

A Corpus-based Research on Synonym Discrimination: Exemplified by “Finish”, “Complete” and “Accomplish”

Pengbiao Ding

Guangdong University of Foreign Studies, Guangzhou 510420, China

Abstract: The emergence of corpus has provided a new perspective for the identification and distinguishing of synonyms, and the linguistic facts contained in the corpus can be relied on to provide a strong support for lexical analysis. Based on *the Corpus of Contemporary American English (COCA)* and its vocabulary retrieval tool *JustTheWord*, this paper adopts a research method combining quantitative and qualitative analysis, and discusses the synonyms “finish”, “complete” and “accomplish” as the research objects from the aspects of registers, word frequency, colligation, collocates, and semantic preference, and explains the significance of this research for vocabulary teaching at the end of the article.

Keywords: Synonyms; Word Discrimination; Corpus Analysis; *COCA*.

1. Introduction

Corpus research was introduced into China in the 1970s and gradually flourished in the mid-1990s. Corpus has become an important means of empirical research in linguistics. The number of corpus-based researches is increasing year by year, and the research fields as well as the methods are constantly expanding, enriched and developed.

Non-native speakers used to make use of English dictionaries as a learning source, but now corpus-based methods can also be applied into language teaching and learning. The definitions of “finish”, “complete” and “accomplish” in the online *Oxford Learner’s Dictionary* (only the most typical senses are selected, that is, the first sense of each word’s definition) are as follows: “1. **finish**: to stop doing something or making something because it is complete. 2. **complete**: to finish making or doing something. 3. **accomplish**: to succeed in doing or completing something.

From the above definitions, it can be seen that “accomplish” contains the meaning of “success”, but in addition, these three verbs all mean “to make something done”, and even such an authoritative dictionary cannot avoid using “complete” to explain “finish” and “accomplish”, and in turn, using “finish” to explain “complete”, which shows the inevitability of circularity in word defining. Therefore, the definition in the dictionary, as is shown above, cannot help users to make a clear and transparent distinction between the three target words. Based on the *Corpus of Contemporary American English (COCA)* and its vocabulary retrieval tool *JustTheWord*, the author adopts a research method combining quantitative and qualitative analysis, and discusses the synonyms “finish”, “complete” and “accomplish” as the research objects. The following questions are discussed and answered:

- 1) What do the target words have in common?
- 2) What are the differences between the target words?
- 3) What kind of enlightenment does this vocabulary research have on vocabulary teaching?

Based on the above questions, this paper conducts an empirical research on synonym discrimination from the aspects of register, word frequency, colligation, collocates, and semantic preference, and explains the significance of this research for vocabulary teaching at the end of the article.

2. Literature Review

Corpus linguistics can lend support to prove the syntagmatic relationship of language, which believes that the meaning of a word exists in its companion relationships, so collocation has become one of the important topics in corpus vocabulary research. In China, especially since 2000, people have paid more and more attention to collocation research, and many researches are combined with foreign language teaching, which has now become the most concentrated topic in vocabulary research.

Francis and H. Kucera began to design and build the first large-scale electronic corpus in the early 1960s, which was later the famous *Brown corpus (BROWN)*. So far, the study of foreign corpus linguistics has experienced 50 years of development.

In terms of the fruits in this field, Rundell (1992) specifically analyzed the frequency difference between “began” and “started” in a spoken corpus and a written corpus. Biber et al. (1999), based on Rundell’s word segmentation, conducted a statistical study on the different forms of the complements of this group of synonyms, and revealed the semantic differences between them. Change et al. (2000) made use of a corpus to analyze the alternation of Mandarin Chinese emotional verbs between different semantic fields. With many previous studies as such, there are many other scholars who have also conducted a large number of researches on synonyms based on corpus. For example, Kennedy (1991) made a delicate analysis of the grammatical and semantic functions of synonymous prepositions. Zhang (2006) once analyzed the linguistic differences between “occur” “happen” and the Chinese phrase “发生(fa-sheng)” from the aspects of word frequency, word distribution and collocation relationships; Wang and Ren (2009) demonstrated the objectivity and feasibility of using the method of corpus linguistics to study the register difference and colligation difference of synonyms, which was a powerful supplement to the traditional lexicological research; Zhang and Ren (2011) analyzed “found” and “establish” based on *Word Sketch Engine* and *British National Corpus (BNC)*.

