



A FACT AND FOUR TYPES: A CORPORA-BASED ANALYSIS

Namkil Kang

Department of English Language, Far East University, South Korea

ARTICLE INFO

Article History:

Received 06<sup>th</sup> July, 2019

Received in revised form 14<sup>th</sup>

August, 2019

Accepted 23<sup>rd</sup> September, 2019

Published online 28<sup>th</sup> October, 2019

Key words:

corpus, token, type, Corpus of Contemporary American English, British National Corpus, the Corpus of Historical American English, Time Magazine Corpus

ABSTRACT

The ultimate goal of this paper is to provide a detailed frequency analysis of *it is said that*, *it is believed that*, *it is alleged that*, and *it is expected that* within the Corpus of Contemporary American English (COCA), the British National Corpus (BNC), the Corpus of Historical American English (COHA), and the Time Magazine Corpus (TMC). The COCA and COHA clearly indicate that the type *it is said that* is the most preferred by American people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. This paper argues that the reason why *it is said that* has the highest frequency may be due to the fact that it carries the highest possibility of a fact and that it is the most suitable for conveying facts and factual information. The BNC clearly shows that *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. This implies that the results from the COCA, BNC, and COHA show exactly the same pattern. A major point of the frequency of the four types in the TMC is that *it is said that* was the most frequently used type by educated American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*, which in turn suggests that *it is said that* was the most preferred by educated American people. When it comes to the frequency of the academic genre, *it is expected that* is the most preferred by Americans, followed by *it is believed that*, *it is said that*, and *it is alleged that*, whereas *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is alleged that*, and *it is expected that*. As for the newspaper genre, *it is said that* is the most frequently used type by American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*, whereas *it is expected that* is the most commonly used type by British people, followed by *it is believed that*, *it is said that*, and *it is alleged that*.

Copyright©2019 Namkil Kang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The main purpose of this paper is to provide a detailed frequency analysis of *it is said that*, *it is believed that*, *it is alleged that*, and *it is expected that* within the Corpus of Contemporary American English, the British National Corpus, the Corpus of Historical American English, and the Time Magazine Corpus. In section 3.1, we argue that the type *it is said that* is the most preferred by American people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. We further argue that the reason why *it is said that* has the highest frequency may be due to the fact that it carries the highest possibility of a fact and that it is the most suitable for conveying something true or correct. In this section, we show that in the academic genre, *it is expected that* is the most preferred by Americans, followed by *it is believed that*, *it is said that*, and *it is alleged that*. Also, we argue that in the newspaper genre, *it is said that* is the most frequently used type by American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*.

In section 3.2, we maintain that *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*, in that order. This indicates that the BNC and COCA show exactly the same pattern. Also, we show that in the academic genre, *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is alleged that*, and *it is expected that* and that in the newspaper genre, *it is expected that* is the most commonly used type by British people, followed by *it is believed that*, *it is said that*, and *it is alleged that*, in that order. In section 4, we contend that *it is said that* is the most frequently used type by Americans for one hundred and ninety years. The COHA crucially shows that among the four types the most preferable type for speakers and writers in America is *it is said that*. In section 5, we show that in the TMC (1920s-2000s), *it is said that* (153 tokens) was the most frequently used type by educated American people, which in turn suggests that *it is said that* was the most preferred by educated American people. In section 6, we argue that *it is said that* was the most frequently used type by Americans from 1810 to 2017, whereas it was the most commonly used type by British people from the 1980s to the early 1990s.

\*Corresponding author: Namkil Kang

Department of English Language, Far East University, South Korea

**METHODOLOGY**

"Corpora are large machine-readable compilations of authentic texts. They can be general or specialized, depending on what kinds of texts are included in the corpus" (Cotos 2017: 249). The ultimate goal of this article is to provide a detailed frequency analysis of the four types *it is said that*, *it is believed that*, *it is alleged that*, and *it is expected that* within the Corpus of Contemporary American English, the British National Corpus, the Corpus of Historical American English, and the Time Magazine Corpus. These four types are closely related to a fact and the following examples (a-d) move from a high possibility to a low possibility with respect to a fact:

- a. It is said that S
- b. It is believed that S
- c. It is alleged that S
- d. It is expected that S

Our discussion focuses on answering the following questions: Which type is the most commonly used one in America and the UK? What does the frequency of the four types in the COCA, BNC, COHA, and TMC stand for? What are the main differences between the COCA and the BNC? What are the characteristics of educated Americans in the use of the four types? What is the correlation among the COCA, BNC, COHA, and TMC?

