

The Feasibility of Parallel Corpus Applied in College Translation Teaching

ZHAO Yushan^{[a]*}; SHI Juan^[b]; LI Haiyan^[c]

^[a]Professor. School of Foreign Languages, North China Electric Power University, Beijing, China.

^[b]Assistant Professor. School of Foreign Languages, Inner Mongolia University of Technology, Hohhot, China.

^[c]Associate Professor. School of Foreign Languages, North China Electric Power University, Beijing, China.

*Corresponding author.

Supported by a grant from Joint Project of Beijing and Central Universities.

Received 25 February 2014; accepted 15 May 2014

Published online 26 May 2014

Abstract

In the field of language teaching and learning, corpus has great potentials as a kind of advanced teaching and learning resource tool. In China corpus-aided translation teaching bears less fruit and there are few empirical studies of Chinese-English translation based on corpus. This paper introduces corpora, concordance, corpus-based translation teaching, and then explains the application as well as the advantages of corpora in translation teaching.

Key words: Translation teaching; Corpus; Feasibility

Zhao, Y. S., Shi, J., & Li, H. Y. (2014). The Feasibility of Parallel Corpus Applied in College Translation Teaching. *Studies in Sociology of Science*, 5(2), 12-17. Available from: URL: <http://www.cscanada.net/index.php/sss/article/view/4937>
DOI: <http://dx.doi.org/10.3968/4937>

INTRODUCTION

In China about three hundred million people are learning English, including twenty-three million students in colleges or universities. Although great achievements have been made in the foreign language teaching, the time-consuming and inefficiency situations have not

changed a lot. With Chinese economy developing fast and the international communication increasing, the translation quality is required. Therefore, students in the translation class should learn many theories and skills about translation and they must get over some difficulties in practice. For example, word stuffing, equivalent and collocation obstacle the students. Usually, students would like to check in the dictionaries, but limited entries in the dictionaries would not offer every context for students to compare and choose, not to mention the situation that most students have only one or two dictionaries, some of which are low-quality and unsatisfactory. Hence, some effective methods should be adopted in translation class.

The approaches based on the corpus have been widely used in research areas such as Second Language Acquisition, Language Teaching, Contrastive Linguistics, Lexicography, Syntax, Semantics, etc. Because of many advantages, people apply it into the research of literature and translation studies as the fast development of corpus linguistics, concordance tools and software also flourish and are constantly being improved.

However, in the area of translation teaching, corpus has not shown its values and advantages completely. Although there are many findings on translation research, fewer people notice its usages in the classroom and put it into use in China. It is meaningful to do some practical research oriented towards assisting Chinese-English translation teaching. On one hand, it could introduce a new approach to teaching by which learners unable to grip the different meanings of the same word in different contexts and gain the proper equivalent for translation. On the other hand, under the influence of rapid development of computer technology, it can improve teaching methods and teaching quality. In this paper, the author intends to explain the feasibility of the corpora used in the translation teaching.

1. CORPUS AND CONCORDANCE

1.1 Corpus

What is corpus? Actually, a corpus can be identified in terms of both its form and its purpose. According to Crystal (1991, p.23), a corpus is “a collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as a starting-point of linguistic description, a means or hypotheses verifying of a language.” Sinclair (1999, p.11) also expounds it as “a collection of naturally occurring language text, selected to characterize a state or variety of a language.”

On the whole, linguists have always used the word corpus to describe a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for morphological study. More recently, the word has been earmarked for collections of texts that are stored and accessed electronically. It is new technology in the study of language, which depends greatly on the use of computer. It is an agency of natural language material stored in computer-readable form. Programs can be drawn up to manipulate the language material in various ways. It is a powerful resource for philological research.

Corpora are ranked on the basis of varied criterions, because corpora can differ in both structure and content according to the purpose for which they were compiled.

