Three Types of Auxiliaries: A Corpus-based Analysis

Namkil Kang

Far East University, Korea
somerville@hanmail.net

Abstract

The Corpus of Contemporary American English (COCA) contains more than one billion words of text from eight genres: spoken, fiction, magazines, newspapers, academic texts, TV and movies subtitles, blogs, and other web pages. The main goal of this article is to provide an in-depth analysis of the frequency of the three modal verb types must be+ing, may be+ing, and might be+ing within the Corpus of Contemporary American English (COCA). The COCA clearly indicates that may be+ing is the most commonly used of these three modal verb types in America. More specifically, may be+ing is the type most preferred by Americans, followed by might be+ing, and must be+ing, in that order. A major point to note is that the type might be+ing is the most widely used in the fiction genre in America. A further point to note is that may be+ing is the most widely used of these three modal verb types in the spoken genre in America. This in turn suggests that Americans are fond of using statements with moderate possibility when they talk. With respect to the magazine genre, may be+ing is the type most preferred by American writers. It is significant to note, however, that American journalists prefer statements with moderate or low possibility to statements with necessity. With respect to the academic genre, it is worth noting that may be+ing is the type most preferred by Americans. The reason why may be+ing is the most widely used in America may be that may be+ing, with its moderate possibility, is the most suitable for conveying knowledge, academic information, and actual facts.

Keywords: COCA, Corpus, Must be+ing, May be+ing, Might be+ing

1. Introduction

The reason why we investigate the frequency of *must be+ing*, *may be+ing*, and *might be+ing* is that we want to find out how they are used by Americans. The necessity of this research is to reveal how the three modal verb types *must be+ing*, *may be+ing*, and *might be+ing* are used by Americans. It is clear that the three modal verb types *must be+ing*, *may be+ing*, and *might be+ing* move from strong conjecture to weak conjecture (must be+ing > may be+ing > might be+ing). The main purpose of this article is to provide a detailed frequency analysis of the three modal verb types within the Corpus of Contemporary American English (COCA) [1]. In section 3, we show that *may be+ing* is the one of these three modal verb types most commonly used in America. More specifically, *may be+ing* is most preferred by Americans, followed by *might be+ing*, and *must be+ing*, in that order. In section 3, we argue that *might be+ing* is the most widely used in the fiction genre in America. We further argue that *may be+ing* is the most widely used in the spoken genre in America. This in turn suggests that

Americans are fond of using statements with moderate possibility when they talk. We contend, on the other hand, that *may be+ing* is the type most preferred by American writers in the magazine genre. We maintain, however, that statements with moderate or low possibility are widely used by American journalists. In addition, we show that *may be+ing* is most preferred by Americans in the academic genre. The reason why *may be+ing* is the most widely used of these three modal verb types in America may be that *may be+ing,* with its moderate possibility, is the most suitable for conveying knowledge, academic information, and actual facts.

2. Methodology

The Corpus of Contemporary American English [1] is probably widely used corpus of English. The main goal of this article is to provide an in-depth frequency analysis of the three modal verb types (*must be+ing, may be+ing, might be+ing*) within the Corpus of Contemporary American English (COCA) [1]. The following modal verb types move from strong conjecture to weak conjecture, as illustrated in (1):

$$must\ be+ing > may\ be+ing > might\ be+ing \qquad (1)$$

In this article, we aim to answer the following main questions: Which of these three modal verb types is the most commonly used in America? Which is most preferred by speakers or writers in America? Which is the most widely used in the five genres of the COCA [1]?

