

Sure of and Sure about in Corpora and ChatGPT

Namkil Kang
Far East University
South Korea

Abstract-The ultimate goal of this paper is to compare *sure of* and *sure about* in the TV Corpus, the Corpus of Contemporary American English, and ChatGPT. A point to note with respect to the TV Corpus is that *sure about* was more frequently used in America, Canada, Australia, and New Zealand, whereas *sure of* was more widely used in the UK and Ireland. A further point to note is that *sure of* and *sure about* were the most preferred by six countries' celebrities in the 2010s. With respect to the COCA, it is worth noting that *sure of it* and *sure of victory* are the most frequently used (12 tokens), followed by *sure of things* (10 tokens), and *sure of success* (9 tokens), in that order, whereas *sure about things* is the most widely used (8 tokens) and followed by *sure about people*. It is also worthwhile noting that only 2 of 42 nouns are the collocations of both *sure of* and *sure about* in the analysis of the COCA (the top 22). Simply put, *sure of* is 4.76% the same as *sure about* in the analysis of the top 22. Finally, it is interesting to point out that 8 of 36 nouns (22.22%) are the collocations of both *sure of* and *sure about* in the data of the ChatGPT (the top 22). It can thus be inferred that *sure of* and *sure about* have a low similarity.

Keywords- TV Corpus, COCA, ChatGPT, *sure of*, *sure about*, collocation

1. Introduction

The main goal of this paper is to compare *sure of* and *sure about* in the TV Corpus, the Corpus of Contemporary American English, and ChatGPT. First, we compare *sure of* and *sure about* in the TV Corpus. The TV Corpus provides the use of *sure of* and *sure about* from the 1950s to the 2010s. Also, the TV Corpus provides information on *sure of* and *sure about* that were used in the TVs of six countries such as America, the UK, Canada, Ireland, Australia, and New Zealand. Additionally, the TV Corpus provides the diachronic use of *sure of* and *sure about* from the 1950s to the 2010s and thereby enables us to see which type is the preferred one. Second, we compare *sure of* and *sure about* by examining

their collocations in the COCA. We provide 22 nouns that commonly follow the phrases *sure of* and *sure about* in the COCA. Third, we attempt to compare *sure of* and *sure about* by probing into their collocations in the ChatGPT. We provide 22 collocations that follow *sure of* and *sure about* and compare them.

2. The TV Corpus

This section is devoted to characterizing the diachronic aspects of *sure of* and *sure about* in the TV Corpus. Table 1 shows the use of *sure of* and *sure about* from the 1950s to the 2010s in the TV Corpus:

Table 1 Diachronic aspects of *sure of* and *sure about* in the TV Corpus

	ALL	1950s	1960s	1970s	1980s	1990s	2000s	2010s
Sure of	4,406	68	331	237	319	461	1,042	1,948
Sure about	6,218	13	110	100	238	495	1,722	3,540

It is important to note that the overall use of *sure of* from the 1950s to the 2000s was 4,406 tokens, while that of *sure about* during the same period was 6,218 tokens. This in turn indicates that *sure about* (6,218 tokens) was preferred over *sure of* (4,406 tokens) by six countries' celebrities. Put differently, they preferred to use

sure about (6,218 tokens) rather than use *sure of* (4,406 tokens) from the 1950s to the 2010s.

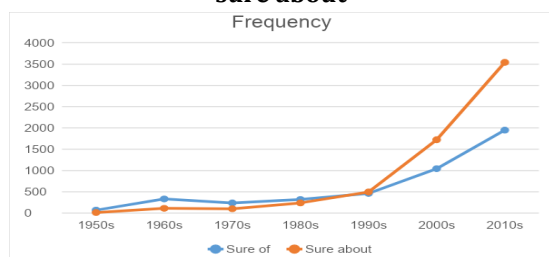
It is interesting to note that *sure of* had the lowest use (68 tokens) in the 1950s. Note, however, that there was a slight rise (263 tokens) in the use of *sure of* in the 1960s. More specifically, there was a rise of 263 tokens in the 1960s. It must be stressed, however, that there was a sudden decrease (94 tokens) in the use of

sure of in the 1970s. After the 1970s, the use of *sure of* continued to increase (from the 1970s to the 2010s). To go into detail, the use of *sure of* increased to 319 tokens in the 1980s. That is to say, there was an increase of 82 tokens in the 1980s. There was a gradual rise (142 tokens) in the use of *sure of* in the 1990s. It is noteworthy that there was a dramatic rise (581 tokens) in the use of *sure of* in the 2000s. That is, there was a sharp increase of 581 tokens in the 2000s. It is significant to note that *sure of* had the highest use (1,948 tokens) in the 2010s. This leads us to assume that six countries' celebrities were keen on using *sure of* (1,948 tokens) in the 2010s.