According to the compiling different purposes, there are general corpora and specialized corpora. General corpora refer to the corpora which consist of a body of texts which linguists analyze to seek answers to particular questions about particular language. Specialized Corpora are corpora with the language samples of a particular field or area.

According to the date of the materials in corpora, synchronic corpora and diachronic corpora are pointed out. A synchronic corpus contains the language or texts of a particular time while a diachronic corpus represents language over a period of time. Such corpora can be relevant to language history study.

Spoken and written corpora are compiled considering the different language channels. Spoken corpora refer to the corpora which consist of texts of spoken language while written corpora refer to those consist of written language. Spoken corpora are more difficult and time-consuming to compile, such as the London-Lund Corpus which consists of about 52,600 words of the standard spoken British English of adults.

According to the involved languages, corpora are divided into monolingual corpora and multilingual corpora. The former refer to those corpora composed of only one language while the latter refers to the corpora containing more than one language. Parallel corpora are those including parallel texts of two languages. Multilingual corpora, especially parallel corpora, are valuable resources in translation research and teaching.

Considering different language users, there are native speaker corpora and learner corpora. The former assemble texts produced by native speakers while the latter are the non-native speakers' language, such as English essays written by Chinese students. These types of corpora are useful pedagogical resources of language teaching and error analysis.

In recent several decades, corpus-based linguistic research is developing in the following several ways: learner corpora construction and their applied studies are becoming one of the important branches of corpus linguistic studies; spoken corpora are attracting more attention; parallel corpora are playing a crucial role in comparative language studies and translation studies; and corpus construction is becoming a popular practice of language teachers and researchers.

The first electronic or digitalized corpus, known as the Brown Corpus was built in 1961, a year considered by most corpus linguists at the time when corpus linguistics or more precisely computer corpus linguistics first took form (ibid.). After that, numerous digitalized corpora have been constructed and used in miscellaneous studies. Beside the Brown Corpus, the most influential ones among these may include the Lancaster-Oslo/Bergen (LOB) corpus in 1970s, the London-Lund Corpus of Spoken English (LLC) corpus in 1980s, the Collins Birmingham University International Language Database (COBUILD) corpus in 1980s, etc. (Leech, 1997). All these corpora can be found either as end products in CD-ROMs, such as the Collins COBUILD Dictionary, or as on-line services through the Internet, usually at some cost.

With the deepening and widening of the research scope, corpora of one million words seemed too small for most kinds of lexical and semantic analysis. Fortunately, developments in technology for text capture and storage came at the right time and made bigger corpora possible, so that by the 1990s corpora of 100 million words or more became available and were called second-generation electronic mega-corpus. Among these there were four representative corpora developed in the 1990s. The first one is the Bank of English under the Cobuild Project. By 1997, its size was said to be over 300 million words and growing. Next one is the Longman Corpus Network, which is composed of three major corpora, the Longman/Lancaster English Language Corpus (LLELC), the Longman Spoken Corpus (LSC), and the Longman Corpus of Learners' English (LCLE). Each of these has different design features and purposes but together they were compiled as a basis from which reliable descriptions of English could be made, especially for the compilation of dictionaries for non-native speakers of English (Summers, 1991). The third representative corpus is the British National Corpus (BNC). It is one-million-word. The last one is the International Corpus of English (ICE). It is an corpus ambitious project for the comparative study of English worldwide (Greenbaum, 1996). These

mega-corpora will be more reliable and accessible for researchers and teachers.

In China, complying with this international trend in the field of language teaching and learning, many researchers and teachers compile corpora for linguistic study and pedagogical purpose, conduct experiments in corpus-based learning environment. Corpus-based study, including corpus compilation, corpus-based linguistic study and corpus-based SLA research in China is becoming a hot topic and bearing fruits. Different types of corpora for various purposes are compiled during the last two decades, for example, the Jiao-Tong University Corpus of English for Science and Technology (JDEST)(1982), the Guangzhou Petroleum English Corpus (GPEC)(1987), HKUST Learner Corpus(1993), Chinese Learner English Corpus (CLEC)(1999), the Corpus for Middle School English Education at the South China Normal University (MSEE)(1999), the Hong Kong Corpus of Conversational English (HKCCE)(1999) and the Beijing University of Foreign Studies Corpus of English Language Teaching Material (BUFS Corpus)(2000)(Wang, 2001)etc.