**RESULTS AND DISCUSSION**

*The Genre Frequency of the Four Types in the COCA*

*Why does it is said that have the highest frequency?*

In what follows, we are concerned with the genre frequency of the four types *it is said that*, *it is believed that*, *it is alleged that*, and *it is expected that* in the COCA (1995-2017).

**Table 1** Genre Frequency of the Four Types

Type	All	Spoken	Fiction	Magazine	Newspaper	Academic
It is said that	421	84	93	70	46	128
It is believed that	349	45	11	53	27	213
It is alleged that	27	10	0	5	2	10
It is expected that	340	14	6	27	31	262

An important question that must be addressed here is "which type is the most commonly used one among the four types?" Table 1 clearly indicates that the type *it is said that* is the most frequently used one among the four types. More specifically, the overall frequency of the type *it is said that* is 421 tokens and it has the highest frequency and the highest proportion (37.02%) among the four types. An immediate question to be asked here is "why does the type *it is said that* have the highest frequency and the highest proportion?" We wish to argue that the reason why it has the highest frequency and the highest proportion is that it carries the highest possibility of a fact among the four types, hence fact-oriented. In addition, when we want to refer to something as a fact, we should find out the most appropriate type related to a fact for communication purposes. That is, when we want to refer to something as a fact, we must find out the relevant type that indicates

something true or correct. The reason why *it is said that* has the highest frequency may be due to the fact that it carries the highest possibility in referring to something true or correct. It is thus reasonable to assume that *it is said that* in America is the most frequently used one in indicating something which is true or correct. Noteworthy is that the overall frequency of the four types shows the effect of the following hierarchy:

- (1) It is said that > It is believed that > It is expected that > It is alleged that

As alluded to in (1), among the four types, the type *it is said that* is the most preferred by American people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*, in that order. What is interesting is that the type *it is expected that* which carries the lowest possibility of a fact is more commonly used than the type *it is alleged that* which carries a low possibility of a fact.

*Why is it is expected that more commonly used than it is alleged that?*

An important question to be asked is "why is *it is expected that* more commonly used than *it is alleged that*?" The reason why *it is expected that* is more frequently used than *it is alleged that* may be due to the fact that the latter is more marked than the former. In linguistics, markedness is the state of standing out as unusual or divergent in comparison to a more common or regular form. Then the type *it is alleged that* is a marked expression and thus rarely used in everyday life. On the other hand, the type *it is expected that* is unmarked in comparison with the latter, hence a more common or regular form. Thus, it seems reasonable to assume that for this reason, *it is expected that* is more commonly used than *it is alleged that*.

*The genre frequency of it is said that*

It is worth noting that in the Table 1 the genre frequency of *it is said that* shows the effect of the following hierarchy:

- (2) Academic > Fiction > Spoken > Magazine > Newspaper

As indicated in (2), the academic genre is mainly influenced by the use of the type *it is said that*. A major question which must be addressed here is "why is *it is said that* in the academic genre the most frequently used one?" The reason why *it is said that* in the academic genre is the most commonly used type may be due to the fact that the academic genre is mainly used to convey factual information in the academic field such as sciences, learning, and academic controversies. In addition, *it is said that* carries the highest possibility of a fact among the four types and is used to indicate something true or correct. More interestingly, the type *it is said that* in the newspaper genre has the lowest frequency among the five genres. This in turn indicates that *it is said that* with the highest possibility of a fact is not the type preferred for American journalists. Even so, *it is said that* has the highest frequency (46 tokens) and the highest proportion (43.39%) in the newspaper genre of the four types.

