

On Keep and Keep on: A Corpora-based Analysis

Namkil Kang

Far East University, South Korea

ABSTRACT

The main goal of this paper is to provide a comparative analysis of *keep* and *keep on* in the TV Corpus and the British National Corpus. With respect to the TV Corpus, it is interesting to note that *keep* was favored over *keep on* from the 1950s to the 2010s. A further point to note is that *keep* and *keep on* reached a peak (28,586 tokens vs. 1,303 tokens) in the 2010s. In addition, it is worth pointing out that *keep* was always preferable to *keep on* from the 1950s to the 2010s. With respect to the BNC, it is probably worthwhile pointing out that *keep* is 71.42% the same as *keep on* in the seven genres of the BNC. When it comes to the Euclidean distance between *keep* and *keep on*, the former is the nearest from the latter in the fiction genre, whereas the former is the furthest from the latter in the spoken genre. Quite interestingly, the BNC clearly shows that *keep going* and *keep on going* are the most preferred (430 tokens vs. 24 tokens) by the British. Finally, it is significant to note that 43.58% of thirty nine gerunds are the collocations of both *keep* and *keep on*.

Keywords: TV Corpus, BNC, *keep*, *keep on*, type, token

INTRODUCTION

According to Murphy (2016, 2019), *keep* and *keep on* are used synonymously, as illustrated in (1):

- (1) Keep or keep on (=do something continuously or repeatedly)
(Murphy 2019: 106)

The main purpose of this paper is to provide a comparative analysis of *keep* and *keep on* in the TV Corpus (TVC) and the British National Corpus (BNC). First, we consider which type is the preferable one in the TV programs of America, the UK, Ireland, Canada, Australia, and New Zealand, observe the diachronic use of *keep* and *keep on*, and compare them. Second, we consider the genre frequency of *keep* and *keep on* in the BNC. By examining the use of *keep* and *keep on* in seven genres, we can see how much alike they are. Additionally, by measuring the Euclidean distance between *keep* and *keep on*, we can see how close they are in each genre. Finally, after considering the collocations of *keep* and *keep on*, we investigate the similarity between *keep* and *keep on*. This paper is organized as follows. In section 2, we try to show that *keep* was favored over *keep on* from the 1950s to the 2010s. We also show that *keep* and *keep on* reached a peak (28,586 tokens vs. 1,303 tokens) in the 2010s. Additionally, we argue that *keep* was always preferable to *keep on* from the 1950s to the 2010s. In section 3, we contend that *keep* is 71.42% the same as *keep on* in the seven genres of the BNC. We also maintain that *keep* is the nearest from *keep on* in the fiction genre, but the former is the furthest from the latter in the spoken genre. In section 4, we argue that *keep going* and *keep on going* are the most

preferred (430 tokens vs. 24 tokens) by the British. We further argue that 43.58% of thirty nine gerunds are the collocations of both *keep* and *keep on*.

THE TV CORPUS

In what follows, we observe the diachronic use of *keep* and *keep on* from the 1950s to the 2010s and compare them. Table 1 shows the use and frequency of *keep* and *keep on* from the 1950s to the 2010s:

Table 1 Frequency of keep and keep on in the TV Corpus

Type	Keep	Keep on
All	49,298	2,467
1950s	158	21
1960s	1,022	109
1970s	1,014	95
1980s	1,677	172
1990s	3,862	197
2000s	12,979	570
2010s	28,586	1,303
US/CA	42,234	2,012
UK/IE	5,845	391
AU/NZ	966	50
MISC	253	14

An important question is “Which type is the preferable one in the TV programs of six countries?” Table 1 clearly indicates that *keep* was preferable to *keep on* from the 1950s to the 2010s. To be more specific, the overall frequency of *keep* is 49,298 tokens, whereas that of *keep on* is 2,467 tokens. This in turn suggests that the type *keep* (49,298 tokens) was preferred over the type *keep on* (2,467 tokens) by the celebrities of six countries. The following graph shows the diachronic use of *keep* and *keep on* from the 1950s to the 2010s:

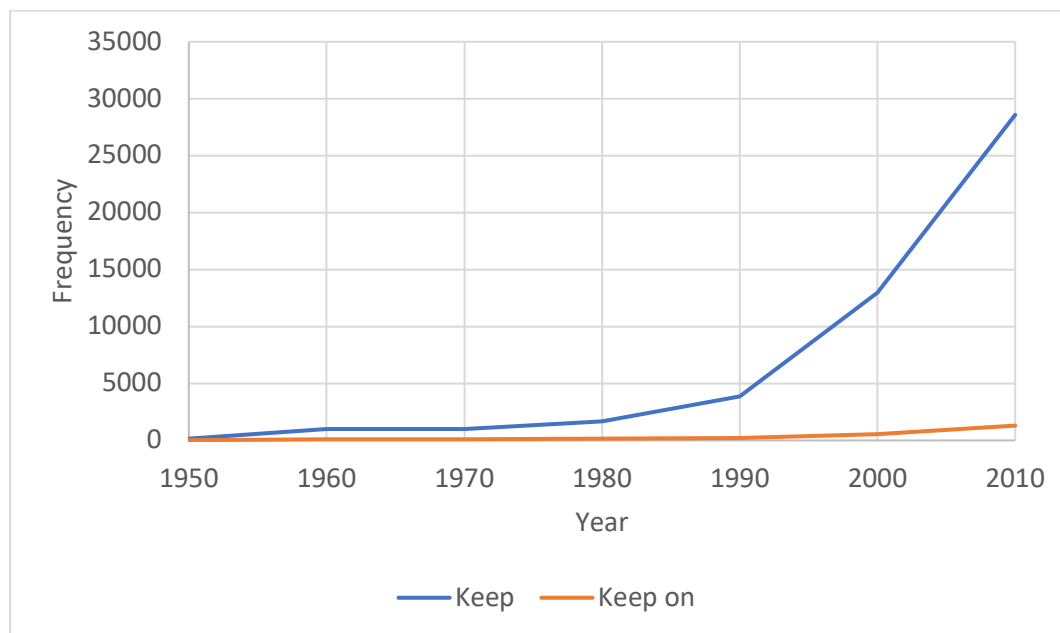


Figure 1 Frequency of keep and keep on from the 1950s to the 2010s

There was a dramatic increase (a rise of 864 tokens) in the figure of *keep* from the 1950s to the 1960s. Interestingly, there was a sudden decline (a fall of 8 tokens) in the figure of *keep* from the 1960s to the 1970s. More interestingly, there was a steady increase (a rise of 663 tokens) in the figure of *keep* from the 1970s to the 1980s. More importantly, there was a sharp rise (an increase of 2,185 tokens) in the figure of *keep* from the 1980s to the 1990s. It is significant to note that there was a dramatic increase (a sharp increase of 24,724 tokens) in the figure of *keep* from the 1990s to the 2010s. It is interesting to point out that *keep* had the highest frequency (28,586 tokens) in the 2010s, whereas it had the lowest frequency (158 tokens) in the 1950s. This in turn indicates that *keep* was the most preferred one (28,586 tokens) in the TV programs of six countries in the 2010s, whereas it was the least preferred one (158 tokens) in the 1950s. Most importantly, *keep* was the most preferred (42,234 tokens) by Americans and Canadian celebrities, followed by British and Irish ones (5,845 tokens), and Australian and New Zealand ones (966 tokens).

It is interesting to note that there was a slight increase (a rise of 88 tokens) in the figure of *keep on* from the 1950s to the 1960s. It should be noted, however, that there was a sudden fall (a decrease of 14 tokens) in the figure of *keep on* from the 1960s to the 1970s. Interestingly, there was a gradual increase (a rise of 102 tokens) in the figure of *keep on* from the 1970s to the 1990s. Quite interestingly, there was a dramatic rise (a rise of 1,106 tokens) in the figure of *keep on* from the 1990s to the 2010s. Most importantly, *keep on* reached a peak (1,303 tokens) in the 2010s, which in turn suggests that it was the most preferred one (the highest frequency) in that period. It is important to note that *keep* was always preferable to *keep on* from the 1950s to the 2010s. Additionally, there was a big difference between the frequency of *keep* and that of *keep on* after the 1990s. When it comes to *keep on*, it was the most preferred (2,012 tokens) by American and Canadian celebrities, followed by British and Irish ones (391 tokens), and Australian and New Zealand ones (50 tokens). We thus conclude that *keep* is preferable to *keep on* in the TV programs of six countries (America, the UK, Canada, Australia, New Zealand, and Ireland) from the 1950s to the 2010s.