Among these corpora, two are worth to mention. The National Research Center for Foreign Language Education in Beijing Foreign Studies University is a pioneer in the research of this area. The center has set up the largest Chinese-English Parallel Corpus in the world (Wang, 2003). Research based on this corpus has resulted in many different findings. Also, Ke (2002) lists four applicable namely, (1) Research on comparative linguistics; (2) Research on translation transformation, translation style and automatic machine translation; (3) Compilation of bilingual dictionaries; (4) Translation teaching and translator education. This accounts for one of the most specific demarcations of the applicable range of bilingual corpus in China. Another important corpus in China is CLEC (Chinese learner English corpus). As the name indicating it is a learner corpus. It collects compositions written by Chinese students. Comparing with BNC, Chinese Learner English Corpus (CLEC) is a rather small corpus, consisting of one million words with all the errors manually tagged according to an error tagging scheme. However, it is useful and more specific, especially for English learners in China. A number of cross-sectional and longitudinal studies of Chinese learner English were carried out basing on CLEC by a group of researchers and teachers. It is part of the important projects in the Ninth Five-Year Plan of the National Foundation of Social Sciences and Humanities. With this corpus, an analysis of the positive transfer and negative transfer of the student's mother tongue (Chinese) in their English learning is provided for the language teachers. The testers make use of the database to check whether the examination is objective and reasonable. The teachers can see that in what points the students are likely to make mistakes by further checking, which will provide as a reference for improving their teaching (Yang, 2002).

1.2 Concordance

A “concordance”, according to the Collins Cobuild English Dictionary, is “an alphabetical list of the words in a book or a set of books which also says where each word can be found and often how it is used.” In the modern linguistics, a concordance is a collection of the occurrences of a word-form, each in its own textual environment (Sinclair, 1999). In its simplest form, it is an index. According to Kennedy (1998), a concordance is a formatted version or display of all the occurrences or tokens of a particular type in a corpus. The type is usually called a keyword but is sometimes known as a target item, node word or search item.

Concordances can be prepared in a number of formats. The most usual form is the Key Word in Context (KWIC) Concordance. There is different kind of concordance, known as parallel concordances. Parallel concordance is typically used in parallel corpus. Once a concordance is performed on a word or type in one language, the translation of that word in other languages within the corpus should be sorted either.

Different sorting within the concordances can bring different results which may attract linguists and language learners. KWIC with concordances unsorted is the most common form of presenting concordances. However, if the context word to the right of the keyword is the basis for sorting, KWIC with Concordance-Right Sorted can be used. In the same way, there is a format where the context word to the left of the keyword is the basis for sorting. People can choose different formats to respond to their needs. Both of the contexts are of high value for collocation and usage of words in foreign language teaching and learning. Moreover, there is a larger contexts format where large contexts or broad contexts of the key words are presented for the wider context research.

2. CORPORA AND TRANSLATION TEACHING

2.1 Translation Teaching

Translation teaching and teaching translation have been distinguished from each other for a long time. It was firstly proposed by Canadian researcher Jean Delisle (1988). Translation teaching is the cultivation of professional translators or interpreters. It gives translation lessons under the guidance of translation theories. It helps improve students' translation skills and strategies. It is the higher stage of English teaching, which is closely linked with certified translation. Teaching translation is a practice to help teaching foreign languages. It is aimed at helping students learn foreign words, structures and grammar and improve their ability of using foreign languages. It requires students to do translation according to certain instances.

