Revista Argentina de Clínica Psicológica
2020, Vol. XXIX, N°5, 1651-1658
DOI: 10.24205/03276716.2020.1161

Research on the Extraction Method of Translation Equivalence in Chinese and English Comparative Corpus

Zifu Liu

Abstract
In English translation teaching, English-Chinese contrast method has important guiding significance. Based on this, this paper focuses on the application of English-Chinese contrasts in translation teaching, starting with the introduction of English-Chinese contrast theory and analyses the application advantages of English-Chinese contrasts in translation teaching. Finally, this paper proposes the specific application content of English-Chinese comparison in translation teaching. I hope that the study of this paper can provide reference suggestions for cultivating students' English thinking habits. In addition, after a simple English-Chinese correspondence, students can understand the connotation and extension of the English language and improve the students' ability to express English correctly.
Keywords: Chinese and English comparable corpora, Translation equivalence, Extraction method

1. INTRODUCTION

Multimedia assistance refers to multimedia teaching refers to the teaching process, according to the teaching objectives and the characteristics of teaching objects, through the teaching design, reasonable selection and use of modern teaching media, and organic combination of traditional teaching methods, to participate in the whole process of teaching, to many This type of media information acts on students, forms a rational teaching process structure, and achieves optimal teaching results. With the development and popularization of science and technology, multimedia computer teaching has gradually applied to various fields of education and replaced traditional teaching methods to play an important role (Fischer, 2007). It is of great significance to use multimedia-assisted teaching in the teaching of junior high school English. Teachers can create lively, lively, funny and realistic life and learning scenarios by combining multimedia approaches and situational teaching methods. Multimedia can use its wide range of features across time and space resources to help English teaching and teaching is conducive to the construction of superior dialogues, thereby creating a good learning atmosphere, expanding the students' learning content and inspiring students' interest in learning. In junior high school English learning, English grammar is a big problem for many students in learning English. The English expression habits are different from those of Chinese, and their corresponding grammars are also different. Students cannot easily and easily grasp the content of knowledge points when learning grammar knowledge points. At this time, teachers can use multimedia assistance to solve difficult situations (Fukumoto & Suzuki, 2007). They will translate grammar points that are more difficult for students into simple dialogue problems or reading, allowing students to grammar in situations. The real purpose of translation teaching is to cultivate students' bilingualism and awareness of bilingualism. To effectively carry out translation teaching, it is necessary to increase the guidance of students' morphology, meaning, and grammar in the transition between the English and Chinese languages, so that students can understand the characteristics of their respective culture, context, and thinking. Therefore, it is of great significance to strengthen the application of English-Chinese contrast in translation teaching.

In the field of English translation teaching and research, the Chinese-English contrast method is one of the important means (Morin & Hazem, 2016). This method began to be applied in the 1930s and it really took off in the 1950s. In 1957,

School of Foreign Studies, Ankang University, Ankang 725000, Shaanxi Province, China

the American linguist Robert put forward the application basis of comparative analysis in the study of language, and laid the foundation for the modern application of contrast method. Since the end of the 20th century, China has increased its emphasis on comparative analysis methods. Many scholars have studied macro and micro aspects based on the internal structure of language. From the macro level, the contrast between English and Chinese reflects the difference between Chinese and Western ways of thinking. From a microscopic perspective, the English-Chinese contrast method is more focused on the comparison and analysis of English and Chinese synchronicity, and is also associated with two languages. Most students have a basic grasp of English grammar and understanding, but also formed a good sense of language in English reading exercises, in certain procedures, you can use English and partners for the most basic exchanges. However, our students still show deficiencies in English practice (Talvensaari, 2008). Especially in the grasp of English vocabulary, as well as the semantic grasp of English there are defects. These deficiencies can directly lead to the misuse or misuse of words in English communication. In fact, in English learning, only by mastering enough vocabulary can we master the core of English. In addition, there is a fundamental gap between Chinese language learners and their English-speaking users. This is the difference in the number of vocabulary mastering. This is also a major problem affecting the quality of English teaching. In the process of translation teaching, English and Chinese contrast methods are applied to enable students to master and correctly use English words (Udupa & Khapra, 2010). In addition, after a simple English-Chinese correspondence, students can understand the connotation and extension of the English language and improve the students' ability to express English correctly.

