

On Speak With and Talk With: A Corpora-Based Analysis

Namkil Kang

Far East University, South Korea



ABSTRACT: The main purpose of this article is to provide a comparative analysis of *Speak with* and *Talk with* in the Corpus of Historical American English (COHA) and the Corpus of Contemporary American English (COCA). With respect to the COHA, it is interesting to note that *Talk with* was always preferred over *Speak with* by Americans except 1820, 1830, and 1840. A further point to note is that *Speak with* had the highest frequency (271 tokens) in 2010, whereas *Talk with* had the highest frequency (633 tokens) in 1950. With respect to the COCA, it is worth pointing out that *Speak with* and *Talk with* show the same property in four genres, whereas they show a different property in the other four genres. Simply put, in the eight genres of the COCA, *Speak with* is 50% the same as *Talk with*. Talking about the distance between *Speak with* and *Talk with*, the former is the nearest to the latter in the web genre. Quite interestingly, the COCA shows that *Talk with people* and *Speak with authority* are the most widely used (126 tokens vs. 56 tokens) in America. Finally, this paper argues that 26.08% of forty six nouns are the collocations of *Speak with* and *Talk with*, thus revealing a low degree of similarity.

KEYWORDS: COHA, COCA, corpus, token, type, speak with, talk with.

1. INTRODUCTION

As Murphy (2016, 2019) points out, *Speak with* and *Talk with* are used interchangeably. The main goal of this article is to provide a comparative analysis of *Speak with* and *Talk with* in two corpora (the Corpus of Historical American English and the Corpus of Contemporary American English). First, we consider the diachronic aspects of *Speak with* and *Talk with* (1820-2010) and compare the two types. To be more specific, we consider whether or not there is a continuing upward trend in the frequency of *Speak with* and *Talk with*. Additionally, we provide time period when the two types had the highest frequency and the lowest frequency. Second, by examining the genre frequency of the COCA, we explore the similarity between *Speak with* and *Talk with*. Simply put, we examine how much similar they are. Third, we attempt to measure the distance between *Speak with* and *Talk with* in the eight genre of the COCA, which tells us whether or not there is a similarity between them in each genre of the COCA. Fourth, we consider the similarity between *Speak with* and *Talk with* by examining their collocations in the top 29. Finally, we use the software package NetMiner by which we can see how much close *Speak with* and *Talk with* are. The organization of this article is as follows. In section 2, we argue that *Talk with* was always preferred over *Speak with* by Americans except 1820, 1830, and 1840. We further argue that *Speak with* had the highest frequency (271 tokens) in 2010, whereas *Talk with* had the highest frequency (633 tokens) in 1950. In section 3, we show that *Speak with* and *Talk with* show the same property in four genres, whereas they show a different property in the other four genres. This in turn implies that in the eight genres of the COCA, *Speak with* is 50% the same as *Talk with*. Also, we show that *Speak with* is the nearest to *Talk with* in the web genre. In section 4, we maintain that *Talk with people* and *Speak with authority* are the most widely used ones (126 tokens vs. 56 tokens) in America. Finally, we argue that 26.08% of forty six nouns are the collocations of *Speak with* and *Talk with*.

2. SPEAK WITH AND TALK WITH IN THE CORPUS OF HISTORICAL AMERICAN ENGLISH

In this section, we aim to observe the diachronic aspects of *Speak with* and *Talk with* from 1820 to 2010. Table 1 shows the frequency of *Speak with* and *Talk with* from 1820 to 2010:

Table 1. Frequency of speak with and talk with

Type	Speak with	Talk with
1820	79	28
1830	162	78
1840	192	131
1850	153	163

On Speak With and Talk With: A Corpora-Based Analysis

1860	139	228
1870	155	306
1880	191	363
1890	175	437
1900	180	460
1910	142	492
1920	131	549
1930	114	588
1940	118	579
1950	148	633
1960	179	574
1970	210	448
1980	213	474
1990	263	346
2000	246	359
2010	271	278
ALL	3,461	7,514

