

Analysis of Synonyms Based on COCA: A Case Study of “Ship” and “Boat”

Sha Wang¹ Xiujun Du² Yajie Zhou³ Ziwei Chang⁴ Wei Zhou⁵

^{1,2,3,4,5} School of Foreign Language Studies, Harbin Engineering University, Harbin, Heilongjiang, China

⁵Corresponding author. Email: zhouwei@hrbeu.edu.cn

ABSTRACT

As an important part of English vocabulary, synonyms are of great significance to English learners. The differentiation and analysis of synonyms is a key and difficult point in English learning, and accurately grasping the meaning of synonyms is of great significance for maritime English learners. In this paper, COCA was chosen as the research tool and the synonyms “ship” and “boat” as the research objects, and compared them in real contexts from three aspects: register, collocation, and semantic prosody, so as to improve learners’ ability in synonyms recognition and to help them in vocabulary learning of marine translation.

Keywords: COCA, Synonym, Register, Collocation, Semantic prosody.

1. INTRODUCTION

The analysis of synonyms has always been a key and difficult point in English learning. Synonym is defined as “a word or expression which means the same as another word or expression” in the Collins Advanced English Chinese Bilingual Learning Dictionary. In fact, there are not two words with completely same meaning and usage in a language, which poses a challenge for English learners in the process of learning a foreign language. Therefore, studying synonyms is of great significance for English learners and translators. With the continuous development of corpora, many scholars have found that corpora provide an efficient and convenient way for learning English vocabulary, especially for distinguishing synonyms. Learners can analyze and judge the usage of synonyms through a large and authentic texts provided by the corpus. Marine translation is characterized by strong professionalism and high requirements for translation accuracy. Therefore, it is very important to analyze synonyms and find words that are more in line with the original text. In the past, it was a challenge to select more accurate vocabulary from synonyms in marine translation, and the development of corpora has made vocabulary selection increasingly easy. (Wang Chunyan, 2009)

2. RESEARCH TOOLS AND ISSUES

2.1 Research Tools

COCA (Corpus of Contemporary American English), developed by Prof. Mark Davies of Brigham Young University in the U.S., consists of one billion words of text and is the latest corpus of contemporary American English and the largest parallel corpus of English in the world today. It is characterized by large scale, new data, and various genres, and is an effective tool for learners to understand the usage and change of American English. (Zhao Yiqing 2022) Meanwhile, COCA is also characterized by the following aspects. Firstly, it includes many genres such as movie subtitles, media dialogues, novels, magazines, newspapers, and academic journals. Secondly, this corpus is able to compare the same vocabulary in different time. Thirdly, it can show the frequency of vocabulary and it has the function of comparing related words. In addition, COCA pays great attention to the timeliness of the content, which is updated and expanded at a rate of 20 million words per year.

Nowadays, the corpus has become an important tool for our research and study of marine translation. For example, a bilingual corpus can provide a large number of texts in the source and target languages, and translators can look for sentences and words similar to the translated texts in the corpus during

the translation, and they can also refer to previous translations. Finding marine terms and marine translation idioms through the corpus can also effectively avoid mistakes and improve the quality and accuracy of translation. Given the professionalism and authority of COCA, this paper compares the synonyms “ship” and “boat” based on the COCA, and analyzes how to do well in marine translation with the help of the corpus.

2.2 Research Issues

Palmer (1981) argues that synonyms may differ in register, collocation, and speaker’s attitude. “Ship” and “boat” are a common set of synonyms in English, and they are also high-frequency words in marine translation, which are easily confused by English learners. Both words have the meaning of “船舶” in Chinese. Webster’s Dictionary explains the word “boat” as “a small vessel for travel on water” and “ship” as “a large sea-going vessel”. While Collins dictionary defines “boat” as “something in which people can travel across water” and “ship” as “a large boat which carries passengers or cargo”. By comparing the word explanations in the two dictionaries, it is still difficult for learners to clarify the difference between the two. Therefore, in order to explore more deeply the subtle differences between the

words “ship” and “boat” in marine translation, this paper will discuss in detail the three aspects of register, collocation and semantic prosody.