*The genre frequency of it is believed that*

When it comes to the genre frequency of *it is believed that*, the academic genre is the most frequently used one by Americans, followed by the magazine genre, the spoken genre, the newspaper genre, and the fiction genre, in that order. The reason why the academic genre is the most commonly used by

Americans may be that *it is believed that* carries a high possibility of a fact and that it is mainly used to convey facts.

**The genre frequency of *it is expected that***

A major point of the genre frequency of *it is expected that* is that the academic genre is the most commonly used one by Americans, followed by the newspaper genre, the magazine genre, the spoken genre, and the fiction genre, in that order. Even though *it is expected that* carries the lowest possibility of a fact among the four types, the academic genre is the most frequently used by Americans since it is used to convey facts and factual information.

**The genre frequency of *it is alleged that***

As for the genre frequency of *it is alleged that*, the academic and spoken genres are the most commonly used genres by Americans. Again, the reason why the academic and spoken genres are the most frequently used genres may be that speakers in the academic and spoken genres say something true and correct with the type *it is alleged that*. It must be noted that *it is alleged that* is the most frequently used type in the spoken and academic genres.

**The spoken genre**

A point to note is that *it is said that* has the highest frequency (84 tokens) and the highest proportion (54.90%) in the spoken genre among the four types. In the spoken genre, *it is said that* (84 tokens) is the most commonly used type by Americans, followed by *it is believed that* (45 tokens), *it is expected that* (14 tokens), and *it is alleged that* (10 tokens), in that order. This in turn suggests that *it is said that* is the most appropriate type which is used to convey factual information in the spoken genre. A further point to note is that *it is alleged that* has the lowest frequency (10 tokens) and the lowest proportion (6.53%) in the spoken genre, which indicates that *it is alleged that* is the least preferred type. This may be due to the fact that *it is alleged that* is a marked type and that it carries a low possibility of a fact among the four types.

**The fiction genre and magazine genres of the four types**

A major point of the fiction genre of the four types is that *it is said that* is the most frequently used type by Americans, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. The reason why *it is said that* has the highest frequency (93 tokens) and the highest proportion (84.54%) among the four types is that fiction books intermingle fact with fiction and that *it is said that* is the most appropriate type which is used to convey facts since it carries the highest possibility of a fact among the four types.

When it comes to the magazine genre, *it is said that* is the most commonly used type by Americans, followed by *it is believed that*, *it is expected that*, and *the alleged that*. A magazine is a publication which is issued regularly, usually every week or every month, and which contains articles and stories. In the magazine genre, the reason why *it is said that* is the most frequently used type by Americans may be that *it is said that* has the highest possibility of a fact among the four types and that it is used to convey facts in articles and stories.

**The newspaper genre and academic genres of the four types**

As for the newspaper genre, *it is said that* is the most preferred by Americans, followed by *it is expected that*, *it is believed that*, and *it is alleged that*, in that order. The reason why *it is*

*said that* has the highest frequency (46 tokens) and the highest proportion (43.39%) in the newspaper genre is that a newspaper is a publication providing news, advertisements, and other information, which is supposed to be true and correct. That is why *it is said that* with the highest possibility of a fact is suitable for the newspaper genre.

It is worth pointing out that in the academic genre, *it is expected that* has the highest frequency (262 tokens) and the highest proportion (42.74%). More specifically, *it is expected that* is the most commonly used type by Americans, followed by *it is believed that*, *it is said that*, and *it is alleged that*. What is interesting is that the type *it is expected that* which carries the lowest possibility of a fact among the four types has the highest frequency (262 tokens). This may be due to the fact that learning deals with issues relating to hypotheses and knowledge which need to be validated. That may be why *it is expected that* is the most frequently used type in the academic genre.

**The Genre Frequency of the Four Types in the BNC**

**Which type is the most frequently used in the UK and America?**

In what follows, greater attention is paid to the genre frequency of the BNC and that of the COCA in order to examine the national variation of the four types. Speakers talk differently, depending on different regions and countries (Biber 1995).