Quite interestingly, the overall frequency of *talk with* (7,514 tokens) is two times higher than that of *speak with* (3,461 tokens). This in turn suggests that *talk with* (7,514 tokens) was preferred over *speak with* (3,461 tokens) by Americans. This amounts saying that Americans preferred using *talk with* rather than using *speak with*. Figure 1 shows the diachronic aspects of *speak with* and *talk with* from 1820 to 2010:

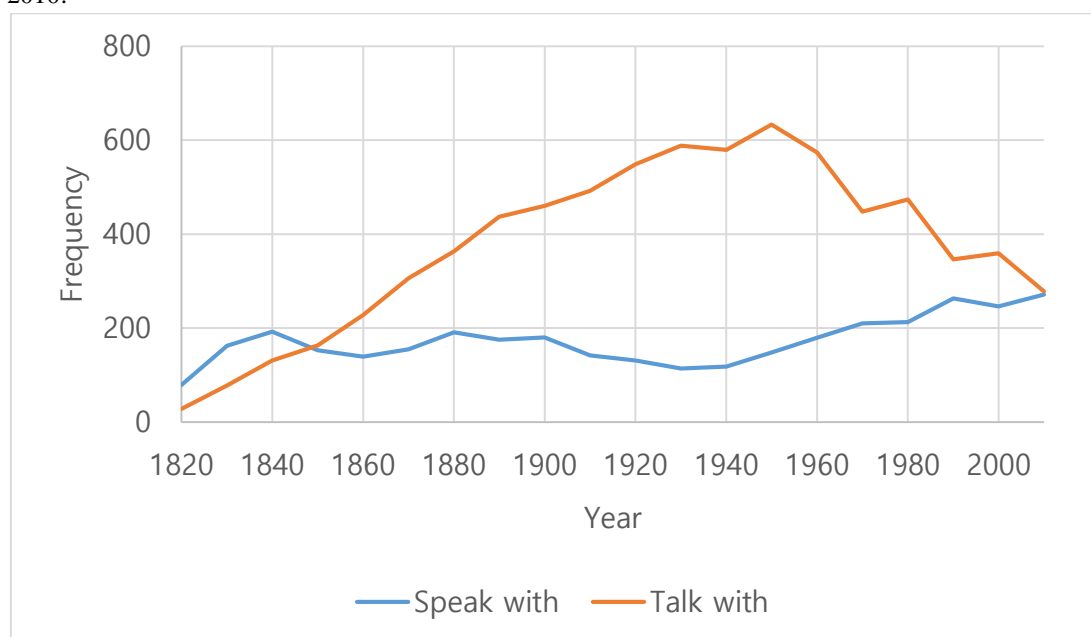


Figure 1. Frequency of speak with and talk with

It is interesting to point out that there was a sharp increase (a rise of 113 tokens) in the figure of *speak with* from 1820 to 1840. It should be noted, however, that there was a minimal drop (a fall of 53 tokens) in the figure of *speak with* from 1840 to 1860. It is worth noting that the frequency of *speak with* increased to 52 tokens from 1860 to 1880. More interestingly, the frequency of *speak with* steadily decreased to 77 tokens from 1880 to 1930 except 1900. It is noteworthy that again, the frequency of *speak with* gradually increased to 65 tokens from 1930 to 1960. It is important to note that the frequency of *speak with* was over 200 tokens from 1970 to 2010. We take this as suggesting that *speak with* is much used after 1970. Additionally, it should be pointed out that there was a slight increase (a rise of 61 tokens) in the figure of *speak with* from 1970 to 2010 except 2000. Most importantly, *speak with* had the highest frequency (271 tokens) in 2010, whereas it had the lowest frequency (79 tokens) in 1820. This in turn indicates that *speak with* was the most preferred one (271 tokens) in 2010, whereas it was the least preferred one (79 tokens) in 1820. It is significant to note that there were slight fluctuations in the frequency of *speak with* from 1820 to 2010, thereby suggesting that there was no continuing upward trend in the figure of *speak with*.