2. TRANSLATION TAECHING APPLICATION ADVANTAGES

2.1. Sampling methods

In international cultural exchanges, Chinese-English translation is one of the most important contents. The process of language translation is the way in which different ethnic languages and cultures permeate and transmit each other. Different nations have great differences in terms of language environment, language model, etc., and they have different ways of organizing languages. This cultural difference increases the complexity of translation. In the process of translation teaching, teachers can choose the translation examples that they think are most suitable for students' language learning. After the students' interpreters, the teacher finally gives the correct answer to allow the students to compare, and then enable the students to fully understand the differences in language and culture. In this paragraph, many words are typical Chinese expressions. But for English readers, their political thinking may be far different from Chinese politics. Therefore, translation is not only a literal conversion of the two languages, but its focus is on human communication. Therefore, in the translation teaching, language contrast methods should be used to strengthen the deep contrast between the languages of the two countries so that the students can use different methods to practice translation in the actual translation. In the past ten years, the study of cultural linguistics has made great progress, but there are also some deficiencies, such as the imbalance between the internal and external research of the discipline, and the depth of the study is not enough. In this regard, we need to rethink and strengthen the study of cultural linguistic theories, and conduct in-depth research to promote its balanced development. Overall, the theoretical foundation of Chinese cultural linguistics is not particularly profound

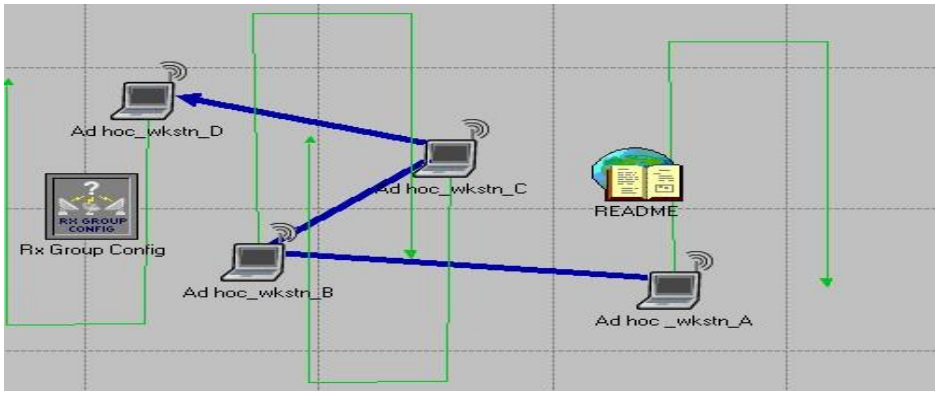

Figure 1. **Network topology**

2.2. Syntactic comparison and translation teaching

Due to the influence of the translation room, the Sino-British parallel corpora are inherently skewed language models. Obviously, the natural language processing systems such as machine translation and cross-language retrieval trained with such a corpus inherit the skewed language model, which seriously affects the performance of the application system. In order to overcome the inherent deficiencies of the parallel corpus, a technical study of constructing and analysing Chinese-English triples comparable corpus is proposed. In this study, a comparative corpora and language techniques were used. Using the combination of statistics and rules, the statistical analysis of the native English and Chinese English of the triple comparable corpora was performed. The research content proposed in this paper not only has practical value for the improvement and development of cross-language processing applications, but also has important significance for foreign language teaching, dictionary compilation, and foreign exchange and cooperation. The triple communicative corpus is the basic resource for carrying out this research. So far, a million-sentence triplet comparable corpora has been built. The original corpus for constructing the corpus mainly comes from dozens of kinds of research reports published by the Institute each year, with a total of more than 2 million English words. In order to ensure the accuracy and readability of the translation of the research report, all English translations of the report must be strictly modified and edited by native language linguists.