Translation teaching studies refer to the study on both teaching theory and practice. The former is generalized from the latter and can be the guidance for the latter. However, translation theory was once neglected by some scholars and teachers of China and other countries. From the last decade of the twentieth century on, more and more teachers began to point out the significance of translation theory in translation teaching. Therefore, they advocated to add some courses about the history of translation development and to introduce some translation theories and basic translation skills or approaches to students. Some scholars think that basic knowledge of translation theory should be part of the course, such as the definition, history, criterion and process of translation, transferability, different types of meaning, the consistency of form and content, context and culture. Other scholars think that the translation theory teaching should focus on pragmatic guiding principles based on a large amount of translation practices. In spite of the disagreements, almost all teachers agree that translation theory should be integrated with translation practice in translation class in order to follow market's needs.

The past ten years witnessed the bloom of translation teaching in China. Many teachers have carried out research on translation teaching theory, translation teaching methodology, translation material design, translation test design, interpreting teaching and so on. More and more these and works on this research have been published in recent years. Furthermore, national forums have been held one after another, discussing translator training and translation teaching. Among them, *Translation Teaching in China* is the first systematic monograph on translation teaching in China (Mu, 1999). After extensive survey, Mu (1999) analyzed translation teaching situation and presented tentative plans for improvement of eight aspects, including discipline construction, curriculum design, textbook design, teacher's training, teaching pedagogy, interpreting teaching, translation assessment and teaching research. Ge (1994) applied the theory of comparative study to the teaching of translating lengthy sentences of English into Chinese. Zhang (2001), another translation teaching expert in China, systematically studied the Chinese- English translation teaching text books from 1949 to 1998, which offered an exact reference to the development of translation textbook in China.

2.2 Corpus-Based Translation Teaching

As more researchers use corpus to investigate translation universals, they also discover that corpus plays an important role in the field of translation teaching. Some scholars tended to lead a group of students to set up their own corpus. For example, Gaviolia and Zanettin (1997) set up a corpus of medical texts together with a group of students. Kübler (2003) set up an LSP (Language for Special Purpose) corpus in the same manner. The two researches show that students have made great progress in

the areas of terminology acquaintance, grammar, subjects field understanding, etc. After training, Kübler (ibid) believed that students could overcome usages of language artificiality through participation in the setting up of the corpus. Also, students who are trained in this area are better recognized by clients in the translation field.

Along with the wider application of computer and computer-assisted tools, corpus to assist translation teaching has become a recent trend. By far, the empirical researches in this area range from the areas of terminology extraction to the establishment of LSP corpus and its applications, such as the corpus of legal texts (Monzó, 2003) and medical texts. It can be well predicted that in a globalize world of information explosion, corpus has become an important resource and method for translation teaching. The ability of using corpus will help translators enhance their competence and text production. In some universities, such as Aston University in Britain and Hong Kong Polytechnic University, corpus-based translation teaching course has been included as a part of the teaching curriculum.

The research of corpus-based translation teaching is blooming as introduced above. However, most of them are a question of French, Spanish, Latin corpora. In *Working with Specialized Language: A Practical Guide to Using Corpora*, the author Lynne (2002) used a French-English corpus to list some practical ways in translation teaching. In China the corpus-based research started much later than that of the Western world. In recent years, some Chinese researchers begin to apply the corpora in the translation teaching. Yu (2004) introduced the application of corpora in analysis of translation ecology, text equivalent probability and translation style. Yang (2007) suggested a translation teaching activity with the KWIC to improve the quality of translation teaching. Wang *et al.* (2007) approached parallel corpus presentation in translation teaching, and find the corpus presentation experiments conducted in the study suggest that the application of parallel corpus to translation teaching contributes to the building of environment for autonomous learning or learning as discovery, and to the formation of steady translation techniques. Xie (2004) showed his researches on the self-compiled small corpus in the teaching and exhibited the positive effect in ESL from different perspectives. In China, however, small corpus is relatively new and the idea of self-compiled small corpus is formally advocated first of all by Liang in his paper *Using Word Pilot to Create a Small-Scale Corpus for EFL Class* published in 2003. China's study on small corpus construction over the past few years is regrettably insufficient in amount and meager in description, few people have actually set out to self-compiled small corpora and the intricate skills of using small corpus in translation teaching were not applied and tested in a typical Chinese classroom setting.

