

# Application of Corpus and Terms Dictionary for Japanese Software Outsourcing Document Translation

Fangting Liu\*

College of Foreign Languages, Bohai University, Jinzhou, 121013, China

liufangting1984@126.com

\*Corresponding author

**Abstract:** The research results of corpus-based translology are constantly enriched, the methods and theories are more mature, and the substantiality and meticulousness of corpus-based translology are increasingly prominent. In order to solve some problems in the document translation of software outsourcing to Japan, this paper studies the application of corpus and term dictionary, and the research content is carried out from three aspects: First, in the field of translation practice, research on auxiliary document translation, provision of language materials, automatic machine translation and online electronic dictionary. The second is the field of translation studies, which studies the characteristics of document translation, the evaluation of document translation, the style of document translator and the comparison of document translation. Third, in the field of talent training, research on the construction of teaching models, translation and collocation teaching, exploring language laws and assisting autonomous learning. The research results are conducive to improving the accuracy of software outsourcing document translation to Japan and creating a new development pattern of software outsourcing services to Japan.

**Keywords:** Software Outsourcing for Japan; Document Translation; Corpus; Terms Dictionary

## 1. Introduction

Software outsourcing service is an important trend in the development of software industry in the world. According to International Data Consulting analysis, the global application software outsourcing market has grown at an average annual rate of 29.2% in recent years. According to the statistics of the US Business Week, one-third of the global software output value needs to be completed by outsourcing [1]. The software outsourcing market is developing rapidly on a global scale, which also provides a good opportunity for the development of China's software outsourcing industry. China is close to Japan, and the cultural background is very close, in these advantaged conditions, our country to undertake software outsourcing mainly to Japan's software outsourcing, this is the software outsourcing to Japan. In China, there are a large number of outsourcing companies to Japan in the cities of Dalian, Shanghai and Beijing.

In the process of software outsourcing to Japan, the mutual translation of software documents is an important task. Sometimes the Japanese documents need to be translated into Chinese, and sometimes the Chinese documents need to be translated into Japanese. At present, there are many problems in the document translation of software outsourcing to Japan, which are mainly reflected in the lack of document translation talents, the low quality of document translation, the slow development of machine translation and the low efficiency of manual translation, which have become an important factor restricting the sustainable development of software outsourcing to Japan. Corpus refers to a large scale electronic text base that has been scientifically sampled and processed. It is the basic resource of corpus linguistics research and the main resource of empirical language research methods. A parallel corpus is a bilingual or multilingual corpus composed of the original text and its parallel counterpart in the target language. In recent years, the research results of corpus-based translology have been continuously enriched, the methods and theories have become more mature, and the substantiality and meticulousness of corpus-based translology have become increasingly prominent [2]. This paper applies the software outsourcing document translation corpus and glossary to translation practice, translation research and talent training, and fundamentally solves the problem of software outsourcing document translation to Japan.

## **2. Field of Translation Practice**

### ***2.1 Assisted Document Translation***

All the data in the corpus have actually appeared in the actual use of language, and the real corpus needs to be processed to ensure that there are no omissions in the actual operation and the real significance of the corpus can be truly played. In addition to drawing on the advantages of traditional translation studies and general linguistics, translation corpora can also actively innovate corpus collection methods to make up for the shortcomings of traditional translation. The software outsourcing document translation corpus and glossary can assist document translation and improve the translation speed and quality. There is no missing translation in the translation, which basically conforms to the style of the original text, the wording is relatively accurate, and basically conforms to the customary translation or the conventional translation. And it can provide the best translation reference between Chinese software developers and Japanese software developers.

### ***2.2 Provide Language Materials***

In fact, through random sampling of natural language use, a corpus represents the overall language use determined in a certain study with a certain size of language samples. The corpus involves a variety of linguistic materials, including dictionaries, manual corpus and text electronic corpus. The collected corpus is basically typical language materials, and certain principles and ideas should be followed in the application process. Based on the acquired corpus data, the analysis and research of vocabulary and the study of syntactic spelling should be combined to improve the accuracy of corpus use. Intuitively, bilingual sentences translated from each other should be similar in length. After tagging the parts of speech of each sentence, we can judge whether two sentences translate each other by the frequency of words that appear in the same part of speech. If the text before the sentence alignment translates each other, then the positions of the two sentences in their respective paragraphs must be similar.