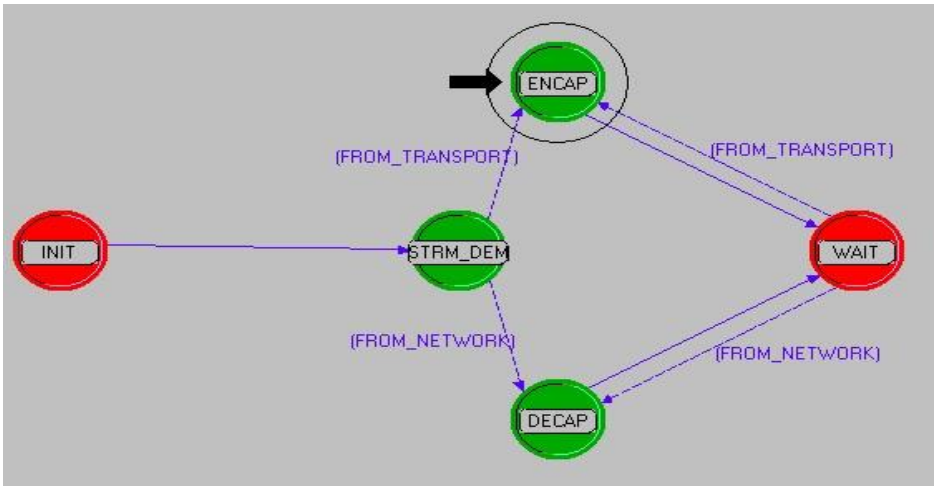

Figure 2. **A comparative study of bilingual corpus term extraction methods**

3. DOMESTRIC RESEARCH

3.1. Automatic clustering of keyword clusters

At present, the research on the construction and analysis of the triple comparable corpora has yielded effective results at the vocabulary level. It overcomes the inherent skewed language model of the parallel Chinese corpora, and builds and mines bilingual lexicon based on native language models. It is of great practical value to improve natural language processing applications such as machine translation. The machine translation system embedded in this research has been widely used at home and abroad. In the future, according to the research method of this article, we can also analyze the significance of differences between the level of speech and the level of semantics. The goal of this study in the future is to extend the study of comparable texts based on keyword and keyword clustering methods to macroscopic studies of comparable texts based on key semantic fields to support content analysis. In this way, the quantitative analysis of the specific triples of comparable corpora can be extended to the qualitative analysis of generalized content-based comparable texts, which effectively expands the research and application of comparable corpora.

The power of the received signal is given by the formula:

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \quad (1)$$

It received signal envelope follows the Rayleigh distribution:

$$f(r) = \frac{r}{\sigma^2} exp\left(\frac{-r^2}{2\sigma^2}\right) \quad r \in [0, \infty) \quad (2)$$

Its probability density function:

$$f(r) = \frac{r}{\sigma^2} exp\left(-\frac{r^2+s^2}{2\sigma^2}\right) J_0\left(\frac{rs}{\sigma^2}\right) \quad r \in [0, \infty) \quad (3)$$

Its time domain channel impulse response function:

$h(t,\tau) = \sum_{l=0}^{L-1} h_l(t) e^{j(2\pi f_{dl} t + \theta_l)} \delta(t - \tau_l(t))$ (4)

Due to the OFDM systems, therefore:

$$x(n) = \sum_{i=0}^{N-1} X_i \, exp(j \frac{2\pi}{N} kn) = \sum_{i=0}^{N-1} X_i \, exp(j2\pi f_k t_n), \quad n = 0,1,2,\dots,N-1 \text{ (5)}$$

Integrating the period T can restore the original signal:

$$Y_l = \frac{1}{T}\int_0^T exp(-j2\pi f_l t) \sum_{i=0}^{N-1} X_i \, exp(j2\pi f_i t_n)\, dt = \frac{1}{T}\sum_{i=0}^{N-1} X_i \int_0^T exp(j2\pi(f_i - f_l)t)\, dt = X_l \text{ (6)}$$

$$\frac{1}{T}\int_0^T exp(j2\pi f_k \tau) \cdot exp(j2\pi f_l \tau) d\tau = 0, \quad k \neq l \text{ (7)}$$

