The Application of the Combined Corpus in C-E Translation Teaching

Ganggui Yang

Abstract—In recent years, corpora have become increasingly sophisticated and are popular among linguists and widely followed by academics because of their powerful functions. This paper investigates the combined use of single and bilingual corpora in teaching translation quality assessment. Using single and bilingual corpora as research objects, the empirical method explores the path of efficient use of combined corpora in translation teaching, and the results show that the combined use of corpora can provide strong support for translation learners in vocabulary selection, comparison of synonyms and conducting translation quality assessment, and this study aims to improve the quality of teaching and learning in translation courses.

Index Terms—Monolingual corpus, bilingual corpus, translation teaching, translation quality, combined use

I. INTRODUCTION

As reform and opening up further, the demand for high-level translators in the international community is increasing, and how can we improve the quality of translation teaching, which should be the focus of teaching in the future. At present, there are 316 general colleges and universities with MTI qualification and generally offer translation-related courses, but due to a variety of practical factors, the translation courses are reflected in the separation of “teaching” and “learning” in terms of quality, and improving the quality of translation teaching and students. However, due to various practical factors, translation courses are reflected in the separation between “teaching” and “learning” in terms of quality.

A combination of monolingual and bilingual corpora is a more effective way of improving the quality of translation teaching. The corpus has a wide range of applications, with many functions and advantages. It can be used in areas such as Chinese language and literature, foreign language and literature, and computer applications. It can be used to find the most common usage of a word, to check the authenticity of an expression in a language, and to look up sentence patterns. Although some dictionaries can give expressions, they are not very authentic. A corpus can give authentic expressions in real foreign languages, helping people to understand the exact meaning of words in real contexts, rather than just giving the Chinese meaning. The potential for the use of corpora in translation teaching is huge. Using corpora effectively in translation teaching can improve the quality of students’ translations.

II. REVIEW OF THE LITERATURE

A. Definition and Classification of Corpus

According to Reppen (2010), a corpus is a large, structured, electronically stored collection of naturally occurring languages [1]. In other words, a corpus is an electronic collection of texts that represent a language or a variant or text category of a language, collected according to certain sampling criteria. Scholars use corpora to explore typical problems of translation, or to study translation language features, translator styles, translation norms, translation teaching, translation quality assessment, etc. [2]. Corpora can be classified by usage (divided into general corpus; special corpus), timeliness (divided into common corpus; ephemeral corpus), style (divided into written corpus; spoken corpus), whether it is native language (divided into native language corpus; foreign language learner corpus), and whether it is annotated (divided into raw language corpus; cooked language corpus). Depending on the language, we can classify corpora into monolingual corpora; bilingual or parallel corpora; and multilingual corpora.

B. The Use of Monolingual Corpora in Translation Teaching

Monolingual corpora, which are relatively easy to build and mature, are more commonly used in translation teaching and are divided into two main categories. They are divided into written and spoken corpora according to their genre, and into general and specialized corpora according to their specialization.

Among the monolingual corpora classified according to language genre, the written language corpus is more numerous. For example, teachers often use two large monolingual corpora of English, the BNC and the COCA, which are rich in resources.

In general, before 2010, it was mostly an introduction to the corpus, and after 2010, it was mostly an empirical study of the corpus. “Research on the application of corpus in translation teaching is basically an introduction to the application of corpus in translation teaching” [3]; after 2010, empirical research on the application of translation teaching gradually becomes mainstream. From the exploration of single corpus in translation teaching [4, 5], to scholars teaching translation based on bilingual corpus, correspondence corpus and multi-model corpus, a view is to further exploring a more appropriate teaching model [6]. With the development, the application of corpus to translation research and teaching has become a phenomenon, and scholars have started to build corpus to optimize translation teaching. For example, Zhu [7] built a small corpus based on the 28th Korean Contest Chinese to English translation entries for translation genre assessment, which has shed some
light on the teaching of translation quality assessment.

C. The Application of Bilingual Corpora in Translation Teaching

Research on the teaching of translation from parallel corpora has also shown a growing trend in recent years, and this is more beneficial to learners with bilingual needs. Bilingual corpora are mainly used for vocabulary searching, bilingual corpus comparison and passage comparison.

