ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 15, No. 1, pp. 30-35, January 2025
DOI: https://doi.org/10.17507/tpls.1501.04

A Comparative Analysis of *Think Over* and *Consider* Through BNC, COCA, and ChatGPT*

Namkil Kang
Far East University, South Korea

***Abstract*—This article aims to provide an in-depth comparative analysis of *think over* and *consider* through the British National Corpus (BNC), the Corpus of Contemporary American English (COCA) and ChatGPT. It is important to note that *consider* and *think over* exhibit identical patterns only in the magazine genre and the miscellaneous genre of the BNC, whereas they share the same pattern only in the newspaper genre of the COCA. This can be taken as confirming evidence that in the BNC, *think over* and *consider* are 28.57% the same, whereas in the COCA, they are 14.28% the same. Simply put, *think over* and *consider* exhibit a low similarity in the BNC and the COCA. A further point to note is that *consider* is most similar to *think over* in the newspaper genre of the BNC, whereas the former is the closest to the latter in the TV/movie genre of the COCA. This, in turn, implies that in the newspaper genre of the BNC and the TV/movie genre of the COCA, *think over* and *consider* exhibit the highest degree of similarity. It is also worth noting that the standard deviation of *think over* and *consider* clearly shows American speakers' preferences. Most importantly, 18 of 30 collocations of *think over* and *consider* are the same, which suggests that *consider* and *think over* share 60% of their collocations.**

***Index Terms*—BNC, COCA, standard deviation, Euclidean distance**

I. Introduction

The main purpose of this article is to provide an in-depth comparative analysis of *think over* and *consider* in the BNC (Online https://corpus.byu.edu/bnc), the COCA (Online https://corpus.byu.edu/coca) and ChatGPT. As is well-known, *think over* and *consider* are often used interchangeably. The so-called corpora (British National Corpus, Corpus of Contemporary American English, Hansard Corpus; Kang, 2022a, 2022b) and the relevant literature (Murphy, 2016, 2019) have provided us with a lot of linguistic insights. First, we aim to explore the difference between *consider* and *think over* through a genre analysis in the BNC (Online https://corpus.byu.edu/bnc) and the COCA (Online https://corpus.byu.edu/coca). By investigating the ranking of these two expressions in the BNC (Online https://corpus.byu.edu/bnc) and the COCA (Online https://corpus.byu.edu/coca), we can observe how similar they are. Second, we aim to measure the difference between *think over* and *consider* in each genre of the BNC (Online https://corpus.byu.edu/bnc) and the COCA (Online https://corpus.byu.edu/coca) in terms of Euclidean distance. By doing so, we can examine how similar two expressions are in each genre. Third, we delve into the difference in the use of *think over* and *consider* in terms of the mean and standard deviation. This will illustrate the use and distribution of the so-called two expressions in eight genres of the COCA (Online https://corpus.byu.edu/coca). Fourth, by using ChatGPT, we attempt to investigate the subtle difference among collocations of *think over* and *consider*. By analyzing their 30 most common collocations, we can observe how similar the collocations of two expressions are.

II. The Difference Between Think Over and Consider in the British National Corpus

The main goal of this section is to examine the difference between *think over* and *consider* in the British National Corpus (Online https://corpus.byu.edu/bnc). By analyzing the ranking of seven genres in the BNC (Online https://corpus.byu.edu/bnc), we will explore the similarity between *think over* and *consider*. Consider the following table:

Table 1
Frequency of Think Over and Consider in the BNC

Genre	ALL	Spoken	Fiction	Magazine	Newspaper	Non-academic	Academic	Miscellaneous
Consider	1,1456	672	709	547	806	1,877	3,609	3,236
Think over	38	6	18	0	1	4	1	8

It is worth noting that *consider* is preferred over *think over* by the British. More specifically, *consider* in the BNC (Online https://corpus.byu.edu/bnc) is used more widely than *think over*. This suggests that the British prefer to use *consider*. It is also worth mentioning that in the academic genre, *consider* ranks first, whereas in the fiction genre, *think*

* This work was supported by the 2023 Far East University Research Grant (FEU2023S01).

over ranks first. Thus, it can be inferred that there is no similarity between *think over* and *consider* at rank one. It must be noted, however, that in the academic genre, British teachers prefer using *consider* to *think over*. Simply put, this may indicate that in the field of British education, teachers favor *consider*. On the other hand, *think over* in the fiction genre has 18 tokens (the highest frequency), but *consider* has 709 tokens, indicating that the latter is preferred over the former by British writers.