Therefore, the research results of corpus-based vocabulary can be said to be fruitful. Thus when it comes to synonym discrimination and vocabulary learning, the corpus-based

approach is undoubtedly applicable and of great efficacy.

3. Research Methods

This paper uses *the Corpus of Contemporary English of America (COCA)* and its retrieval tool *JustTheWord* to analyze the target words.

COCA is a free online American English corpus developed by Professor Mark Davies of Brigham Young University, and it is also the largest balanced English corpus in the world. *COCA* contains five types of sub-corpora, which cover spoken language, novels, magazine, newspaper and academic journal, and these five types of sub-corpora are evenly distributed. Users can search in the entire database or limit the query to any one of these sub-databases, or a combination of several sub-databases.

JustTheWord is a new tool to help users write and choose words in English. It can provide the common mistakes people from different native language backgrounds use in English. It can also describe the collocation of the target word in detail, that is, what kinds of collocation the target word has in modern English, and which words the target word often matches with in terms of a specific type of collocation.

First, the author retrieved the node words in *COCA* (that is, the sample synonyms “finish”, “complete” and “accomplish”), obtained the observed frequency for each of the node words and their distribution in different sub-corpora, and then analyzed the sample synonyms’ similarities and differences in these aspects. Second, with the help of *JustTheWord*, the author derived the representative colligations when the sample synonyms “finish”, “complete” and “accomplish” are used as verbs. Then the corresponding frequency of each colligation in *COCA* can be accessed separately. Later, the semantic preference of the sample synonyms was analyzed by observing the retrieved lines extracted from the corpus. Finally, the author discussed and analyzed the similarities and differences of the sample synonyms in the four linguistic features such as register and word frequency, colligation, collocation and semantic preference.

4. Discussion and Analysis

(1) Analysis on word frequency and registers

In *COCA*’s consultation interface, the author typed in the search box the three node words “finish”, “complete” and “accomplish” respectively, and obtained their register distribution in *COCA*, as shown in Figure 1:



Figure 1. Register distribution of “finish” in *COCA*



Figure 2. Register distribution of “complete” in *COCA*



Figure 3. Register distribution of “accomplish” in *COCA*

The study of frequency distribution based on different registers can give learners a macroscopic guidance in distinguishing synonyms. This method helps learners distinguish and analyze synonyms to a certain extent, and cultivates learners’ overall understanding of the target words.

The word frequencies of the three node words “finish”, “complete” and “accomplish” in *COCA* are 111,094 times, 71,752 times and 33,992 times each, and these words appear in 13.6%, 9.2% and 5.3% of the ranged texts respectively. It can be seen from the frequency and the percentage that people use “finish” more frequently in daily life. The frequency of this word in *COCA* is more than the sum of the other two target words, and it can be applied in a wider textual scope.

However, Figure 1-3 show that “finish” has the lowest occurrence in academic texts. Instead, it has a relatively high frequency of use in texts of other registers. Its frequency in TV, novels, and news texts is particularly prominent, and the frequency in blogs, the Internet, magazines, and spoken language are not low, either. Therefore, it can be known that when learners use the word “finish”, they should generally avoid using it in academic texts. In contrast to “finish”, when people use “complete”, they apply it most frequently to academic texts, well beyond the word’s usage in other registers. The distribution of the word in blogs, Internet, magazines, and newspapers are rather even, and this word is less used in TV, spoken language, and novels. In a sense, “complete” presents a complementary distribution to “finish”. Finally, the distribution of “accomplish” is a bit more complicated. The most commonly used registers of the word are academic texts, blogs, and websites. It is also used frequently in spoken language, magazines, and news, too. However, we can see its lowest frequency when it is used in TV and novel texts. Considering that the overall frequency of this word is the lowest among the three target words, it can be said that this word is basically not used in TV and novel texts, but its distribution in other text types (such as academic and non-academic texts, spoken and non-spoken) is quite balanced.