It is interesting to note that the frequency of *talk with* gradually increased to 103 tokens from 1820 to 1840. After this period, the

On Speak With and Talk With: A Corpora-Based Analysis

frequency of *talk with* steadily increased to 457 tokens from 1840 to 1930. It is significant to note that there was an upward trend in the figure of *talk with* from 1820 to 1930. It must be noted, however, that there was a sudden decline (a drop of 9 tokens) in the figure of *talk with* from 1930 to 1940. More importantly, there were slight fluctuations in the figure of *talk with* from 1940 to 2010. It is worthwhile pointing out that there was a downward trend in the figure of *talk with* from 1950 to 2010. More interestingly, *talk with* had the highest frequency (633 tokens) in 1950, whereas it had the lowest frequency (28 tokens) in 1820. This in turn suggests that *talk with* was the most preferable one (633 tokens) for Americans in 1950. It is important to note that *talk with* was always preferred over *speak with* by Americans except 1820, 1830, and 1840. As can be seen from Figure 1, *speak with* was preferable to *talk with* from 1820 to 1840.

3. SPEAK WITH AND TALK WITH IN THE COCA

In this section, we aim to observe the similarity between *speak with* and *talk with* by examining their frequency in the eight genre of the COCA. Table 2 shows the genre frequency of *speak with* and *talk with* in the COCA:

Table 2. Genre Frequency of speak with and talk with

GENRE	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Speak with	7,346	722	814	2,170	1,307	1,095	430	500	308
Talk With	10,996	950	1,170	1,583	3,465	1,448	898	896	586

It is significant to note that the overall frequency of *talk with* (10,966 tokens) is far higher than that of *speak with* (7,346 tokens). This in turn implies that *talk with* (10,966 tokens) is preferred over *speak with* (7,346 tokens) by Americans.

It is worth pointing out that *speak with* ranks first (2,170 tokens) in the TV/movie genre, whereas *talk with* ranks first (3,465 tokens) in the spoken genre, thus showing no similarity in rank-one. It is interesting to note that the frequency of *speak with* (2,170 tokens) in the TV/movie genre is even higher than that of *talk with* (1,583 tokens). We take this as implying that Americans celebrities prefer using *speak with* (2,170 tokens) rather than using *talk with* (1,583 tokens). It should be pointed out, on the other hand, that the frequency of *talk with* (3,465 tokens) in the spoken genre is two times higher than that of *speak with* (1,307 tokens). This in turn indicates that Americans prefer to use *talk with* rather than use *speak with* in daily conversation.

It is worth observing that *speak with* ranks second (1,307 tokens) in the spoken genre, whereas *talk with* ranks second (1,583 tokens) in the TV/movie genre, thus implying that there is no similarity between *speak with* and *talk to* in rank-two. Note that *speak with* and *talk with* are used interchangeably, but they show no similarity in rank-two.

It is worthwhile noting that *speak with* and *talk with* rank third (1,095 tokens vs. 1,448 tokens) in the fiction genre, hence showing a close similarity in rank-three. It should be noted, however, that the frequency of *talk with* (1,448 tokens) is far higher than that of *speak with* (1,095 tokens) in the fiction genre. We take this fact as indicating that American writers prefer to use *talk with* rather than use *speak with*.

Noteworthy is that *speak with* and *talk with* rank fourth (814 tokens vs. 1,170 tokens) in the web genre, thereby showing a deep similarity in rank-four. It must be pointed out, however, that *talk with* (1,170 tokens) is preferable to *speak with* (814 tokens) in the web genre. We infer this from the fact that the frequency of *talk with* (1,170 tokens) is much higher than that of *speak with* (814 tokens) in the web genre.

It is worthwhile pointing out that *speak with* and *talk with* rank fifth (722 tokens vs. 950 tokens) in the blog genre, hence showing a high degree of similarity in rank-five. It should be pointed out, however, that the frequency of *talk with* (950 tokens) is even higher than that of *speak with* (722 tokens). From this, it is clear that Americans prefer using *talk with* to using *speak with* in the blog genre.

