Comparative analysis of English synonyms based on COCA corpus: Taking "value, worth, merit" as an example

Qin Meiyan[a], Zhang Xin

Xi'an Shiyou University, Xi'an, China
[a]qinmeiyan71@gmail.com

Abstract: *Vocabulary has always been the most important part of foreign language learning, and the analysis of near-synonyms has always been the most difficult part of English learning. In this article, the author uses the COCA corpus to analyze the synonyms of value, worth, and merit, mainly from the aspects of domain, semantic rhyme, and collocation. The differences in each of the three terms are identified and conclusions are drawn to inform the use of the three terms.*

Keywords: *COCA corpus; proxemics; domain; semantic rhyme; collocation*

1. Introduction

In the process of learning and studying English near-synonyms, the ability to accurately distinguish the differences between words and achieve the best expression in the appropriate context is the value of near-synonym identification. The corpus facilitates the analysis of proxemics by collecting real corpus situations, which makes it easier for learners to identify the different collocations of different words and their usage in different contexts more clearly. Generally speaking, people analyze near-synonyms only by consulting dictionaries and by memorizing the sense of language, which has a lot of shortcomings. A large number of articles on corpus research have proved that the emergence of corpora has provided a new choice for the analysis of English near-synonyms, which greatly improves the accurate usage of English near-synonyms. With the help of the COCA corpus, the author intends to analyze the group of synonyms "value, worth, merit".

2. Research methods

2.1 Research questions

This study focuses on the domain distribution, semantic rhyme, and collocation of the synonym value, worth, and merit in the corpus, and the specific problems are as follows:

(1) What are the similarities and differences between value, worth, and merit in the distribution of registers?

(2) What are the differences in the semantic rhyme of value, worth, and merit?

(3) What are the differences in the collocation of value, worth, and merit?

2.2 Tools

This study mainly uses the Chart and Collocates in COCA to analyze the differences in the domain, collocation, and semantic rhyme of the three synonyms value, worth, and merit. First of all, the author makes use of the functions in the corpus to search these three terms respectively, to obtain the basic information of the transverse register distribution and longitudinal diachronic distribution of the word, including the distribution frequency (Table 1) and time distribution (Table 2) of each word in each domain. Secondly, the author uses the Collocate to obtain the collocation information of each word and studies the significant collocation of synonyms, which can help learners to be more clear about the specific context of each word. The author sets the span to -5/+5, that is, the filter result is set to relevance, and the minimum frequency value is set to 20. The strength of collocation between words is usually expressed

by MI, and larger values indicate that words have a greater impact on their lexical environment. In corpus studies, words with MI values equal to or greater are usually defined as significant collocations. This article combines frequency distribution and MI value to study the typical collocation of these three synonyms and selects the top 10 collocation words with a value of ≥3 (see Table 3)

3. Results and discussion

3.1 A Comparative Study of the Domain Distribution of Value, Worth, and Merit in the COCA Corpus

Domain refers to a kind of language variants produced by people in the actual language activities in the need of communication, or due to their different occupations and interests, as well as due to the different situations, objects, places, and topics of their discourse, which are manifested in different styles of speech and tone of terminology, etc.[1] Therefore, the study of the domain is essential in the comparative study of near-synonyms, which plays a very important role in guiding the identification of near-synonyms. There are significant differences in usage in different registers, and if we use the corpus to compare and analyze the registers, we can clearly and accurately grasp the differences in the usage of the related synonyms.

By searching the chart of the COCA corpus, the word frequency distribution of value, worth, and merit in eight different linguistic domains can be obtained (see Table 1). Frequency is one of the most important concepts in corpus linguistics, which refers to the number of occurrences of a word or phrase per million words in a corpus, also called standardized frequency. Since there are subtle differences between the near-synonyms, they often show different distributional characteristics in different linguistic domains, so it is necessary to count their frequency differences in different linguistic domains for English learners to learn near-synonyms better.

Table 1 Word frequency distribution of value, worth, and merit in different domains

Table 1		BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
value	Freq	22441	26499	2990	5969	3191	17539	13737	33240
	Per mil	174.48	174.48	23.35	47.32	26.97	139.1	112.84	277.48
worth	Freq	20524	17103	14042	10834	9282	17221	15442	7007
	Per mil	159.98	137.65	109.64	85.89	78.45	136.58	126.84	58.49
merit	Freq	1692	1791	346	658	498	1261	1312	2090
	Per mil	13.16	14.41	2.7	5.22	4.21	10	10.78	17.45

From Table 1, we can see that the word "value" appears most frequently in academic journals, and less frequently in TV and movies as well as novels, which means that value is mainly used in text writing, and its frequency in blogs and the Internet is also higher, and value has a higher usage in both virtual and real life. The word "worth" has the highest frequency of usage in spoken language, which is nearly 17 times higher than that of merit in spoken language, and nearly 2 times higher than that of value. The word "worth" is the most frequently used in spoken language, nearly 17 times more frequently than merit, and nearly 2 times more frequently than value, which shows that the word "worth" is dominant in our daily life. Value and worth are more frequently used than merit in each domain, but the word worth is more frequently used than the other two words in movies and TV, novels, and magazines. To summarize, merit is relatively infrequently used, value is more applicable to drama, and worth is more relevant to life.