And there is only one non-zero element:

$$s_i = [s_i(0), s_i(1), \dots, s_i(T-1)]^T = \left[\underbrace{0,0,\dots,0}_{d-1}, e^{j\phi}, 0, \dots, 0\right]^T \text{ (8)}$$

Receiving end forms a received signal vector:

$$\overbrace{\begin{bmatrix} y_{ks+N_c-T+2} \\ \vdots \\ y_{(k+1)s} \end{bmatrix}}^{\boldsymbol{y}_k} = \left[\boldsymbol{O}_{(1)} \middle| \begin{matrix} \boldsymbol{h} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{h} \end{matrix} \middle| \boldsymbol{O}_{(2)}\right] \cdot \begin{bmatrix} \boldsymbol{P} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{P} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{P} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{IF}_N & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{IF}_N & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{IF}_N \end{bmatrix} \cdot \begin{bmatrix} x_{1:N}(k-1) \\ x_{1:N}(k) \\ x_{1:N}(k-1) \end{bmatrix} + \overbrace{\begin{bmatrix} n_{ks+N_c-T+2} \\ \vdots \\ n_{(k+1)s} \end{bmatrix}}^{\boldsymbol{n}} = \boldsymbol{H} \cdot \boldsymbol{X} + \boldsymbol{n} \text{ (9)}$$

1) Chinese and English color meaning comparison

English and Chinese color words not only have the function of expressing color, but also express people's inner activities, thoughts and emotions. In the translation process, translators cannot stay on the surface of the original text, but also grasp the specific historical background and social customs behind the color words. In recent years, traditional Chinese medicine and pharmacy have been widely accepted by their own advantages, and overseas influence has been increasing. To fundamentally promote the long-term healthy development of Chinese medicine in overseas countries, the key is to achieve accurate and complete English translation of the basic theory of Chinese medicine abroad and its widespread dissemination. The basic theories and thinking methods involved in the basic theories of traditional Chinese medicine are the basis for learning and understanding other disciplines of traditional Chinese medicine. It is precisely this type of discipline that determines its important position in the study and clinical application of traditional Chinese medicine and the external communication of traditional Chinese medicine.

2) The translation method of color words in different contexts

This article selects Chinese and English original classical Chinese and English textbooks as the source of corpus. It explores and constructs Chinese and English comparable corpus of TCM basic theories. It provides new ideas for studying the accuracy and completeness of English translation of TCM basic theories and overseas dissemination. On this basis, taking the yin and yang theory chapter as an example, the differences and similarities between Chinese and English materials will be compared and analyzed from the four aspects of TCM terminology, information elements, concept definition and interpretation methods, and the current status of TCM basic theory in foreign countries will be studied. It also provides reference for Chinese translation of Chinese medicine. At the same time, from the point of view of accepting aesthetics from the perspective of cross-cultural communication, the differences in the methods of interpretation between Chinese and English are discussed in order to increase the effectiveness of Chinese medicine's external communication and make Chinese medicine more acceptable to more foreign audiences.

3.2. Research tools and methods

The theoretical knowledge and cultural connotation contained in TCM mean that English translation and external communication of Chinese medicine not only include the transformation of Chinese-English language itself, but also face severe challenges that span Chinese and Western cultural differences and communicate tradition and modernity. How to use English accurately and effectively to disseminate TCM academic ideology and theoretical connotation is the key and difficult problem to be solved urgently. However, the current translation practice and external communication of TCM basic theories are not ideal. On the one hand, English-language textbooks published in China are mainly translated by English language professionals in the country according to the abbreviated version of the relevant materials. They are mostly published in bilingual Chinese and

English versions. The quality of the translation is often influenced by the professional knowledge of the translator. Corpus Translation Studies is a corpus-based translation study that can provide a large amount of real corpus and statistical data for translation studies, enhance the persuasiveness of conclusions, and can be applied to translation commonality, translator style, translation practice, and teaching. In recent years, with the expansion of corpus application scope.