It is worth mentioning that *speak with* ranks sixth (500 tokens) in the newspaper genre, whereas *talk with* ranks sixth (898 tokens) in the magazine genre, thereby indicating no similarity in rank-six. With respect to the newspaper genre, it is interesting to note that the frequency of *talk with* (896 tokens) is far higher than that of *speak with* (500 tokens). From this, it is evident that *talk with* (896 tokens) is favored over *speak with* (500 tokens) in the newspaper genre. When it comes to the magazine genre, *talk with* (898 tokens) is preferable to *speak with* (430 tokens). This stems from the fact that the frequency of *talk with* (898 tokens) is much higher than that of *speak with* (430 tokens).

It is interesting to point out that *speak with* ranks seventh (430 tokens) in the magazine genre, whereas *talk with* ranks seventh (896 tokens) in the newspaper genre, hence implying that there is no similarity between *speak with* and *talk with*.

Finally, it is worthwhile noting that *speak with* and *talk with* rank eighth (308 tokens vs. 586 tokens) in the academic genre, thus showing a high degree of similarity. It should be noted, however, that there is a difference of 278 tokens between the frequency of *talk with* (586 tokens) and that of *speak with* (308 tokens) in the academic genre. This in turn indicates that Americans prefer to use *talk with* rather than use *speak with* in the academic field. To sum up, *speak with* and *talk with* show the same property in four genres, while they show a different property in the other four genres. From all of this, it is clear that in the eight genres of the COCA, *speak*

On Speak With and Talk With: A Corpora-Based Analysis

with is 50% the same as *talk with*.

Figure 2 and Figure 3 show the use of *speak with* and *talk with* in eight genres:

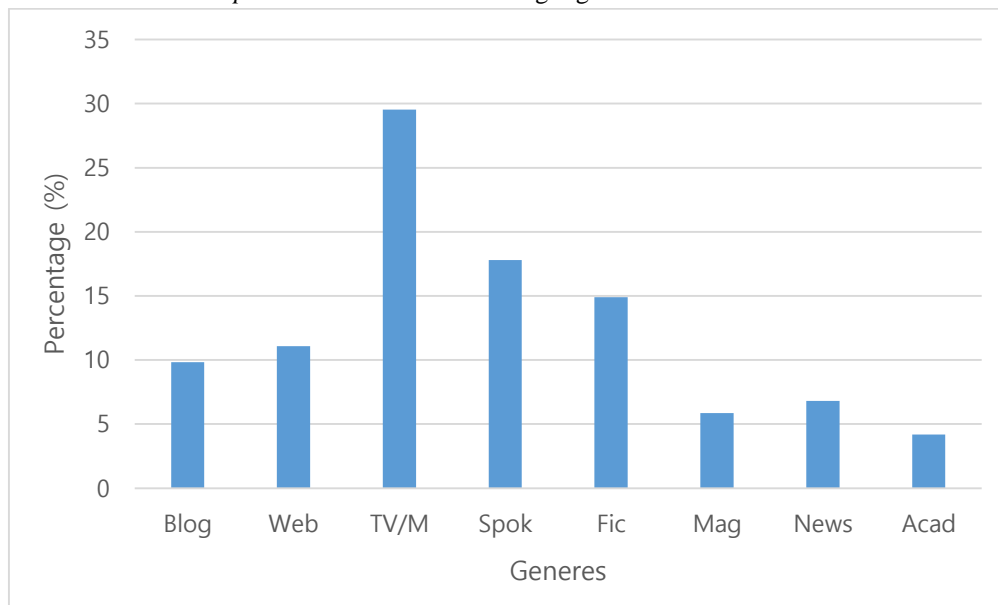


Figure 2. Use of *speak with* in eight genres

As can be seen from Figure 2, the TV/movie genre is the most influenced by *speak with*, followed by the spoken genre, the fiction genre, the web genre, the blog genre, the newspaper genre, the magazine genre, and the academic genre, in that order.