KEEP AND KEEP ON IN THE BNC

In the following, we provide a frequency analysis of *keep* and *keep on* in the seven genres of the BNC. Table 2 shows the use and frequency of *keep* and *keep on* in the BNC:

Table 2 Frequency of *keep* and *keep on* in the BNC

GENRE	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-ACAD	ACADEMIC	MISC
Keep	2,671	995	607	216	294	165	53	341
Keep on	333	91	72	21	51	27	11	60

An important question is “Which type is the preferable one for the British?” Table 2 clearly indicates that the type *keep* is the preferred one in the UK. More specifically, the overall frequency of *keep* is 2,671 tokens, whereas that of *keep on* is 333 tokens. The frequency of *keep* (2,671 tokens) is eight times higher than that of *keep on* (333 tokens). This in turn implies that *keep* is preferable to *keep on* in the UK.

It is significant to note that *keep* and *keep on* rank first (995 tokens vs. 91 tokens) in the spoken genre of the BNC. Quite interestingly, *keep* and *keep on* show the same property in rank-one,

hence revealing a high similarity in the spoken genre. It should be noted, however, that *keep* is preferable to *keep on* in the spoken genre. The frequency of *keep* is 995 tokens, whereas that of *keep on* is 91 tokens (the frequency of *keep* is ten times higher). This in turn implies that the British prefer using *keep* (995 tokens) rather than *keep on* (91 tokens) in daily conversation.

It is worth mentioning that *keep* and *keep on* rank second (607 tokens vs. 72 tokens) in the fiction genre. Interestingly, *keep* shows the same property as *keep on*, thus revealing a high similarity in the fiction genre. It should be pointed out, however, that the frequency of *keep* (607 tokens) is eight times higher than that of *keep on* (72 tokens) in the fiction genre. This in turn suggests that British writers prefer using *keep* to using *keep on* in their novels.

It is worthwhile noting that *keep* and *keep on* rank third (341 tokens vs. 60 tokens) in the misc genre. Again, *keep* and *keep on* show the same pattern in rank-three, hence revealing a high similarity in the misc genre. It should be noted, however, that *keep* (341 tokens) is favored over *keep on* (60 tokens) in the misc genre.

It is interesting to point out that *keep* and *keep on* rank fourth (294 tokens vs. 51 tokens) in the newspaper genre. More interestingly, *keep* shows the same property as *keep on* in rank-four, thus revealing a high similarity in the newspaper genre. It must be noted, however, that the frequency of *keep* (294 tokens) is five times higher than that of *keep on* (51 tokens). We take this as indicating that *keep* is favored over *keep on* in the newspaper genre. This in turn indicates that British journalists prefer using *keep* rather than using *keep on* in their newspapers.

It is noteworthy that *keep* ranks fifth (216 tokens) in the magazine genre, whereas *keep on* ranks fifth (27 tokens) in the non-academic genre. Interestingly, *keep* does not show the same pattern as *keep on* in rank-five. It must be noted that the frequency of *keep* (216 tokens) is ten times higher than that of *keep on* (21 tokens) in the magazine genre. This in turn indicates that British journalists prefer using *keep* to using *keep on* in their magazines. Also, it should be pointed out that *keep* (165 tokens) is preferable to *keep on* (27 tokens) in the non-academic genre (the frequency of *keep* is six times higher). It is thus reasonable to assume that the British prefer using *keep* (165 tokens) rather than using *keep on* (27 tokens) in the non-academic genre.

It is worth observing that *keep* ranks sixth (165 tokens) in the non-academic genre, whereas *keep on* ranks sixth (21 tokens) in the magazine genre. Interestingly, *keep* (165 tokens) does not show the same property as *keep on* (21 tokens) in rank-six, thus revealing a low similarity in two genres.

It is probably worthwhile pointing out that *keep* and *keep on* rank seventh (53 tokens vs. 11 tokens) in the academic genre. Interestingly, *keep* (53 tokens) shows the same pattern as *keep on* (11 tokens) in rank-seven, hence revealing a high similarity in the academic genre. It must be noted, however, that the frequency of *keep* (53 tokens) is four times higher than that of *keep on* (11 tokens). We take this fact as indicating that *keep* (53 tokens) is preferable to *keep on* (11 tokens) in the academic genre. To sum up, *keep* and *keep on* show the same pattern in the spoken, misc, and academic genres and the newspaper and fiction genres, but they show a different pattern in the other genres (two genres). From this, it is clear that *keep* is 71.42% the

same as *keep on* in seven genres. It can thus be inferred that *keep* and *keep on* show a high similarity in the seven genres of the BNC.