(2) Analysis on colligations

With the help of *JustTheWord*, the author obtained the representative colligations of the sample synonyms when they are used as verbs. Then the corresponding frequencies of each colligation could be retrieved separately from *COCA*. The following table summarizes the colligations and their corresponding frequencies. The asterisk (*) in the table indicates the inclusion of all inflections of the verb.

The co-occurrence relationship between words has both grammatical hints and semantic hints, and that relationship between parts of speech and grammatical categories is colligation. There are usually multiple colligation types for each one of the sample synonyms. Table 1 only lists the commonly used colligation types. It can be seen that “ADV+V*” and “V*+PREP” are the common and most frequent forms of the sample synonyms’ colligation; In addition, a word followed by a noun as an object is also a common form of the colligation among the three. It can also

be seen from the above table that “complete” has the most types of colligation (7), followed by “finish” (6), and the least is “accomplish” (3). From the point of view of the colligation types, the colligation of “finish” includes all that of “accomplish”, and “complete” includes “finish”. In terms of the grammatical co-occurrence relationship with words of other parts of speech, “complete” is the most widely used one, followed by “finish”, while “accomplish” is relatively limited.

Table 1. Colligations and their frequency in COCA

Word	Colligations & frequencies
finish	FINISH*+N(6137); ADV+FINISH*(13126); FINISH*+ADV(8551); FINISH*+PREP(13075); V and FINISH*(656); FINISH* and V (496).
complete	COMPLETE*+N(5536); N+COMPLETE*(10831); ADV+COMPLETE*(7600); COMPLETE* + ADV (1984); COMPLETE*+PREP(7506); V and COMPLETE*(350); COMPLETE* and V (648).
accomplish	ACCOMPLISH*+N(739); ADV+ACCOMPLISH*(2509); ACCOMPLISH*+PREP (6382).

Colligations can not only provide the basis for target words’ discrimination from the perspective of grammatical co-occurrence, but also serve as the support for further research on collocation features, semantic preference of sample synonyms.

(3) Analysis on collocation features and semantic preference

Here the author mainly studied the colligation “V*+N”, that is, the nominal collocations standing after the sample synonyms, and uses the mutual information value (MI) to measure the relation between collocations and the node words’ collocational strength. The mutual information value is used to measure the collocational strength between words, and it will not be affected by the size of the corpus. When $MI > 3$, it indicates that there is a strong collocation between the words, so the noun can be regarded as a significant collocation word of the node word. This paper analyzes 100 randomly retrieved lines of each of the three target words, and offers a summary on that.

After searching, the author finds that the typical collocation words of “finish” mainly include: (1) words related to study, such as the terms expressing meanings of education, majors, learning activities and so on. For example, “school”, “college”, “university”, “law”, “graduate”, “homework”, “Harvard”, “writing”, “reading”, etc.; (2) words related to human’s jobs, such as “work”, “projects”, “building”, “painting”, “construction”, “tasks”, etc.; (3) diet-related words expressing the meals of a day, names of the food that are cooked, etc., such as “breakfast”, “lunch”, “supper”, “eating”, “baking”, “drinking”, etc.; (4) words related to competitions, often indicating rankings, types of competition, etc., for example, “games”, “second”, “shooting”, “training, plays”, “races”, etc.; (5) words indicates other daily activities, mostly relating to housework, dates, etc. The examples are “cleaning”, “chemotherapy”, “Sunday”, “recording”, “packing”, etc. The picture below shows the randomly selected high-frequency nominal collocations of “finish”:

HELP	ALL FORMS (SAMPLE: 100 200 500)	FREQ	ALL	%	MI
1	FINISH SCHOOL	327	527285	0.06	6.65
2	FINISHED COLLEGE	212	199386	0.11	7.43
3	FINISH COLLEGE	207	199386	0.10	7.39
4	FINISHED EATING	190	59792	0.32	9.01
5	FINISHED SCHOOL	187	527285	0.04	5.84
6	FINISH SECOND	164	336745	0.05	6.30
7	FINISHED WORK	103	813581	0.01	4.36
8	FINISHED READING	99	157861	0.06	6.66
9	FINISH WORK	79	813581	0.01	3.97
10	FINISHED DINNER	77	85097	0.09	7.19

Figure 4. Nominal collocations of “finish” with high frequency

The typical collocation words of “complete” mainly include: (1) a large number of words related to study, mostly expressing education, courses, academic tasks, etc., such as “college”, “studies”, “assignments”, “courses”, “homework”, “degrees”, etc.; (2) words related to scientific research: “surveys”, “research”, etc.; (3) words denoting certain written documents: “questionnaires”, “forms”, etc.; (4) abstract nouns representing stages, periods, etc.: “level”, “phase”, “stage”, etc.; (5) nouns denoting certain activities in the workplace: “probation”, “interviews”, “negotiations”, “work”, “missions”, “projects”, “constructions”, etc. The picture below shows the randomly selected high-frequency nominal collocations of “complete”:

HELP	ALL FORMS (SAMPLE: 100 200 500)	FREQ	ALL	%	MI
1	COMPLETED QUESTIONNAIRES	94	3885	2.42	12.08
2	COMPLETE TASKS	87	23217	0.37	9.39
3	COMPLETE SET	73	326717	0.02	5.32
4	COMPLETE COLLEGE	66	199386	0.03	5.89
5	COMPLETED WORK	66	813581	0.01	3.86
6	COMPLETED COLLEGE	56	199386	0.03	5.65
7	COMPLETE ASSIGNMENTS	52	8593	0.61	10.08
8	COMPLETE WORK	52	813581	0.01	3.52
9	COMPLETING TASKS	46	23217	0.20	8.47
10	COMPLETING ASSIGNMENTS	42	8593	0.49	9.77

Figure 5. Nominal collocations of “complete” with high frequency

HELP	ALL FORMS (SAMPLE: 100 200 500)	FREQ	ALL	%	MI
1	ACCOMPLISH THINGS	138	631346	0.02	8.20
2	ACCOMPLISH TASKS	50	23217	0.22	11.50
3	ACCOMPLISH GOALS	34	54981	0.06	9.70
4	ACCOMPLISHING THINGS	32	631346	0.01	6.09
5	ACCOMPLISHED THINGS	19	631346	0.00	5.34
6	ACCOMPLISHING TASKS	17	23217	0.07	9.94
7	ACCOMPLISHING GOALS	16	54981	0.03	8.61
8	ACCOMPLISH CHANGE	9	327055	0.00	5.21
9	ACCOMPLISH GOD	8	405503	0.00	4.73
10	ACCOMPLISH OBJECTIVES	6	14775	0.04	9.09

Figure 6. Nominal collocations of “accomplish” with high frequency

The typical collocations of “accomplish” mainly include: (1) non-referential nouns representing an abstract concept: “things”, “part”, “work”, “something”, etc.; (2) abstract nouns related to human imagination or other mental activities: “wonders”, “peace”, “miracles”, “virtue”, “purposes”, etc.; (3) words indicating some abstract actions: “salvation”, “innovation”, “implantation”, “standardization”, “echolocation”, etc.; (4) nouns expressing the attributes of things: “interoperability”, “convertibility”, “macro-scale”, etc.; (5) a small number of specific referential nouns: “tax”, “policy”, “tasks”, “activities”, etc. Figure 6. shows the randomly selected high-frequency nominal collocations of “accomplish”.

From the above discussion, it can be found that the sample synonyms “finish”, “complete” and “accomplish” have

different semantic preferences. “finish” can be used to express: a. the completion of study, assignments, tasks, etc.; b. the completion a specific task, such as “finish cleaning/recording/painting”; c. the fact that the meal is finished or the cooking is over; d. the indication of the end of an event or the ranking achieved in the event.