3. TEXT ANALYSIS

3.1 Register

The concept of register was first introduced by Raid in 1956, and was further investigated in 1964 by M.A.K. Halliday et al. in their study of “Framework of Language Planning”. They believed that language will change according to its function, and this kind of language variants which are distinguished by their usage are register. For example, the language used in church is different from the language used in school lessons, and they belong to different registers. Halliday also argued that the differences between registers are mainly formal, i.e. in lexis and grammar, with lexical distinctions being the most obvious. (Halliday, 1973) Therefore, according to the distribution of synonyms in different registers, it can help learners understand synonyms better.

Using the Chart module of the COCA corpus to search for “ship” and “boat”, the distribution of the two words in different registers is shown in “Table 1”.

Table 1. The distribution frequency of “ship” and “boat” in different registers of COCA

Word		ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
Ship	freq	63012	4598	6918	14277	4509	17202	6935	5401	3712
	per mil	63.45	35.75	55.68	111.47	35.75	145.38	55.00	44.36	26.48
Boat	freq	49411	3202	3925	9485	3979	10815	10944	5337	1754
	per mil	49.79	24.90	31.59	74.06	31.55	91.40	86.79	43.84	14.64

The following conclusions can be drawn from “Table 1”: (1) In real contexts, the frequency of “ship” is significantly higher than that of “boat”, approximately 1.3 times that of “boat”. The statistics shows that “ship” is more flexible and has significant advantages in various registers. (2) “ship” and “boat” have the highest frequency in novel and magazine, accounting for 27% and 22% of the total respectively, indicating that these two words are more applicable in literary registers. (3) Overall, the top three registers of “ship” and “boat” are both novel, magazine, and television, while the bottom three registers include academia and blogs, indicating a high degree of similarity in the distribution of these two words in register.

3.2 Words Collocation

Firth pointed out that “collocation is the habitual companion usage of lexical items.” That is to say, the co-occurrence of some words with others constitutes collocation. (Firth, 1957) Since word collocation depends on the speaker’s natural or habitual intuition of the language, which is overlap of vocabulary and grammar. It is difficult for traditional grammatical studies to observe a large number of word collocations directly, which is made possible by corpus methods. (Hu Xianyao and Zeng Jia 2011) In corpus linguistic research, all words that are within the span of a node word are

regarded as collocations of that node word, and these words co-occur with the node word with a certain frequency. (Chunyu Hu, 2011)

Mutual Information Value (MI) can be used to judge the collocation strength between a node word and its collocations, when the MI value is greater than 3, it means that the collocation is obvious between the words. The greater the MI value, the more obvious the collocation between the node word and its collocations, and the stronger the collocation strength. (ZhuYanqiong, 2021)

The two node words are retrieved through the Collocate module of COCA, the left and right spans are set to -4~4, and the words with MI value greater than 3 are sorted according to the frequency from the highest to the lowest (in case of the same frequency, the word with the larger MI value is ranked in the first place, and all the following sequences are the same). The top 15 words with the highest frequency collocating to “ship” are: cruise, sail, captain, cargo, sink, navy, sea, crew, jump, pirate, port, abandon, sinking, board, and passenger.