The following graphs show the percentage of the frequency of *keep* and *keep on* in seven genres:

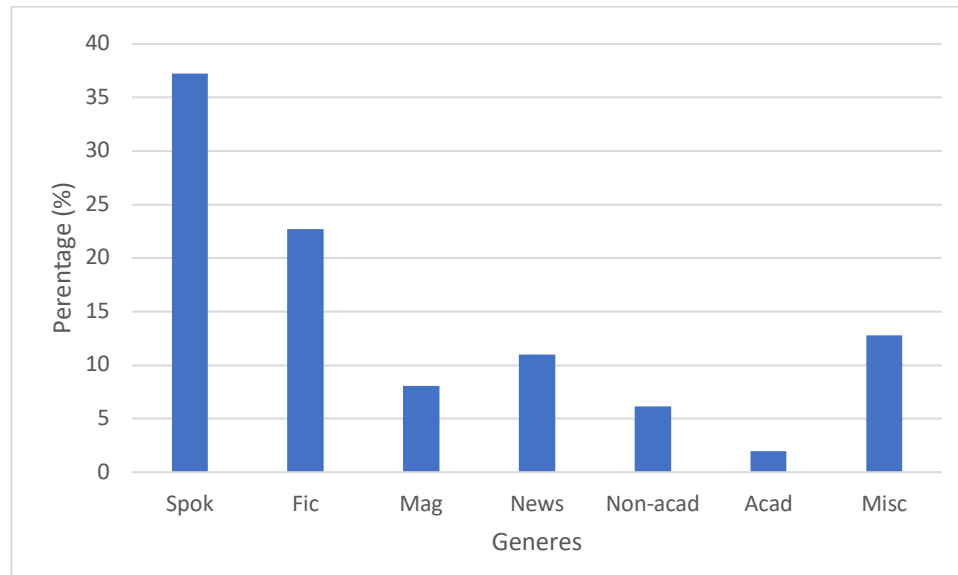


Figure 2 Percentage of the frequency of *keep* in the BNC

As exemplified in Figure 2, the spoken genre is the most influenced by the type *keep*, followed by the fiction genre, the misc genre, the newspaper genre, the magazine genre, the non-academic genre, and the academic genre, in that order.

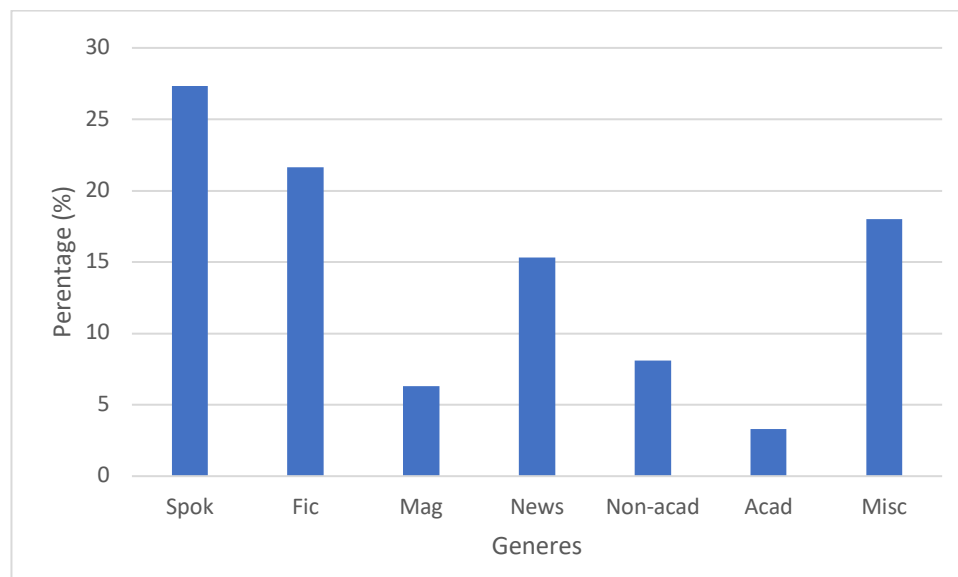


Figure 3 Percentage of the frequency of *keep on* in the BNC

As indicated in Figure 3, the spoken genre is the most influenced by the type *keep on*, followed by the fiction genre, the misc genre, the newspaper genre, the non-academic genre, the magazine genre, and the academic genre, in descending order.