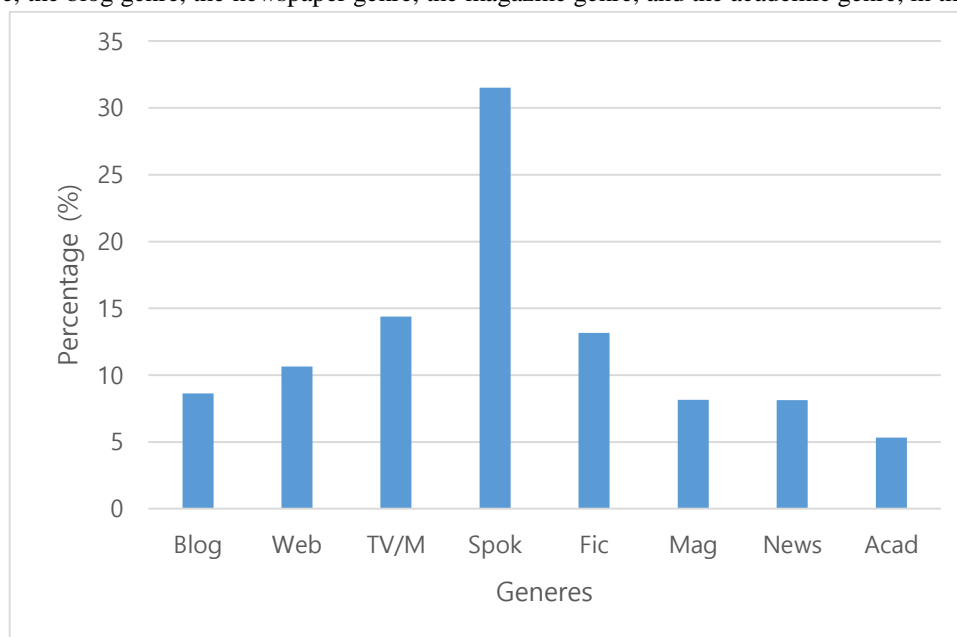


Figure 3. Use of *talk with* in eight genres

As exemplified in Figure 3, the spoken genre is the most influenced by *talk with*, followed by the TV/movie genre, the fiction genre, the web genre, the blog genre, the magazine genre, the newspaper genre, and the academic genre, in descending order.

Now we aim to measure the distance between *speak with* and *talk with* in eight genres. Note that the Euclidean distance we are adopting here provides the distance between *speak with* and *talk with* in each genre of the COCA:

(1) Euclidean distance:

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

The following table shows the Euclidean distance between *speak with* and *talk with* in eight genres:

On Speak With and Talk With: A Corpora-Based Analysis

Table 3. Euclidean distance between speak with and talk with

GENRE	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Percentage of speak with	9.82	11.08	29.53	17.79	14.9	5.85	6.8	4.19
Percentage of talk with	8.63	10.64	14.39	31.51	13.16	8.16	8.14	5.32
Euclidean distance	1.19	0.44	15.14	13.72	1.74	2.31	1.34	1.13

More interestingly, *speak with* is the furthest from *talk with* in the TV/movie genre. To be more specific, the Euclidean distance in that genre is 15.14, which is the highest. It is thus reasonable to assume that *speak with* and *talk with* show a low degree of similarity. Most importantly, *speak with* is the nearest to *talk with* in the web genre. More specifically, the Euclidean distance in that genre is 0.44, which is the lowest. This in turn shows that *speak with* and *talk with* indicate a deep similarity in the web genre. To sum up, the web genre indicates the highest similarity and the academic genre, the blog genre, the newspaper genre, the fiction genre, the magazine genre, the spoken genre, and the TV/movie genre follow. We thus conclude that *speak with* is the nearest to *talk with* in the web genre.