**Table 2** The Genre Frequency of the Four Types in the BNC (from the 1980s to the early 1990s)

Type	All	Spoken	Fiction	Magazine	Newspaper	Non-acad	Academic	Misc
It is said that	248	4	18	15	12	62	64	73
It is believed that	154	1	7	7	25	52	20	42
It is alleged that	41	4	0	1	7	4	13	12
It is expected that	150	1	0	8	28	43	10	60

Table 2 indicates the use and genre frequency of the four types in the BNC. An important question to be asked is "which type is the most frequently used one in the UK?" The BNC clearly shows that the type *it is said that* has the highest frequency (248 tokens) and the highest proportion (41.82%). More specifically, *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*, in that order. This indicates that the BNC and COCA show exactly the same pattern. Simply put, the BNC and COCA show the effect of the following hierarchy:

It is said that > It is believed that > It is expected that > It is alleged that

When it comes to the overall frequency of the four types, there is practically no difference between the two corpora. It is thus reasonable to hypothesize that *it is said that* among the four types is the most preferred by American people and British people. Now an immediate question to be asked is "why does the type *it is said that* show the highest frequency (248 tokens) and the highest proportion (41.82%) among the four types?" In the UK as well as America, the type *it is said that* is suitable for conveying facts and factual information since it carries the highest possibility of a fact among the four types.

**The frequency of the three types**

Interestingly, the type *it is believed that* ranks second among the four types. The reason why it ranks second is that *it is said*

that carries the highest possibility of a fact, whereas *it is believed that* a high possibility of a fact. What is interesting is that *it is expected that* ranks third even though it carries the lowest possibility of a fact, whereas *it is alleged that* ranks fourth even though it has a low possibility of a fact. As observed earlier, the reason why *it is expected that* is more commonly used than *it is alleged that* may be that the latter is a marked type, but the former is an unmarked type in the UK and America. Thus, it is reasonable to assume that *it is expected that* with the lowest possibility of a fact is more frequently used than *it is alleged that* with a low possibility of a fact since the latter is marked, whereas the former is unmarked.

#### ***The spoken genre of the four types in the BNC and COCA***

As for the spoken genre of the four types, *it is said that* and *it is alleged that* have the highest frequency (4 tokens) and the highest proportion (40%) among the four types. More specifically, *it is said that* (4 tokens) and *it is alleged that* (4 tokens) are the most preferred by British people, and followed by *it is believed that* (1 token) and *it is expected that* (1 token). On the other hand, when it comes to the spoken genre of the four types, *it is said that* is the most preferred by Americans, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. The BNC and COCA exactly show the same pattern with respect to the frequency of *it is said that* in the spoken genre. That is to say, in the spoken genre, *it is said that* has the highest frequency (the BNC (4 tokens), the COCA (84 tokens)) and the highest proportion (the BNC (40%), the COCA (54.90%)) among the four types in the UK and America. This may be due to the fact that *it is said that* carries the highest possibility of a fact and that speakers in the conversation tend to use *it is said that* in order to convey something true and correct. It is thus reasonable to hypothesize that Americans and British people prefer using the highest possibility of a fact in the spoken genre. Most interestingly, the result of the COCA is different from that of the BNC in that in the spoken genre of the four types, *it is alleged that* ranks fourth in America, whereas it ranks first in the UK. This leads us to hypothesize that Americans do not prefer using a low possibility of a fact in the spoken genre, whereas British people prefer using a low possibility of a fact in the spoken genre.

#### ***The fiction genre of the four types in the BNC and COCA***

With regard to the fiction genre of the four types, *it is said that* is the most commonly used by British people, followed by *it is believed that*, and *it is alleged that* and *it is expected that*. On the other hand, in the same genre of the four types, *it is said that* is the most frequently used type by Americans, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. The common characteristic of the fiction genre of the four types is that *it is said that* is the most preferred by British people and Americans. This may be due to the fact that fiction books intermingle fact and fiction and that *it is said that* is the most suitable for conveying facts since it carries the highest possibility of a fact. A point to note about the fiction genre of the four types is that *it is believed that* ranks second among the four types in the UK and America. It happened due to the fact that *it is believed that* carries a high possibility of a fact. In the fiction genre, however, *it is alleged that* and *it is expected that* have no frequency (0 token) in the UK. On the other hand, the number of the token of *it is alleged that* and *it is expected that*

in America is 0 and 6, respectively. This in turn indicates that *it is alleged that* with a low possibility of a fact and *it is expected that* with the lowest possibility of a fact are rarely used in the fiction genre. Therefore, it seems reasonable to conclude that *it is alleged that* and *it is expected that* in the fiction genre are the least preferred type in the UK and America.

