

## Research on Design and Construction of Simple Chinese-English Parallel Corpus of Foreign Business Correspondence

Suping Jiang<sup>1</sup>, Biqing Li<sup>2,\*</sup>

<sup>1</sup> Guangxi Talent International College, Qinzhou Guangxi 535000, China;

<sup>2</sup> School of Information and Communication Engineering, Hezhou University, Hezhou Guangxi 542899, China;

\*janliful@163.com

### Abstract

Proficiency in writing and translating foreign business correspondence will play a great role in the process of imports and exports business for students majoring in business English, customs declarations and international trade. The study and application of a parallel Chinese- English simple corpus of foreign business correspondence makes the teaching effect very obvious. This paper introduces the purpose of the establishment of Chinese-English bilingual parallel corpus and discusses the contents, classification, corpus collection and the characteristics of the corpus.

### Keywords

Foreign Business Correspondence; Bilingual; Corpus.

### 1. Introduction

Foreign business English correspondence plays an important role in the employment of students majoring in business English, customs declarations and international trade. But due to the influence and limitation of the traditional teaching idea of foreign business English correspondence, most of the classroom teaching is dominated by the terminative test and the evaluation criterion is mechanical, ignoring the learning process, so it is not conducive to the improvement of students' ability of writing foreign trade letters. Since China's entry into the WTO, international trade activity is more frequent. The demand of translation of foreign business documents, such as commercial letters is increasing. Graduates in the workplace should not only be familiar with the import and export business process, but also be able to communicate in writing with foreign businessmen in a very appropriate way, write a variety of letters, and make reasonable responses to foreign correspondents to express their intentions. To provide real language materials and objectively reflect the practical use of target language in the real language environment is the greatest advantage of the corpus.

Taking the language materials provided by corpus data driven technology as an example of foreign business English correspondence writing and translation teaching is more direct, more natural, more vivid, more real and more capable of expressing the natural essence of language than the examples written by intuition. For teachers, the corpus makes a great increase in the efficiency of acquiring teaching materials. From the students' perspective, the learning process becomes easy and relaxed by emphasizing the teaching method of self-learning and the enthusiasm of learning is obviously enhanced.

### 2. The purpose of creation of Chinese-English bilingual parallel foreign business correspondence corpus

It is necessary to build some small and professional corpuses at the same time in building large corpus. It is not difficult to find, with the rapid development of "Internet plus" technology, the collection of language materials will never be as difficult as it used to be, with the aid of electronic products like scanners, electronic publications and the dazzling online resources. For the purpose of further research, researchers are likely to build a corpus useful to in-depth research. After deciding to build a

perfectly satisfactory corpus, the scale, standard and level of construction is truly controlled by ourselves and it will be used for teaching and research. We should do a good job of preprocessing before the establishment of the corpus, considering the channels and ways of obtaining the language materials, the format of the language materials, the format of text encoding and the index software [1]. On the base of random sampling, we collect language materials of representative authentic foreign business English correspondence as samples. After the corpus is built, we make statistics and analysis of frequency characteristics of some language programs of English correspondence and apply the results to the teaching of writing and translation of foreign business English correspondence, so that students can learn correspondence directly in a “top-down” way.