Now let us turn our attention to the Euclidean distance between *keep* and *keep on*. The Euclidean distance provides an index about the similarity between *keep* and *keep on* in seven genres:

(2) The Euclidean distance:

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

The following table indicates the Euclidean distance between *keep* and *keep on* in the seven genres of the BNC:

Table 3 Euclidean distance between keep and keep on in seven genres

Genre	Spok	Fic	Mag	News	Non-acad	Acad	Misc
Percentage of keep	37.25	22.72	8.08	11	6.17	1.98	12.76
Percentage of keep on	27.32	21.62	6.30	15.31	8.10	3.30	18.01
Euclidean distance	9.93	1.1	1.78	4.31	1.93	1.32	5.25

More interestingly, *keep* is the furthest from *keep on* in the spoken genre. The Euclidean distance between *keep* and *keep on* in the spoken genre is 9.93, which is the highest. This in turn indicates that *keep* and *keep on* reveal a low degree of similarity in the spoken genre. On the other hand, *keep* is the nearest to *keep on* in the fiction genre. The Euclidean distance between *keep* and *keep on* in the fiction genre is 1.1, which is the lowest. This in turn suggests that *keep* and *keep on* show a high similarity in the fiction genre. Similarly, the Euclidean distance between *keep* and *keep on* in the academic genre is 1.32, which is the second lowest. This in turn implies that *keep* and *keep on* show a high degree of similarity in the academic genre. We thus conclude that *keep* is the nearest from *keep on* in the fiction genre, but the former is the furthest from the latter in the spoken genre.

THE COLLOCATIONS OF KEEP AND KEEP ON IN THE BNC

In what follows, we consider the collocations of *keep* and *keep on* in the BNC. Also, by using the software NetMiner, we try to compare *keep* with *keep on* in the top 28. Table 4 shows the collocation of *keep* in the BNC:

Table 4 Collocation of keep in the BNC

Number	Collocation	Frequency
1	keep going	430
2	keep saying	134
3	keep telling	102
4	keep coming	101
5	keep trying	89
6	keep getting	75
7	keep looking	61
8	keep moving	60
9	keep talking	57
10	keep doing	55
11	keep thinking	51
12	keep asking	48
13	keep forgetting	42
14	keep working	32
15	keep putting	29
16	keep sending	29
17	keep taking	29
18	keep changing	28
19	keep reminding	28
20	keep playing	25
21	keep having	24
22	keep making	21
23	keep calling	20
24	keep pushing	20
25	keep running	18
26	keep hearing	17
27	keep using	17
28	keep watching	17

An important question is “Which expression is the most preferred by the British?” Table 4 clearly shows that the expression *keep going* is the most frequently used one (430 tokens) in the UK. This in turn implies that *keep going* is the most preferable one (430 tokens) for the British. As illustrated in Table 4, *keep going* is the most preferred (430 tokens) by the British, followed by *keep saying*, *keep telling*, *keep coming*, *keep trying*, *keep getting*, *keep looking*, and *keep moving*, in that order. It is interesting to note that *keep trying* ranks fifth (89 tokens) in the BNC. It is noteworthy, on the other hand, that *keep talking* ranks ninth (57 tokens) in the BNC. Quite interestingly, *keep thinking* ranks eleventh (51 tokens) in the BNC. Finally, *keep working* ranks fourteenth (32 tokens) in the BNC. We thus conclude that the everyday expression *keep going* is the most preferable one (430 tokens) for the British.

Now attention is paid to the collocation of *keep on* in the top 28:

Table 5 Collocation of keep on in the BNC

Number	Collocation	Frequency
1	keep on going	24
2	keep on doing	19
3	keep on trying	15
4	keep on saying	13
5	keep on running	13
6	keep on walking	10
7	keep on playing	8
8	keep on using	7
9	keep on asking	6
10	keep on looking	6
11	keep on forgetting	6
12	keep on talking	6
13	keep on thinking	6
14	keep on getting	5
15	keep on coming	5
16	keep on being	5
17	keep on telling	5
18	keep on fighting	4
19	keep on growing	4
20	keep on having	4
21	keep on paying	4
22	keep on rising	4
23	keep on winning	4
24	keep on repeating	3
25	keep on making	3
26	keep on learning	3
27	keep on increasing	3
28	keep on giving	3

An immediate question is “Which expression is the most preferred one in the UK?” Table 5 clearly indicates that *keep on going* is the most frequently used one (24 tokens) in the UK. This in turn suggests that *keep on going* is the most preferable one (24 tokens) for the British. As alluded to in Table 5, *keep on going* is the most preferred (24 tokens) by the British, followed by *keep on doing*, *keep on trying*, *keep on saying*, *keep on running*, *keep on walking*, *keep on playing*, *keep on using*, and *keep on asking*, in descending order. Most importantly, *keep on going* and *keep going* rank first (24 tokens vs. 430 tokens) in the BNC. It is interesting to point out that *keep on doing* ranks second (19 tokens) in the BNC, whereas *keep doing* ranks tenth (55 tokens). More interestingly, *keep on trying* ranks third (15 tokens) in the BNC, whereas *keep trying* ranks fifth (89 tokens). Finally, it is worth observing that *keep on talking* ranks twelfth (6 tokens) in the BNC, whereas *keep talking* ranks ninth (57 tokens). We thus conclude that *keep going* and *keep on going* are the most preferable ones (430 tokens vs. 24 tokens) for the British.