### ***2.3 Machine Automatic Translation***

Machine translation technology involves computer science, cognitive science and linguistics, and has been recognized by the scientific community as one of the most difficult topics in the field of artificial intelligence. The machine translation system has a large number of dictionaries and translation models that can cover multiple fields and industries, and can translate in different scenarios. With the development of artificial intelligence technology and the deepening of brain science research, machine translation will replace human translators to undertake most of the translation work. The biggest advantage of machine translation is the accuracy and ease of understanding of the translated content [3]. In the field of Japanese software outsourcing document translation, case-based machine translation is widely used, and translation knowledge is expressed in the form of examples and dictionaries, which is easy to add or delete. If a large library of translation examples is used and accurate comparative analysis is carried out, high-quality translations may be generated, and the difficulty of deep linguistic analysis must be avoided in traditional machine translation methods [4].

### ***2.4 Online Electronic Dictionary***

With the popularization of computers and the Internet, people are getting more and more accustomed to using electronic documents. The core function of the dictionary is to provide an online application. When translators translate documents, they can refer to the online electronic dictionary at any time. Compared with paper dictionaries, electronic dictionaries have a huge advantage [5]: they have a rich vocabulary, and if the same vocabulary appears in paper dictionaries, the length will increase significantly. The revision cycle is short, the revision cycle of paper dictionaries is very long, and the revision cycle of electronic dictionaries is relatively short, and it can be revised at any time. Ease of use, can be instant translation, instant pronunciation, fast response, these are unmatched by paper dictionaries. Artificial intelligence technology, electronic dictionary makes full use of the convenience provided by artificial intelligence, can carry out fuzzy query, and provide interactive functions for users. Multimedia technology, electronic dictionary can use audio-visual, animation, color and other multimedia means to visually explain the vocabulary.

### **3. Field of Translation Studies**

#### ***3.1 Document Translation Features***

The translation characteristics of a document refer to the linguistic characteristics of the translated text that are different from the original text due to a series of factors such as the differences between the source text, translator, reader, source language and target language, as well as social and cultural factors, including the lexical, syntactic and textual aspects of the translated text, which mainly refer to the universal characteristics of translation. The universal features of translation are some typical, cross-linguistic and universal features presented in the target text that are different from the source text, mainly simplification, normalization and manifestation. Simplification refers to the simplification of the information in the source language text in the target language text, the use of fewer types of vocabulary, the reuse of common words, the reduction of substantive words, the increase of function words, and the language difficulty of the translated text is generally reduced. Normalization refers to the tendency of the translated text to follow the tradition of the target text in terms of punctuation, vocabulary choice, style, sentence structure and text structure. Manifestation is to interpret the linguistic information of the source text from the perspective of discourse, so as to make the cohesive relationship of the translation clearer.

#### ***3.2 Document Translation Evaluation***

Terminology dictionaries can not only query the meaning of words, word collocation and grammatical functions, but also provide specific information such as the frequency of use of words, the style of the text in which the words are located, the specific context and the source of the text, and can be compared with other words in parallel. Compared with the paper dictionary, the glossary of terms is timelier. In the field of software outsourcing to Japan, new words are constantly being created, and paper dictionaries cannot be updated in a timely manner due to compilation and printing restrictions. The term dictionary stored in electronic information can be updated in real time and queried in real time. The advantages of translation evaluation based on corpus or glossary are simple retrieval, fast query speed and accurate query results. Fuzzy matching, sorting and wildcard characters can be used to quickly and accurately locate the content to be queried [6]. The use of corpora can significantly reduce the subjective factors in translation evaluation, provide objective and effective constructive feedback, and help translators identify and correct errors in translation more accurately, quickly and efficiently.