Table 2. Nouns that can be collocated with “ship” and “boat”

Ship				Boat			
	collocates	freq	MI		collocates	freq	MI
1	Cruise	1934	7.62	1	Fishing	1760	6.84
2	Captain	1156	4.90	2	River	515	3.32
3	Cargo	953	7.10	3	Ride	495	4.43
4	Navy	838	5.47	4	Dock	452	6.32
5	Sea	823	3.89	5	Sea	433	3.28
6	Crew	775	4.63	6	Captain	384	3.63
7	Pirate	567	6.05	7	Owner	382	3.05
8	Port	543	5.07	8	Lake	343	3.23
9	Sinking	495	8.15	9	Tour	303	3.44
10	Passenger	461	4.39	10	Charter	296	5.22

From “Table 2”, it can be seen that there is partial overlap in the nouns collocated with the words “ship” and “boat”. The two words can be collocated with both “captain” and “sea”, and the frequency is relatively high. However, MI of “ship” with these two words is significantly higher than that of “boat” with them. For location nouns, “ship” and “sea” have a higher frequency of collocation, while “boat” and “river” have a higher frequency of collocation, followed by “sea” and “lake”. This may indicate that “ship” often navigates in the sea, while “boat” often navigates in river and lake.

Among them, nouns and verbs appear the most frequently, with 11 and 4 occurrences respectively, indicating that the collocation of “ship” with nouns and verbs is the most common. After searching for “boat”, it was found that the top 15 words with the highest frequency collocating to “boat” are: fishing, rock, river, ride, dock, sail, sea, sink, row, capture, owner, lake, tour, chart, plane. Among them, nouns and verbs appear the most frequently, with 11 and 4 times respectively, indicating that “boat” is also the most common combination with nouns and verbs. Therefore, the following is a detailed analysis of nouns and verbs to explore their usage.

3.2.1 Nouns

In the Collocate module of COCA, the span is set to -4~4. Nouns collocated with “ship” and “boat” are searched for respectively, and the collocations of the two were ranked according to the frequency from the highest to the lowest after eliminating the words with the MI value less than 3. The top ten ranked words are shown in “Table 2”.

3.2.2 Verbs

In the Collocate module of COCA, the span is set to -4~4, and verbs collocated with “ship” and “boat” are searched for respectively, and after eliminating words with MI value less than 3, the collocated words of the two are ranked according to the frequency from the highest to the lowest, and the top ten words are shown in “Table 3”.

Table 3. Nouns that can be collocated with “ship” and “boat”

Ship				Boat			
	collocates	freq	MI		collocates	freq	MI
1	Sail	1211	7.15	1	Rock	748	6.46
2	Sink	951	5.94	2	Sail	439	6.00
3	Jump	690	3.94	3	Sink	416	5.06
4	Abandon	530	4.65	4	Row	411	7.81
5	Board	469	6.14	5	Lift	254	3.16
6	Destroy	460	3.22	6	Float	228	4.43
7	Cruise	328	6.13	7	Board	214	5.32
8	Dock	286	7.35	8	Steer	185	4.99
9	Land	221	3.10	9	Dock	164	6.87

From “Table 3”, it can be concluded that both “ship” and “boat” can be collocated with “sail”, “sink”, “board”, “dock”, and “steer”. Except for “steer”, the other four words have a stronger MI with “ship” than with “boat”, with “sail” having the strongest MI with the two words. This indicates that when describing hull motion, the words “sail”, “sink”, “board”, and “dock” are all more suitable. In terms of manipulation, “ship” is usually collocated with “steer”, while “boat” is usually collocated with “row”.

3.3 Collocational Semantic Prosody

In 1987, British linguist Sinclair first proposed the concept of semantic prosody, which he believed indicates the speaker’s attitude in a particular discourse situation. (Sinclair, 1991) Stubbs (1996) classified semantic prosody into three categories: positive semantic prosody, neutral semantic prosody and negative semantic prosody. The semantic prosody of a word can only be seen through the words it is collocated with, and a large number of examples in the corpus can be used to understand the semantic prosody of words, which is important for lexical research. (Zhao Yiqing, 2022) Wei Nai-Xing (2002) suggests that the index rows of semantic prosody should be greater than or equal to 100 rows in order to have research value. Combined with the corpus analysis of the collocation above, the nouns and verbs collocated with “ship” and “boat” were searched in COCA, and the top 100 words with the highest frequency were analyzed respectively, and the following conclusions were drawn.