It is interesting to note that in the miscellaneous genre, *consider* and *think over* have 3,236 and 8 tokens, respectively. Notably, these two expressions rank second in the same genre, revealing a high similarity. It should be pointed out that in the miscellaneous genre, *consider* is used more widely than *think over*. In other words, *consider* is favored over *think over* in the miscellaneous genre.

It is worth emphasizing that in the non-academic genre, *consider* ranks third, while in the spoken genre, *think over* ranks third. This, in turn, reveals that these two expressions do not exhibit similarity at rank three. As shown in Table 1, even in the non-academic and spoken genres, *consider* is favored over *think over*, suggesting that the British prefer *consider* over *think over* in the non-academic contexts and daily conversation.

It is also interesting to note that in the newspaper genre, *consider* ranks fourth (806 tokens), whereas in the non-academic genre, *think over* ranks fourth (4 tokens). Again, these two expressions do not share the same characteristics at rank four, indicating that they are not alike in this rank (no similarity). It should be noted that British generalists prefer *consider* over *think over* in their newspapers, as evidenced by the higher frequency of *consider* in the newspaper genre. What is interesting is that in the fiction genre, *consider* ranks fifth. It is particularly noteworthy that in both the newspaper genre and the academic genre, *think over* ranks fifth. This suggests that *consider* and *think over* do not share the same pattern, exhibiting no similarity. It would be misleading not to mention that in the fiction genre, *consider* is favored over *think over*. This implies that British writers are inclined to use *consider* rather than *think over* in their novels. Additionally, it is worth pointing out that *consider* is preferred over *think over* by British generalists and British teachers. This can be inferred from the frequency of *think over* and *consider*.

It is very interesting that in the spoken genre, *consider* has 672 tokens and ranks sixth. More interestingly, *think over* has no rank-six position. It seems appropriate to conclude that *consider and think over* show no similarity in the spoken genre. Particularly noteworthy is the fact that *consider* is used more frequently than *think over* in the spoken genre, suggesting that the British prefer *consider* in daily conversation.

It must be stressed that in the magazine genre, *consider* has 547 tokens and ranks seventh. Interestingly, *think over* also ranks seventh but has no frequency (0 token). However, they are similar in that they share the same ranking in this genre. This indicates a high similarity between the two expressions in the magazine genre. It is important to mention that in magazine genre, the use of *consider* is far more frequent than that of *think over*. This suggests that British generalists prefer *consider* over *think over*. To summarize, while *consider* and *think* over exhibit the same pattern only in the magazine genre and miscellaneous genres, they do not share the same pattern in the other five genres. This suggests that *think over* and *consider* are 28.57% similar in terms of their ranking patterns.

III. The Difference Between Think Over and Consider in the COCA

In the following, we explore the difference between *think over* and *consider* through a genre analysis of the COCA (Online https://corpus.byu.edu/coca). More specifically, by analyzing the ranking of eight genres, we can examine how closely *think over* and *consider* are related. Consider Table 2:

Table 2
Frequency of Think Over and Consider in the COCA

Genre	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Consider	110,624	21,050	19,478	6,662	8,892	5,533	16,059	12,535	20,415
Think over	333	36	38	26	134	38	20	30	11

It is worth noting that in America, *consider* is preferred over *think over*, as illustrated in Table 1. This may be because the use of *consider* in the COCA (Online https://corpus.byu.edu/coca) is far higher than that of *think over*. More specifically, while *consider* appears 110,624 times in the COCA (Online https://corpus.byu.edu/coca), *think over* appears only333 times. This suggests that Americans prefer using *consider* to *think over*.