#### ***3.3 Document Translator Style***

The study of translator style in corpus translatology describes the regular linguistic features of the translated text or translation process through corpus analysis and quantitative statistics, and explores and explains the causes behind them. The study of translator style can be extended to the study of the regularity of a certain linguistic feature in the source language or text in different translations. Corpora provide a rational quantitative analysis method for large-scale data statistics of translatology, which can more intuitively reveal the translator's translation language style. Most of the existing corpus studies on translator's style are satisfied with the description of the specific representation of the translator's style, but do not summarize the translator's style as a whole. The translator's local style in translation is not equal to the overall style, and there are great differences between the two. The investigation of translation style of translators should not be limited to traditional parameters such as average word length, average sentence length and standardized category, but should be expanded to semantic, pragmatic, rhetorical and sociocultural parameters, and the research methods of corpus stylistics, quantitative linguistics and computational linguistics should be used for reference, and quantitative statistics and qualitative analysis should be combined to broaden the research scope of translation style or style [7].

#### ***3.4 Document Translation Comparison***

Taking large-scale bilingual corpus as the object, the diachronic or consensus interpretation of translation phenomena is carried out by combining intralingual contrast and interlingual contrast. Specifically, it includes the following indicators [8] : First, vocabulary richness, which is usually represented by "type token ratio". The higher the ratio, the richer the vocabulary of the text. However, this index is easily affected by the length of the text, and the longer the text, the smaller the value. Second, lexical density, reflecting translation style and text readability, is the ratio of the number of real word characters in the text to the number of total shape characters in the text. The number of character types reflects the amount of text information. With the same amount of information, the less content words, the more concise the text. Third, the activity of the text, the proportion of the total number of

verbs in the text to the sum of the total number of descriptive verbs, reveals the balance of the text between description and narrative. The purpose of examining the mobility of two translated documents is to explore the language usage tendency of different translated documents. Fourth, the average sentence length, expressed by the total number of symbols divided by the period mark, reflects the complexity of the sentence, the longer the average sentence, the higher the complexity. The shorter the average sentence, the less complex it is.

#### **4. Field of Personnel Training**

Translation is not only an important language skill, but also an important basis for measuring learners' language application ability. The development of corpus-based translation teaching is of great value to training Japanese document translation talents.

##### ***4.1 Construction of Teaching Model***

To construct a teaching model of corpus translation and promote the rational use of corpus in Japanese translation teaching. Corpus-based translation teaching is a learner-centered teaching model that attaches particular importance to students' participation and interaction, and teachers play more roles as guides and organizers [9]. First of all, before class preparation, teachers set up a corpus teaching mode according to the actual situation of students, so that students can master the knowledge of Japanese translation through the corpus and set up Japanese classroom teaching activities. Secondly, for students, before class teaching, according to the teacher's guidance, they learn relevant knowledge in the corpus, translate relevant Japanese texts, and complete the preparation work before class. Thirdly, in classroom teaching, the teacher showed the corpus to the students, strengthened the students' grasp of translation knowledge, divided the students into several groups, guided the students to find problems, and asked the students to explore the corpus through collaborative learning based on their Japanese knowledge level. Finally, after class, students should be guided to reflect on teaching. After class, they can use the corpus to learn knowledge again and carry out discussion activities, so as to enhance their further mastery of Japanese translation knowledge and give play to the value of the corpus in Japanese translation teaching.

##### ***4.2 Translation Collocation Teaching***

Collocation is the combination of words in language expression, collocation is the focus of translation teaching, relying on teachers to explain the conventional usage, to provide students with limited forms of translation collocation. Modern corpus lexical retrieval research shows that mastering appropriate collocation is an important sign to distinguish native speakers from non-native speakers, and has become an important field of corpus linguistics research. In translation teaching, students' collocation knowledge of target language is not rich enough, and due to the influence of the collocation mode of the source language, inappropriate collocation often occurs in the translation. Collocation has an implicit feature, and it is difficult for students to directly observe different collocation forms, so it is necessary to use a large corpus to broaden students' vision of collocation translation, gradually improve their intuitive understanding of translation in the search and observation, and find the best collocation form.