Tognini Bonelli believes that the greatest use of a parallel corpus is to show the finished translation products to people. From these products, people can understand similarities and differences between different languages, which is not provided by a monolingual corpus. And it can provide a standardized, professionalized, specialized and honest Chinese-English bilingual service to specialized persons engaged in foreign trade, such as documentary clerks, merchandisers, freight forwarders, e-business members and purchasers. At the same time, it accords with the requirements of bilingual teaching, E-C translation and instant communication, and provides direct and abundant research materials for comparative analysis of Chinese-English business correspondence. The Chinese-English bilingual parallel corpora has high value of theoretical innovation and strong generalization. It will play a role of throwing away a brick in order to get a gem in the popularization of corpus data driven technology of translation teaching of other languages. The tailored parallel corpus is also suitable for the study of writing and translation in many other fields. The depth, breadth and strength of the teaching reform will be extended indefinitely and the use value and practicability of its can be extended to the extreme. The parallel corpus parallels the original and the translation with a large amount of information and a high material authenticity. It can be both referable and comparable and the value of context and the value of computer technology are immeasurable. The traditional method of translation with the help of dictionaries can not be compared with the translation methods based on corpora. It will improve the students' ability to write and translate foreign business English correspondence. It has brought strong conditions for the bilingual correspondence and translation study of foreign business correspondence. Its creation and subsequent research are fully prepared for the future expansion of the corpus. The establishment and utilization of the library make the terms and translation of the foreign trade correspondence amazing and synergetic, providing a large number of selectable corpus for exploring the characteristics of the English and foreign trade correspondence and the translation and communication of the foreign trade correspondence. The establishment and utilization of the library make the terms and translation of the foreign trade correspondence amazing and synergetic, providing a large number of selectable corpus for exploring the characteristics of the English and foreign trade correspondence and the translation and communication of the foreign trade correspondence. It will provide standard and normalized, harmonious and professional language material platform for learners, translators and researchers of foreign business correspondence, so that all platform users can quickly and easily lock the words, phrases, names, rules and chapters they are looking for.

### **3. The content and classification of Chinese-English bilingual parallel foreign business correspondence corpus**

In view of the mentioned above, in order to meet the needs of users and to achieve the purpose of building corpus, we should maximize the coverage and cover the contents of foreign business correspondence. This corpus is designed to cover all the contents of the concept of the Foreign Trade Correspondence. That is, we should firstly clarify the purpose and intention of the research, and then find out what the concept of the Chinese-English parallel corpus is and which is of great significance for the expansion of the corpus. There is no doubt that the selection of the corpus, the way of alignment and the tagging of the corpus are determined by the purpose and intention of the study.

Special-purpose corpus for professional research is often designed to maximize the satisfaction of users and researchers of foreign business correspondence. In addition to the contents and classifications of the above table, it is necessary to list a detailed list of documents, instruments and textbooks for foreign business correspondence, the important foreign business correspondence organizations and their statutes at home and abroad, the expression of commonly used terms in international trade, the currencies of some countries or regions, the names of major banks in the world and main import and export companies and institutions of China's. For this reason, in terms of corpus alignment, in view of the characteristics of the source text of the foreign business correspondence, the text alignment is mainly adopted. In terms of corpus annotation, the emphasis is placed on the comparison of commonly used terms in international trade. Other research in the future will be further developed and perfected.

#### **4. Languages materials collection of Chinese-English bilingual parallel corpus of foreign business correspondence**

The key point of corpus collection lies in the representativeness and canon of the corpus. It is clear that it will be the best policy to collect the text of the foreign business correspondence and the text of the translated text. On the basis of delineating the scope of corpus, there is a clear object in collecting corpus. That is, according to the four major classifications and their supplementary contents, to collect in groups, each of the categories of the specific curriculum system will be resolutely implemented in every group. Methods and research group in accordance with the specific data collection standards and principles, based on the data of the "three nature" of the data platform, namely standardization, specialization, standardization, consistency, consistency and accuracy of the platform. In practice, according to the needs, define the reasonable scale of research related to business correspondence, including writing and translation studies. The key is the first Chinese English two languages correspond exactly to those text input computer, followed by a comparative analysis shows that the relationship between the two. The data collection and sorting according to the complete program: the equivalent of a series of actions or directs the computer to execute judgment instructions, rules and processes, principles of data collection data collection personnel, the official training corpus collection, preservation, classification and recovery data processing software will be converted into a text file using the corpus text, and text that removed clean spaces, enter extra; usually the original set of foreign trade correspondence program, source text and target text design preparation, running on the writing and translation teaching target System structure. That is to say, in determining a good range of text selection, translation of text to acquisition, began in the English input. Part of the corpus can be obtained through the existing electronic text, such as the original version and download directly from the network. People often say that attitude determines the height of the pattern decided to end. The key to the success of the self built corpus, attitude is more important than ability, so we must carefully collect widely. There are the following channels: the first is the corpus source website of domestic companies and product introduction. From the text corpus download may be wrong, some Chinese and English side by side control, the finishing process is relatively smooth, the majority of Chinese text translation work. Facing great challenges, because the translation is complete and then the corresponding data can be used only in English speaking countries. Second is the website of the company and product information, and they only need two steps, namely English translation and bilingual parallel correspondence. Of course these data can also be compiled at home and abroad website is English size monolingual corpus, its purpose is to introduce similar products for comparison and analysis, find out the differences of English national language habits and non English speaking countries have much.