It is worth examining the blog genre and the spoken genre. Interestingly, while *consider* ranks first in the blog genre, *think over* ranks first in the spoken genre. As shown in Table 2, these two expressions permit a different pattern at rank one, revealing that *consider* and *think over* have a low similarity in rank one. It should be noted, however, that *consider* is preferred over *think over* by Americans in both the spoken genre and in the blog genre. This indicates that Americans prefer *consider* in both genres.

It is worth looking at these two expressions at rank two. To begin with, it is worth considering *consider* in the academic genre. As expected, *consider* obtained 20,415 tokens in the academic genre (the second highest). In contrast, *think over* obtained 38 tokens in both the fiction and web genres (the second highest in those genres). This suggests that the two expressions permit different patterns in these genres, implying a low similarity between them. However, it must be pointed out that Americans prefer using *consider* in all three genres. More specifically, *consider* is used more

frequently than *think over* in the academic, fiction and web genres, further suggesting that Americans prefer *consider* over *think over*.

It is worthwhile examining *consider* in the web genre. Interestingly, while *consider* obtained 19,478 tokens in the web genre (the third highest), *think over* does not have a rank-three position. This suggests that the two expressions have a low similarity in this genre.

It should be pointed out that while *consider* ranks fourth in the magazine genre, *think over* ranks fourth in the blog genre. Again, the two expressions do not share the same ranking at rank four, suggesting no similarity. However, it must be noted that in both the magazine and blog genres, *consider* is used more frequently. This suggests that American generalists and bloggers prefer *consider* in their magazines and blogs.

It is worth examining *think over* and *consider* in the newspaper genre. In the newspaper genre, *think over* obtained 30 tokens and *consider* obtained 12,535 tokens. Both expressions rank fifth in this genre, revealing that they share the same ranking. This suggests that the two expressions exhibit a high similarity in the newspaper genre. It must also be pointed out that *consider* is preferred over *think over* by American generalists. This is likely due to the significantly higher frequency of *consider* in the newspaper genre.

It must be stressed that while *consider* ranks sixth in the spoken genre, *think over* ranks sixth in the TV/movie genre. Once again, the two expressions do not share the same ranking in these genres, revealing no similarity. In conclusion, these two expressions behave like distinct expressions in these genres. However, it must be noted that Americans prefer to use *consider* in daily conversation. The same can be said for the TV/movie genre, where *consider* obtained 6,662 tokens, while *think over* obtained only 26 tokens. What this suggests is that American celebrities prefer using *consider*.

It is interesting to note that while *consider* obtained 6,662 tokens in the TV/movie genre (the seven highest), *think over* obtained 20 tokens in the magazine genre (the seven highest). Similarly, these two expressions do not reveal the same pattern, exhibiting no similarity. More importantly, in both the TV/movie and magazine genres, *consider* is used more frequently. This, in turn, indicates that American generalists and celebrities prefer *consider* over *think over* in these genres.

It must be emphasized that while *consider* obtained 5,533 tokens in the fiction genre, *think over* obtained 11 tokens in the academic genre. Likewise, these two expressions do not share same ranking in these genres, implying no similarity. However. it should be stressed that in both of these genres, *consider* is used more widely than *think over*. This, in turn, suggests that American generalists and teachers prefer using *consider* over *think over*. To sum up, *think over* and *consider* exhibit the same pattern only in the newspaper genre, while they do not share the same pattern in the other seven genres. This suggests that *think over* and *consider* are 12.5% similar based on this genre analysis. Put differently, *think over* and *consider* are likely to have a low similarity overall.

