

A Comparative Analysis of Last and Final: A Corpus-based Analysis

Namkil Kang

Far East University
South Korea

Abstract

The main goal of this article is to provide a comparative analysis of the adjectives *last* and *final* in the Corpus of Contemporary American English (COCA). A major point to note is that the adjectives *last* and *final* indicate the same property in two genres, whereas they indicate a different property in six genres. This in turn implies that *last* is 33.33% the same as *final* in the genre analysis of the COCA. Talking about the distance between *last* and *final*, the former is the nearest to the latter in the blog genre, thereby indicating a high degree of similarity in the blog genre. The COCA shows that *last year* is the most widely used one in America, followed by *last week*, *last night*, *last time*, and *last month*, in that order. The COCA further shows that *final decision* is the most frequently used one in America, followed by *final word*, *final days*, and *final report*, in descending order. Finally, this article argues that 5.88% of sixty eight nouns occur with both *last* and *final*. It can thus be concluded that the adjectives *last* and *final* exhibit a very low similarity.

Keywords: last, final, COCA, type, token, similarity

1. Introduction

The main purpose of this article is to provide a comparative analysis of the adjectives *last* and *final* in the Corpus of Contemporary American English (COCA). As pointed out by Kang (2022a, 2022b, 2022c, 2022d), the so-called corpus is a database containing more than millions of words and the COCA is a heavily used corpus for English learning. First, we aim at providing an in-depth analysis of the COCA with regard to the adjectives *last* and *final*. More specifically, we aim to provide the genre analysis of two types and thereby capture the degree of the similarity between the two types. By doing so, we can see how similar they are in the COCA. Also, we aim to capture the similarity between *last* and *final* by using the Euclidean distance. The so-called Euclidean distance provides us with the degree of the similarity between *last* and *final* in eight genres. Additionally, we look into the use and frequency of nouns occurring with the collocations *last* and *final* and consider how similar they are. Finally, we provide the visualization of which nouns are linked to the adjectives *last* and *final* and thereby capture the similarity between them. This article is organized as follows. In section 2, we argue that the adjectives *last* and *final* indicate the same property in two genres, whereas they indicate a different property in six genres. The COCA clearly shows that *last* is 33.33% the same as *final* in the genre analysis of the COCA. We also show that *last* is the nearest to *final* in the blog genre, thus indicating a high degree of similarity in the blog genre. In section 3, we maintain that *last year* is the most preferred one (108,365 tokens) among Americans, followed by *last week*, *last night*, *last time*, and *last month*, in that order. We contend, on the other hand, that *final decision* is the most preferred by Americans, followed by *final word*, *final days*, and *final report*, in descending order. Finally, this paper argues that 5.88% of sixty eight nouns occur with both *last* and *final*.

2. The COCA

In what follows, we aim at providing a comparative analysis of *last* and *final* in the COCA. Table 1 shows the use and frequency of *last* and *final* in the COCA:

Table 1 Frequency of last and final in the COCA

GENRE	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Last	741,474	91,327	84,279	92,572	111,619	80,669	90,706	160,607	29,695
Final	120,564	14,669	18,033	7,836	12,792	9,490	15,072	23,613	19,059

As illustrated in Table 1, the overall frequency of *last* is 741,474 tokens, whereas that of *final* is 120,564 tokens. Simply put, *last* (741,474 tokens) is preferred over *final* (120,564 tokens) by Americans. This in turn implies that Americans prefer using *last* to using *final*. It is also worthwhile noting that *last* has the highest frequency (160,607 tokens) in the newspaper genre. Exactly the same can be said of *final*. *Final* also has the highest frequency (23,613 tokens) in the newspaper genre. Quite interestingly, *last* and *final* indicate the same property (rank-one in the newspaper genre) with regard to their ranking. This in turn indicates that they have a high similarity in rank-one. It should be noted, however, that *last* ranks second (111,619 tokens) in the spoken genre, whereas *final* ranks second (19,059 tokens) in the academic genre, thus indicating no similarity in rank-two. It must be pointed out, however, that in the spoken genre, *last* (111,619 tokens) is preferable to *final* (12,792 tokens).