As for the comparison of bilingual corpora, some scholars have made a systematic overview of the resources that a bilingual corpus can support. For example, Bernardini provides a systematic overview of the teaching resources that a bilingual parallel corpus may provide in translation teaching [8, 9].

Another advantage of a bilingual corpus is the ability to see the linguistic differences of bilingualism more visually through paragraph comparisons, as Gilquin [10] shows through examples how the application of cognitive corpus linguistics can combine the advantages of corpus data with the advantages of cognitive principles to help learners construct language systems.

Bilingual corpora also provide a wealth of data for featured translations, which facilitates the assessment of translation quality of professional texts. Some institutions have built their own bilingual corpora, for example, Wang [11] points out that the “English-Chinese Chinese-English Translation Distance Learning System” developed by the City University of Hong Kong has greatly facilitated bilingual distance learning by relying on the platform. Then later, the scale of bilingual corpora was further enhanced, such as the large-scale 100-million-word large-scale English-Chinese parallel corpus retrieval platform V2.0 [12], and the Chinese monolingual corpora that have been completed and put into use include CLL developed by Peking University and BCC developed by Beijing Language and Culture University. These corpora not only have a large capacity and a wide range of topics, but also are available for free use and access to them in teaching. The corpus is more convenient to access in teaching [13].

In conclusion, some studies have shown that most of them mainly use a single corpus to teach translation, but there exists a certain research gap. The combined use of corpus is not very extensive in teaching translation quality assessment, and there are few studies on the combined use of various corpora to teach translation quality assessment. The combined use of corpora can not only stimulate students’ enthusiasm and interest in independent learning, but also break the barrier between ‘classroom teaching’ and ‘acquisition quality’ and enhance the learning effect.

III. COMBINED APPLICATION OF SINGLE AND BILINGUAL CORPORA IN TEACHING TRANSLATION QUALITY

Building on existing research, the use of a combined bilingual corpus explores powerful initiatives for teaching translation quality assessment. The use of monolingual and bilingual corpora has become increasingly common and their role in teaching should not be overlooked. The world’s largest free English corpus is the Corpus of Contemporary English of the United States (COCA) which contains a wide range of corpus data, in addition to the British National Corpus (BNC), a large monolingual corpus developed by Oxford House Publishers and others in the UK, with a predominantly British English corpus, both spoken and written.

There are also many mainstream bilingual corpora, mainly the English-Chinese parallel “Dream of the Red Chamber” corpus, the “Chinese-English bilingual tourism corpus”; the “Applied Text Chinese-English bilingual parallel corpus”; the bilingual parallel corpus retrieval software (mainly ParaConc in the study; AntConc, WordSmith) [14].

This paper explores the combined use of corpora in translation teaching in terms of vocabulary selection, comparison of synonyms and contextual comparison. In the teaching of translation quality assessment, teachers and students can use a combination of monolingual and bilingual corpora to translate more effectively or to assess the quality of appreciated translations.

A. Vocabulary Selection

Teachers and students can assess the quality of translations in terms of vocabulary, which is a very small unit of language but has a huge impact on the meaning of a sentence and a whole chapter. The most common problem that students have is precisely their choice of vocabulary. When teaching translation, it is common for students to be unsure of the vocabulary they are translating. If they cannot work out or do not understand to express a specific word, they can be inspired to conduct a vocabulary search in a monolingual or bilingual corpus, as in the example below.

Unlike other corpora, the corpus of Global Web-based English (GloWbE) is unique in that it allows you to compare between different variants of English, and it is also associated with other corpora. GloWbE contains around 1.9 billion words of text from 20 different countries. This makes it approximately 100 times larger than other corpora (such as the International English Corpus), and it also supports offline use after download, which is very student-friendly. This corpus pays particular attention to comparisons between countries and virtual corpora, and it allows the creation of personalized collections of texts related to a specific area of interest, facilitating the habits of individual users.

Example 1:

ST: There is a sentence in the 2022 government work report: In the past year, in the face of the complex and severe situation at home and abroad and many risk challenges, the whole country worked together to co-ordinate epidemic prevention and control and economic and social development, and the main objectives and tasks of the year were better accomplished.