The word “complete” also shows its unique semantic preference: a. the completion of study, courses, assignments, etc.; b. the completion of a scientific study; c. the act of filling up a written document; d. the experience of a certain stage; e. the completion of a formal task.

The semantic preferences of “accomplish” are as follows: a. the completion of abstract concepts or behaviors; b. the fact that things have reached a certain quality; c. a small number of specific nouns, such as “accomplish tax/policy”.

There are always many special cases in language learning. Similarly, in this article, it cannot be ruled out that some nouns are shared by the sample synonyms, such as “finish tasks”, “complete tasks” and “accomplish tasks”. We cannot arbitrarily decide whether one or another expression is wrong. This also shows the similarity among the three target words, that is, they can all be used to express “to make a specific and concrete task done”.

5. Conclusion

Based on *COCA* and *JustTheWord*, this paper distinguishes the synonyms “finish”, “complete” and “accomplish” from each other from the aspects of register and word frequency, colligation, collocation features, and semantic preference. This empirical study of lexical discrimination, through the analysis of the linguistic facts in the corpus, shows that the three target words all have the meaning of “to make something done”, and can be followed by objects indicating work, such as “tasks” and “work”, etc. However, the differences among them are also salient.

First of all, in terms of register and word frequency, “complete” and “accomplish” show a significant tendency for written language, and “complete” presents the most prominent academic feature; The most colloquial one is “finish”, and it also appears frequently in registers like TV, novels, and news; In addition, “accomplish” shows a combination of oral English features and academic features, with the latter being its most prominent feature. Secondly, in terms of colligation types, “ADV+V*” and “V*+PREP” are the common and frequent colligation forms of sample synonyms; Also, being followed by a noun as an object is also

a mutual colligation of the three. Besides, from the perspective of colligation types, “finish” (6) includes all the colligation forms of “accomplish” (3), and is included in “complete” (7). As for the collocation features and semantic preference of the sample synonyms, the author finds that there are significant differences among the three. When the synonyms “finish”, “complete” and “accomplish” are followed by an object, the possibility of the object being an abstract noun rises one by one, while that of the object being a concrete noun becomes less, as detailed above.

Traditional vocabulary teaching only emphasizes the teaching of basic pronunciation, spelling, parts of speech and meaning of a word, but is not carried out from the perspective of teaching words' usage in depth. Vocabulary teaching should also include the teaching of vocabulary grammar, meaning, and registers. Teachers can provide English learners with some knowledge of registers, colligation, and semantic preference to promote their deep understanding of vocabulary. For example, in order to distinguish the target words in this article, teachers can first present examples of their collocations with objects, and effectively distinguish them by analyzing the number of abstract nouns as objects.

References

- [1] Biber et al. (1999). *Longman Grammar of Spoken and Written English* [M]. Pearson education Limited.
- [2] Change, Li-li, Keh-jiann Chen, & Huang Chu-Ren. (2000). Alternation across semantic fields: A study on Mandarin verbs of emotion. *International Journal of Computational Linguistics and Chinese Language Processing*, 61-80.
- [3] Kennedy, G. (1991). Between and through: The company they keep and the functions they serve. In J. Startvik (ed.). *English Corpus Linguistics*. Singapore: Longman.
- [4] Rundell, Michael, & P. Stocks. (1992). The corpus revolution. *English Today*, 8(2), 9-14.
- [5] Wang Xuan, Ren Peihong. (2009). A Corpus-based Differentiation of English Synonyms. *Journal of Henan Polytechnic University*, 1, 128-132.
- [6] Zhang Jidong, Liu Ping. (2006). A Corpus-based Study of the Differences between the Three Synonyms: happen, occur & “fasheng”. *Foreign Languages Research*, 5, 19-22.
- [7] Zhang Xiaoning, Ren Peihong. (2011). Differentiating English Synonyms Based on the Online Sketch Engine--Taking “Found” and “Establish” as Examples. *Journal of Jilin Provincial Institute of Education*, 2, 84.