Now attention is paid to the use of *sure about* from the 1950s to the 2010s. It is interesting to observe that *sure about* had the lowest use (13 tokens) in the 1950s. There was a slight rise (97 tokens) in the use of *sure about* in the 1960s. Notice, however, that there was a sudden decline (10 tokens) in the use of *sure about* in the 1970s. Quite interestingly, the use of *sure about* steadily increased from the 1970s to the 2010s. To go into detail, there was a steady increase (138 tokens) in the use of *sure about* in the 1980s. There was a gradual rise (257 tokens) in the use of *sure about* in the 1990s. Most importantly, there was a dramatic rise (1,227 tokens) in the use of *sure about* in the 2000s. More interestingly, the use of *sure about* reached a peak (3,540 tokens) in the 2010s. Simply put, *sure about* had the highest use (3,540 tokens) in the 2010s. This in turn implies that *sure about* was the most preferred by six countries' celebrities in the 2010s.

To sum up, *sure of* and *sure about* both continued to show an upward trend in their use from the 1950s to the 2010s except for only the 1970s. Let us have a look at the diachronic aspects of *sure of* and *sure about* from the 1950s to the 2010s:

Figure 1 Diachronic aspects of *sure of* and *sure about*



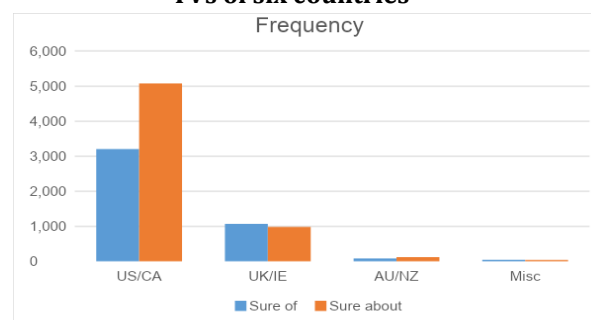
Most interestingly, the use of *sure of* was slightly higher (at least 55 tokens) than that of *sure about* from the 1950s to the 1980s. This in turn suggests that *sure of* was preferred over *sure about* by six countries' celebrities from the 1950s to the 1980s. Note, however, that *sure about* was more commonly used by six countries' celebrities compared to *sure of* from the 1990s to the 2010s. As exemplified in Figure 1, the use of *sure about* was by far higher (at least 34 tokens) than that of *sure of* during that time.

Now let us take a look at Table 2:

Table 2 Use of *sure of* and *sure about* in the TVs of six countries

	US/CA	UK/IE	AU/NZ	Misc
Sure of	3,207	1,068	84	47
Sure about	5,074	985	118	41

Figure 2 Use of *sure of* and *sure about* in the TVs of six countries



It is worthwhile noting that *sure of* was the most widely used (3,207 tokens) in America and Canada. Notice that it was the second most frequently used (1,068 tokens) in the UK and Ireland. Interestingly, it was the third most widely used (84 tokens) in Australia and New Zealand. Exactly the same can be said about *sure about*. It was the most frequently used (5,074 tokens) in America and Canada. Interestingly, it was less frequently used in the UK and Ireland. Most importantly, *sure about* was preferred over *sure of* by American and Canadian celebrities, as illustrated in Figure 2. Quite interestingly, *sure of* was favored over *sure about* in the UK and Ireland, as exemplified in Figure 2. Note that *sure about* was favored over *sure of* in Australia and New Zealand. We thus conclude that *sure about* was more frequently used in America,

Canada, Australia, and New Zealand, while *sure of* was more widely used in the UK and Ireland.