#### ***The Magazine genre of the four types in the BNC and COCA***

When it comes to the magazine genre, *it is said that* is the most commonly used type by British people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*. On the other hand, in the magazine genre, *it is said that* is the most frequently used type by Americans, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. The common feature of the magazine genre in the UK and America is that *it is said that* has the highest frequency (the BNC (15 tokens), the COCA (70 tokens)), and the highest proportion (the BNC (48.38%), the COCA (45.16%)). The reason why *it is said that* has the highest frequency may be that it is the most suitable for conveying something true and correct since it carries the highest possibility of a fact. Interestingly, in the magazine genre, *it is expected that* ranks second in the UK, whereas *it is believed that* ranks second in America. This suggests that British people prefer using the lowest possibility of a fact in the magazine genre, whereas American people prefer using a high possibility of a fact in the magazine genre. What is interesting is that *it is believed that* ranks third in the UK, whereas *it is expected that* ranks third in America. Another common feature of the magazine genre is that *it is alleged that* is the least preferred type in the UK and America. The reason why *it is alleged that* is the least preferred type in the magazine may be that it is marked and that it carries a low possibility of a fact.

#### ***The newspaper genre in the BNC and COCA***

With respect to the newspaper genre, *it is expected that* is the most preferred by British people, followed by *it is believed that*, *it is said that*, and *it is alleged that*, in that order. On the other hand, in the newspaper genre, *it is said that* is the most preferred by American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*. Most interestingly, *it is expected that* has the highest frequency (28 tokens) in the UK, whereas *it is said that* the highest frequency (46 tokens) in America. This in turn suggests that British people prefer using the lowest possibility of a fact in the newspaper genre, whereas Americans prefer using the highest possibility of a fact in the newspaper genre. Therefore, it seems reasonable to assume that Americans are emphasizing a fact and delivering a fact in the newspaper genre. Interestingly, *it is believed that* in the UK ranks second in the newspaper genre, whereas *it is expected that* in America ranks second. This indicates that British people prefer a high possibility of a fact after the lowest possibility of a fact, whereas Americans prefer the lowest possibility after the highest possibility of a fact. Finally, *it is alleged that* in the newspaper genre is the least preferred type in the UK and America. This may be due to the fact that it is marked and that it carries a low possibility of a fact among the four types.

#### ***The non-academic and academic genres***

The BNC clearly indicates that *it is said that* in the non-academic genre has the highest frequency (62 tokens) and the

highest proportion (38.50%) among the four types. More specifically, *it is said that* (62 tokens) is the most commonly used type by British people, followed by *it is believed that* (52 tokens), *it is expected that* (43 tokens), and *it is alleged that* (4 tokens). This suggests that British people prefer using *it is said that* with the highest possibility of a fact in the non-academic genre. Also, British people prefer using *it is believed that* with a high possibility of a fact, which in turn suggests that they like conveying something true or correct. The reason why *it is alleged that* is the least preferred type may be because it is marked.

When it comes to the frequency of the academic genre, *it is said that* is the most frequently used type by British people, followed by *it is believed that*, *it is alleged that*, and *it is expected that*, in that order. On the other hand, in the academic genre, *it is expected that* is the most commonly used type by Americans, followed by *it is believed that*, *it is said that*, and *it is alleged that*. This indicates that in order to convey facts British people prefer using *it is said that* with the highest possibility of a fact in the academic genre, whereas Americans prefer using *it is expected that* with a lowest possibility of a fact. It must be noted that *it is expected that* is the least preferred type in the UK, whereas *it is alleged that* is the least preferred type in America. This in turn suggests that in the UK, the lowest possibility of a fact is the least preferred in the academic genre, whereas in America, a low possibility of a fact is the least preferred.

