

A Comparative Analysis of What Is Going On and What Is Happening In Three Corpora

Namkil Kang

Far East University, South Korea

Abstract: The ultimate goal of this paper is to compare what is going on with what is happening in the TV Corpus (TVC), the Movie Corpus (MC), and the Corpus of Contemporary American English (COCA). With respect to the TV Corpus, it is interesting to note that what is going on was preferable to what is happening. A further point to note is that what is going on (5,326 tokens) was always favored over what is happening (1,904 tokens) from the 1950s to the 2010s. When it comes to the Movie Corpus, it is worth noting that what is going on was preferable to what is happening in the movies of six countries. Additionally, it is worthwhile noting that what is going on and what is happening reached a peak (1,123 tokens vs. 569 tokens) in the 2010s. Quite interestingly, the COCA clearly shows that what is going on is favored over what is happening in America. With respect to the COCA, it is interesting to point out that what is going on is 12.5% the same as what is happening in eight genres. The COCA further shows that what is going on is the nearest to what is happening in the fiction genre, but the former is the furthest from the latter in the TV/movie genre.

Keywords: TV Corpus, Movie Corpus, COCA, what is going on, what is happening, token.

I. Introduction

As pointed out by Murphy (2016, 2019), *what is going on* and *what is happening* are used synonymously. The main goal of this paper is to compare *what is going on* with *what is happening* in the TV Corpus, the Movie Corpus, and the Corpus of Contemporary American English. First, we consider which expression is the preferable one in six countries (America, the UK, Canada, Australia, New Zealand, and Ireland). More specifically, we consider the diachronic use of *what is going on* and *what is happening* in the TV Corpus and the Movie Corpus. Second, we observe the similarity between *what is going on* and *what is happening* in the eight genres of the COCA. Third, we also observe the distance between *what is going on* and *what is happening* in the eight genres of the COCA. Simply put, we consider the Euclidean distance between *what is going on* and *what is happening*. This paper is organized as follows. In section 2, we argue that *what is going on* (5,326 tokens) was preferable to *what is happening* (1,904 tokens) in the TV Corpus. We further argue that *what is going on* (5,326 tokens) was always favored over *what is happening* (1,904 tokens) from the 1950s to the 2010s. In section 3, we maintain that *what is going on* (2,487 tokens) was preferable to *what is happening* (1,186 tokens) in six countries' movies. We also show that *what is going on* and *what is happening* reached a peak (1,123 tokens vs. 569 tokens) in the 2010s. In section 4, we contend that *what is going on* (6,837 tokens) is favored over *what is happening* (6,164 tokens) in America. We argue that *what is going on* is 12.5% the same as *what is happening* in the eight genres of the COCA. Finally, we further argue that *what is going on* is the nearest to *what is happening* in the fiction genre, but the former is the furthest from the latter in the TV/movie genre.

II. The TV Corpus

In what follows, we aim to consider the diachronic use of *what is going on* and *what is happening* in the TV Corpus (1950s-2010s). Table 1 shows the use of *what is going on* and *what is happening* from the 1950s to the 2010s:

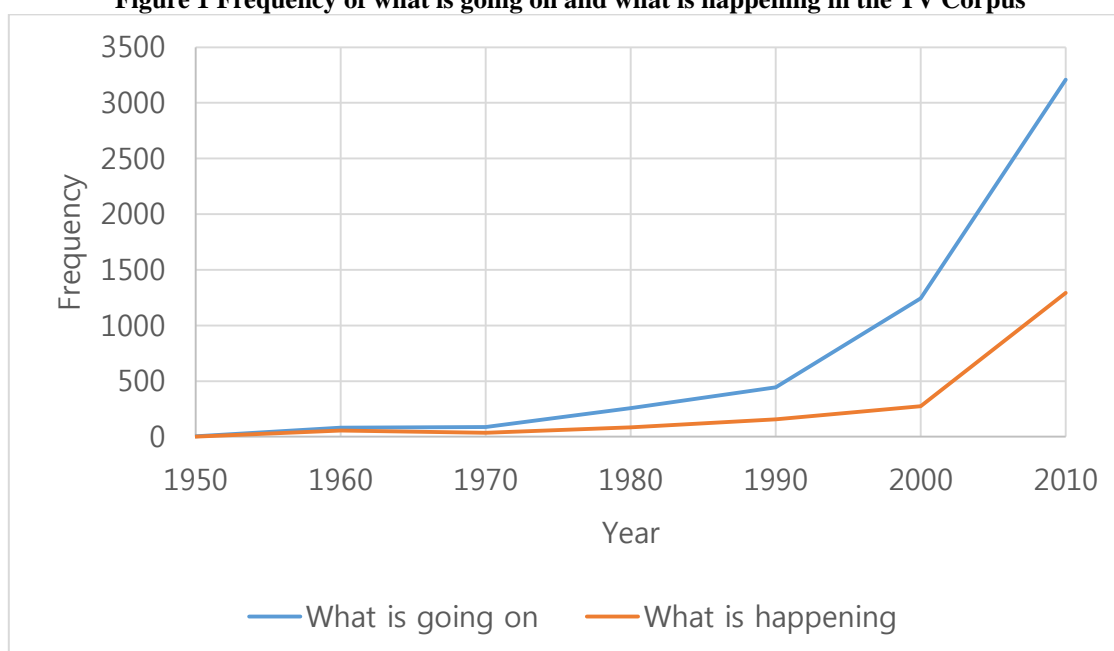
Table 1 Frequency of what is going on and what is happening in the TV Corpus