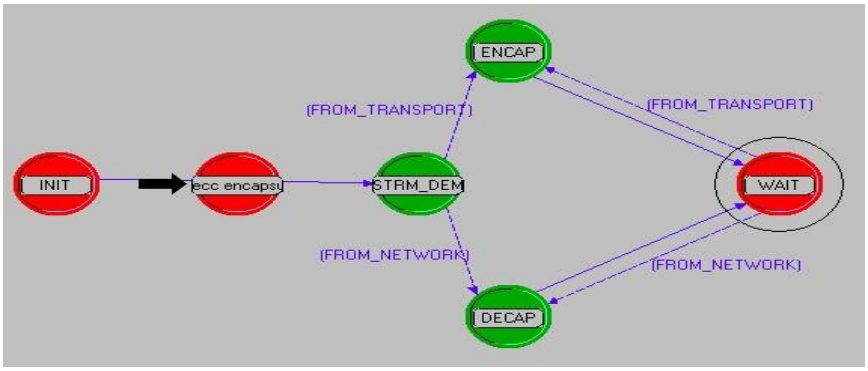

Figure 3. **Construction of comparable corpora**

3.3. Corpus translation studies provides a new method for English translation studies

The corpus can be divided into different types. The parallel corpus is composed of original texts and parallel translated texts. It can be used to compare the original texts with the target texts in a certain language. Therefore, it is widely used in translation research. However, in the field of Chinese-English translation of basic theories of Chinese medicine, the Chinese-British basic books on traditional Chinese medicine published in the country conform to the basic conditions for constructing Chinese-English parallel corpora. However, influenced by the length of the text and the language level of the translator, it is difficult for both the content integrity and the language itself to meet the requirements for constructing a high-level corpus for corpus. The books on basic theories of traditional Chinese medicine published abroad are mainly English original texts. Therefore, in view of the current status of the English translation of basic theories of traditional Chinese medicine, the conditions for constructing parallel corpus are not available for the time being, and the construction of Chinese-English comparable corpus is a more feasible and ideal choice. In some literatures, corpora made up of translated texts and original texts in the same language with similar subject matter are called comparable corpora.

