ENGLISH LEARNING IN BIG DATA: A KEYWORD ANALYSIS IN 80 YOUTUBE VIDEOS

Namkil Kang 1

1College of Liberal Arts, Far East University, South Korea

ABSTRACT

The main goal of this paper is to analyze 80 YouTube videos in connection with English Learning. With respect to word length, it is interesting to note that the four-word expression has the highest frequency (159 tokens) and the highest proportion (0.14). A major point to note is that YouTubers think of the so-called word as an essential one for English learning. A further point to note is that topic 10 was the most widely used by YouTubers, followed by topic 2 (topic 7), topic 5, and topic 9, in that order. Talking about the frequency of 80 YouTube videos, the word English was the most widely used one, followed by video, shorts, practice (sentence, word), and learning (vocabulary), in that order. Finally, this paper argues that the words education, video, practice, word, speaking, news, class, study, vocabulary, lesson, sentence, etc. are linked to English and learning. It is concluded that these words linked to English and learning indicate essential prerequisites for English learning.

Keywords: English Learning, Topic, Keyword, Youtube, Big Data, Visualization

1. INTRODUCTION

The main purpose of this paper is to analyze 80 YouTube videos in connection with English learning. We collected 80 YouTube videos (on 12, 10) in terms of the YouTube data collector and analyzed them in terms of the software package NetMiner. First, we provide information on the frequency of word length. Second, we look into the frequency of words related to English learning. Third, we provide 10 topics which were much used in 80 YouTube videos. Each topic is constituted by 5 keywords used frequently in 80 YouTube videos. By analyzing 10 topics and their keywords, one can see what YouTubers think about English learning. Fourth, we consider how many times a particular word appear in 80 YouTube videos. That is to say, we examine the frequency of documents in which a word occurs. Finally, we provide the visualization of 26 words related with English learning. The

How to cite this article (APA): Kang, N. (2022). English Learning in Big Data: A Keyword Analysis in 80 Youtube Videos. International Journal of Research - GRANTHAALAYAH, 10(11), 1-9. doi: 10.29121/granthaalayah.v10.i11.2022.4864
organization of this paper is as follows. In section 3.1, we argue that the four-word expression has the highest frequency (159 tokens) and the highest proportion (0.14). In section 3.2, we further argue that YouTubers think of the so-called word as an indispensable keyword for English learning. In section 3.3, we contend that topic 10 was the most widely used by YouTubers, followed by topic 2 (topic 7), topic 5, and topic 9, in that order. In section 3.4, we maintain that the word English was the most widely used one, followed by video, shorts, practice (sentence, word), and learning (vocabulary), in that order. In section 3.5, we show that the words education, video, practice, word, speaking, news, class, study, vocabulary, lesson, sentence, etc. are linked to English and learning. This in turn suggests that they are all indispensable factors for English learning.

2. METHODS

The main goal of this paper is to analyze 80 YouTube videos collected on 12, 10, 2022 in connection with English learning. We collected them in terms of the YouTube data collector and analyzed them in terms of NetMiner. The main purpose of this paper is to answer the following questions: Can we provide the frequency of word length? Can we provide the frequency of words related with English learning? What are topics which are formed by main keywords? Can we provide information on the frequency of documents? Finally, can we provide the visualization of words related to English learning?

3. RESULTS

3.1. WORD LENGTH

The goal of this section is to provide the frequency of word length. Table 1 shows word length, its frequency, its proportion, and its cumulative proportion:

| Value | Frequency | Proportion | Cumulative Proportion |
|-------|-----------|------------|-----------------------|
| 2.0   | 27        | 0.024      | 0.024                 |
| 3.0   | 71        | 0.063      | 0.086                 |
| 4.0   | 159       | 0.14       | 0.227                 |
| 5.0   | 158       | 0.139      | 0.366                 |
| 6.0   | 133       | 0.117      | 0.483                 |
| 7.0   | 117       | 0.103      | 0.586                 |
| 8.0   | 83        | 0.073      | 0.66                  |
| 9.0   | 72        | 0.063      | 0.723                 |
| 10.0  | 37        | 0.033      | 0.756                 |
| 11.0  | 31        | 0.027      | 0.783                 |
| 12.0  | 20        | 0.018      | 0.801                 |
| 13.0  | 37        | 0.033      | 0.833                 |
| 14.0  | 19        | 0.017      | 0.85                  |
| 15.0  | 23        | 0.02       | 0.87                  |
| 16.0  | 9         | 0.008      | 0.878                 |
| 17.0  | 11        | 0.01       | 0.888                 |
| Value | Count | Probability | Cumulative Probability |
|-------|-------|-------------|------------------------|
| 18.0  | 12    | 0.011       | 0.899                  |
| 19.0  | 15    | 0.013       | 0.912                  |
| 20.0  | 11    | 0.01        | 0.922                  |
| 21.0  | 9     | 0.008       | 0.929                  |
| 22.0  | 11    | 0.01        | 0.939                  |
| 23.0  | 10    | 0.009       | 0.948                  |
| 24.0  | 7     | 0.006       | 0.954                  |
| 25.0  | 10    | 0.009       | 0.963                  |
| 26.0  | 5     | 0.004       | 0.967                  |
| 27.0  | 4     | 0.004       | 0.971                  |
| 28.0  | 3     | 0.003       | 0.974                  |
| 29.0  | 6     | 0.005       | 0.979                  |
| 30.0  | 1     | 0.001       | 0.98                   |
| 31.0  | 1     | 0.001       | 0.981                  |
| 33.0  | 1     | 0.001       | 0.981                  |
| 35.0  | 2     | 0.002       | 0.983                  |
| 36.0  | 1     | 0.001       | 0.984                  |
| 37.0  | 6     | 0.005       | 0.989                  |
| 38.0  | 4     | 0.004       | 0.993                  |
| 41.0  | 1     | 0.001       | 0.994                  |
| 46.0  | 1     | 0.001       | 0.995                  |
| 47.0  | 1     | 0.001       | 0.996                  |
| 49.0  | 1     | 0.001       | 0.996                  |
| 54.0  | 1     | 0.001       | 0.997                  |
| 58.0  | 1     | 0.001       | 0.998                  |
| 65.0  | 1     | 0.001       | 0.999                  |
| 86.0  | 1     | 0.001       | 1                      |
| Total | 1134  | 0.001       | 1                      |

It is interesting to note that the four-word expression has the highest frequency (159 tokens) and the highest proportion. More interestingly, its proportion and their cumulative proportion is 0.14 and 0.227, respectively. It is also interesting to point out that the five-word expression is the second highest (158 tokens). Its proportion is 0.139 and its cumulative proportion is 0.366. It should be pointed out, on the other hand, that the six-word expression ranks third (133 tokens). Additionally, the seven-word expression ranks fourth (117 tokens). Its proportion is 0.103 and its cumulative proportion is 0.586. It is worthwhile noting that the eight-word expression is the fifth highest (83 tokens). Finally, it must be noted that the nine-word expression ranks sixth (72 tokens). Its proportion and its cumulative proportion is 0.063 and 0.723, respectively. We thus conclude that the four-word expression has the highest frequency (159 tokens) and the highest proportion (0.14).
3.2. FREQUENCY OF WORDS RELATED TO ENGLISH LEARNING

In this section, we aim to examine the frequency of words which are closely related to English learning. Table 2 shows the frequency of main words related to English learning:

Table 2

| Words             | Part of Speech | Frequency |
|-------------------|----------------|-----------|
| Channel           | Noun           | 10        |
| Daily             | Adjective      | 41        |
| English           | Noun           | 364       |
| Learn             | Noun           | 34        |
| Learning          | Noun           | 19        |
| Use               | Noun           | 11        |
| Channel           | Noun           | 19        |
| Class             | Noun           | 74        |
| Education         | Noun           | 16        |
| English           | Noun           | 56        |
| Grammar           | Noun           | 17        |
| Language          | Noun           | 21        |
| Learning          | Noun           | 28        |
| Lesson            | Noun           | 11        |
| Meaning           | Noun           | 51        |
| News              | Noun           | 10        |
| Practice          | Noun           | 59        |
| Sentence          | Noun           | 114       |
| Shorts            | Noun           | 38        |
| Speaking          | Noun           | 21        |
| Study             | Noun           | 13        |
| Use               | Noun           | 113       |
| Video             | Noun           | 79        |
| Vocabulary        | Noun           | 71        |
| Word              | Noun           | 196       |
| Youtube shorts    | Noun           | 10        |