Type	What is going on	What is happening
1950s	3	2
1960s	83	56
1970s	87	36
1980s	258	85
1990s	443	157

2000s	1,244	274
2010s	3,208	1,294
All	5,326	1,904
US/CA	4,671	1,567
UK/IE	526	283
AU/NZ	92	29
MISC	37	25

An important question is “Which expression was the preferable one in the TV Corpus?” Table 1 clearly indicates that *what is going on* was the preferable one (5,326 tokens) in the TV Corpus. The overall frequency of *what is going on* is 5,326 tokens, whereas that of *what is happening* is 1,904 tokens. That is to say, the frequency of *what is going on* (5,326 tokens) is two times higher than that of *what is happening* (1,904 tokens). This in turn suggests that *what is going on* was preferable to *what is happening* from the 1950s to the 2010s. The following graph shows the diachronic use of *what is going on* and *what is happening* for sixty years:

Figure 1 Frequency of what is going on and what is happening in the TV Corpus



There was a gradual rise (an increase of 84 tokens) in the figure of *what is going on* from the 1950s to the 1970s. More interestingly, there was a steady increase (a rise of 185 tokens) in the figure of *what is going on* from the 1980s to the 1990s. Quite interestingly, there was a dramatic increase (an increase of 2,765 tokens) in the figure of *what is going on* from the 1990s to the 2010s. More importantly, *what is going on* had the highest frequency (3,208 tokens) in the 2010s, which in turn indicates that *what is going on* was the most preferred one (3,208 tokens) in that period. It is worth noting that *what is going on* was the most preferred (4,671 tokens) by American and Canadian celebs, followed by British and Irish ones (526 tokens), and Australian and New Zealand ones (92 tokens).

It is worth noting that there was a slight rise (an increase of 54 tokens) in the figure of *what is happening* from the 1950s to the 1960s. Interestingly, there was a sudden decline (a decrease of 20 tokens) in the figure of *what is happening* in the 1970s. Quite interestingly, there was a gradual rise (an increase of 121 tokens) in the figure of *what is happening* from the 1970s to the 1990s. More importantly, there was a dramatic rise (a sharp rise of 1,020 tokens) in the figure of *what is happening* from the 2000s to the 2010s. It is significant to note that *what is happening* reached a peak (1,294 tokens) in the 2010s. This in turn suggests that *what is happening* was the most preferred one (1,294 tokens) in the TV programs of six countries in the 2010s. It is important to note that *what is going on* (5,326 tokens) was always preferable to *what is happening* (1,904 tokens) in six countries' TV programs from the 1950s to the 2010s. It would be worth mentioning that *what is going on* and *what is happening* have such a big difference in their frequency after the 1990s. Most importantly, *what is happening* was the most preferred by American and Canadian celebs (1,567 tokens), followed by British and

Irish ones (283 tokens), and Australian and New Zealand ones (29 tokens), in that order. We thus conclude that *what is going on* is favored over *what is happening* in six countries' TV programs from the 1950s to the 2010s.