TT: Machine translation result: Over the past year, in the face of complex and grim domestic and international situations and many risks and challenges, the whole country has worked together to coordinate epidemic prevention and control with economic and social development work well done. The Government of the People’s Republic of China (PRC) has been working together to coordinate epidemic prevention and control with economic and social development work well done.

TT: Xinhua Translation: Last year, facing complicated and challenging circumstances both in and outside of China, as
well as many risks and challenges, all of us the last year, facing complicated and challenging circumstances both in and outside of China, as well as many risks and challenges, all of us throughout the country worked hard as one.

Analysis: In the process of teaching translation, the original government report is shown first. “severe” is a difficult point in this translation, and words such as “severe” and “austere” are literally the words “severe” and “austere” can be used literally, but attention should be paid to the word combinations. The Corpus of Global Web-Based English (GloWbE) shows that the phrase “challenging circumstances” has a frequency of 239 times, including 64 times in the UK, 23 times in the US and 19 times in Canada. The word “severe” is usually used with punishment or physical conditions. The phrase “severe circumstances” has a frequency of 25 times, the phrase “grim situations” has a frequency of 8 times and the phrase “austere circumstances” is even less frequent, at 5 times. It is clear that the phrase “challenging circumstances” is reasonably accurate. The search can also be done in the analysis text function, which will give a linked analysis of the corpus (as shown Figs. 1 and 2).

B. Comparison of Synonyms/Near-Synonyms

In the process of teaching translation, the author find that students are often hesitant to choose the synonyms of a certain word in the source language. In view of this, students who are unsure of the difference between synonyms when translating can consider conducting a vocabulary search in a single or double corpus to find the most appropriate word for the target language, as in the example below.

The British National Corpus (BNC) is a good choice for synonym/near-synonym comparison options, and its Simple Search is a free online service that allows users to search the full BNC World corpus, and not only that, but to view the frequency of search terms and 50 randomly selected instances. To use this feature, the author simply go to the BNC’s search page and then enter the characters you want to search for in the search box. (See in Fig. 3)

Example 3:

It is worth noting that BNC has produced various word frequency lists and comparative lists, which can be viewed directly through the links on the website, as shown in Fig. 4 (some extracts), and which help students to determine whether the target language words to be translated are in the spoken or written form.

The BYU-BNCm has a function to look up synonyms and near-synonyms, simply by adding the “*” sign in front of the word you are looking up, and the settings box below allows you to select the corresponding source for more advanced and subtle qualification. For example, you can look up the usage

(See in Fig. 3)
of the synonyms about “as to and as for”. (See in Fig. 5)

In addition, a comparison search between as to and as for is also carried out in COCA (see Figs. 6 and 7), and we can see that the corresponding example sentences are given to help teachers and students to deepen their understanding in context and to make reasonable results, and other words and phrases can also be compared in this way.

**Example 4:**

**ST:** When an insurance contract is formed, the policyholder pays the premium as agreed.

**TT:** Once an insurance contract is formed, the applicant shall pay the premium ________ the terms and conditions of the contract.

Analysis: This is a representative example of a monolingual and bilingual corpus in which synonyms or near-synonyms are compared. When students saw “according to”, most of them used “according to” as their first response, while others gave “according with”. If you are not sure which word to use, try looking up “according to” and “according with” in the legal corpus to compare and test your hypothesis, as shown in Fig. 8.

When looking up “according with”, it is often followed by “terms, conditions, law, principles and other legal terms”.

In addition, a comparison search between as to and as for is also carried out in COCA (see Figs. 6 and 7), and we can see that the corresponding example sentences are given to help teachers and students to deepen their understanding in context and to make reasonable results, and other words and phrases can also be compared in this way.

In addition, a comparison search between as to and as for is also carried out in COCA (see Figs. 6 and 7), and we can see that the corresponding example sentences are given to help teachers and students to deepen their understanding in context and to make reasonable results, and other words and phrases can also be compared in this way.
regulations corpus [15].