IV. The Euclidean Distance

In what follows, we attempt to measure the difference between *think over* and *consider* in the BNC (Online https://corpus.byu.edu/bnc) and COCA (Online https://corpus.byu.edu/coca) by employing Euclidean distance. By doing so, we can assess the degree of similarity between *think over* and *consider* in both corpora. The Euclidean distance, in this context, refer to the degree of similarity between the two expressions. We define it as follows:

(1) Euclidean Distance

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

Now attention is paid to British National Corpus (Online https://corpus.byu.edu/bnc):

Table 3
Euclidean Distance Between Think Over and Consider in the BNC

Genre	Spoken	Fiction	Magazine	Newspaper	Non-academic	Academic	Miscellaneous
Consider (%)	5.86	6.18	4.77	7.03	16.38	31.5	28.24
Think over (%)	15.78	47.36	0	2.63	10.52	2.63	21.05
Euclidean distance	9.92	41.18	4.77	4.4	5.86	28.87	7.19

Most interestingly, *consider* is the furthest from *think over* in the fiction genre, as shown in Table 3. More specifically, the distance between *think over* and *consider* in the fiction genre is 41.18, which is the highest. What this implies is that they exhibit the lowest similarity. It is worth considering the academic genre. In this genre, the distance between *think over* and *consider* is 28.87, which is the second highest, indicating a low similarity. Notably, *consider* is the closest to *think over* in the newspaper genre. The distance between them is 4.4, which is the lowest, suggesting that they exhibit the highest similarity in this genre. It is interesting to observe that the newspaper genre is followed by the magazine

genre. In the magazine genre, the distance between *consider* and *think over* is 4.77, which is the second lowest. This, in turn, implies that these two expressions reveal a high similarity. To summarize, while *consider* is the furthest from *think over* in the fiction genre, it is the closest to *think over* in the newspaper genre.

Now, attention shifts to the Corpus of Contemporary American English (Online https://corpus.byu.edu/coca):

TABLE 4
EUCLIDEAN DISTANCE BETWEEN THINK OVER AND CONSIDER IN THE COCA

Genre	Blog	Web	TV/Movie	Spok	Fic	Mag	News	Acad
Consider (%)	19.02	17.6	6.02	8.03	5	14.51	11.33	18.45
Think over (%)	10.81	11.41	7.8	40.24	11.41	6	9	3.3
Euclidean distance	8.21	15.42	1.78	32.21	6.41	8.51	2.33	15.15

Most importantly, in the spoken genre, *consider* is the furthest from *think over*. Specifically, in the spoken genre, the Euclidean distance between *think over* and *consider* is 32.21, which is the highest. This, in turn, implies that in this genre, *think over* and *consider* exhibit the lowest similarity. It is also worth noting that the spoken genre is followed by the web genre. In the web genre, the distance between *think over* and *consider* is 15.42, which is the second highest. This suggests that in the web genre, *think over* and *consider* have a low similarity. Next, let's consider the TV/movie genre. Interestingly, *think over* is the closest to *consider* in this genre, with a distance of 1.78, which is the lowest. What this suggests is that in the TV/movie genre, *think over* and *consider* have the highest similarity. It is also noteworthy that the TV/movie genre is followed by the newspaper genre. In the newspaper genre, the distance between *think over* and *consider* is 2.33, which is the second lowest. This indicates that in the newspaper genre, *think over* and *consider* exhibit a high similarity. In conclusion, while in the spoken genre, *consider* is the furthest from *think over*, in the TV/movie genre, the former is the closest to the latter.

V. STANDARD DEVIATION

In what follows, we aim to examine the use of *think over* and *consider* in terms of standard deviation and mean. Additionally, we argue that, in support of American preferences, while in only some genres, *think over* and *consider* are used more frequently, they are used less frequently in others. The standard deviation allows us to capture the maximum and minimum frequencies of *think over* and *consider*. The mean frequency lies between these values, enabling us to compare the use of *think over* and that of *consider*.