Table 2 Word frequency distribution of value, worth, and merit at different time

Table 2	Value		Worth		Merit	
	Freq	Per mil	Freq	Per mil	Freq	Per mil
1990-94	13647	112.69	12973	107.12	1179	9.74
1995-99	13285	106.10	13030	104.06	1278	10.21
2000-04	12275	98.50	12144	97.44	1147	9.20
2005-09	12022	97.70	12065	98.05	899	7.31
2010-14	11068	89.73	12000	97.29	871	7.06
2015-19	14369	117.06	11616	94.63	791	6.44

The COCA corpus not only shows the frequency of use of a word but also the historical development of a word (see Table 2). From the point of view of historical trajectory, the words "value "and worth have remained relatively stable with a decreasing trend. However, people have increased the frequency of using these words in the last few years, which shows that the use of value and worth was decreasing before 2015, and people started to pay more attention to the use of these two words after 2015. Merit, on

the other hand, has been on the decline, especially since 2005. The word "merit" has been used less frequently, suggesting that it has been used less frequently in recent years.

3.2 A Comparative Study of the Semantic Rhyme of Value, Worth, and Merit in the COCA Corpus

Semantic rhyme as a linguistic phenomenon was first proposed by Sinclair, a British linguist, in 1987. Sinclair (1987) believes that lexical items with the same or similar characteristics of semantic rhyme appear in the text habitually and repeatedly with keywords, to form a special semantic atmosphere within the language.[2] The real meaning of words is not confined to the words themselves, but is expressed through the association of words, such as praise or blame, good or bad, and so on. Stubbs (1996) classifies semantic rhymes into positive, neutral, and negative rhymes. "The study of semantic rhyme can enrich people's knowledge of the rules of language, deepen their understanding of the mechanism of language operation, and promote people to pay attention to the behavioral patterns of words."[3] Positive rhyme means that the words with which the word is used are positive and righteous so that the context is expressed with positive color. Negative rhymes are the opposite of positive rhymes, and the context is negative. Neutral rhymes refer to words with both positive and negative connotations, or words that do not have a definite positive or negative connotation, but convey an ambiguous context. In other words, if we can not see the semantic rhyme of a word from a single word, we must combine its collocation words for comprehensive analysis. The COCA corpus provides a convenient way for us to search the collocations of the target word to find out the semantic rhyme of the word.

3.3 Collocation

Table 3: Word collocation of worth, value, and merit

	WORDS	FREQ	MI
worth	MILLION	637	4.44
	POUNDS	476	5.00
	NOTING	339	8.15
	THOUSAND	208	4.06
	CONSIDERING	200	5.28
	BILLION	166	4.88
	STOLEN	134	5.31
	REMEMBERING	133	6.26
	MENTIONING	121	7.69
	EFFORT	119	3.69
value	MONEY	837	3.68
	MARKET	475	3.15
	ADDED	403	3.92
	VALUE	266	3.07
	TAX	265	3.19
	NET	230	4.27
	PROPERTY	224	3.32
	SHARES	211	3.82
	ASSETS	164	4.42
	ASSET	136	5.22
merit	ORDER	90	4.48
	AWARD	31	4.10
	ARTISTIC	25	7.09
	ATTENTION	24	3.92
	PROJECT	21	3.55
	TABLE	21	3.20
	MERIT	16	6.85
	AWARDED	15	5.67
	CERTIFICATE	15	5.46
	LITERARY	15	5.25

Firth (1957) believes that collocation is a linguistic phenomenon in which words and words are used together.[4] Lu Jun(2010) thinks that studying the semantic features of high-frequency collocation words is an important basis for examining semantic rhyme.