III. The Movie Corpus

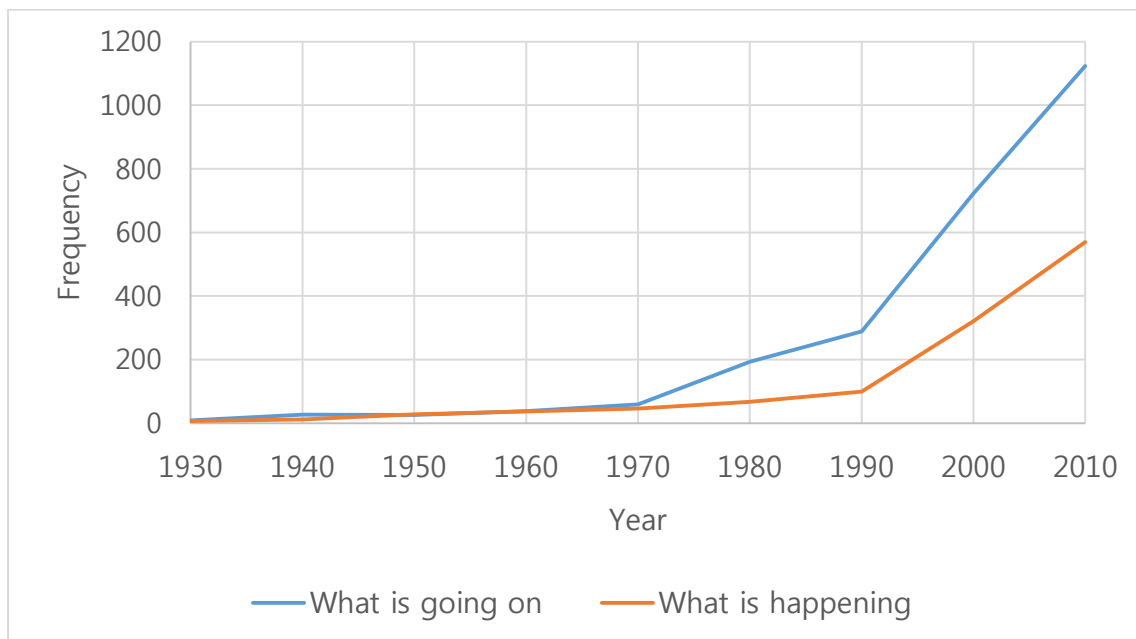
In what follows, we aim to consider the diachronic use of *what is going on* and *what is happening* from the 1930s to the 2010s. Table 2 shows the use of *what is going on* and *what is happening* in six countries' movies from the 1930s to the 2010s:

Table 2 Frequency of what is going on and what is happening in the Movie Corpus

Type	What is going on	What is happening
1930s	9	7
1940s	27	12
1950s	26	28
1960s	38	37
1970s	59	46
1980s	193	67
1990s	289	99
2000s	723	321
2010s	1,123	569
All	2,487	1,186
US/CA	2,005	886
UK/IE	248	184
AU/NZ	44	18
MISC	190	98

An important question is “Which expression was the preferable one in six countries' movies?” Table 2 clearly indicates that *what is going on* was preferable to *what is happening* in the movies of six countries. More specifically, the overall frequency of *what is going on* is 2,487 tokens, whereas that of *what is happening* is 1,186 tokens. This in turn indicates that *what is going on* was preferred over *what is happening* by the movie writers of six countries. The following graph shows the diachronic use of *what is going on* and *what is happening* from the 1930s to the 2010s:

Figure 2 Frequency of what is going on and what is happening in the Movie Corpus



It is worth noting that there was a slight increase (a rise of 18 tokens) in the figure of *what is going on* from the 1930s to the 1940s. It should be noted, however, that there was a sudden fall (a decline of 1 token) in the figure of *what is going on* in the 1950s. Interestingly, there was a steady increase (an increase of 155 tokens) in the figure of *what is going on* from the 1960s to the 1980s. Quite interestingly, there was a dramatic rise (a rise of 834 tokens) in the figure of *what is going on* from the 1990s to the 2010s. It is important to note that *what is going on* reached a peak (1,123 tokens) in the 2010s. This in turn suggests that *what is going on* was the most preferred one in the movies of six countries in the 2010s. It is worth pointing out that *what is going on* was the most preferred by American and Canadian movie writers (2,005 tokens), followed by British and Irish ones (248 tokens), and Australian and New Zealand ones (44 tokens), in that order.