Among the 100 collocational nouns of “ship”, there are no words with positive semantic prosody; and 4 words with negative prosody, including “rebel”, “fool”, “sink”, “sinking”; the rest 96 are

with neutral prosody. Among the top 100 verbs with the highest frequency of occurrence in the collocations of “ship”, there are no words with positive prosody; 10 words with negative prosody, including “sabotage”, “desert”, “hijack”, etc.; and the rest 90 are with neutral prosody.

Among the top 100 nouns with the highest frequency of occurrence in the collocations of “boat”, there is 1 word with positive prosody, which is “pleasure”; no words with negative prosody; and the rest 99 are with neutral prosody. Among the 100 verbs of “boat”, there are no words with positive prosody; and 4 words with negative prosody, including “swamp”, “crash”, “miss” and “hijack”; the rest 96 are with neutral prosody.

From the above statistics, it can be seen that “ship” and “boat” both exhibit a neutral prosody overall, but the negative prosody presented by “ship” is more pronounced than that of “boat”, indicating that in reality, “ship” is more common in shipwreck situations than “boat”. In addition, as one of the most common synonym in marine translation, it is necessary to avoid collocate “ship” and “boat” with emotionally strong words and use neutral and objective word more often when using these two words.

4. CONCLUSION

This paper studies the synonyms “ship” and “boat” from three aspects: register, collocation, and semantic prosody. It has found that in terms of register, the use of “ship” in real contexts is more widespread than that of “boat”, which is more in line with the language habits of native English speakers. In terms of collocation, both “ship” and “boat” is more likely to be collocated with nouns and verbs, and the frequency of collocation with nouns is higher than that of verbs. In terms of

semantic prosody, although the negative prosody presented by “ship” is more pronounced than that of “boat”, both “ship” and “boat” overall exhibit neutral prosody. Therefore, when choosing word collocating with them, it is important to pay attention to the neutrality and objectivity of the vocabulary. The study also indicates that COCA provides a large number of authentic language materials for English learners and provides a convenient and effective channel for distinguishing synonyms in marine translation. Translators and learners of marine English can use COCA to deepen their learning and better understanding of synonyms, thereby assisting in the translation and learning of marine English.

REFERENCES

- [1] Firth J R. Papers in linguistics[M]. London: Oxford University Press, 1957: 12.
- [2] Halliday, M.A.K.1973, Explorations in the Function of Language, London, Edward Arnold.
- [3] Palmer, F.R. Semantics [M].Cambridge: Cambridge University Press, 1981.
- [4] Sinclair, J. Corpus. Concordance and Collocation[M]. Oxford: Oxford University Press, 1991: 12.
- [5] Stubbs, M. 1996. Text and Corpus Analysis. [M]. Oxford: Blackwell Publishers.
- [6] Chunyu Hu. Corpus and Business English Vocabulary Research. [J] Journal of Guangdong University of Foreign Studies and Business, 2011 (2).
- [7] Chunyan Wang. A Corpus-based Approach to Distinguishing English Synonyms for Chinese Learners [J]. Foreign Language and Foreign Language Teaching, 2009 (6): 27-31.
- [8] Naixing Wei. The General Methods of Semantic Rhyme Research.[J] Foreign Language Teaching and Research, 2002 (4).
- [9] Xianyao Hu, Jia Zeng. A New Trend in Corpus-based Research on Translation Commonalities. [J] Journal of the PLA Foreign Language Institute, 2011 (1): 56-62+127-128.
- [10] Yiqing Zhao. Vocabulary Research Based on COCA Corpus—Taking Synonyms "related", "related", and "related" as Examples. [J]. Writer Tian-di, 2022, No.626 (29): 111-113.
- [11] Yanqiong Zhu. Analysis of English Synonym Collocation Based on Sketch Engine Online BNC Corpus: Taking “effect” and “influence” as Examples. [J] Overseas English, 2021 (4): 117-118.