The top ten collocations for the word "value" are money, market, added, value, tax, net, property, shares, assets, and asset, with the highest number of collocations for value being money, a word related to finance. Market, added, value, tax, net, property, shares, assets, and asset are also related to finance. As with worth, the author chooses words with an MI value greater than 5, such as nominal, rateable, numeric, resale, asset, etc. It can be seen that the number of financial-related nouns collocated with value is on the high side, and none of the collocated nouns have strong emotions, which can be concluded that the word value has a neutral rhyme. The top ten collocations for the word "worth" are million, pounds, nothing, thousand, considering, billion, stolen, remembering, mentioning, and effort, with the highest number of collocations with worth being "million", which is a word for quantity. The rest of the nouns are also related to quantity, such as thousands, pounds, billion, and so on are words indicating quantity. In addition to the ten words given in Table 3, the author also counts the words with MI values greater than 5. Words with MI value greater than 5 that are collocated with worth, such as nothing, mentioning, pounds, considering, stolen, remembering, exploring, vouchers, quoting, and bothering. In both cases, it can be seen that most of the words that go with the word "worth" include words that express quantity and the v-ing, and it is a regular practice to use the v-ing after worth. The top ten collocations of the word "merit" are order, award, artistic, attention, project, table, merit, awarded, and it can be seen that the word merit is mainly followed by a noun, which is similar to value. The words "merit" and "value" are to some extent interchangeable and represent virtual material or spiritual values. The author counts the words with an MI value greater than 5, such as artistic, award, literary, legion, raises, consideration, badge, etc. The collocations of the above three words do not have strong negative emotions, and almost all of them are nouns with no clear positive or negative emotions, so it can be deduced that the above three words have a neutral rhyme.

3.4 Differences in Colligation

Colligation is a grammatical feature of a word or a class of words that represents a combination of words in a grammatical or lexical category, not a collocation with a specific word, e.g. adj + n is a colligation. Since value, worth and merit occur more frequently, 100 randomly selected records from the COCA corpus were used to summarise the three words and calculate the percentages.

Value is often followed by nouns, verbs, and adjectives, worth is mostly followed by nouns and verbs, and merit is followed by nouns, adjectives, and adverbs. In terms of modifying nouns, the category of worth + n is the highest among all the colligations of worth (21.6%), the category of merit + n (20.3%), and the category of value + n (11%), which means that the category of worth and merit followed by a noun is more common. In terms of verbs, value + v has the highest percentage, the second is worth + v, and the lowest is merit + v (10% > 9% > 3.3%), suggesting that the addition of a verb to these three words is not common. For more specific analyses, in the case of "worth", the word "worth" is often followed by words related to money, e.g. "dollar", "cents", "price" and so on. The word "value" is often followed by words related to the economy and the market, such as "market", "stocks", "shareholder", and so on. The word "merit" is often used with words related to the criteria of judgement, such as "scholarships", "awards", and "norms". The results of the study show that, compared with the traditional approach of analysing near-synonyms from the dictionary or the sense of language alone, the use of the COCA corpus as a research tool for near-synonym identification is more accurate and persuasive, and that the corpus can provide more accurate data support and basis for near-synonym comparisons and analyses and improve the accuracy of language use.

4. Implications and Reflections on the Comparative Study of Value, worth, and merit in the COCA Corpus

Based on the COCA corpus, this article analyzes the synonyms of value, worth, and merit from two dimensions: domain and semantic. It is found that the introduction of the corpus helps researchers to find out the subtle differences of English near-synonyms, and to understand the knowledge and the usage of related words more deeply. Of course, only using the COCA corpus to analyze value, worth and merit is a relatively superficial and one-dimensional study. In the future, we can take other corpora such as BMC and COHA to make a deeper comparative analysis of near-synonyms, to find out more lexical patterns, which will be convenient for daily English learning.[5]

5. Conclusion

Through the traditional qualitative analysis and the quantitative analysis of the COCA corpus, this article compares and contrasts the near-synonyms value, worth, and merit in terms of domain and semantic prosody. It shows the importance and convenience of using the corpus in English vocabulary learning. Therefore, in daily English learning, English learners should use the corpus as a tool to look up relevant vocabulary, to be more comfortable with the use of near-synonyms, and to be able to express words accurately in appropriate contexts.

References

[1] Wei Naixing. (2002). A General Approach to Semantic Rhyme Research. Foreign Language Teaching and Research (4), 300-307.

[2] Lan Lizhen. (2017). A study of English near-synonyms based on the COCA corpus: careful and cautious as examples. Journal of Inner Mongolia University of Finance and Economics 15(06), 107-110.

[3] Liu Yang. (2020). The analysis of English near-synonyms destroy, damage, and break based on COCA corpus. Communication Research 4 (16), 146+149.

[4] Liu Hui, Gong Fangxia. (2020). Comparative analysis of English near-synonyms based on COCA corpus--Taking "vice" and "associate" as an example[J]. Shengli Journal of China University of Petroleum 34(04),63-66.

[5] Liu Qiaoyi. (2020). A corpus-based analysis of business English near-synonyms: licensing and franchising. Journal of Quanzhou Normal University (05), 39-41+66.