**The Frequency of the COHA**

Now attention is paid to the frequency of the four types in the COHA (1810-2000):

**Table 3** The Frequency of the four types in the COHA

Types	It is said that	It is believed that	It is alleged that	It is expected that
1810	2	3	0	0
1820	31	13	0	1
1830	105	40	5	6
1840	114	38	4	4
1850	140	29	8	7
1860	124	37	11	2
1870	175	44	9	16
1880	225	84	15	28
1890	233	75	10	57
1900	234	93	16	37
1910	165	73	18	55
1920	180	130	12	107
1930	92	93	6	97
1940	88	60	2	68
1950	61	44	1	57
1960	58	29	3	26
1970	68	34	2	22
1980	57	18	6	12
1990	35	7	2	10
2000	27	11	0	11
Total	2214	955	130	623

It is worth pointing out that *it is said that* is the most frequently used type by Americans for one hundred and ninety years. Table 3 crucially shows that among the four types the most preferable type for speakers and writers in America is *it is said that*. This may be due to the fact that *it is said that*

(2214 tokens) with the highest possibility of a fact is the most suitable for conveying something that is true or correct. It is also worth noting that *it is believed that* ranks second among the four types. The type *it is believed that* ranks second, but Americans prefer using *it is believed that* (955 tokens) after *it is said that* (2214 tokens). This may be due to the fact that *it is believed that* has a high possibility of a fact. On the other hand, *it is alleged that* ranks fourth among the four types. It is the least preferred type for one hundred and ninety years, which may be because *it is alleged that* is marked and has a low possibility of a fact.

As can be seen from Table 3, a major point of *it is said that* is that its frequency continued to increase to 138 tokens for forty years from 1810 to 1850. After this period, the frequency of *it is said that* decreased to 16 tokens for ten years from 1850 to 1860. From 1860 to 1900 the frequency of *it is said that* continued to increase to 110 tokens and the figure of *it is said that* reached the peak (234 tokens) in the 1900s. This in turn indicates that *it is said that* was the most preferred by Americans in the 1900s. After this period, there was a dramatic decrease in the figure from 1920 to 1960 and then there was a steady decline in the figure from 1970 to 2000. We thus conclude that *it is said that* was the most frequently used type by Americans in the 1900s.

As alluded to in Table 3, the overall frequency of *it is believed that* is 955 tokens for one hundred and ninety years and *it is believed that* ranks second among the four types. The frequency of *it is believed that* increased to 37 tokens for 20 years from 1810 to 1830. After this period, there were slight fluctuations in the frequency of *it is believed that* from 1840 to 1910. One crucial point of *it is believed that* is that its figure (130 tokens) reached the peak in the 1920s, which in turn suggests that *it is believed that* was the most preferred by Americans in the 1920s. A further point to note is that there was a steady decline in the figure for forty years from 1920 to 1960 and then there was a sudden rise in the frequency of *it is believed that* in the 1970s. After this period, the frequency of *it is believed that* continued to decrease from 1970 to 2000. We thus conclude that *it is believed that* was the most frequently used type by Americans in the 1920s.

When it comes to the frequency of *it is alleged that*, there was no frequency for twenty years from 1810 to 1820. After this period, there were slight fluctuations in the figure of *it is alleged that* from 1830 to 1920. The crucial point to note is that the figure (18 tokens) of *it is alleged that* reached the peak in the 1910s, which suggests that *it is alleged that* was the most commonly used type by Americans in the 1910s even though it ranked fourth among the four types. After year 1910 there were slight fluctuations in the figure of *it is alleged that*.

With respect to the frequency of *it is expected that*, there were slight fluctuations in its figure from 1810 to 1850. From 1860 to 1890 the figure of *it is expected that* continued to increase to 55 tokens and then it reached the peak (107 tokens) in 1920. Most importantly, there was a dramatic decline in the figure of *it is expected that* from 1920 to 2000. We thus conclude that the type *it is expected that* was the most preferred by Americans in the 1920s.