3. The Collocations of *sure of* and *sure about* in the COCA

This section focuses on comparing the collocations of *sure of* and those of *sure about* in the Corpus of Contemporary American English. Table 3 shows the collocations of *sure of* and *sure about* in the COCA:

Table 3 Collocations of *sure of* and *sure about* in the COCA

Number	Collocations of <i>Sure of</i>	Frequency	Collocations of <i>sure about</i>	Frequency
1	<i>sure of it</i>	12	<i>sure about things</i>	8
2	<i>sure of victory</i>	12	<i>sure about people</i>	3
3	<i>sure of things</i>	10	<i>sure about lapse</i>	3
4	<i>sure of success</i>	9	<i>sure about grapes</i>	3
5	<i>sure of hand</i>	4	<i>sure about dessert</i>	2
6	<i>sure of consent</i>	4	<i>sure about chartreuse</i>	2
7	<i>sure of spelling</i>	3	<i>sure about Americans</i>	2
8	<i>sure of people</i>	3	<i>sure about lunch</i>	2
9	<i>sure of defeat</i>	3	<i>sure about touring</i>	2
10	<i>sure of effect</i>	2	<i>sure about twitter</i>	2
11	<i>sure of</i>	2	<i>sure</i>	1

	events		about patients	
12	<i>sure of love</i>	2	<i>sure about parking</i>	1
13	<i>sure of right</i>	2	<i>sure about panda</i>	1
14	<i>sure of forgiveness</i>	2	<i>sure about olds</i>	1
15	<i>sure of foot</i>	2	<i>sure about Mrs</i>	1
16	<i>sure of salvation</i>	2	<i>sure about Mr</i>	1
17	<i>sure of sounds</i>	2	<i>sure about mom</i>	1
18	<i>sure of security</i>	1	<i>sure about Miss</i>	1
19	<i>sure of women</i>	1	<i>sure about mice</i>	1
20	<i>sure of wisdom</i>	1	<i>sure about me</i>	1
21	<i>sure of win</i>	1	<i>sure about MB</i>	1
22	<i>sure of will</i>	1	<i>sure about matters</i>	1

It is important to mention that *sure of it* obtains the highest frequency (12 tokens) in the COCA, while *sure about things* obtains the highest frequency (8 tokens). This in turn indicates that *sure of it* and *sure about things* are the most preferable ones (12 tokens vs. 8 tokens) for Americans. What is interesting about *sure of victory* is that it also ranks first (12 tokens) in the COCA. This may imply that this expression is the most preferred one (12 tokens) for Americans. It should also be pointed out that *sure about people* ranks second (3 tokens) in the COCA. This could imply that this expression is the second most widely used (3 tokens). It is worthwhile to consider *sure of things*. It ranks

third (10 tokens), which in turn suggests that *sure of things* is the third most widely used (10 tokens) in the COCA. It is worthwhile to consider rank-four. *Sure of success* ranks fourth in the COCA, which in turn means that it is the fourth most preferred one in the COCA. It can thus be inferred that *sure of it* and *sure of victory* are the most preferred ones (12 tokens) for Americans, followed by *sure of things* (10 tokens), and *sure of success* (9 tokens), in that order, while *sure about things* is the most preferable one (8 tokens) for Americans and followed by *sure about people* (*sure about lapse*, *sure about grapes*). It is significant to note that *sure of things* ranks third (10 tokens) in the COCA, whereas *sure about things* ranks first (8 tokens). It is also worth noting that *sure of people* ranks seventh (3 tokens) in the COCA, while *sure about people* ranks second (3 tokens). Most importantly, as exemplified in Table 3, only 2 of 42 nouns are the collocations of both *sure of* and *sure about* in the analysis of the top 22. From this it seems clear that *sure of* is 4.76% the same as *sure about*. We thus conclude that *sure of* and *sure about* exhibit a low similarity in the collocation analysis of the top 22.