Now let us turn our attention to the visualization of the collocations of *keep* and *keep on* in the BNC:

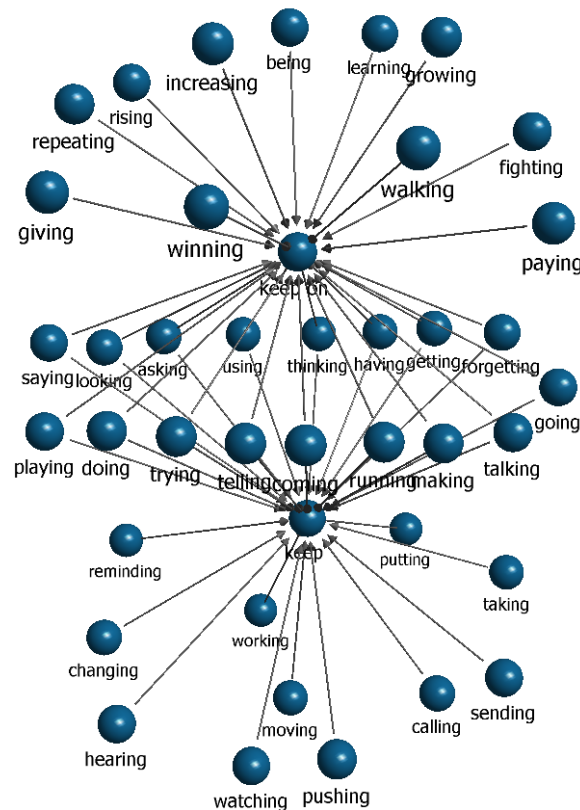


Figure 4 Visualization of the collocations of *keep* and *keep on* in the BNC

As exemplified in Figure 4, eleven gerunds are linked to *keep* and *keep on*, respectively, which indicates that these gerunds have a collocation relationship with *keep* and *keep on*, respectively. Most importantly, seventeen gerunds are linked to both *keep* and *keep on*, which indicates that these gerunds are the collocations of both *keep* and *keep on*. The gerunds linked to two types are *going*, *doing*, *trying*, *saying*, *running*, *playing*, *using*, *asking*, *looking*, *forgetting*, *talking*, *thinking*, *getting*, *coming*, *telling*, *having*, and *making*. From all of this, it is clear that 43.58% of thirty nine gerunds are the collocations of both *keep* and *keep on*. This in turn indicates that *keep* and *keep on* are used synonymously, but they show a low degree of similarity.

CONCLUSION

To sum up, we have provided a comparative analysis of *keep* and *keep on* in the TV Corpus and the British National Corpus. In section 2, we have shown that *keep* was preferable to *keep on* from the 1950s to the 2010s. We have also shown that *keep* and *keep on* had the highest frequency (28,586 tokens vs. 1,303 tokens) in the 2010s. Additionally, we have argued that *keep* was always preferable to *keep on* from the 1950s to the 2010s. In section 3, we have maintained that *keep* is 71.42% the same as *keep on* in the seven genres of the BNC. We have also contended that *keep* is the nearest from *keep on* in the fiction genre, but the former is the furthest from the latter in the spoken genre. In section 4, we have argued that *keep going* and *keep on going* are the most preferable ones (430 tokens vs. 24 tokens) for the British. We have further argued

that 43.58% of thirty nine gerunds are the collocations of both *keep* and *keep on*. This may imply that *keep* and *keep on* are used synonymously, but they show a low degree of similarity.

References

- British National Corpus (BNC). 1, March 2022. Online <https://corpus.byu.edu/bnc>
- Murphy, R. (2016). *Grammar in Use*. Cambridge University Press.
- Murphy, R. (2019). *English Grammar in Use*. Cambridge University Press.
- TV Corpus (TVC). 1, March 2022. Online <https://english-corpora.org/tv/>