The translation of the phrase “in accordance with the contract” can be searched in the Chinese-English parallel corpus of Chinese laws and regulations (mainland), as this corpus collects the Chinese-English parallel text corpus of the corresponding laws and regulations, which allows for a quick search of the corpus of various laws and regulations, as can be seen in Fig. 10, the context-specific search results given below (part It is clear that the use of the phrase) It is clear that the use of the phrase “according with” is more in line with the specific context of the “according to the agreement” context.

Fig. 10. Bilingual corpus of Chinese laws and regulations (PCCLD) to aid search.

C. Quality Assessment of Translations

In addition to the combined use of single and double corpora in vocabulary selection and the comparison of synonyms or near-synonyms, teachers and students can also use them in translation teaching to assess the quality of translations. Translation quality assessment is a rigorous and complex matter, which can be grasped separately at the macro and micro levels. At the macro level, it is mainly in the discourse and paragraph dimensions; at the micro level, it involves smaller aspects of words and sentences. When assessing the quality of a translation, if you are unsure of the accuracy of the words used in a sentence or do not know what the corresponding authentic expressions, idioms or slang in the source language are, you can conduct a phrase search in the corpus, as in the example below.

Example 5:
ST: Apart from computers, films and television are also spreading bad information. In this respect, my attitude is firm: I am in favour of tighter regulation. First of all, foreign films and TV works are not in line with the national situation and should be banned. Secondly, domestic film and television professionals are of different standards and produce many bad productions.

TTT1: Besides computer, film and TV also spread unhealthy information. (Source: Wang Xiaobo, The Silent Majority)

TTT2: The bad information is also disseminated by the movie and television besides computer.

TTT3: The film and television programs are spreading harmful information. (Source: Chinese Learners’ Corpus of Interpreters and Translators)

Analysis: In this example, the teacher gave three comparative translations based on the original text, so that students could discuss which translation was better and more in line with the original expression. The majority of students felt that the second translation was more in line with the original meaning. Students generally thought that the word ‘unhealthy information’ was not accurate, but they were not sure whether ‘harmful information’ was reasonable. “The teacher and students worked together to verify this in the Corpus of Contemporary American English corpus. As shown in the diagram below, it is possible to look up the specific words and word frequencies that go with it, and also to grasp its meaning in context.

As shown in Fig. 11, we can see that the corpus can look up the frequency of words that correspond to the collocation of “information”. The top three verbs with the word “disperse” are “DISSEMINATE, DISSEMINATION, DISSEMINATING” with 228, 216 and 176 times in addition, the top three verbs with the word “spread” are “spread, spreading, spreadsheet”, with a frequency of 102, 56 and 14 times respectively. The top three adjectives with the word ‘information’ are ‘new’, ‘additional’ and ‘personal’, with a frequency of 2552, 1531 and 1438 respectively.

Next, we can use the relevant corpus to verify the contextual collocations of “bad information”. Information is 314, 238 times, and we find that only the collocation “bad information” is used, while “unhealthy information” and “harmful information” are not generally used in English. The reason is that “unhealthy” is biased towards physical unhealthiness, while “harmful” refers more to causing injury. The corpus search results are as Fig. 12.

In addition, during the translation quality assessment, students were led through the following example, which enabled them to appreciate further the advantages of a corpus at the segmental level - the ability to give contextual information about the text, which is an advantage, but also the limitations of the corpus itself, as in the example below.

Example 6:
ST: My computer is not yet connected to the Internet, and I have thought about connecting to the Internet. I was told that
the Internet was awash with pornography and some reactionary stuff was spreading, and these claims scared me. Some time ago someone suggested that the Internet should be restricted, and I’m all for that...

Revised TT: Besides computer, film and TV also spread information that contains obscene and pornographic content.

The following are the corresponding contexts.

The Internet is a tool for transmitting information. There are also the tools for processing information, which are the various personal computers. If you think about it, without computers, you can’t connect to the Internet even if you have one. Besides, disks and CDs are enough to traffic in order to achieve an effective assessment, education, and technical translation teaching and students’ initiative, encourage them to explore the various functions of the corpus, and develop their ability to use various corpora in cross combinations.

**CONFLICT OF INTEREST**

The author declares no conflict of interest.

**ACKNOWLEDGMENTS**

I wished to thank the anonymous reviewers for reading my manuscript and giving valuable suggestions.

**REFERENCES**


Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).