3. The corpus of contemporary american english and the frequency of the three modal verb types

In the following, we provide an in-depth analysis of the use and frequency of *must be+ing*, *may be+ing*, and *might be+ing* in the COCA (1990-2019) [1]:

Table 1. Frequency of the three types in the COCA

Type	Spoken	Fiction	Magazine	Newspaper	Academic	Total
must be+ing	503	1443	408	334	133	2,821
may be+ing	3247	433	2613	2193	1749	10,235
might be+ing	1866	1941	1276	1184	471	6,738

An important question to be asked is "which type of modal verb is most frequently used in America?" [Table 1] clearly shows that *may be+ing* is the most widely used of these three modal verb types in America. As alluded to in [Table 1], the most preferable type for Americans is *may be+ing* since it has the highest frequency (10,235 tokens) and the highest proportion (51.74%) among the three modal verb types. Then, why is *may be+ing* the one most commonly used in America? We wish to argue that *must be+ing* carries strong conjecture, so it conveys necessity, while *might be+ing* carries weak conjecture, so it conveys low possibility. We wish to argue, on the other hand, that *may be+ing* carries moderate conjecture, so it conveys moderate possibility. That may be why Americans prefer *may be+ing* to *must be+ing* and *might be+ing*. In addition, the reason why *must be+ing* is used less frequently than *may be+ing* is that *must be+ing* is semantically marked. It is evident that a type which is semantically marked is used less frequently than a type which is not semantically marked. Thus, Americans may prefer *may be+ing* with its moderate possibility

to *must be+ing* with its necessity. Another reason why *must be+ing* is used less frequently than *may be+ing* is that speakers or writers can use *must* only when they have evidence of some kind (Palmer 1987: 122) [2]. Let us observe the following sentences (Palmer 1987: 97).

a. John may be in his office.

b. John must be in his office. (2)

As pointed out by Palmer (1987: 97), *may* can be interpreted in terms of some kind of possibility, whereas *must* expresses some kind of necessity [2]. (2b) makes a judgment that John is bound to be in his office. Palmer (1987: 122) points out that "*must* essentially makes a conclusive judgment, usually from evidence of some kind" (Palmer 1987: 122) [2]. Thus, it is clear that the modal verb type *must be+ing* cannot be used without any evidence. This is why Americans prefer using *may be+ing,* with its moderate possibility, to *must be+ing,* with its necessity.

Now, a question to be asked is "which of the three modal verb types is most commonly used in the spoken genre?" [Table 1] clearly indicates that of the three types, *may be+ing* is the one most widely used in the spoken genre. Why is *may be+ing* the one most frequently used in the spoken genre? It is more common for people to talk with moderate possibility than with necessity. [Table 1] clearly shows that Americans do not prefer *must be+ing,* since it carries some kind of necessity. A further point to note is that Americans prefer *might be+ing* to *must be+ing*. This may imply that Americans like speaking to people not with necessity but with low possibility. As pointed out by Palmer (1987: 119), "*might* and *could* are used as tentative forms of epistemic *may* in all its possible environments, to express a lower degree of possibility" (Palmer 1987: 119) [2].

Now an important question to be asked is "which of the three modal verb types is the one most widely used in the fiction genre?" [Table 1] indicates that *might be+ing* is the one most frequently used in the fiction genre. More specifically, *might be+ing* has the highest frequency (1941 tokens) and the highest proportion (50.85%) among the three types. This suggests that Americans prefer using weak conjecture to using strong conjecture or moderate conjecture in the fiction genre. Then, why do Americans prefer *might be+ing* to *must be+ing* and *may be+ing* in the fiction genre? The reason why *might be+ing* is the type most preferred by Americans in the fiction genre may be that to use *might be +ing* with its low possibility in fiction can excite readers' curiosity. In addition, writers can develop stories by giving clues about the hidden themes. Thus, *must be+ing,* with its necessity, can be used less frequently in the fiction genre than *might be+ing,* with its low possibility.