Another part is from foreign trade letters. Most of these from the course of foreign trade correspondence textbooks, guidance information, foreign trade business website for netizens to learn the letters. The paper needs to be scanned first, then sorted into electronic version, while the ones from the website only need to carry on some simple treatment to be used. The self-built corpus,

homework, usually translated works online information, e-books, manuscript processing language material.

According to the teaching contents, students level, teaching habits or different emphasis and ever found to update the data, after nearly a year of construction period, initially built a small foreign trade English correspondence writing translation corpus has its own teaching characteristics. The parallel corpus and solve the difficulties and problems of many effect in writing teaching of translation collaborative application. Such as word stuffing, translation equivalence effect and collocation function, which is helpful to improve the students' writing and translation results and explore the related learning strategies. There are many parts from each graduates graduation, students in the foreign trade company internship and customer business correspondence, text natural and lively, give people a sense of reality. Of course, there are some unable to obtain electronic text, forced to manual input or scan the way into the electronic version of the paper version. In order to improve the scientific quality of corpus and the research results, these essential burdens has to use manual proofreading carefully updated. A part of pre-editing data is also important. The follow-up study was conducted on corpus tagging, text filling head, namely to the corpus, the author of the text field, style, age, publishing information, text words. One patient and meticulous processing can be regarded as the corpus with use of corpus.

## **5. Characteristics of a simple parallel corpus of foreign business Chinese-English correspondence**

As mentioned above, the creation of a parallel corpus is often based on two purposes: the first is the study of language and translation, and the second is the course teaching, that is, the correspondence of foreign business English and it is teaching of international business correspondence in some schools. For parallel corpora, the first step is to align the corpus, which is also the most basic requirement of a variety of applications for parallel corpus. Alignment refers to the establishment of a correspondence between the original and the translation. There are many types of alignment, such as text alignment, segment alignment, sentence alignment, and word alignment. Text alignment and segment alignment always differ from each other in word length. It is seldom used by scholars to translate. Maybe word alignment is the most attractive, especially in English Chinese translation. However, the difficulty coefficient of word alignment is high because of the great difference in the arrangement of sentence order between English and Chinese. The way to get people to agree is sentence alignment. In view of the characteristics of the original text of Foreign Trade Correspondence, the self-construction of this corpus selects text alignment. There is no doubt that the automatic alignment of Chinese and English bilingual text must be aided by some specialized corpus software when dealing with large quantities of corpus. The powerful parallel corpus retrieval software for Chinese and various bilingual control searches is designed by Michael Barlow, a famous corpus expert. Corpus processing is inseparable from the support of application software, and the selection and use of data processing, storage and retrieval software is a great challenge for self-built corpora.