As illustrated in Table 2, the word English was the most widely used one (364 tokens). Quite rightly, the word English has the highest frequency (364 tokens) and the highest proportion. It is worthwhile pointing out that word is the second most widely used one (196 tokens). This in turn suggests that YouTubers think of words as the most important keyword for English learning. It is natural that the word sentence ranks third (114 tokens), which implies that YouTubers think of the word sentence as essential. Quite interestingly, YouTubers believe that videos for English learning are also indispensable. Thus, the word video is the fifth highest among
keywords. It should be noted, on the other hand, that the word *vocabulary* is the seventh highest. This in turn suggests that many YouTubers also think of vocabularies as important for English learning. That’s why the words *vocabulary* and *word* rank high. It is worthwhile pointing out that the word *class* ranks sixth (74 tokens). This in turn implies that many YouTubers believe that the so-called *class* is necessary for English learning. Finally, it should be pointed out that the word *practice* is the eighth highest, which in turn suggests that many YouTubers also judge it as necessary. We thus conclude that many YouTubers think of words as the most important for English learning.

### 3.3. TOPICS AND THEIR KEYWORDS

In this section, we provide ten topics and their keywords:

| Table 3 Topic Information |
|----------------------------|
| **1st Keyword** | **2nd Keyword** | **3rd Keyword** | **4th Keyword** | **5th Keyword** |
|-------------------|-----------------|-----------------|-----------------|-----------------|
| **Topic-1**       | Question        | Gk              | Answer          | Fluency         | Exam            |
| **Topic-2**       | Complaylist     | Learning        | Level           | Shorts          | Day             |
| **Topic-3**       | Practice        | English         | Conversation    | Beginner        | Language        |
| **Topic-4**       | English         | Odia            | Use             | Odia            | Class           |
| **Topic-5**       | English         | Short           | Tamil           | Speaking        | Youtube         |
| **Topic-6**       | Education       | Instagram       | India           | Art             | Motivation      |
| **Topic-7**       | Word            | English         | Meaning         | Shorts          | Use             |
| **Topic-8**       | Video           | Learning        | Kid             | Learn           | Skill           |
| **Topic-9**       | Sentence        | Kaise           | Use             | Practice        | Video           |
| **Topic-10**      | English         | Course          | Bengali         | Spoken          | Learn           |

As exemplified in Table 3, there are ten topics that were much used by YouTubers. It is important to note that topic 3 is constituted by 5 keywords such as *practice, English, conversation, beginner, and language*. This in turn implies that many YouTubers judge *practice* as the most important. Note that as can be seen from Table 3, the 1st keyword is *practice*. It is interesting to point out that in topic 1, the 1st keyword is the word *question*. This may indicate that many YouTubers think of it as the most important. Quite interestingly, five keywords such as *video, learning, kid, Learn, and skill* constitute topic 8. In this topic, the 1st keyword is *video*, which suggests that many YouTubers judge it as the most necessary. It is significant to note that as the 1st keyword, the word *English* was the most widely used by YouTubers, whereas the 2nd keyword, *learning and English* were equally the most used ones. It should be pointed out, on the other hand, that as the 3rd keyword, the word *use* was the most used one, whereas the 4th keyword, the word *shorts* was the most used one.

Now let us turn to the frequency of documents:

| Table 4 Frequency of Documents |
|-------------------------------|
| **# of documents** |
| **Topic-1**  | 5 |
| **Topic-2**  | 11 |
It is important to note that topic 10 was the most widely used one. More specifically, it occurred in 20 YouTube videos. As observed earlier, topic 10 is constituted by the keywords *English*, *course*, *Bengali*, *Spoken*, and *learn*. It is worth pointing out that topic 2 and topic 7 were the second most frequently used ones. They appeared in 11 YouTube videos. Topic 2 is formed by the keywords such as *complaylist*, *learning*, *level*, *shorts*, and *day*, whereas topic 7 is constituted by *word*, *English*, *meaning*, *shorts*, and *use*. It is noteworthy that topic 5 was the third most widely used one. That is to say, it occurred in 9 YouTube videos. Finally, topic 9 occurred in 8 YouTube videos. It ranks fourth among 10 topics. Note that topic 9 include the keywords *sentence*, *kaise*, *use*, *practice*, and *video*. It can thus be concluded that topic 10 was the most widely used one, followed by topic 2 (topic 7), topic 5, and topic 9, in that order.