Now, let us consider the magazine genre and the newspaper genre. An immediate question to be asked is "which modal verb type in these genres is the type most preferred by Americans?" [Table 1] clearly shows that *may be+ing* is the one most widely used in these genres. More specifically, of the three types *may be+ing* has the highest frequency (2613 tokens) and the highest proportion (60.80%) in the magazine genre, and it has the highest frequency (2193 tokens) and the highest proportion (59.09%) in the newspaper genre. So, why does *may be+ing* have the highest frequency in these two genres? Note that *may be+ing,* with its moderate possibility, is neither a strong statement nor a weak statement. Thus, the reason why *may be+ing* obtains the highest frequency in these genres is that *may be+ing,* with its moderate possibility, may be suitable for conveying factual information.

Let us turn our attention to the academic genre. Which type of modal verb is the one most commonly used in America? [Table 1] indicates that of the three types, *may be+ing* is the type most preferred by Americans in the academic genre. *May be+ing* has the highest

frequency (1749 tokens) in the academic genre, while *must be+ing* has the lowest frequency (133 tokens). Why does this happen? We wish to argue that *may be+ing* is the most suitable among the three types for conveying factual information, since it carries moderate possibility. Exactly the same thing can be said about the magazine genre and the newspaper genre. *May be+ing* is the most preferred among the three types by American writers since it maintains a neutral position. Just as in the magazine genre and the newspaper genre, *may be+ing* is the most preferred among the three types by Americans in the academic genre since it is neutral and thus is suitable for conveying knowledge, academic information, and actual facts.

Consider the ranking of the type *must be+ing* in the COCA. As indicated in Table 1, *must be+ing* has the highest frequency (1443 tokens) and the highest proportion (51.15%) in the fiction genre. On the other hand, it has the lowest frequency (133 tokens) and the lowest proportion (4.71%) in the academic genre. This indicates that *must be+ing* is the modal verb type most preferred by American writers in the fiction genre, whereas it is not preferred in the academic genre. Why is *must be+ing* the type most preferred by American writers in the fiction genre? We wish to argue that American writers use *must be+ing,* with its necessity, in order to give their fiction a dynamic impact. Also, it must be noted that writers often give clues about the hidden themes by using statements with low possibility. Thus, it is reasonable to conclude that both *might be+ing* with its low possibility and *must be+ing* with its necessity are widely used by Americans in the fiction genre.

Now, let us consider *must be+ing* in the academic genre. As indicated in [Table 1], *must be+ing* has the lowest frequency (133 tokens) and the lowest proportion (4.71%) in the academic genre. Why does *must be+ing* have its lowest frequency in the academic genre? This may be due to the fact that the progressive combined with *must* cannot be widely used in the academic genre. The progressive essentially indicates dynamic situations (Quick & Greenbaum [3], Saeed [4]). In addition, "the progressive indicates a temporary state, whereas the simple present indicates a permanent state" [5]. The academic genre can be characterized as a knowledge database, and knowledge in the database has a permanent property, but the progressive indicates a temporary state, which leads to the infrequency of *must be+ing* in the academic genre. Furthermore, *must* can be used only when speakers or writers have evidence of some kind. Koster (1997: 244) argues that a form A is more specialized than B if A can fulfill fewer functions than B [6]. As such, *must be+ing* is a specialized form since it can be used only when speakers have evidence of some kind. Simply put, *must be+ing* fulfills fewer functions than *may be+ing* or *might be+ing* in the academic genre. Thus, it is reasonable to conclude that for these reasons, *must be+ing* has its lowest frequency in the academic genre.

Now let us observe the frequency of *may be+ing* in the COCA. A major point to note is that *may be+ing* has its highest frequency (3247 tokens) and highest proportion (31.72%) in the spoken genre. Why does *may be+ing* has the highest frequency in the spoken genre? This may be due to the fact that *may be+ing* carries moderate possibility. *Must be+ing,* with its necessity, is semantically marked to be used in daily conversation, so *may be+ing* or *might be+ing* may be preferred over *must be+ing* in the spoken genre. It should be noted that *may be+ing* is the type most preferred by Americans, followed by *might be+ing*, and *must be+ing*, in that order.