4. THE COLLOCATIONS OF SPEAK WITH AND TALK WITH IN THE COCA

In this section, we aim to compare *speak with* and *talk with* by examining their collocations. Table 4 shows the collocation of *speak with* in the top 29:

Table 4. Collocation of speak with in the COCA

Number	Collocation	Frequency
1	Speak with authority	56
2	Speak with Mr	39
3	Speak with reporters	37
4	Speak with people	35
5	Speak with president	18
6	Speak with Dr	16
7	Speak with confidence	14
8	Speak with students	13
9	Speak with Miss	11
10	Speak with tongues	11
11	Speak with us	11
12	Speak with conviction	10
13	Speak with police	10
14	Speak with clarity	8
15	Speak with agent	8
16	Speak with detective	8
17	Speak with media	8
18	Speak with passion	8
19	Speak with senator	8
20	Speak with members	7
21	Speak with parents	7
22	Speak with Americans	6
23	Speak with pride	6
24	Speak with congressman	5
25	Speak with captain	5
26	Speak with experts	5
27	Speak with father	5
28	Speak with investigators	5
29	Speak with journalists	5

On Speak With and Talk With: A Corpora-Based Analysis

As alluded to in Table 4, *speaking with authority* is the most frequently used one (56 tokens) in America. This in turn suggests that *speaking with authority* is the most preferable one (56 tokens) for Americans. As illustrated in Table 4, *speaking with authority* is the most preferred (56 tokens) by Americans, followed by *speaking with Mr*, *speaking with reporters*, *speaking with people*, *speaking with president*, and *speaking with Dr*, in that order. It is interesting to note that *speaking with reporters* ranks third (37 tokens) in the COCA and *speaking with president* ranks fifth (18 tokens). It is worth mentioning, on the other hand, that *speaking with police* ranks twelfth (10 tokens) in the COCA. Quite interestingly, *speaking with media* ranks fourteenth (8 tokens) in the COCA and *speaking with congressman* ranks twenty fourth (5 tokens).

Table 5 shows the collocation of *talk with* in the top 29:

Table 5. Collocation of talk with in the COCA

Number	Collocation	Frequency
1	Talk with people	126
2	Talk with us	58
3	Talk with students	49
4	Talk with Dr	40
5	Talk with friends	40
6	Talk with reporters	26
7	Talk with children	24
8	Talk with Mr	23
9	Talk with president	21
10	Talk with senator	20
11	Talk with parents	17
12	Talk with patients	16
13	Talk with strangers	16
14	Talk with women	13
15	Talk with Mrs	12
16	Talk with police	12
17	Talk with members	11
18	Talk with teachers	11
19	Talk with kids	10
20	Talk with republican	10
21	Talk with family	9
22	Talk with folks	9
23	Talk with congressman	8
24	Talk with author	7
25	Talk with adults	7
26	Talk with colleagues	7
27	Talk with mom	7
28	Talk with neighbors	7
29	Talk with voters	7

As illustrated in Table 5, *talk with people* is the most widely used one (126 tokens) in America. This in turn indicates that *talk with people* is the most preferable one (126 tokens) among Americans. As indicated in Table 5, *talk with people* is the most preferred one (126 tokens), followed by *talk with us*, *talk with students*, *talk with Dr* (*talk with friends*), *talk with reporters*, and *talk with children*, in descending order. It is significant to note that *talk with people* ranks first (126 tokens) in the COCA, whereas *speaking with people* ranks fourth (35 tokens). It must be pointed out, on the other hand, that *talk with reporters* ranks sixth (26 tokens) in the COCA, whereas *speaking with reporters* ranks third (37 tokens). More interestingly, *talk with president* ranks ninth (21 tokens) in the COCA, whereas *speaking with president* ranks fifth (18 tokens). Finally, it is worth observing that *talk with congressman* ranks twenty third (8 tokens) in the COCA, whereas *speaking with congressman* ranks twenty fourth (5 tokens). We thus conclude that *talk with people* and *speaking with authority* are the most widely used ones (126 tokens vs. 56 tokens) in America.