4. The Collocations of *sure of* and *sure about* in the ChatGPT

This section is devoted to comparing the collocations of *sure of* and those of *sure about* in the ChatGPT. Table 4 shows the collocations of *sure of* and *sure about* in the ChatGPT:

Table 4 Collocations of *sure of* and *sure about* in the ChatGPT

Number	Collocations of <i>sure of</i>	Collocations of <i>sure about</i>
1	outcome	decision
2	result	choice
3	answer	fact
4	decision	idea
5	choice	plan
6	direction	theory
7	path	concept
8	course	statement
9	purpose	opinion
10	goal	belief
11	destination	understanding
12	route	reasoning

13	reason	outcome
14	explanation	result
15	meaning	prediction
16	understanding	hypothesis
17	fact	assumption
18	theory	assertion
19	concept	conclusion
20	method	judgement
21	approach	interpretation
22	solution	intuition

The first collocations of *sure of* and *sure about* that the ChatGPT recommended are the nouns *outcome* and *decision*, respectively, as illustrated in Table 4. This in turn indicates that there is a high probability that these collocations, both spoken and written, will be used. This probability stems from both train data and webs since the ChatGPT acquires a wealth of knowledge from its train data and webs. It must also be stressed that the second recommendation of the ChatGPT with respect to the collocations of *sure of* and *sure about* is the nouns *result* and *choice*, respectively. It should be noted, on the other hand, that the third collocations of *sure of* and *sure about* that the ChatGPT recommended are the nouns *answer* and *fact*, respectively. This in turn implies that train data also include these collocations. It is noteworthy that *sure of* and *sure about* have the noun *decision* in common as their collocation. Also, *sure of* and *sure about* have a commonality with respect to the collocation *choice*. The noun *choice* was also used as the collocation of both *sure of* and *sure about*. Likewise, the noun *fact* turned up as the collocation of both *sure of* and *sure about*. The same can be said of the noun *theory*. It occurred as the collocation of both *sure of* and *sure about*. As exemplified in Table 4, 8 of 36 nouns are the collocations of both *sure of* and *sure about*. The collocations of both *sure of* and *sure about* in the top 22 are the nouns *decision*, *choice*, *fact*, *theory*, *concept*, *understanding*, *outcome*, and *result*. From all of this, it seems clear that *sure of* is 22.22% the same as *sure about* in the top 22 where we obtained 8 common collocations from the ChatGPT. We thus conclude that the first collocations of *sure of* and *sure about* that the ChatGPT recommended are *outcome* and *decision*, respectively. It is vital that they are the

collocations of both *sure of* and *sure about* that the ChatGPT recommended. For the analysis of synonyms and big data, see Kang (2022a, 2022b, 2022c, 2022d, 2023a, 2023b).

5. Conclusion

To sum up, we have compared *sure of* and *sure about* in the TV Corpus, the Corpus of Contemporary American English, and ChatGPT. In section 2, we have argued that *sure about* was more frequently used in America, Canada, Australia, and New Zealand, while *sure of* was more widely used in the UK and Ireland. We have further argued that *sure of* and *sure about* were the most preferred by six countries' celebrities in the 2010s. In section 3, we have maintained that *sure of it* and *sure of victory* are the most preferred ones (12 tokens) for Americans, followed by *sure of things* (10 tokens), and *sure of success* (9 tokens), in that order, while *sure about things* is the most preferable one (8 tokens) for Americans and followed by *sure about people* (*sure about lapse*, *sure about grapes*). We have also contended that only 2 of 42 nouns are the collocations of both *sure of* and *sure about* in the analysis of the COCA (the top 22). That is to say, *sure of* is 4.76% the same as *sure about* in the analysis of the top 22. Finally, we have maintained that 8 of 36 nouns (22.22%) are the collocations of both *sure of* and *sure about* in the data of the ChatGPT (the top 22). It can thus be inferred that *sure of* and *sure about* have a low similarity.

References

- [1] 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.
- [2] 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.
- [3] Kang, N. (2022c). On Speak to and Talk to: A Corpora-based Analysis. *Theory and Practice in Language Studies*, 12(7), 1262-1270.
- [4] 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.

- [5] Kang, N. (2023a). K-Pop in BBC News: A Big Data Analysis. *Advances in Social Sciences Research Journal*, 10(2), 156-169.
- [6] Kang, N. (2023b). K-Dramas in Google: A NetMiner Analysis. *Transaction on Engineering and Computing Sciences*, 11(1), 193-216.