3
16	Sorry for delay	3
17	Sorry for papa	3
18	Sorry for typos	3
19	Sorry for Bustin	2
20	Sorry for Americans	2
21	Sorry for doctors	2
22	Sorry for eavesdropping	2
23	Sorry for folks	2
24	Sorry for flight	2
25	Sorry for harm	2

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

An important question is “Which expression is the most preferred by Americans?” Table 5 clearly shows that the expression *Sorry for people* is the most frequently used. This in turn indicates that the expression *Sorry for people* is the most preferable one for Americans. As illustrated in Table 5, *Sorry for people* is the most preferable one among Americans, followed by *Sorry for Mr.*, *Sorry for me*, *Sorry for Mrs.*, *Sorry for kids (Sorry for swearing)*, *Sorry for guys*, and *Sorry for rambling (Sorry for things)*, in descending order. It is important to point out that the everyday expression *Sorry for delay* ranks sixteenth in the COCA. It is interesting to note, on the other hand, that *Sorry for things* is the eighth most preferred one in America. It should be pointed out that the expression *Sorry for typos* ranks eighteenth in the COCA.

Now we visualize the collocations of the type *Sorry for* in the top twenty five:

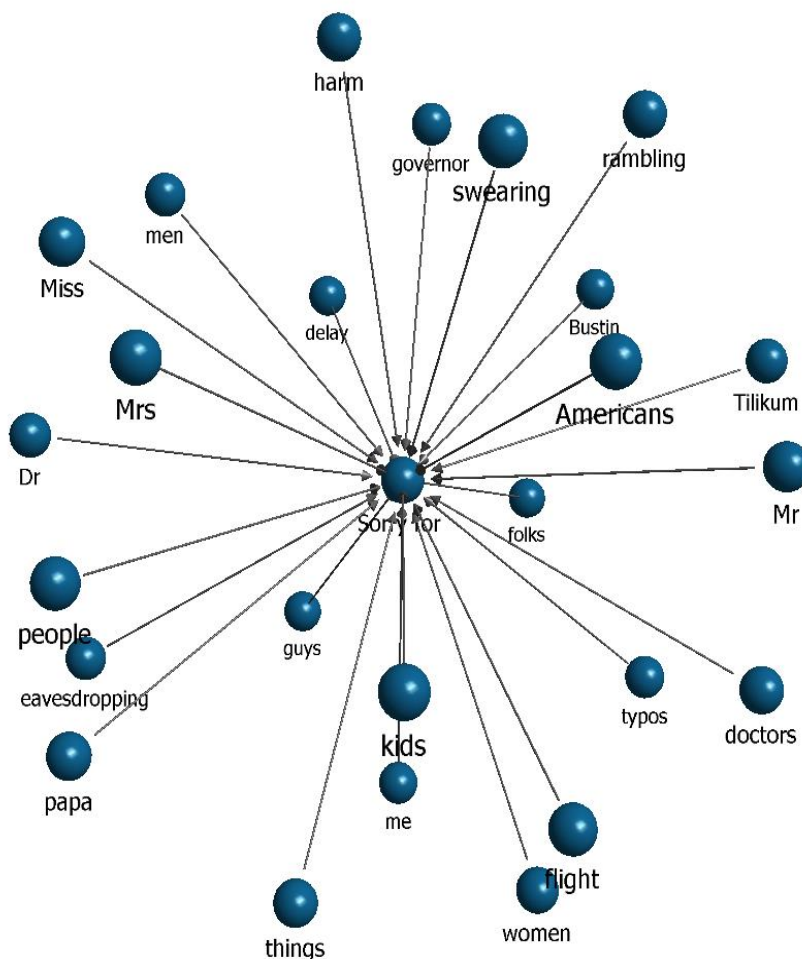


Figure 5. Visualization of the collocation of *Sorry for*

This 3-D visualization of the collocations of *Sorry for* was performed by the software NetMiner. Particular nouns are linked to the type *Sorry for*, which indicates that these nouns have a collocation relationship with the type *Sorry for* and that they are much used with the type *Sorry for*.