To enable the user to focus on English Chinese parallel corpus of two text vocabulary, syntactic and textual differences, for example, to identify two text length which is longer, the corresponding relationship between words, and to summarize the principles and rules in translation between English and Chinese, and find out the similarity between the two parts and the corresponding characteristic and reason analysis. CHINGLISH. The group's Foreign Trade Correspondence in Chinese English parallel corpus according to the needs of the two free text separately or in tandem, a English Chinese corpus correspond to any version of the bilingual parallel corpora, or as the translations of monolingual corpora, each will be able to control for small or large Chinese English and English. After each text sentence alignment and annotation on the corpus to be rich in content, good extension, data sources to improve the quality of the language Automatic link type database statement is likely to parallel retrieval, can also be used as the corpus research and statistical analysis effectively, such as statistics and the type and token frequency data, data comparison and analysis of the text of the word length, sentence length, sentence number, sentence patterns, stylistic features. The Chinese context

emphasizes the function of discourse, if the meaning of sense, are not concerned about how its form. Chinese language and logical thinking of foreign trade correspondence inevitable lack of information problems in foreign language understanding, logic processing based on. At this time, automatic link type parallel retrieval can help solve, because the comparative sentence patterns, stylistic analysis helps to find the reason, phenomenon of discontinuity or contradictory on can alleviate or get to solve.

## 6. Conclusion

Today, with the rapid development of information technology, the research group has processed, aligned, tagged, and retrieved Chinese and English bilingual texts of Foreign Trade Correspondence, and finally built a bilingual parallel corpus of Foreign Trade Correspondence with a certain scale. The use of parallel corpus retrieval function superiority in the construction process, but due to constraints, and the time is not allowed, eventually unable to avoid a lot of technical difficulty is too high, only a small simple Foreign Trade Correspondence bilingual parallel corpus, its construction is still the initial stage. But no matter how hard in the beginning, through it, many of the original, with comparison of characteristics of corpus, help scholars provide objective basis for the detection, for writing and translation studies. This study established the Foreign Trade Correspondence corresponding text corpus, for Chinese English, English comparative study, it promotes the collaboration with foreign trade English correspondence translation study to a higher level, we should continue to develop and use.

This work is supported the following fund :

2017 The student work subject program of Hezhou university:” college students' network addiction of the student work service research” (No,hzxysz201707).

2016 The project of improving the basic ability of young teachers in Colleges and universities in Guangxi:”An empirical study of Guangxi Beibu Gulf Economic Zone Foreign Trade Company foreign trade English correspondence ” (No,KY2016YB886).

2016 The higher education research subject project project of Hezhou university:” between collaborative innovation research of entrepreneurial talent cultivation system”.

2017 college students' innovative entrepreneurial project:” The agricultural robot autonomous operations research based on machine vision”(No,201711838096) &&”Multi-functional autonomous agriculture robot development”(No,201711838121) &&”Research and development of navigation system for agricultural robot”(No,201711838008) &&”Research and development of data acquisition system and servo control system for mobile platform of agricultural robot”(No,201711838075) &&”Research and development of visual navigation system for agricultural robot test platform”(No,201711838076).

2017 Guangxi education science 'ten-three-five' planning project:” Based on zte's ICT education platform of fusion, work-integrated learning education training mode research”(No,2017B107).

## References

- [1] B.Q LI, Y.F LING, H.Y ZHANG, S.Y ZHENG: The Design and Realization of Cherry Tomato Harvesting Robot Based on IOT. International Journal of Online Engineering, 12(12), 23, (2016).
- [2] B.Q LI, W.L GUAN, S.Y Zheng, X.G Yue: OPTIMISATION DESIGN OF CORN PRECISION SEEDER BASED ON MULTI-ROUTE AND MULTI-CHANNEL CONTROL. JOURNAL OF THE BALKAN TRIBOLOGICAL ASSOCIATION, 21(4A), 1215, (2015).
- [3] S.Y Zheng, et al, Social Work in Teen Addiction Correction Services Research Under the New Situation, in: PROCEEDINGS OF THE 2015 4TH NATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS AND COMPUTER ENGINEERING(NCEECE 2015), Xian, China, 2015, PP.252-255.