Interestingly, there was a gradual increase (a rise of 21 tokens) in the figure of *what is happening* from the 1930s to the 1950s. It is interesting to point out that there was a steady rise (an increase of 30 tokens) in the figure of *what is happening* from the 1960s to the 1980s. Quite interestingly, there was a gradual increase (a rise of 470 tokens) in the figure of *what is happening* from the 1990s to the 2010s. More importantly, *what is happening* had the highest frequency (569 tokens) in the 2010s, which in turn indicates that *what is happening* was the most preferred by six countries' movie writers in that period. It is interesting to note that *what is going on* was always preferable to *what is happening* except the 1950s. In the 1950s, *what is happening* was favored over *what is going on* in the movies of six countries. Finally, it should be pointed out that *what is happening* was the most preferred by American and Canadian movie writers (886 tokens), followed by British and Irish ones (184 tokens), and Australian and New Zealand ones (18 tokens), in that order. We thus conclude that *what is going on* was preferable to *what is happening* in six countries' movies.

IV. What is going on and what is happening in the COCA

In what follows, we aim to consider the genre frequency of *what is going on* and *what is happening* in the COCA. Also, we compare the ranking of *what is going on* and that of *what is happening* in the eight genres of the COCA. Table 3 shows the use of *what is going on* and *what is happening* in eight genres:

Table 3 Frequency of what is going on and what is happening in the COCA

Genre	All	Blog	Web	TV/M	Spok	Fic	Mag	News	Acad
What is going on	6,837	1,204	859	2088	1412	291	341	330	312
What is happening	6,164	1,244	952	762	1,389	319	461	554	483

An important question is “Which expression is the preferable one for Americans?” Table 3 clearly shows that *what is going on* is preferable to *what is happening* in America. More specifically, the overall frequency of *what is going on* is 6,837 tokens, whereas that of *what is happening* is 6,164 tokens. This in turn suggests that *what is going on* is preferred over *what is happening* by Americans.

It is significant to note that *what is going on* ranks first (2,088 tokens) in the TV/movie genre, whereas *what is happening* ranks first (1,389 tokens) in the spoken genre. *What is going on* and *what is happening* show a different pattern about rank-one, hence showing a low degree of similarity. With respect to the TV/movie genre, it is interesting to point out that *what is going on* is favored over *what is happening*. That is to say, the frequency of *what is going on* (2,088 tokens) is almost three times higher than that of *what is happening* (762 tokens). We take this as indicating that American celebs prefer using *what is going on* to using *what is happening*. Likewise, the frequency of *what is going on* (1,412 tokens) is much higher than that of *what is happening* (1,389 tokens) in the spoken genre. This in turn implies that Americans prefer using *what is going on* rather than using *what is happening* in daily conversation.

It is worth mentioning that *what is going on* ranks second (1,412 tokens) in the spoken genre, whereas *what is happening* ranks second (1,244 tokens) in the blog genre. Again, *what is going on* and *what is happening* show a different pattern about rank-two, thus showing a low degree of similarity in two genres. Additionally, it should be pointed out that *what is happening* is preferable to *what is going on* in the blog genre. To be more specific, the frequency of *what is happening* (1,244 tokens) is higher than that of *what is going on* (1,204 tokens) in the blog genre. This in turn indicates that American bloggers prefer using *what is happening* rather than using *what is going on*.

It is worthwhile noting that *what is going on* ranks third (1,204 tokens) in the blog genre, whereas *what is happening* ranks third (952 tokens) in the web genre. Again, *what is going on* and *what is happening* show a different property in rank-three, hence showing a low degree of similarity. It must be noted that the frequency of *what is happening* (952 tokens) is higher than that of *what is going on* (859 tokens) in the web

genre. We take this fact as indicating that *what is happening* is preferable to *what is going on* in the web genre.

It is noteworthy that *what is going on* ranks fourth (859 tokens) in the web genre, whereas *what is happening* ranks fourth (762 tokens) in the TV/movie genre. Interestingly, *what is going on* and *what is happening* do not show the same pattern in rank-four, thus showing a low degree of similarity.