##### ***4.3 Exploring Language Rules***

From the study of translation universals and characteristics in specific languages to the study of translators' evolving styles, from the analysis of translation norms to the exploration of translation practices and strategies, corpora play an active role in all fields of translation, which not only enriches the diversity of translation studies, but also provides unlimited possibilities for the new development of translation. Corpus translation teaching focuses on the recipient's perception and experience of translation in the context of social culture and linguistic communication, and improves the learner's cognitive ability of the relationship between translation subject, translation object and recipient. Bilingual parallel corpus has the closest relationship with translation teaching. Through retrieval software, dynamic context can be realized, and collocation of words or structures can be observed in specific contexts, as well as semantic features in different contexts [10]. In parallel corpora, students can form their own translation views and translation strategies.

##### ***4.4 Assisted Autonomous Learning***

Autonomous learning ability refers to the ability to adjust thoughts, emotions and actions in a

planned way in order to achieve personal goals, including the ability to set learning goals, focus on learning, use resources effectively, manage time effectively, hold positive beliefs about personal abilities, and cognitive abilities including belief, perception and understanding. The corpus-based Japanese translation teaching model can cultivate students' self-learning ability, as well as their ability to independently recognize and solve problems. In the process of using the corpus to learn, students also complete the task. They must set the appropriate search conditions according to the task requirements and the characteristics of the corpus, and summarize and analyze the relevant information. Through thinking, discussion, debate, mutual revelation, correction, retrieval of corpus, analysis of results and publication of conclusions, students not only acquire relevant knowledge, but also cultivate independent learning ability and team cooperation ability. Emphasis is placed on the cultivation of learning environment and teaching environment conducive to students' independent learning, and the supply of various resources required to assist students' self-study and group learning [11].

## 5. Conclusions

Software documents are materials related to the development, maintenance and use of software, and are technical data and information in a form that people can read. Software documentation specifies software design details, describes software functions, and describes how to use the software. In the process of software outsourcing to Japan, the mutual translation of software documents is an important task. Sometimes the Japanese documents need to be translated into Chinese, and sometimes the Chinese documents need to be translated into Japanese. In order to solve the problems existing in the translation of Japanese software outsourcing documents, this topic studies the application of Japanese software outsourcing document translation corpus and glossary, improves the accuracy of Japanese software outsourcing document translation, and creates a new development pattern of software outsourcing services to Japan.

## Acknowledgements

This work is supported by 2023 annual social science planning fund project of Liaoning province (L23BYY014): Compilation and application of terminological dictionary for Japanese software outsourcing document translation based on corpus.

## References

- [1] S. J. Lu. *Research on college talent training model for software outsourcing industry*[J]. *University Education*, 2015, 4(10): 24-26.
- [2] X. M. Tong. *Manifest and implicit analysis of corpus-based translation*[J]. *Journal of Shenyang Normal University(Social Science Edition)*, 2015, 39(05): 125-127.
- [3] J. Zhu, M. Gu. *Corpus-based machine translation*[J]. *Modern Communication*, 2019, 33(17): 100-101.
- [4] Z. W. Feng. *Corpus-based machine translation system*[J]. *Terminology Standardization & Information Technology*, 2010, 15(01): 28-35.
- [5] H. F. Gu. *A brief introduction to the status quo of bilingual electronic dictionaries in China*[J]. *China Science and Technology Information*, 2010, (22):154-156.
- [6] H. M. Wang. *An Empirical Study of Evaluation Corpus and Translation Quality Evaluation Based on Translation Teaching*[J]. *Journal of Wenzhou University(Social Science Edition)*, 2020, 33(04): 93-101.
- [7] L. B. Huang. *Reflections on the corpus studies of translator's style*[J]. *Foreign Language Education*, 2018, 39(01): 77-81.
- [8] D. Han, Z. J. Zhang. *A Corpus-Based Comparative Study of the Translation Styles in the Russian Versions of The True Story of Ah O*[J]. *Foreign Languages and Their Teaching*, 2023, (01): 105-116+148.
- [9] Y. G. Cai, X. Yu. *Corpus-based University English-Chinese Translation Teaching*[J]. *Journal of Guangzhou City Polytechnic*, 2021, 15(03): 11-15.
- [10] L. L. Yu. *Innovative Study of Corpus Applied in College Translation Courses*[J]. *Journal of Chengdu Aeronautic Polytechnic*, 2020, 36(03): 37-39+49.
- [11] X. Chen. *Translation teaching design based on parallel corpora*[J]. *Overseas English*, 2020, 21(04): 12-13.