Now, attention is drawn to the following graph:

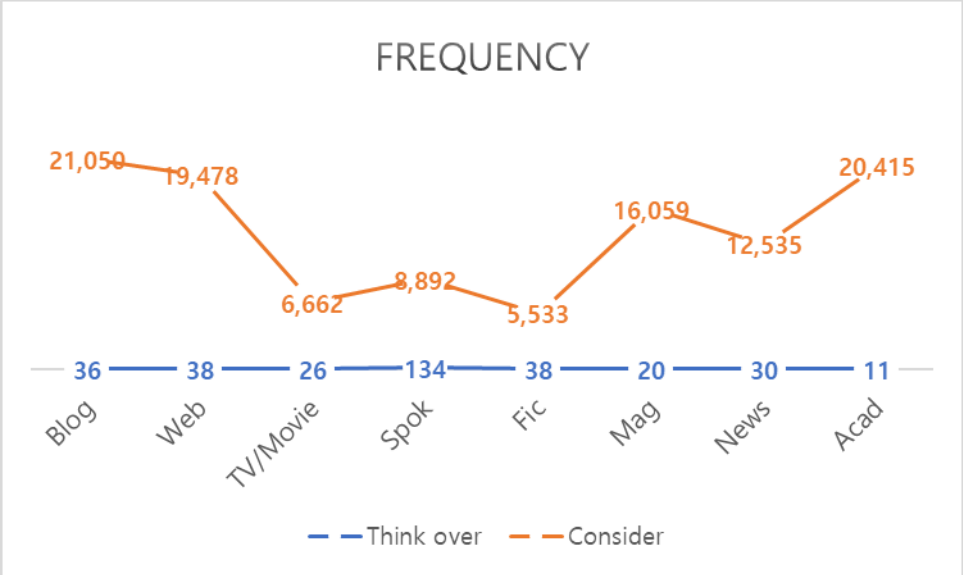

Figure 1. Frequency of Think Over and Consider in the COCA

Now take a look at the following table:

TABLE 5
MEAN, VARIANCE, AND STANDARD DEVIATION IN THE COCA

Type	Mean	Variance	Standard Deviation
Think over	41.65	9,553.8853125	97.744
Consider	13,828	241,128,977.125	15,528.328

It is interesting to observe that the mean frequency of *think over* is 41.65, as illustrated in Table 5. Perhaps it is worthwhile noting that only in the spoken genre, is the frequency of *think over* higher than 41.65, as indicated in Figure 1. This, in turn, indicates that *think over* is primarily used in daily conversation. On the other hand, the mean frequency of *consider* is 13,828, as shown in Table 3. It is worth noticing that in the blog, web, magazine, and academic genres, *consider* is used more widely, as exemplified in Figure 1. More specifically, in these genres, the frequency of *consider* exceeds 13,828. From this, it seems evident that *think over* and *consider* is used differently. Most importantly, the standard deviation of *think over* is 97.744, which means that the frequency of *think over* is roughly between 41.65 - 91.744 and 41.65 + 97.744. On the other hand, the standard deviation of *consider* is 15,528, meaning that its frequency is roughly between 13,828 - 15528 and 13,828 + 15,528. This suggests that American speakers prefer using *consider* over *think over*. In other words, the standard deviations of *think over* and *consider* clearly reflect the preference of American speakers for *consider*.

VI. Collocations of Think Over and Consider

In this section, we had ChatGPT generate 30 collocations for *consider* and *think over*. We aim to examine these 30 collocations and compare those of *consider* and those of *think over*. This method allows us to assess how similar their collocations are:

Table 6
Collocations of Think Over and Consider

Number	Collocations of Consider	Collocations of Think Over
1	options	options
2	alternatives	choices
3	outcomes	strategies
4	implications	proposals
5	feedback	decisions
6	perspectives	possibilities
7	risks	challenges
8	benefits	risks
9	challenges	priorities
10	resources	solutions
11	priorities	implications
12	strategies	feedback
13	constraints	commitments
14	timelines	goals
15	impacts	factors
16	methods	outcomes
17	trends	ideas
18	needs	timelines
19	values	resources
20	solutions	assumptions
21	data	perspectives
22	assumptions	scenarios
23	roles	values
24	objectives	responsibilities
25	dynamics	dynamics
26	practices	methods
27	regulations	regulations
28	stakeholders	trends
29	environments	environments
30	technologies	commitments