Now attention is paid to the collocations of the type *Sorry about* in the COCA:

Table 6. Collocation of *Sorry about* in the COCA

Number	Collocation of <i>Sorry for</i>	Frequency
1	Sorry about dinner	16
2	Sorry about lunch	7
3	Sorry about Mr	5
4	Sorry about Ure	5
5	Sorry about agent	4
6	Sorry about that	3
7	Sorry about things	3
8	Sorry about breakfast	2
9	Sorry about back	2

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

and *Sorry for* do not show very close similarities. It can thus be inferred that the types *Sorry for* and *Sorry about* are interchangeably used, but they are slightly different from each other in their use.

5. THE COLLOCATIONS OF SORRY FOR AND SORRY ABOUT IN THE COCA

In what follows, we aim to provide a comparative analysis of the collocations of *Sorry for* and *Sorry about* in the COCA. Also, we provide the visualization of the collocations of *Sorry for* and *Sorry about* in terms of the software NetMiner:

Table 7. Collocation of Sorry for in the COCA

Number	Collocation of Sorry for	Frequency
1	Sorry for being	215
2	Sorry for interrupting	67
3	Sorry for having	55
4	Sorry for getting	48
5	Sorry for making	35
6	Sorry for taking	35
7	Sorry for calling	28
8	Sorry for saying	25
9	Sorry for coming	24
10	Sorry for bothering	24
11	Sorry for wasting	23
12	Sorry for doing	22
13	Sorry for leaving	20
14	Sorry for trying	20
15	Sorry for hurting	19
16	Sorry for putting	18
17	Sorry for bringing	17
18	Sorry for killing	17
19	Sorry for disturbing	16
20	Sorry for doubting	16
21	Sorry for yelling	16
22	Sorry for barging	15
23	Sorry for waking	15
24	Sorry for keeping	14
25	Sorry for ruining	14

An important question is “Which collocation is the most preferred by Americans?” Table 7 clearly shows that the expression *Sorry for being* is the most commonly used. This in turn implies that *Sorry for being* is the most preferable one for Americans. As exemplified in Table 7, the expression *Sorry for being* is the most preferable one among Americans, followed by *Sorry for interrupting*, *Sorry for having*, *Sorry for getting*, *Sorry for making* (*Sorry for taking*), *Sorry for calling*, and *Sorry for saying*, in descending order. It is important to point out that the everyday expression *Sorry for interrupting* is the second most preferred one in America. It is interesting to point out, on the other hand, that the expression *Sorry for saying* is the eighth most preferred one in America. It should be noted that the everyday expression *Sorry for bothering* is the ninth most preferred one in America. We thus conclude that in America, the collocation *being* is the most widely used along with *Sorry for*.

Now let us turn to the visualization of the collocations of *Sorry for*:

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

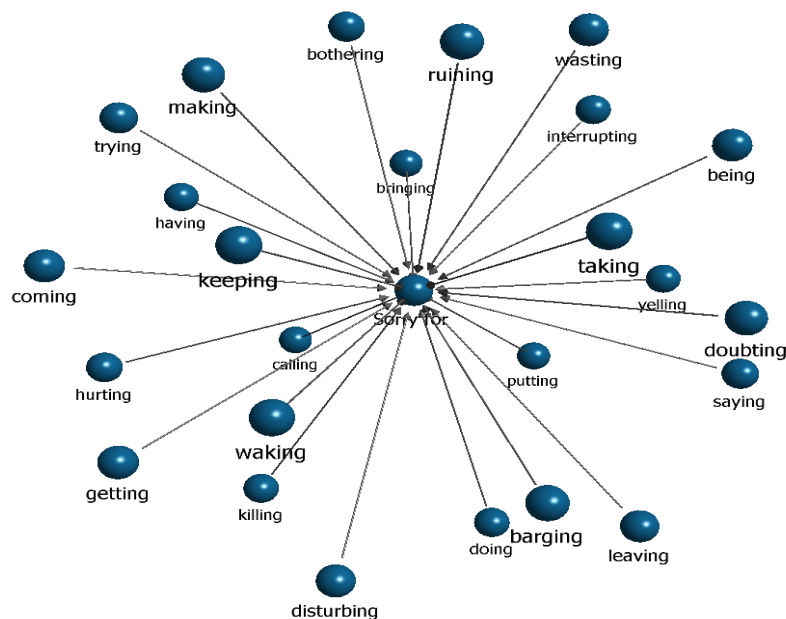


Figure 7. Visualization of the collocations of Sorry for in the COCA

As shown in Figure 7, twenty five gerunds are linked to the type *Sorry for*. Figure 7 indicates that these gerunds are frequently used with the type *Sorry for*. These gerunds that are linked to *Sorry for* are the collocations of *Sorry for*.

Now let us turn our attention to the collocations of *Sorry about* in the top twenty five. Table 8 indicates the collocation of *Sorry about* in the COCA:

Table 8. Collocation of Sorry about in the COCA

Number	Collocation of Sorry for	Frequency
1	Sorry about being	30
2	Sorry about getting	15
3	Sorry about calling	12
4	Sorry about leaving	9
5	Sorry about having	9
6	Sorry about coming	6
7	Sorry about posting	5
8	Sorry about hitting	5
9	Sorry about making	5
10	Sorry about asking	5
11	Sorry about taking	5
12	Sorry about bringing	4
13	Sorry about breaking	4
14	Sorry about lying	4
15	Sorry about messing	4
16	Sorry about putting	4
17	Sorry about throwing	4
18	Sorry about using	4
19	Sorry about missing	3
20	Sorry about running	3
21	Sorry about ruining	3
22	Sorry about killing	3
23	Sorry about losing	3
24	Sorry about keeping	3
25	Sorry about freaking	3

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

An immediate question is “Which gerund is the most preferred by Americans?” Table 8 clearly shows that the expression *Sorry about being* is the most commonly used. This in turn shows that *Sorry about being* is the most preferred by Americans. As alluded to in Table 8, the expression *Sorry about being* is the most preferable one for Americans, followed by *Sorry about getting*, *Sorry about calling*, *Sorry about leaving (Sorry about having)*, *Sorry about coming*, and *Sorry about posting*, in that order. It is significant to note that the expression *Sorry about calling* is the third most preferred one in America, whereas *Sorry for calling* is the seventh most preferred one. It is interesting to note that the expression *Sorry about coming* is the sixth most preferred one in America, whereas *Sorry for coming* is the ninth most preferred one. Interestingly, *Sorry about bringing* ranks twelfth in the COCA, whereas *Sorry for bringing* ranks seventeenth.

Now we visualize the collocations of *Sorry about* and *Sorry for* in the top twenty five:

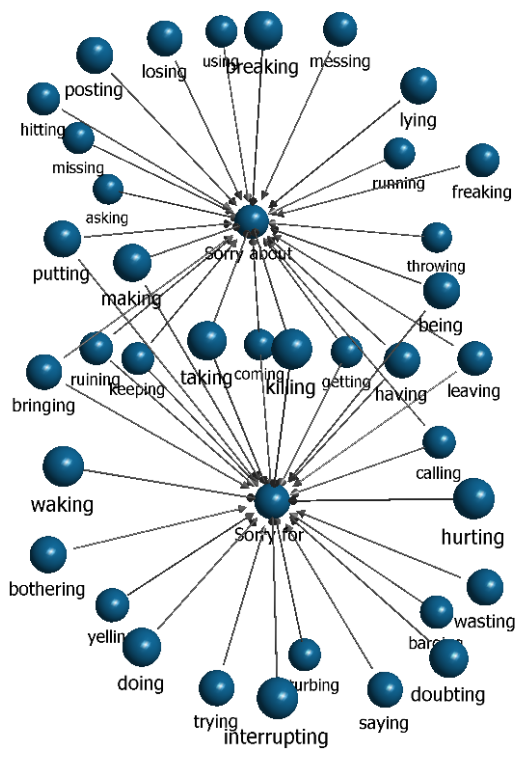


Figure 8. Visualization of the Collocations of Sorry about and Sorry for

As indicated in Figure 8, thirty seven gerunds are linked to *Sorry about* and *Sorry for*. Most importantly, thirteen gerunds are linked to both *Sorry about* and *Sorry for*. This indicates that these gerunds are the collocations of *Sorry about* and *Sorry for*. The gerunds are *being*, *getting*, *calling*, *putting*, *leaving*, *having*, *coming*, *making*, *taking*, *bringing*, *killing*, *ruining*, and *keeping*. Thus, eleven of thirteen seven gerunds are linked to both *Sorry about* and *Sorry for*. This in turn implies that *Sorry about* and *Sorry for* are interchangeably used, but they are somewhat different from each other in their use.

6. CONCLUSION

To sum up, we have provided a comparative analysis of *Sorry for* and *Sorry about* in the COCA and BNC. In section 2.1, we have argued that the types *Sorry for* and *Sorry about* show the same pattern in the TV/movie and fiction genres and the academic genre, whereas they show a different pattern in the blog, web, magazine, newspaper genres and the spoken genre. In section 2.2, we have maintained that *Sorry for* is the furthest type from *Sorry about* in the TV/movie genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. In section 3.1, we have contended that *Sorry for* and *Sorry about* show the same pattern in the fiction genre and the spoken genre, whereas they show a different pattern in the newspaper and magazine genres and the misc, non-academic and academic genres. This in turn indicates that these two types are interchangeably used, but they are somewhat different from each other in their use. In section 3.2, we have argued that *Sorry for* is the furthest type from *Sorry about* in the spoken genre, whereas *Sorry for* is the nearest type to *Sorry about* in the academic genre. This in turn indicates that these two types show deep similarities in the academic genre. In section 4, we have argued that *Sorry for people* is the most preferable one among Americans, followed by *Sorry for Mr.*, *Sorry for me*, *Sorry for Mrs.*, *Sorry for kids (Sorry for swearing)*, *Sorry for guys*, and *Sorry for rambling (Sorry for things)*, in descending order. We have further argued that *Sorry about dinner* is the most preferable one for Americans,

A Comparative Analysis of Sorry for and Sorry about in Two Corpora

followed by *Sorry about lunch*, *Sorry about Mr.*, *Sorry about agent*, *Sorry about that (Sorry about things)*, and *Sorry about breakfast*, in that order. We have shown that only two nouns (*Mr.* and *things*) can be the collocations of both *Sorry about* and *Sorry for*. This in turn implies that these two types are slightly different from each other in their use. In section 5, we have maintained that the expression *Sorry for being* is the most preferable one among Americans, followed by *Sorry for interrupting*, *Sorry for having*, *Sorry for getting*, *Sorry for making (Sorry for taking)*, *Sorry for calling*, and *Sorry for saying*, in descending order. We have also maintained that the expression *Sorry about being* is the most preferable one for Americans, followed by *Sorry about getting*, *Sorry about calling*, *Sorry about leaving (Sorry about having)*, *Sorry about coming*, and *Sorry about posting*, in that order. Finally, we have shown that eleven of thirteen seven gerunds are linked to both *Sorry about* and *Sorry for*. This in turn implies that *Sorry about* and *Sorry for* are interchangeably used, but they are somewhat different from each other in their use.

REFERENCES

- 1) British National Corpus (BNC). 20, October 2021. Online <https://corpus.byu.edu/bnc>
- 2) Corpus of Contemporary American English (COCA). 20, October 2021. Online <https://corpus.byu.edu/coca>
- 3) Murphy, R. (2016). *Grammar in Use*. Cambridge University Press.
- 4) Murphy, R. (2019). *English Grammar in Use*. Cambridge University Press.