2.3 The Advantages of Corpus-Based Translation Teaching

Bilingual corpora offer other advantages. Firstly, corpora have the potential to be more extensive and inclusive than other resources: hundreds of thousands of words of running texts can be stored on a diskette and millions can fit easily onto a hard drive. In addition, their electronic forms mean that they are easier to be updated and thus the linguistic information contained is fresh and new. Because of the space restrictions imposed on printed dictionaries, dictionaries do not provide enough contextual and usage information of lexical units besides, most of dictionaries do not contain frequency information about the use of a particular lexical unit. Contrarily, in a corpus a wealth of authentic usage information is contained and a concordance software user can easily retrieve it and thus observe actual uses and frequencies.

Secondly, in a corpus numerous texts composed by specialists are available and students can consult them as often and handy as possible. Lastly, the unrestricted volume of information in the corpus means that students can consult the advice of many different specialists. Another advantage of the corpus is that many terms in it are provided by specialists and therefore they are the professionally accepted forms, which mean that they are useful to every translator.

2.4 Corpus-based Translation Teaching for Students

Many teachers of translation have integrated suitable corpora with translation class in order to raise students' translation accuracy and efficiency and arouse their interest as well. Silvia (1993), a Canadian scholar has tried out corpus-based translation teaching with advanced learners of English in their last years of studies as undergraduates at the School for interpreters and translators of the University of Bologna at Forli. These students are required to complete the translation task with the help of the British National Corpus and the software like WordSmith Tools and some other resources available on the Internet. They are given more freedom in choosing resources, designing queries and deciding translation works. Results show that corpus-based translation teaching is very encouraging and motivating. With the experience of obtaining the rich authentic data from the corpora, students are much more confident in translation and can get serendipitous finding during learning.

As we have noticed, classes in concordance with ESL (English as second language) has been proved to be promising since it emphasizes context-bound regularities and much information, so that corpora can be used to teach students to interpret instances of language production as samples rather than examples. The easy access to instances of language performance allows students to depend less on one or two people with their idiosyncrasies and their limited intuition. In other words,

students browse corpora to search for information they need in order to complete a translation task, analyze the results and choose a solution that can satisfy their needs. In this way, they are more creative and more daring in their choice.

Students can also try with a corpus in their native language which may convince them of some intuition or assumption of their mother tongue. In this way, it changes students' role from "learner" to "researcher" so that they can have the same interests or competence as teachers or linguists. Therefore, they need to be more interested and confident in finding better and more accurate search that makes the translation results more precise. Encouraged to become more autonomous in their study, students are required to do the searching seriously and are encouraged to combine observation with participation. They may discuss findings in pairs or in small groups before taking final written works which can stimulate them to consider more actively and help them learn more reasonably.

2.5 Corpus-Based Translation Teaching for Teachers

Corpus-based translation teaching is a useful tool, not only for students, but also for teachers, especially for those who are non-native speakers of the language. Teachers have learned the foreign language for a longer time. Although their students can learn many learning strategies and experience of difficulties and success from them, but their limited intuition concerning acceptability and appropriateness is still problems for teaching.

In corpus-based translation class, a teacher is not artificially setting up tasks and requiring students to give information he/she already has, but everyone in the classroom is actively trying to find the solution to a problem, to guess the meaning of an expression or to find the appropriate equivalent of a word, a term or a sentence in a foreign language. In this way, the teacher acts as a learning expert rather than a language expert which can motivate students and improve their learning efficiency. Teachers can also raise foreign language awareness at the same time. In spite of that, corpus can function as an assistant to help teachers discover additional problems or issues that have never been cared which would improve teaching and enlighten further research.