We take this as showing that *last* is favored over *final* in the spoken genre. It must be mentioned, on the other hand, that in the academic genre, *last* (29,695 tokens) is used more frequently than *final* (19,059 tokens). This in turn indicates that Americans are keen on using *last* (29,695 tokens) rather than *final* (19,059 tokens) in the academic genre.

It is particularly noteworthy that *last* ranks third (92,572 tokens) in the TV/movie genre, whereas *final* ranks third (18,033 tokens) in the web genre. Again, the types *last* and *final* show no similarity regarding rank-three. It is probably worthwhile pointing out, however, that in the TV/movie genre, *last* (92,572 tokens) is used more widely than *final* (7,836 tokens). We take this as meaning that *last* (92,572 tokens) is preferable to *final* (7,836 tokens) in the TV/movie genre. Perhaps it is worthwhile saying, on the other hand, that the frequency of *last* is 84,279 tokens in the web genre, whereas that of *final* is 18,033 tokens. This in turn indicates that in the web genre, *last* (84,279 tokens) is used more frequently than *final* (18,033 tokens).

It must be stressed that *last* ranks fourth (91,327 tokens) in the blog genre, while *final* ranks fourth (15,072 tokens) in the magazine genre. Again, *last* and *final* indicate no similarity in rank-four. It must be emphasized, however, that *last* (91,327 tokens) is preferred over *final* (14,669 tokens) by American bloggers. It is interesting to observe, on the other hand, that in the magazine genre, *last* (90,706 tokens) is used more widely than *final* (15,072 tokens). From this, it seems evident that American journalists prefer to use *last* in their magazines.

It is worth saying that *last* ranks fifth (90,706 tokens) in the magazine genre, while *final* ranks fifth (14,669 tokens) in the blog genre. Quite interestingly, *last* and *final* indicate a different property in rank-five, thereby meaning that there is no similarity between them in rank-five.

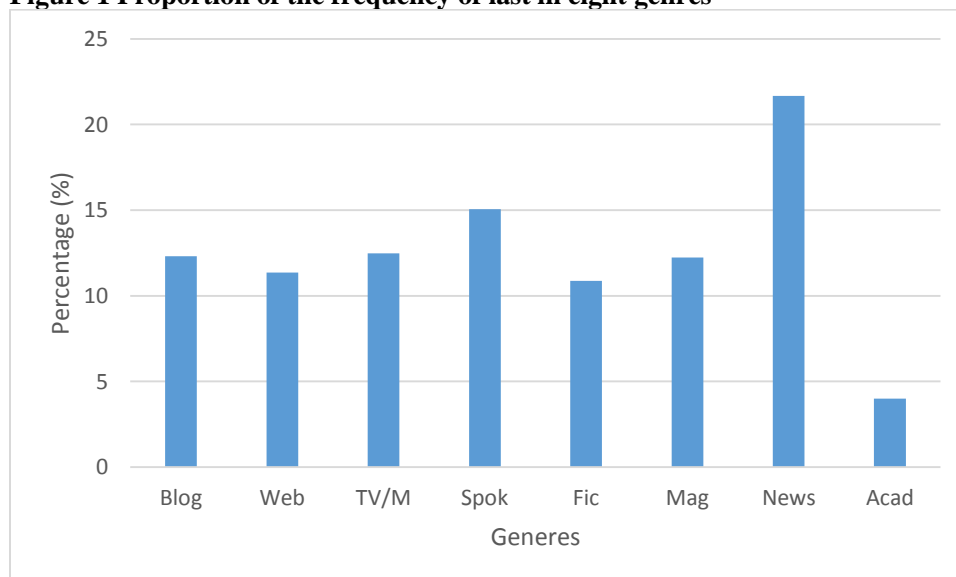
It should also be pointed out that *last* ranks sixth (84,279 tokens) in the web genre, whereas *final* ranks sixth (12,792 tokens) in the spoken genre. Again, *last* and *final* exhibit no similarity in rank-six.

It must be said that *last* and *final* rank seventh (80,669 tokens vs. 9,490 tokens) in the fiction genre. As can be seen from Table 1, they indicate the same property in the fiction genre, hence meaning that there is a high similarity between them regarding rank-seven. It should also be pointed out that the frequency of *last* (80,669 tokens) is eight times higher than that of *final* (9,490 tokens). This in turn indicates that American writers prefer to use *last* (80,669 tokens) in their novels.