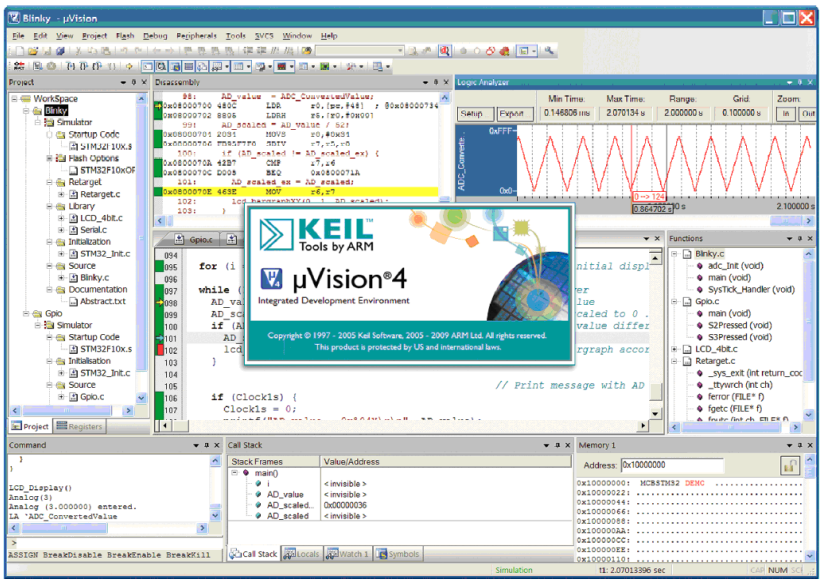

Figure 4. **Overview of comparable corpus and basic topics**

4. EMPIRICAL ANALYSIS

4.1. Corpus analysis

Chinese medicine translators in Chinese medicine began to explore the construction of TCM English-language corpus to carry out relevant research. On the whole, the current research on the English-Chinese translation corpus remains mainly in the stage of theoretical study and exploration. There is less research on specific practical applications, and the focus is on building a small bilingual corpus of ancient Chinese medicine. TCM terminology is the core unit of TCM theory, and it is also the most important aspect of TCM's external communication. Compared with previous comparison studies based on dictionaries or translation standards, research on terminology based on comparable corpora can provide reference for terminology translation based on the actual context of real corpus and ensure authentic translation and use of context. In view of the status of the classic textbooks of English corpus in Western Chinese medicine education, the English expression of medical terms has considerable influence. It can be said that the use of English expressions in TCM terminology is actually reflected in a certain extent. By comparing the Chinese and English terms used in the English-language terminology, we can provide reference for domestic terminology translation. In order to further compare the English translation of Chinese terminology, this study also lists the English translation of the terms used in the appendix of Chinese textbooks.

Table 1. **Chinese and English comparable corpus section subject pairing table**

Number	X	Y	Z
500	89.190%	90.283%	89.737%
600	89.578%	90.738%	90.158%
700	89.609%	90.485%	90.047%
800	89.797%	90.940%	90.369%
900	89.728%	90.576%	90.152%
1000	89.701%	91.009%	90.355%
50	87.684%	89.521%	88.603%
100	88.637%	90.070%	89.354%
150	89.053%	90.375%	89.714%
200	89.701%	91.009%	90.355%
250	90.524%	91.284%	90.904%
300	90.987%	91.578%	91.283%
1100	71.987%	78.654%	75.320%
1200	73.594%	77.364%	75.479%
1300	74.984%	76.489%	75.736%

4.2. The term Sino-British expression control study

In addition to the pairing analysis of TCM terminology in Chinese English, a further comparative analysis of the Chinese and English versions of TCM concepts represented by the term will also help further the study of TCM theories in English translation and external communication. The focus of this part of the study is to translate the concepts expressed in English corpus back into Chinese and compare them with Chinese language materials. If the back translation is better, it reflects the consistency of the understanding of the theory itself, and can be used as a reference and reference for the English translation of traditional Chinese medicine. It is necessary to further analyze the reasons for the poor translation back-to-back expression, whether it is caused by the different interpretation of the concept of Chinese medicine from the two perspectives or even different understandings. The bilingual corpus's two-language texts are entirely different descriptions of events by writers or reporters in their native language. For sporting events, they are likely to contain reports of the same event and are therefore comparable. The comparable corpus does not have the shortcomings of the parallel text corpora that are limited by the original text, and it is very hopeful to extract the true corresponding bilingual word pairs from the bilingual comparable corpora. At the same time, since this extraction can be performed immediately after acquiring a new bilingual comparative corpus, the extraction and application of bilingual new words can be well solved.

Table 2. **Yin and Yang theoretical terms Chinese and English**

Month	Nominal exchange rate	Real exchange rate
1994	862.12	74.84
1995	834.90	80.43
1996	831.43	87.46
1997	828.97	98.1
1998	827.91	92.64
1999	827.83	90.2
2000	827.84	95.41
2001	827.70	98.25
2002	827.70	91.91
2003	827.70	86.01
2004	827.68	81.89
2005	819.17	87.66
2006	797.18	86.65
2007	760.40	90.59
2008	694.51	102.69
2009	683.10	97.19
2010	665.15	101.47
2011	632.81	107.77
2012	629.31	110.11
2013	611.6	118.75

4.3. Effective of evaluation index

The most research work around the comparable corpus is the acquisition of new terminology and terminology for bilingual dictionaries. Many researchers in the United States, Britain, France, Japan, Germany, Singapore, and Hong Kong have conducted pioneering research. The method of bilingual vocabulary extraction based on comparable corpora is different from the translation dictionary acquisition based on parallel corpora. Translation vocabulary extraction in parallel corpora generally uses techniques such as paragraph alignment, sentence alignment and word alignment. The extracted translation dictionary is based on sentence pairs and can rely on the alignment of sentence pairs. The basic idea of bilingual vocabulary extraction based on comparable corpora is based on the basic assumption that when a word in a language corresponds to another language, the co-occurrence collocation between the word and its surrounding words is still maintained. There is similarity between the context of a word and its corresponding word in different languages.