CONCLUSION

The significance of the corpus-based translation teaching is demonstrated. The corpus and concordance, advantages of corpus-based translation teaching for teachers and students are explored. Corpus is based on DDL (Data-Driven Learning). In data-driven learning, learners are supposed to gain new insights into or gain knowledge through learning more or less autonomously. This advantage can greatly help students to do translation and act as useful supplementary materials to the conventional

tools. It offers a specific method to improve the instruction manner of conventional teaching and conduct a constructivist translation teaching class. The corpus will be a powerful tool for translation and it will enable students to reduce the word-stuffing, choose the suitable equivalent and collocations and shorten the time of the translation.

REFERENCES

- Bernardini, S. A. (1997). *Trainee' translator's perspective on corpora*. Paper Presented at Corpus Use and Learning to Translate Held in November, Bertinoro.
- Crystal, D. (1991). Stylistic linguistics: Studies in honour profiling (Aijmer & Altenberg, Eds.). English Corpus of Jan Svartvik.
- Delisle, J. (1988). *Translation: An interpretative approach*. Ottawa, Canada/London, England: University of Ottawa Press.
- Ge, X. Q. (1994). On teaching E-C long sentences. *Chinese Translators Journal*, (2).
- Greenbaum, S. (1996). *Comparing English*. Oxford: Clarendon Press.
- Ke, F. (2002). Bilingual corpora: New approach for translation studies. *Foreign Languages and Their Teaching*, (9).
- Kennedy, G. (1998). *An introduction to corpus linguistics*. Beijing: Tsinghua University Press.
- Kübler, N., & Stewart. (2003). Corpora and LSP translation. In F. Zanettin & S. D. Bernardini (Eds.), *Corpora in Translator Education*. Manchester: St. Jerome.
- Leech, G. (1997). Introducing corpus annotation. In R. Garside, G. Leech & Tony McEnery (Eds.), *Corpus annotation*. London: Longman.
- Lynne, B. (2002). Working with specialized language: A practical guide to using corpora. In F. Zanettin, S. Bernardini & D. Stewart (Eds.), *Corpora in translator education*. Manchester: St. Jerome.
- Monzó, E. (2003). Corpus-based teaching: The use of original and translated texts in the training of legal translators. *Translation Journal*, 7, 4.
- Mu, L. (1999). *Translation teaching in China*. Shanghai: Shanghai Foreign Language Education Press.
- Sinclair, J. (1999). *Corpus, concordance, collection*. Shanghai: Shanghai Foreign.
- Summers, D. (1991). Longman/Lancaster *English language corpus: Criteria and design*. Harlow: Longman.
- Wang, J. X. (2001). Recent progress in international journal of corpus linguistics. *Corpus Linguistics in China*, (6).
- Wang, K. F. (2003). *The construction and application of bilingual corpora*. Beijing: Foreign Language Teaching and Research Press.
- Wang, K. F., Qin, H. W., & Wang, H. X. (2007). The application of translation teaching platform based on bilingual corpora. *Educational Technology for Foreign Language Teaching*, (12).
- Xie, J. C. (2004). Construction and application of small English and Chinese parallel corpora. *Journal of PLA University of Foreign Languages*, (3).
- Yang, H. Z. (2002). *An introduction to corpus linguistics*. Shanghai: Shanghai Foreign Language Education Press.
- Yang, X. J. (2007). Corpus-based study on translation and translator education. *Foreign Languages and Their Teaching*, (10).
- Yu, L. J. (2004). Corpus-based study on translation teaching. *Educational Technology for Foreign Language Teaching*, (4).
- Zanettin, F. (1997). Bilingual comparable corpora and the training of translators. *Meta*, 43(4).
- Zhang, M. F. (2001). *English-Chinese translation textbooks (1949-1998)*. Shanghai: Shanghai Foreign Language Education Press.