Now attention is paid to the visualization of the collocations of *speaking with* and *talk with*:

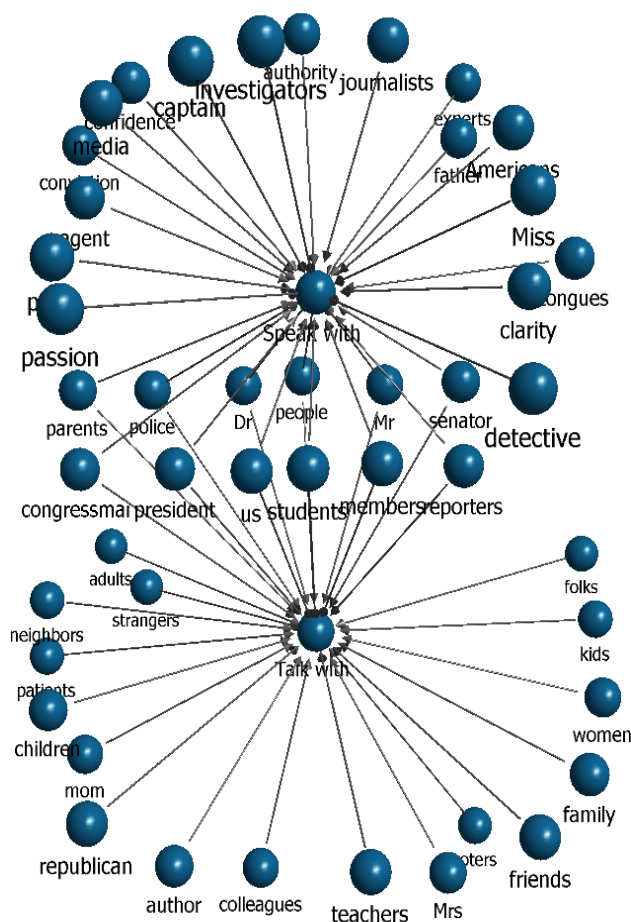


Figure 4. Visualization of the collocations of *Speak with* and *Talk with*

As exemplified in Figure 4, twenty nine nouns are linked to *Speak with* and *Talk with*, respectively. Most importantly, twelve nouns are linked to both *Speak with* and *Talk with*. The twelve nouns are the collocations of both *Speak with* and *Talk with* and they include *people, students, Dr, reporters, us, Mr, president, senator, parents, police, members, and congressman*. This in turn suggests that 26.08% of forty six nouns are the collocations of *Speak with* and *Talk with*, thereby implying that they show a low degree of similarity.

5. CONCLUSION

To sum up, we have provided a comparative analysis of *Speak with* and *Talk with* in two corpora. In section 2, we have argued that *Talk with* (7,514 tokens) was preferred over *Speak with* (3,461 tokens) by Americans. We have further argued that *Speak with* had the highest frequency (271 tokens) in 2010, whereas *Talk with* had the highest frequency (633 tokens) in 1950. Additionally, we have shown that *Talk with* was always preferred over *Speak with* by Americans except 1820, 1830, and 1840. In section 3, we have maintained that *Speak with* and *Talk with* show the same property in four genres, whereas they show a different property in the other four genres. This suggests that in the eight genres of the COCA, *Speak with* is 50% the same as *Talk with*. We have contended, on the other hand, that *Speak with* is the nearest to *Talk with* in the web genre. In section 4, we have shown that *Talk with people* and *Speak with authority* are the most widely used ones (126 tokens vs. 56 tokens) in America. We have also shown that 26.08% of forty six nouns are the collocations of *Speak with* and *Talk with*.

REFERENCES

- 1) Corpus of Contemporary American English (COCA). 1, May 2022. <https://corpus.byu.edu/coca>.
- 2) Corpus of Historical American English (COHA). 1, May 2022. <https://corpus.byu.edu/coha>.
- 3) Murphy, R. (2016). *Grammar in Use*. Cambridge University Press.
- 4) Murphy, R. (2019). *English Grammar in Use*. Cambridge University Press.



There is an Open Access article, distributed under the term of the Creative Commons Attribution – Non Commercial 4.0 International (CC BY-NC 4.0) (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits remixing, adapting and building upon the work for non-commercial use, provided the original work is properly cited.