**The Frequency of the Four Types in the TMC**

Now attention is paid to the frequency of the four types in the Time Magazine Corpus.

**Table 4** Frequency of the Four Types (TMC)

Types	It is said that	It is believed that	It is alleged that	It is expected that
1920s	70	32	11	53
1930s	20	4	1	14
1940s	6	6	1	7
1950s	8	4	1	3
1960s	14	2	0	14
1970s	12	4	0	3
1980s	11	8	0	2
1990s	10	3	0	4
2000s	2	1	0	0
Total	153	64	14	100

A major point of the frequency of the four types in the TMC (1920s-2000s) is that *it is said that* (153 tokens) was the most frequently used type by educated American people, followed by *it is expected that* (100 tokens), *it is believed that* (64 tokens), and *it is alleged that* (14 tokens). This suggests that *it is said that* was the most preferred by educated American people. An important question to be asked is “why was *it is said that* the most preferred by educated American people?” The reason why it was the most preferred by educated American people may be that it has the highest possibility of a fact and that it is the most suitable for conveying something true or correct. Despite the fact that educated American people prefer using the highest possibility of a fact, they also prefer using the lowest possibility of a fact. Simply put, *it is expected that* ranks second among the four types. The overall frequency of *it is expected that* is 100 tokens and the frequency difference between *it is said that* and *it is expected that* is 53 tokens. Interestingly, the type *it is believed that* ranks third and the overall frequency of it is 64 tokens. This indicates that educated American people do not like using *it is believed that*. This may be due to the fact that it has a low possibility of a fact among the types. It is worth pointing out that *it is alleged that* was the least preferred type among the four types. This in turn suggests that educated American people tend to avoid *it is alleged that*. The reason may be that it has a low possibility of a fact and that it is marked.

The figure of *it is said that* reached the peak (70 tokens) in the 1920s, which indicates that *it is said that* was the most preferred by educated American people. Interestingly, there was a dramatic decline in the frequency of *it is said that* from the 1920s to the 1940s. After this period, there was a gradual fall in the frequency of *it is said that* from the 1960s to the 2000s. Most interestingly, the figure of *it is expected that* (53 tokens), *it is believed that* (32 tokens), and *it is alleged that* (11 tokens) reached the peak in the 1920s. After this period, there were slight fluctuations in the frequency of *it is expected that*, *it is believed that*, and *it is alleged that* and then there was a gradual decline in their frequency.

**A Comparative Analysis of the Four Corpora**

The main purpose of this section is to provide a comparative analysis of the frequency of the four types in the COHA, BNC, COHA, and TMC. The COCA clearly indicates that *it is said that* is the most frequently used type by American people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. On the other hand, the BNC shows that *it is said that* is the most commonly used type by British people, followed by *it is believed that*, *it is expected that*, and *it is*

*alleged that*, in that order. Also, the COHA shows that *it is said that* is the most preferred by Americans, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. Interestingly, *it is said that* (153 tokens) was the most commonly used type by educated American people, followed by *it is expected that* (100 tokens), *it is believed that* (64 tokens), and *it is alleged that* (14 tokens). These indicate that the COCA, BNC and COCA show exactly the same pattern with respect to the overall frequency of the four types. However, the result of the TMC is different from that of the other corpora in that *it is expected that* is the second most frequently used type by educated American people. Educated American people prefer *it is expected that* to *it is believed that*. This in turn suggests that educated American people prefer the lowest possibility of a fact to a high possibility of a fact. However, educated people prefer the highest possibility of a fact to the lowest possibility of a fact.

As for the frequency of the academic genre, *it is expected that* is the most preferred by Americans, followed by *it is believed that*, *it is said that*, and *it is alleged that*. On the other hand, in the academic genre, *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is alleged that*, and *it is expected that*, in that order. This indicates that Americans prefer using the lowest possibility of a fact in the academic genre, whereas British people prefer using the highest possibility of a fact. Interestingly, *it is alleged that* in America is the least preferred type in the academic genre, whereas *it is expected that* in the UK is the least preferred type. This implies that Americans tend to avoid a low possibility of a fact, whereas British people tend to avoid the lowest possibility of a fact.