Now let us turn our attention to the status of *may be+ing* in the fiction genre in the COCA [1]. [Table 1] clearly indicates that *may be+ing* has the lowest frequency (433 tokens) and the lowest proportion (4.23%) in the fiction genre. Why does *may be+ing* have its lowest frequency in the fiction genre? As observed earlier, *might be+ing* is most preferred in the fiction genre by Americans, followed by *must be+ing*, and *may be+ing*, in that order. The reason why *may be+ing* has its lowest frequency in the fiction genre may be that *may be+ing*

carries moderate possibility. It is evident that writers encourage readers to read a book with excitement. In order to excite readers, writers can use inclusive strategies for readers by using *might be+ing* with its low possibility or *must be+ing* with its necessity. However, half possibility (*may be+ing*) would neither astonish nor excite readers, since *may be+ing* carries moderate possibility.

Now, let us consider *might be+ing* in the fiction genre in the COCA [1]. [Table 1] clearly shows that *might be+ing* has its highest frequency (1941 tokens) and highest proportion (28.80%) in the fiction genre. An important question to be asked is "why does *might be+ing* obtain its highest frequency in the fiction genre?" We wish to argue that fiction can become more emotional, intense, and vivid if the progressive is used. Cowan (2010: 363) argues that "the progressive expresses much more emotion than the simple present" (Cowan 2010: 363) [7](Cowan 2010: 363).

<div align="center">

a. This operation is really costing a lot of money.

b. This operation costs a lot of money. (3)

</div>

In addition, as observed earlier, by using statements with low possibility, writers can excite readers' curiosity.

It is worth noting that *might be+ing* has its lowest frequency (471 tokens) and lowest proportion (6.99%) in the academic genre. This may be due to the fact that *might be+ing* carries low possibility. Academic fields require statements with moderate possibility or necessity since they deal with knowledge, scientific, technological, and academic information, and actual facts. Thus, it is reasonable to conclude that for this reason, *might be+ing* has its lowest frequency in the academic genre.

4. Conclusion

To sum up, we have provided a detailed frequency analysis of the three types of modal verbs within the Corpus of Contemporary American English (COCA). In section 3, we have shown that the most preferable type for Americans among the three modal verb types is *may be+ing*. More specifically, *may be+ing* is the type most preferred by Americans, followed by *might be+ing*, and *must be+ing*, in that order. Also, we have argued that *might be+ing* is the one most widely used in the fiction genre in America. We have further argued that *may be+ing* the most frequently used one in the spoken genre in America. This in turn suggests that Americans are fond of using statements with moderate possibility when they talk. We have contended, on the other hand, that *may be+ing* is the type most preferred by American writers in the magazine genre. We have maintained, however, that American journalists prefer statements with moderate possibility or low possibility to statements with necessity. Additionally, we have shown that in the academic genre, *may be+ing* is the most preferred of these modal verb types by Americans, which may reflect the fact that *may be+ing,* with its moderate possibility, is the most suitable for conveying knowledge, academic information, and actual facts.

References

[1] Corpus of Contemporary American English (COCA), https://corpus.byu.edu/coca.

[2] Palmer F. R., "The English Verb," New York: Longman Publishers, (**1987**)

[3] Quirk R. and Greenbaum S., "University grammar of English," Longman, (**1993**)

[4] Saeed J. "Semantics," Oxford. Blackwell Publishers, (**2009**)

[5] Kang N-K., "On non-progressive verbs: a corpora-based analysis," International Journal of Language and Linguistics, vol.5, no.4, pp.166-177, **(2018)**

[6] Koster Jan, "Anaphora and the uniformity of grammar," In H. Bennis, P. Pica, and J. Rooryck, eds., Atomism and Binding, pp.235-254, Foris Publications, **(1997)**

[7] Cowan R., "The teacher's grammar of English," Cambridge: Cambridge University Press, **(2010)**