Finally, it is also worth mentioning that *last* ranks eighth (29,695 tokens) in the academic genre, while *final* ranks eighth (7,836 tokens) in the TV/movie genre. Again, there is no similarity between *last* and *final* in rank-eight. To sum up, as observed earlier, *last* and *final* indicate the same property in two genres, whereas they indicate a different property in six genres. This in turn implies that *last* is 33.33% the same as *final* in the genre analysis of the COCA.

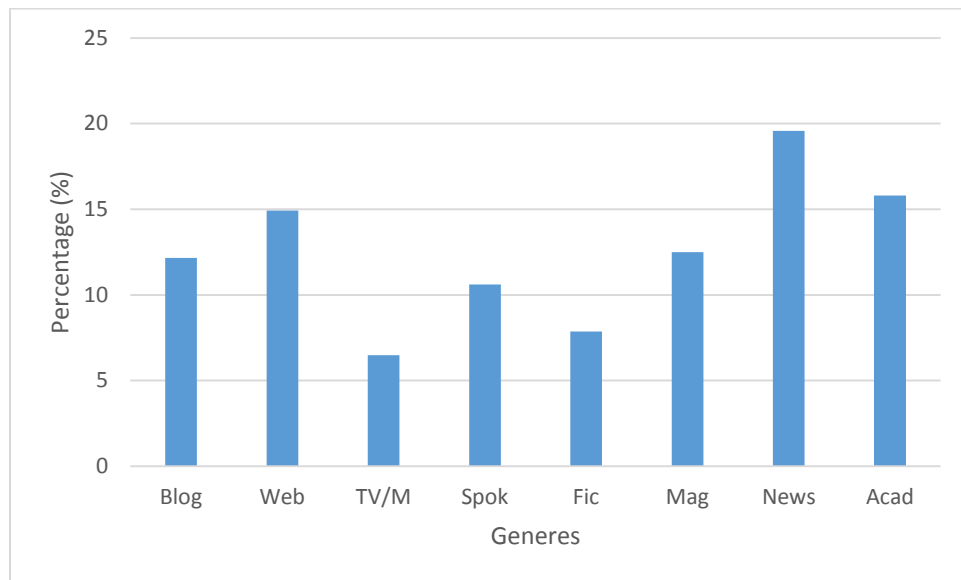
Now attention is paid to Figure 1. Figure 1 shows the proportion of the frequency of *last* and *final* in eight genres:

Figure 1 Proportion of the frequency of last in eight genres



As indicated in Figure 1, the newspaper genre is the most influenced by *last*, followed by the spoken genre, the TV/movie genre, and the blog genre, in that order.

Figure 2 Proportion of the frequency of final in eight genres



As exemplified in Figure 2, the newspaper genre is the most influenced by *final*, followed by the academic genre, the web genre, and the magazine genre, in that order.

Now let us observe the similarity between *last* and *final* in eight genres. Note that the Euclidean distance provides the distance between *last* and *final* in eight genres. More specifically, the distance between them refers to the degree of the similarity between them:

(1) The Euclidean distance:

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

Table 2 The distance between last and final in eight genres:

GENRE	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Percentage of last	12.31	11.36	12.48	15.05	10.87	12.23	21.66	4
Percentage of final	12.16	14.93	6.49	10.61	7.87	12.5	19.58	15.8
Euclidean distance	0.15	3.57	5.99	4.44	3	0.27	2.08	11.8

Quite interestingly, the Euclidean distance between *last* and *final* in the academic genre is 11.8, which is the highest. This in turn indicates that *last* is the furthest from *final* in the academic genre. It can thus be inferred that *last* and *final* indicate a low degree of similarity in the academic genre. More importantly, the Euclidean distance between *last* and *final* in the blog genre is 0.15, which is the lowest. This in turn implies that *last* is the nearest to *final* in the blog genre. It can thus be concluded that *last* and *final* indicate a high degree of similarity in the blog genre.