### 3.4. DEGREE

The goal of this section is to provide information on degree (the frequency of videos):

| Number | Word       | Frequency |
|--------|------------|-----------|
| 1      | English    | 65        |
| 2      | Video      | 29        |
| 3      | Shorts     | 26        |
| 4      | Practice   | 22        |
| 5      | Sentence   | 22        |
| 6      | Word       | 22        |
| 7      | Learning   | 21        |
| 8      | Vocabulary | 21        |
| 9      | English    | 19        |
| 10     | Use        | 18        |
| 11     | Learn      | 17        |
| 12     | Speaking   | 16        |
| 13     | Daily      | 15        |
| 14     | Course     | 15        |
| 15     | Meaning    | 14        |
Table 5 indicates the frequency of videos in which a particular word appear. It is significant to note that the word English appeared 65 YouTube videos. This in turn indicates that it was the most widely used one in 65 YouTube videos. It is interesting to note, on the other hand, that the word video was the second most widely used one. Quite interestingly, it appeared in 29 YouTube videos. This in turn indicates that many YouTubers believe that videos are an effective way to learn English. It is
worth pointing out that the word practice occurred in 22 YouTube videos, which in turn indicates that many YouTubers judge it as essential. It must be pointed out, on the other hand, that the word sentence was the fourth most widely used one. Quite interestingly, it appeared in 22 YouTube videos. Likewise, word occurred in 22 YouTube videos and was the fourth most frequently used one. This in turn suggests that the so-called word is considered as essential by YouTubers. The word vocabulary is more or less the same as word. It occurred in 21 YouTube videos and was the seventh most widely used one. To sum up, the word English was the most widely used one, followed by video, shorts, practice (sentence, word), and learning (vocabulary), in that order. It is worthwhile noting that the word Practice occurred in 10 YouTube videos, that the word conversation occurred in 9 YouTube videos, and that the word news appeared in seven YouTube videos. From all of this, it is evident that they are all necessary for English learning.

### 3.5. VISUALIZATION OF WORDS

The main goal of this section is to provide the visualization of which words are closely related to English learning. Figure 1 shows the visualization of words related to English learning:

![Visualization of English Learning](image1)

As exemplified in Figure 1, 26 words are closely related to one another. Words linked to English and learning are education, video, practice, word, speaking, news, class, study, vocabulary, lesson, sentence, etc. This in turn implies that they are closely related to English learning and important factors for it. For the visualization of synonyms, see Kang (2022a), Kang (2022b), Kang (2022c), Kang (2022d). To sum up, Figure 1 provides us with the picture of which factors are closely related to English learning.
4. CONCLUSION

To sum up, we have analyzed 80 YouTube videos in connection with English learning. In section 3.1, we have shown that the four-word expression has the highest frequency (159 tokens) and the highest proportion (0.14). In section 3.2, we have argued that YouTubers think of the so-called word as the most important keyword for English learning. In section 3.3, we have further argued that topic 10 was the most widely used by YouTubers, followed by topic 2 (topic 7), topic 5, and topic 9, in that order. In section 3.4, we have maintained that the word English was the most widely used one, followed by video, shorts, practice (sentence, word), and learning (vocabulary), in that order. In section 3.5, we have shown that the words education, video, practice, word, speaking, news, class, study, vocabulary, lesson, sentence, etc. are linked to English and learning. This in turn implies that they are indispensable factors for English learning.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Kang, N. (2022a). A Comparative Analysis of Search for and Look for in Four Corpora. Advances in Social Sciences Research Journal 9 (3), 168-178. https://doi.org/10.14738/assrj.93.11980.

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. https://doi.org/10.17507/tpls.1205.01.

Kang, N. (2022c). On Speak to and Talk to: A Corpora-based Analysis. Theory and Practice in Language Studies 12 (7), 1262-1270. https://doi.org/10.17507/tpls.1207.03.

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.