With regard to the newspaper genre, *it is said that* is the most frequently used type by American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*. On the other hand, *it is expected that* is the most commonly used type by British people, followed by *it is believed that*, *it is said that*, and *it is alleged that*, in that order. This implies that American people prefer using the highest possibility of a fact in the newspaper genre, whereas British people prefer using the lowest possibility of a fact in the newspaper genre. What is interesting is that *it is alleged that* is the least preferred type in America and in the UK, which suggests that Americans and British people tend to avoid a low possibility of a fact in the newspaper genre.

Now an important question that must be addressed here is “when is *it is said that* the most commonly used type?” The COCA clearly indicates that *it is said that* was the most frequently used type by Americans from 1995 to 2017. On the other hand, the BNC shows that *it is said that* was the most commonly used type by British people from the 1980s to the early 1990s. The COHA indicates that *it is said that* was the most preferred by Americans from 1810 to 2000. Finally, the TMC shows that it was the most preferred by Americans from the 1920s to the 2000s. These suggest that *it is said that* was the most frequently used type by Americans from 1810 to 2017, whereas it was the most commonly used type by British people from the 1980s to the early 1990s.

**CONCLUSION**

To sum up, we have provided a detailed frequency analysis of *it is said that*, *it is believed that*, *it is alleged that*, and *it is expected that* within the COCA, BNC, COHA, and TMC. In

section 3.1, we have shown that the type *it is said that* is the most preferred by American people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. The reason why *it is said that* has the highest frequency may be due to the fact that it carries the highest possibility in referring to something true or correct. Also, we have shown that in the academic genre, *it is expected that* is the most preferred by Americans, followed by *it is believed that*, *it is said that*, and *it is alleged that*. With regard to the newspaper genre, *it is said that* is the most frequently used type by American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*. In section 3.2, we have argued that *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is expected that*, and *it is alleged that*. This indicates that the BNC and COCA show exactly the same pattern. We have maintained that in the academic genre, *it is said that* is the most preferred by British people, followed by *it is believed that*, *it is alleged that*, and *it is expected that*, in that order. We have contended that *it is expected that* is the most commonly used type by British people, followed by *it is believed that*, *it is said that*, and *it is alleged that*. In section 4, we have shown that *it is said that* is the most frequently used type by Americans for one hundred and ninety years. The COHA crucially shows that among the four types the most preferable type for speakers and writers in America is *it is said that*. In section 5, we have argued that *it is said that* (153 tokens) was the most frequently used type by educated American people, followed by *it is expected that*, *it is believed that*, and *it is alleged that*. This suggests that *it is said that* was the most preferred by educated American people. Finally, in section 6, we have shown that *it is said that* was the most frequently used type by Americans from 1810 to 2017, whereas it was the most commonly used type by British people from the 1980s to the early 1990s.

## References

- Biber, D. (1995). On the role of computational, statistical, and interpretive techniques in multi-dimensional analysis of register variation. *Text* 15 (3): 314-370.
- British National Corpus (BNC). <https://corpus.byu.edu/bnc>.
- Corpus of Contemporary American English (COCA). <https://corpus.byu.edu/coca>
- Corpus of Historical American English (COHA). <https://corpus.byu.edu/coha>
- Cotos, E. (2017). Language for specific purposes and corpus-based pedagogy: In *the handbook of technology and the second language teaching and learning*. Edited by Carol A. Chapelle and Shannon Sauro, 248-264. Wiley Blackwell.
- Time Magazine Corpus (TMC). <https://corpus.byu.edu/time>.

### How to cite this article:

Namkil Kang (2019) 'A Fact and Four Types: A Corpora-Based Analysis', *International Journal of Current Advanced Research*, 08(10), pp. 20297-20303. DOI: <http://dx.doi.org/10.24327/ijcar.2019.20303.3961>

\*\*\*\*\*