3. The Collocations last and final in the COCA

In what follows, we aim at providing the frequency of nouns occurring with the collocations *last* and *final* in the COCA. Table 3 shows the frequency of nouns occurring with the collocation *last*. The list was cut off in the top 36:

Table 3 Frequency of nouns occurring with the collocation last

Number	Collocation	Frequency
1	Last year	108,365
2	Last week	64,703
3	Last night	54,998
4	Last time	33,309
5	Last month	24,601
6	Last season	17,213
7	Last thing	10,131
8	Last summer	9,265
9	Last fall	6,249
10	Last couple	5,939
11	Last decade	5,537
12	Last Name	5,434
13	Last minute	5,191
14	Last day	5,066
15	Last weekend	4,266
16	Last spring	4,140
17	Last word	3,598
18	Last chance	3,183
19	Last resort	2,838
20	Last century	2,397
21	Last winter	2,172
22	Last days	2,079
23	Last person	2,020
24	Last words	1,762
25	Last question	1,598
26	Last election	1,489
27	Last part	1,437
28	Last years	1,365
29	Last place	1,344
30	Last quarter	1,344
31	Last moment	1,321
32	Last game	1,264
33	Last line	1,181
34	Last man	1,133
35	Last post	1,076
36	Last hour	1,067

It is significant to note that the expression *last year* has the highest frequency (108,365 tokens). This in turn indicates that the expression *last year* is the most preferred one (108,365 tokens) for Americans. It is worth mentioning, on the other hand, that the expression *last week* ranks second (64,703 tokens). Quite interestingly, the expression *last night* ranks third (54,998 tokens). It should also be pointed out that *last time* ranks fourth (33,309 tokens) in the COCA. It is also interesting to point out that the expression *last month* ranks fifth (24,601 tokens). From all of this, it seems clear that *last year* is the most preferred one (108,365 tokens) for Americans, followed by *last week*, *last night*, *last time*, and *last month*, in that order. More interestingly, the expression *last chance* ranks eighteenth (3,183 tokens) in the COCA. Also, it is interesting to note that the expression *last resort* ranks nineteenth (2,838 tokens) in the COCA. Finally, the expression *last part* ranks twenty seventh (1,437) in the COCA. We thus conclude that the expression *last year* is the most preferable one (108,365 tokens) for Americans.

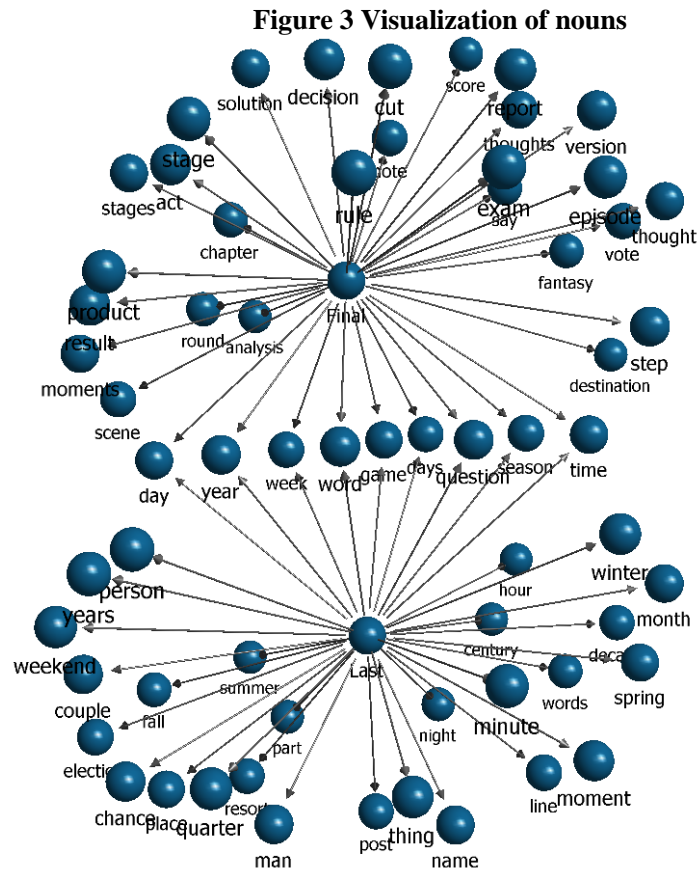
Now attention is paid to the frequency of nouns occurring with the collocation *final* in the COCA:

Table 3 Frequency of nouns occurring with the collocation final

Number	Collocation	Frequency
1	Final decision	1,656
2	Final word	1,346
3	Final days	1,338
4	Final report	1,309
5	Final season	1,298
6	Final round	1,271
7	Final day	1,116
8	Final analysis	1,097
9	Final product	1,025
10	Final year	1,016
11	Final question	934
12	Final version	883
13	Final game	834
14	Final step	726
15	Final scene	700
16	Final fantasy	698
17	Final stage	680
18	Final time	656
19	Final moments	649
20	Final score	626
21	Final destination	611
22	Final chapter	606
23	Final rule	589
24	Final stages	588
25	Final say	580
26	Final act	576
27	Final vote	563
28	Final result	558
29	Final thoughts	557
30	Final week	519
31	Final exam	516
32	Final cut	500
33	Final solution	493
34	Final thought	493
35	Final note	490
36	Final episode	488

Perhaps it is worthwhile saying that the expression *final decision* has the highest frequency (1,656 tokens). This in turn suggests that *final decision* is the most preferable one (1,656 tokens) among Americans. It is probably worthwhile pointing out, on the other hand, that the expression *final word* ranks second (1,346 tokens) in the COCA. Quite interestingly, the expression *final days* rank third (1,338 tokens) in the COCA. It should also be noted that the expression *final report* ranks fourth (1,309 tokens) in the COCA. From all of this, it seems evident that *final decision* is the most preferred one for Americans, followed by *final word*, *final days*, and *final report*, in descending order. It is noteworthy that *final destination* ranks twenty first (611 tokens) in the COCA. Also, it must be pointed out, on the other hand, that *final note* ranks thirty fifth (490 tokens) in the COCA. We thus conclude that the expression *final decision* is the most preferred (1,656 tokens) by Americans.

In what follows, we aim to provide the visualization of nouns and the collocations *last* and *final*. This visualization provides the degree of the similarity between *last* and *final* in the COCA. Figure 3 shows the picture of which nouns are linked to *last* and *final* and both of them:



As exemplified in Figure 3, nine nouns are linked to both *last* and *final*. Nine of sixty three nouns are linked to both *last* and *final*. However, only four nouns linked to both *last* and *final* show a similar meaning, namely *last word* vs. *final word*, *last question* vs. *final question*, *last day* vs. *final day*, *last days* vs. *final days*. This in turn indicates that 5.88% of sixty eight nouns occur with both *last* and *final*. It can thus be inferred that *last* and *final* exhibit a very low similarity.

4. Conclusion

To sum up, we have provided a comparative analysis of the adjectives *last* and *final* in the COCA. In section 2, we have argued that *last* and *final* indicate the same property in two genres, whereas they indicate a different property in six genres. This in turn shows that *last* is 33.33% the same as *final* in the genre analysis of the COCA. We have shown that *last* is the nearest to *final* in the blog genre. This in turn implies that *last* and *final* indicate a high degree of similarity in the blog genre. In section 3, we have maintained that *last year* is the most preferred one (108,365 tokens) for Americans, followed by *last week*, *last night*, *last time*, and *last month*, in that order. We have contended, on the other hand, that *final decision* is the most preferred one for Americans, followed by *final word*, *final days*, and *final report*, in descending order. Finally, we have shown that 5.88% of sixty eight nouns occur with both *last* and *final*. From this, it seems clear that the adjectives *last* and *final* exhibit a very low similarity.

References

- Kang, N. (2022a). A Comparative Analysis of Search for and Look for in Four Corpora. *Advances in Social Sciences Research Journal* 9 (3): 168-178.
- Kang, N. (2022b). A Comparative Analysis of Impressed by and Impressed with in Two Corpora. *Theory and Practice in Language Studies* 12 (5): 819-827.
- Kang, N. (2022c). On Speak to and Talk to: A Corpora-based Analysis. *Theory and Practice in Language Studies* 12 (7):1262-1270.
- Kang, N. (2022d). On Speak with and Talk with: A Corpora-based Analysis. *International Journal of Social Science and Human Research* 5 (8): 3354-3360.