In the algorithm implementation, experiments can be performed by setting different options. For example, the context vector value may be modified by controlling whether only the content word is calculated and the word class information of the seed word is considered. You can also choose to give different weights to the Chinese-English dictionary when there are multiple pairs of translated English expressions to distinguish the differences in their contribution to similarity. The referential corpus referred to in this article refers to a corpus composed of original texts in different languages with highly similar topics, and the former is called an analog corpus for distinction. Translation-oriented bilingual corpus that can embody the macro-text structure and micro-language features can provide a theoretical reference for translation practice and teaching.

5. CONCLUSION

The above experimental results show that the extraction method of translation equivalence pairs in the Chinese-English comparable corpus is effective in this paper, which can help people to extract equivalence pairs of vocabulary from comparable corpora. This provides valuable resources for applications such as machine translation. At the same time, it should also be noted that the scale of the experimental corpus used in this paper is still relatively small and the problems reflected are not necessarily comprehensive. In the next work, it is planned to further expand the scale of the experimental corpora, and at the same time try to apply the above-mentioned extraction algorithm to the automatic extraction of translation equivalent pairs in different professional fields, and further improve according to the processing effect. Considering the relationship of computational complexity, this experiment focuses on the automatic extraction effect of a Chinese word and an English word translation equivalent pair, ignoring other conditions during calculation. Since the emphasis is on the case where one Chinese word corresponds to one English word in the experiment, and the longer Chinese word generally corresponds to multiple English words, Chinese words longer than four Chinese characters are also ignored. Among them, stop words are excluded from the calculation of co-occurrence values. In this experiment, the selection criteria of stop words are mainly based on part of speech. Functional words are included in the stop word list, and co-occurrence values are calculated for real words. On the one hand, it can simplify the selection criteria of stop words, and it can also enhance the degree of use of language knowledge in the judgment.

Acknowledgement

The research in this paper was supported by Shaanxi Provincial Education Department Project: Research on the Study on External Communication of Hanjiang River Culture in the Belt and Road Environment.

References

[1] Fischer, K. (2007). “The role of users’ concepts of the robot in human-robot spatial instruction”. *Lecture Notes in Computer Science, 4387*, 76-89.

[2] Fukumoto, F., & Suzuki, Y. (2007). “Topic tracking based on bilingual comparable corpora and semisupervised clustering”. *Acm Transactions on Asian Language Information Processing, 6*(3), 11.

[3] Fung, P., Zweigenbaum, P., & Rapp, R. (1998). “Proceedings of the 2nd workshop on building and using comparable corpora: from parallel to non-parallel corpora”. *Lecture Notes in Computer Science, 1529*, 1-17.

[4] García, N. R. (2006). “Using comparable corpora for english-spanish contrasts: implications and applications in traslation”. *Computer Graphics Forum, 32*(6), 1–23.

[5] Morin, E., & Hazem, A. (2016). “Exploiting

unbalanced specialized comparable corpora for bilingual lexicon extraction". *Natural Language Engineering, 22*(4), 575-601.

[6] Rahimi, Z., & Shakery, A. (2011). "Topic based creation of a persian-english comparable corpus". *Lecture Notes in Computer Science, 7097*, 458-469.

[7] Rauf, S. A., & Schwenk, H. (2011). "Parallel sentence generation from comparable corpora for improved smt". *Machine Translation, 25*(4), 341-375.

[8] Sanjika, H., & Stephan, V. (2016). "Extracting parallel phrases from comparable data for machine translation, â". *Natural Language Engineering, 22*(4), 549-573.

[9] Talvensaari, T. (2008). "Comparable corpora in cross-language information retrieval". *Acm Transactions on Information Systems, 25*(1), 79-82.

[10] Udupa, R., & Khapra, M. M. (2010). "Transliteration equivalence using canonical correlation analysis". *Lecture Notes in Computer Science,5993*, 75-86.

[11] Yoo, & Jeongju. (2013). "Using diy corpus in legal translation: english translation of causative verbs in korean statues". *Mathematical Problems in Engineering, 2013*(1), 133-174.