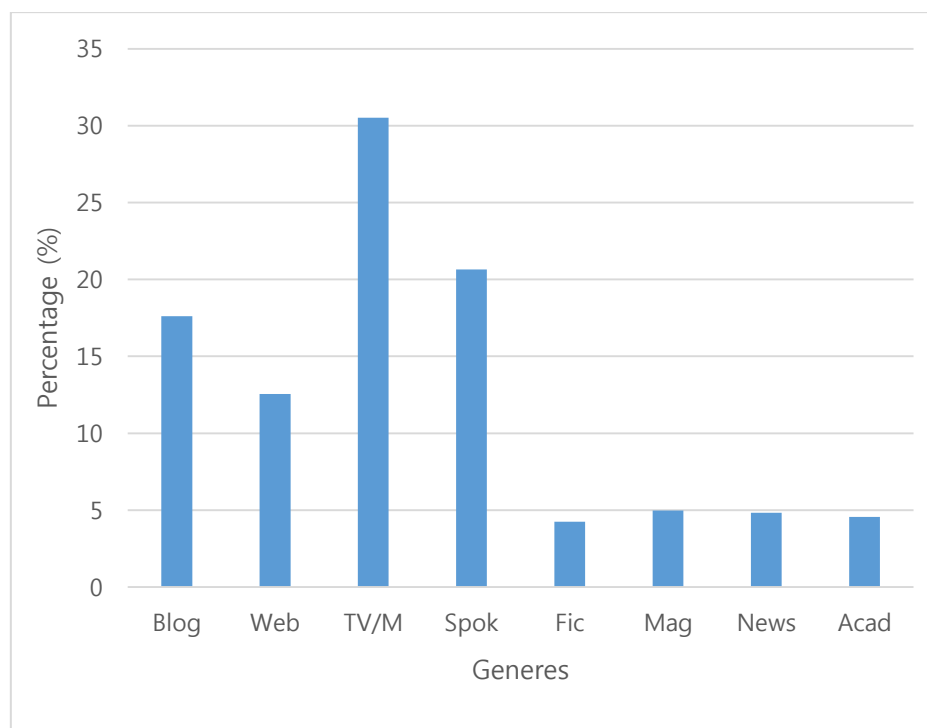
It is worthwhile pointing out that *what is going on* ranks fifth (341 tokens) in the magazine genre, whereas *what is happening* ranks fifth (554 tokens) in the newspaper genre. Again, *what is going on* does not show the same pattern as *what is happening* in rank-five, hence implying that they are used interchangeably, but they show a low degree of similarity. Additionally, it must be noted that *what is happening* (554 tokens) is favored over *what is going on* (330 tokens) in the newspaper genre. This in turn implies that American journalists prefer using *what is happening* (554 tokens) rather than using *what is going on* (330 tokens). Likewise, *what is happening* (461 tokens) is preferable to *what is going on* (341 tokens) in the magazine genre.

It would be interesting to note that *what is going on* ranks sixth (330 tokens) in the newspaper genre, whereas *what is happening* ranks sixth (483 tokens) in the academic genre. Interestingly, *what is going on* does not show the same property as *what is happening*. That is to say, they reveal a different pattern in rank-five. More interestingly, in the academic genre, the frequency of *what is happening* (483 tokens) is much higher than that of *what is going on* (312 tokens). This in turn shows that *what is happening* (483 tokens) is favored over *what is going on* (312 tokens) in the academic genre.

It is worth observing that *what is going on* ranks seventh (312 tokens) in the academic genre, whereas *what is happening* ranks seventh (461 tokens) in the magazine genre. Interestingly, *what is going on* and *what is happening* do not have a commonality in two genres. Finally, *what is going on* (291 tokens) and *what is happening* (319 tokens) rank eighth in the fiction genre. It should be pointed out, however, that *what is happening* (319 tokens) is preferred over *what is going on* (291 tokens) by American writers. To sum up, *what is going on* and *what is happening* show the same pattern in the fiction genre, but they show a different pattern in the other genres (seven genres). This in turn suggests that *what is going on* is 12.5% the same as *what is happening*, hence implying that they show a low degree of similarity.

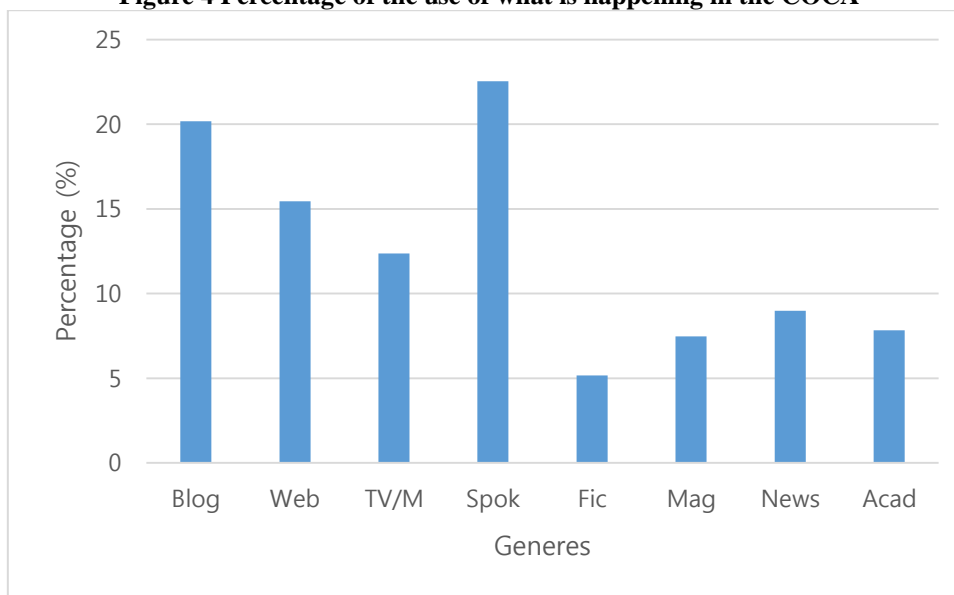
The following graph shows the use of *what is going on* and *what is happening* in eight genres:

Figure 3 Percentage of the use of what is going on in the COCA



As indicated in Figure 3, the TV/movie genre is the most influenced by *what is going on*, followed by the spoken genre, the blog genre, the web genre, the magazine genre, the newspaper genre, the academic genre, and the fiction genre, in that order.

Figure 4 Percentage of the use of what is happening in the COCA



As exemplified in Figure 4, the spoken genre is the most influenced by *what is happening*, followed by the blog genre, the web genre, the TV/movie genre, the newspaper genre, the academic genre, the magazine genre, and the fiction genre, in that order.

Finally, attention is paid to the Euclidean distance between *what is going on* and *what is happening*. The Euclidean distance is as follows:

(1) The Euclidean distance:

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

Table 4 shows the Euclidean distance between *what is going on* and *what is happening* in eight genres:

Table 4 Euclidean distance between what is going on and what is happening

Genre	Blog	Web	TV/M	Spok	Fic	Mag	News	Acad
Percentage of what is going on	17.61	12.56	30.53	20.65	4.25	4.98	4.82	4.56
Percentage of what is happening	20.18	15.44	12.36	22.53	5.17	7.47	8.98	7.83
Euclidean distance	2.57	2.88	18.17	1.88	0.92	2.49	4.16	3.27

As shown in Table 4, *what is going on* is the furthest from *what is happening* in the TV/movie genre. The Euclidean distance between *what is going on* and *what is happening* in the TV/movie genre is 18.17, which is the highest. On the other hand, *what is going on* is the nearest to *what is happening* in the fiction genre. The Euclidean distance between *what is going on* and *what is happening* is 0.92, which is the lowest. This in turn suggests that *what is going on* and *what is happening* show deep similarities in the fiction genre. Interestingly, the Euclidean distance between *what is going on* and *what is happening* in the spoken genre is 1.88, which is the second lowest. This in turn indicates that *what is going on* and *what is happening* show a high degree of similarity in the spoken genre. We thus conclude that *what is going on* is the nearest to *what is happening* in the fiction genre.

V. Conclusion

To sum up, we have compared *what is going on* with *what is happening* in the TV Corpus, the Movie Corpus, and the Corpus of Contemporary American English. In section 2, we have argued that *what is going on* was preferable to *what is happening* from the 1950s to the 2010s. We have further argued that *what is going on* (5,326 tokens) was always preferable to *what is happening* (1,904 tokens) in six countries' TV programs from the 1950s to the 2010s. In section 3, we have maintained that *what is going on* (2,487 tokens) was preferred over *what is happening* (1,186 tokens) by the movie writers of six countries. We have also shown that *what is going on* (1,123 tokens) and *what is happening* (569 tokens) had the highest frequency in the 2010s. In section 4, we have contended that *what is going on* (6,837 tokens) is preferred over *what is happening* (6,164 tokens) by Americans. We have also shown that *what is going on* is 12.5% the same as *what is happening* in the eight genres of the COCA. Finally, we have argued that *what is going on* is the nearest to *what is happening* in the fiction genre, whereas the former is the furthest from the latter in the TV/movie genre.

References

- [1]. Corpus of Contemporary American English (COCA). 1, March 2022. Online <https://corpus.byu.edu/coca>
- [2]. Movie Corpus (MC). 1, March 2022. Online <https://english-corpora.org/movies/>
- [3]. Murphy, R. (2016). *Grammar in Use*. Cambridge University Press.
- [4]. Murphy, R. (2019). *English Grammar in Use*. Cambridge University Press.
- [5]. TV Corpus (TVC). 1, March 2022. Online <https://english-corpora.org/tv/>