- 
- [4] B.Q Li, et al, Intelligent Control Management System and Its Application, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON ECONOMICS AND MANAGEMENT INNOVATIONS, Wuhan, China, 2016, PP.68-71.
- [5] B.Q Li, et al, Design and Implementation of Tanks War Game Based on the Android Platform, in: PROCEEDINGS OF THE 2016 2ND WROKSHOP ON ADVANCED RESEARCH AND TECHNOLOGY IN INDUSTRY APPLICATIONS, Dalian, China, 2016, PP.963-966.
- [6] B.Q Li, et al, Design of a Tea Garden Antifreezing Control System, in: PROCEEDINGS OF THE 2016 6TH INTERNATIONAL CONFERENCE ON MACHINERY, MATERIALS, ENVIRONMENT, BIOTECHNOLOGY AND COMPUTER(MMEBC), Tianjin, China, 2016, PP.736-738.
- [7] B.Q Li, et al, Design of Electronic Compass, in: PROCEEDINGS OF THE 2016 6TH INTERNATIONAL CONFERENCE ON MACHINERY, MATERIALS, ENVIRONMENT, BIOTECHNOLOGY AND COMPUTER(MMEBC), Tianjin, China, 2016, PP.1240-1243.
- [8] B.Q Li, et al, Research of Automatically Light-Adjusting Lamp, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING, INFORMATION SCIENCE & APPLICATION TECHNOLOGY (ICCIA 2016), Guilin, China, 2016, PP.249-252.
- [9] B.Q Li, et al, The Design and Realization of Fruit Harvesting Robot Based on IOT, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON COMPUTER ENGINEERING, INFORMATION SCIENCE & APPLICATION TECHNOLOGY (ICCIA 2016), Guilin, China, 2016, PP.261-264.
- [10] B.Q Li, et al, A New Type of Automatic Opening and Closing Light-Operated Curtain, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON MECHATRONICS ENGINEERING AND INFORMATION TECHNOLOGY(ICMEIT), Xian, China, 2016, PP.66-69.
- [11] B.Q Li, et al, Design of the Intelligent Air Humidifier, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON MECHATRONICS ENGINEERING AND INFORMATION TECHNOLOGY(ICMEIT), Xian, China, 2016, PP.201-203.
- [12] B.Q Li, et al, The Design Implementation of the APP of Experiencing Guangxi Folk Custom, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON ECONOMICS AND MANAGEMENT INNOVATIONS, Wuhan, China, 2016, PP.47-50.
- [13] S.Y Zheng, et al, Digital Display Design of Ethnic Clothing of Nanling, in: PROCEEDINGS OF THE 4TH INTERNATIONAL CONFERENCE ON MECHATRONICS, MATERIALS, CHEMISTRY AND COMPUTER ENGINEERING 2015(ICMMCCE 2015), Xian, China, 2015, PP.2805-2808.
- [14] S.Y Zheng, et al, Design and Implementation of Supermarket Personnel Management System Based On Java, in: PROCEEDINGS OF THE 2015 INTERNATIONAL CONFERENCE ON EDUCATION, MANAGEMENT AND COMPUTING TECHNOLOGY, Tianjin, China, 2015, PP.1724-1727.
- [15] S.Y Zheng, et al, Analysis of Internet of Things Talent Training and Curriculum System Innovation, in: PROCEEDINGS OF THE 2016 INTERNATIONAL CONFERENCE ON EDUCATION, MANAGEMENT AND COMPUTING TECHNOLOGY(ICEMCT-16), Hangzhou, China, 2016, PP.957-960.
- [16] S.Y Zheng, et al, Brief analysis on “HeYuanTong” Campus Mobile Phone APP Design, in: PROCEEDINGS OF THE 2015 5TH INTERNATIONAL CONFERENCE ON COMPUTER SCIENCES AND AUTOMATION ENGINEERING, Sanya, China, 2015, PP.151-154.