It is worth taking a look at the collocation *options*. Interestingly, the word *options* is the first collocation that ChatGPT provided. This suggests that the word *options* is the most widely used collocation for *consider* and *think over*, since ChatGPT's data are derived from web sources. Thus, it seems appropriate to conclude that the word *options* is the most commonly used collocation for both expressions. It is worth mentioning that while the word *alternatives* is the second collocation for *consider*, the word *choices* is the second for *think over*. This suggests that these collocations are the second most widely used in web sources. What is interesting is that while the word *outcomes* is the third collocation of *consider*, the word *strategies* is the third for *think over*. These two collocations are assumed to be the third most widely used in web sources. It is also worth noting the collocations *implications* and *proposals.* As shown in Table 6, these are the fourth most widely used collocations for *consider* and *think over*. Additionally, while the word *feedback* is the fifth collocation for *consider*, the word *decisions* is the fifth for *think over*. This suggests that these two collocations are the fifth most widely used in web sources. Most importantly, 18 of 30 collocations overlap between *consider* and *think over*. Specifically, the following words are collocations for both expressions: *options*, *outcomes*, *implications*, *feedback*, *perspectives*, *risks*, *challenges*, *resources*, *priorities*, *strategies*, *timelines*, *methods*, *trends*, *values*, *solutions*, *assumptions*, *regulations*, and *environments*. It therefore seems plausible to conclude that the collocations of *consider*

and *think over* are 60% the same among the top 30.

VII. Conclusion

To sum up, we have provided an in-depth comparative analysis of *think over* and *consider* using the BNC (Online https://corpus.byu.edu/bnc), the COCA (Online https://corpus.byu.edu/coca) and ChatGPT. We have argued in this paper that in the BNC, *think over* and *consider* are 28.57% similar, whereas in the COCA (Online https://corpus.byu.edu/coca), they are 14.28% similar. Simply put, *think over* and *consider* exhibit a low degree of similarity. We have further argued that *consider* is the closest to *think over* in the newspaper genre of the BNC, whereas the former is the closest to the latter in the TV/movie genre of the COCA (Online https://corpus.byu.edu/coca).This, in turn, implies that in the newspaper genre of the BNC (Online https://corpus.byu.edu/bnc) and the TV/movie genre of the COCA (Online https://corpus.byu.edu/coca), *think over* and *consider* have the highest similarity. Additionally, we have maintained that American speakers tend to prefer *consider*. The standard deviation of *think over* and *consider* clearly reflects this preference. We have shown that the following words such as *options*, *outcomes*, *implications*, *feedback*, *perspectives*, *risks*, *challenges*, *resources*, *priorities*, *strategies*, *timelines*, *methods*, *trends*, *values*, *solutions*, *assumptions*, *regulations*, and *environments* are the collocations of both *consider* and *think over*. Moreover, we have argued that 18 of the top 30 collocations overlap for both expressions. From this, it seems evident that 60% of the top 30 collocations for *consider* and *think* are the same.

References

[1] *British National Corpus* (BNC). 6, October, 2024, Online https://corpus.byu.edu/bnc.
[2] *Corpus of Contemporary American English* (COCA). 6, October, 2024, Online https://corpus.byu.edu/coca.
[3] *Hansard Corpus* (HC). 6, October, 2024, Online https://english-corpora.org /hansard/.
[4] Kang, N. (2022a). A Comparative Analysis of Impressed by and impressed with in Two Corpora. *Theory and Practice in Language Studies*, *11*(11), 819-827.
[5] Kang, N. (2022b). On Speak to and talk to: A Corpora-based Analysis. *Theory and Practice in Language Studies*, *12*(7), 1262-1270.
[6] Murphy, R. (2016). *Grammar in Use*. Cambridge University Press. Singapore.
[7] 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 140 articles and three books. He is a recent ex-chairman of the Joongwon Linguistic Society of South Korea and the Jungang English Language and Literature Association of South Korea.