Identification of Synonyms (CANCEL ABOLISH ELIMINATE) Based on the COCA Corpus

Yuting Huang

Fujian Normal University, Fuzhou 350007, Fujian, China

Abstract: A corpus provides the most accurate and understandable results when comparing synonyms. English learners struggle with synonyms. Based on word frequencies, collocation, and semantic rhyme, a corpus can define synonyms. Using COCA corpus data, this study compares decide with determine. This study offers a new viewpoint on synonyms and references for vocabulary and second language development. This study uses word frequency, collocation, and semantic rhyme distributions to distinguish these two terms that are commonly used interchangeably. Data suggests that verb collocation and register distinguish the two synonyms. The research is one of numerous that shows how corpora may help differentiate synonyms. Comparing data and utilizing frequency as a statistic helps clarify the synonyms' subtle differences. COCA is the world's largest free online balanced English corpus. It helps linguists and language learners compare words, phrases, and sentence patterns and understand their frequency. The COCA corpus of Contemporary American English gathers data from spoken language, literature, newspapers, academic study, and other sources of reliable information. This article examines synonyms from three perspectives: domain, collocation, and meaning.

Keywords: Domain; Synonyms; Corpus

1. Introduction

When comparing a group of synonyms, using a corpus for analysis is the method that yields the most accurate and comprehensible results. Distinguishing between synonyms is a continuous issue for students of the English language. The technique of classifying synonyms using a corpus can provide a comprehensive definition of synonyms based on the distribution of word frequencies, collocation, and semantic rhyme. This research investigates the differences between two English synonyms, decide and determine, by using data from the COCA corpus. The purpose of this investigation is to present a fresh perspective on the differentiation of synonyms and references for vocabulary education and second language learning. This research makes a distinction between these two words that are often used interchangeably by analyzing the word frequency distribution, the collocation distribution, and the semantic rhyme distribution. According to the data, the fundamental distinction between the two synonyms appears to lie in the verb collocation as well as the register. The study is also one of several instances that demonstrate how corpora may be a useful technique for distinguishing between synonyms. The nuanced distinctions between the synonyms can be better understood by comparing the data and using the frequency of usage as a metric. The Corpus of Contemporary American English (COCA) is the biggest free online corpus of English that is considered to be balanced in the world. It is intended for people who are learning a language as well as linguists so that they may compare information about words, phrases, and sentence structures and comprehend the frequency with which they occur. The COCA corpus of Current American English collects data from a wide variety of sources, including spoken language, fiction, newspapers, academic research, and other areas providing vast amounts of trustworthy information. In this article, I will examine synonyms from three different points of view: the domain, the collocation, and the meaning.

2. Strategy for Research

2.1 The Goal of the Research

This study looked at the use patterns, collocations, and semantic rhymes of the words decide and dest-mine in the COCA corpus,

which is a database of native speakers. The purpose of this research is to discover the similarities and differences between these lexical items in distribution, collocation, semantic rhyme, and other aspects. [1]This will demonstrate that the corpus is one of the most powerful tools for distinguishing synonyms, provides an effective synonym discrimination method for learners of second languages, and provides new insight into the process of teaching English vocabulary.

2.2 Methods and Tools

Throughout the course of this research, the Corpus of Contemporary American English was drawn upon (COCA). It is the only major corpus that has 20 million words for each year from 1990 to 2015 and is the largest online corpus that is free to access. The American English Corpus was the one responsible for its creation.

This corpus is significant not just for academics, but also for teachers of English and students of the language.

Throughout this experiment, the online retrieval tool that COCA provides was utilized. Distribution, collocation, and semantic rhyme are the three categories of data that need to be gathered in order to differentiate between the two words that are synonymous with each other.

To get started, go to "Chart Options" in COCA and search for each word individually. After that, display the frequency of each term in different parts of the chart based on where it appears. Collocation data collection is the second phase in the process. The Collocates option in COCA is used to search each word in order to determine which terms are most commonly matched. In the last step of this process, we select relevant collocation phrases in order to investigate the semantic rhyme of each synonym.

3. Analysis of the difference

3.1 Analysis of the difference in the domain of the occurrence of cancel abolish eliminate

While discussing a set of synonyms, it is common to hear each of them employed in a variety of settings and fields. Because of inconsistencies in the place of language usage or in the manner of language expression, words that have the same meaning might contain certain linguistic tendencies of their own. For instance, among a group of synonyms, some are more likely to be used in spoken language, whilst others are more likely to be used in written language on the internet. Certain terms could be considered particularly formal or professional, and they might be found more commonly in academic or literary contexts. [2]Some terms could be considered particularly formal or professional, and they might be found more often in literary works or academic works. The most common domains may be recognized by analyzing the frequency with which the various words that comprise the synonyms are used.

The frequency of CANCEL, ABOLISH and ELIMINATE in the corpus is 22480, 5980 and 42246 respectively. It can be seen that ELIMINATE is the most frequently used word, covering 5.9 percent of the articles. The word ABOLISH is the least frequently used, covering 0.9 percent of the articles. CANCEL is in between. Comparing the tree chart, we can see that CANCEL is mostly found in TV news and blogs, ABOLISH is mostly found in academic and web pages, and ELIMINATE is mostly found in magazines and academic. It is important to note that ABOLISH and ELIMINATE appear very infrequently in fiction and television. It can be seen that ABOLISH and ELIMINATE appear more frequently in academics than CANCEL, indicating that the vocabulary is more specialized.

3.2 Difference between the collocations of cancel abolish eliminate

We can discriminate and analyze synonyms by starting with different words collocated with different words.

Using the corpus collocates function, the collocation search of the three words was conducted, and the following three sets of data were obtained.

COLLOCATES **CANCEL** **VERB**

+ NOUN		NEW WORD	?
852	6.20	flight	
645	3.98	show	
463	4.80	trip	
435	3.37	plan	
378	4.19	meeting	
329	3.70	order	
328	7.59	subscription	
301	4.52	contract	
296	2.90	class	
296	3.07	event	
285	5.97	appointment	

COLLOCATES **ABOLISH** **VERB**

+ NOUN		NEW WORD	?
397	9.04	slavery	
219	4.29	death	
199	7.15	penalty	
181	3.40	law	
180	2.55	state	
180	4.10	tax	
150	2.84	government	
127	2.60	system	
107	3.62	department	
107	5.63	amendment	
96	4.67	ice	

COLLOCATES **ELIMINATE** **VERB**

+ NOUN		NEW WORD	
1410	4.56	need	
1064	3.85	tax	
826	2.49	job	
621	2.28	program	
431	4.19	possibility	
341	2.79	risk	
336	2.09	plan	
330	3.50	threat	
311	2.37	position	
273	2.27	cost	
268	2.12	effort	
243	4.25	waste	
243	6.50	deduction	
228	3.53	competition	
223	4.18	deficit	
221	4.56	barrier	
213	3.10	debt	

The most frequent collocation with CANCEL was flight, which appeared 852 times. The next words are show, trip, plan, and meeting, which are all temporary activity plans that are variable. For example, a temporary appointment or order, a variable contract, etc. The most frequent collocation with ABOLISH was slavery, death and penalty. These words are all about the legal system and customs. These are relatively important and fixed activities, which are relatively unchangeable in the grand scheme of things, such as slavery or tax laws. The most frequent words with ELIMINATE are need, tax, and job, where need and tax appear most frequently and with almost equal frequency. They mostly refer to things that are needed in life, such as jobs, desires, etc. These words are mostly abstract things, such as effort or threat.

4. Conclusion

CANCEL is used for the cancelation of a prearranged activity (trip, meeting, plan), which is highly flexible, such as "The meeting has been cancelled because of the flu". ABOLISH is used for the abolition of a law, custom, or institution, which is used in a more formal context, such as "The government abolished the tax on alcohol". ELIMINATE is used for the elimination of something that already exists but is not needed now. Mostly used for abstract things, such as "The losing team was eliminated from further competition./ The doctor helped him eliminate toxins from the intestine." Thus, it is important to use different words in different genres or in different texts according to the situation so that there is no ambiguity. Use synonyms flexibly in order to express ideas more accurately and appropriately.

References:

[1] Biber D. Longman Grammar of Spoken and Written English [M]. Pearson education Limited, 1999.

[2] Kennedy G. An Introduction to Corpus Linguistics[M]. Bei- jing: Foreign Language Teaching and Research Press, 2000: 23-59.