Research on part-of-speech relationship between English-Chinese translation based on corpus

Zhaoxia Liu[*]

University of Sanya, Sanya, Hainan, China
[*]*Corresponding author*

Abstract: Based on the self-built English and Chinese translation corpus, this paper compares the differences between Chinese translation and English original texts, including English translation and Chinese original, makes clear the scope of common parts of speech to distinguish the category of text, analyzes and thinks about the differentiation of text language classes and the correlation degree of various language characteristics, and analyzes and studies the verb changes in English and Chinese translation.

1. Introduction

In the process of translating language, it is inevitable that there will be traces left after translation, and these traces are often regarded as a special language body. In many studies, this language body is referred to as a third language code independent of the pre-translation language and the translated language. The development and emergence of corpus linguistics has helped many linguists discover the characteristics of translation languages, that is, translation commonality, which proves that translation languages are not affected by specific languages. As the basic grammatical scope, the use of parts of speech is an important basis for linguistic analysis, which is not only a change in form, but also a reversal of cognitive interpretation, so it has semantic significance.

2. The Difference in the Distribution of Part-of-speech

First, the author uses the T-test to investigate whether there are differences between the comparative texts between English to Chinese and Chinese to English. According to the criteria of the conditions used for the t-test, the distribution of relevant data can first be tested in a visual way [1]. According to the research, it can be found that the percentage values of different parts of speech are basically normally distributed, and there are also certain differences between various parts of speech. In the three types of text, the distribution of conjunctions, adjectives, and prepositions will be more concentrated, and nouns and analogues will be more scattered than words, which means that the text content will be very different in these two characteristics. In almost all projects, the Chinese translation will be closer to the original Chinese text than the original English text, even if there is a correspondence between the content translated from Chinese to English and the original English text. This means that in the translation process, when the relevant person uses the word class, the target

language is far more than the source language. Not only that, among the two types and low-frequency words with richer representative words, the Chinese to English translation is obviously higher than the original Chinese and the original English text, which is contrary to the concept that the translation language is far less abundant than the original language. It can be inferred that this situation is related to the Chinese language system and the type of translation content. First of all, Chinese has strong flexibility, and in the process of translation, in order to accurately convey the meaning of the original text, relevant personnel will often change or adjust its structure and organizational form, so as to solve the situation that the meaning of words is not exactly the same as the original text. Secondly, Chinese is more receptive to emerging words or new combination structures, and if the relevant staff encounter new concepts in the process of translation, they will also carry out word creation behavior. Other than that, the average value of the Chinese translation and the original content in adjectives and adverbs is not much different. Between the Chinese translation and the original text, in addition to adverbs and related adjectives, other types of words have great differences and differences. However, between Chinese to English and English original texts, there is basically no close mean of parts of speech, and there are basically obvious differences. Therefore, from the analysis of parts of speech, Chinese translation is not only different from the original Chinese text, but also different from the original English text, so can the Chinese translation be called the third type of code? This cannot be rushed to conclusion. As far as parts of speech are concerned, in addition to the differences in language systems, the influencing factors include the influence of the times, the type of text, and even the writer. Among these factors, the difference between language genres can first be ruled out, and the factors of language type and writer are completely out of control, then it is obvious that it is a matter of the times. After certain practice and research, it can be found that whether it is an English original text or a Chinese translation, in the process of comparison, there is no obvious difference between various parts of speech, but in the overall comparison, the differences in most of the important parts of speech are very significant. Based on this situation, the most logical reason is that the text content belongs to a different era.

3. The Degree of Differentiation between the Category of Parts of Speech and the Text

After analyzing the differences and feature differences of different language texts, the three linguistic features that reflect word density, syntactic complexity, and vocabulary richness are now studied to see if these three languages can effectively distinguish between the three types of texts. After research and analysis, it can be found that the distribution of parts of speech between Chinese translation and the original Chinese text is very similar, but if it is related to the characteristics of parts of speech, then the similarity between the two texts is very high. Secondly, whether in the original English text and the two types of Chinese text, or in the Chinese translation and the original Chinese, the degree of differentiation between the connecting words that prove the complexity of syntax, the characteristics of the two parts of speech are synthesized, and the degree of distinction between prepositions is very high. Finally, the similarity between the type ratio and the less frequently used words is also relatively high, which means that the richness of words is not related to the distinguishing characteristics between words. Finally, among the Chinese translations and English original texts, the combination of the three types of feature sets has the lowest similarity and obvious differences [2].

4. Correlation Analysis of Parts of Speech

In addition to the differences in the distribution of parts of speech in different texts, in the same text content, and between the translated and translated texts, various parts of speech will also be closely related because they can influence each other. Correlation analysis can help those involved

identify such links. For example, with the help of all the information data in the English-Chinese corresponding text, and then combining it with the original Chinese information in the same quantity and without a correspondence between the two contents, the correlation coefficient in various texts is obtained. After research, it is not difficult to see a few points. First, there is no very obvious correlation between Chinese texts and Chinese translations and English original texts. This is normal, because the original Chinese itself has no correspondence with either text. Second, there is a very high positive correlation between less frequently used words and type ratios, even if the number of low-frequency words is directly related to the proportion of class characters or form proportions. Third, words that have a clear positive correlation with low-frequency words are mostly nouns or adjectives, while parts of speech that are obviously negatively related to low-frequency words are mostly adverbs and verbs [3]. From this conclusion, it can be inferred that most low-frequency words are adjectives related to new nouns or new characteristics, although sometimes new feature adjectives are specially created, but this cannot be denied. Fourth, in Chinese translation and English original texts, there are positively correlated parts of speech range sorted according to the correlation coefficient, from low to high are prepositions, the correlation coefficient is 0.29, conjunctions, the correlation coefficient is 0.39, adverbs, the correlation coefficient is 0.53, verbs, the correlation coefficient is 0.57, adjectives, the correlation coefficient is 0.63, nouns, the correlation coefficient is 0.78, low-frequency words, the correlation coefficient is 0.84, and the type ratio occupies the highest correlation coefficient, 0.93. In other words, in all the differences between parts of speech, there is a very obvious competition between Chinese translation and English original articles, and the transmission effect of the original article content on the part of speech is also more prominent. The fifth and final point is that Chinese translation and verbs in the original English text have a very high positive correlation coefficient, and at the same time, Chinese translation and adverbs in the original English also have a high positive correlation coefficient, but nouns, prepositions, and low-frequency words in Chinese translation. Various parts of speech, such as adjectives, have a negative correlation coefficient with their English equivalents. Therefore, relevant researchers can think and assume that the verbs in Chinese translation and the word classes in the original English article have a certain functional relationship, which is not zero or absurd, and relevant staff can conduct regression analysis on relevant data information to discuss.

5. Regression Analysis of Verbs Translated from Chinese

In order to carry out the regression analysis of Chinese-translated verbs, relevant personnel must first integrate the characteristics of multiple English parts of speech and analyze the influence of these English word class characteristics on Chinese translation verbs. Therefore, multiple linear regression analysis methods can be used by interested personnel. First of all, the percentage of verbs translated in Chinese can be regarded as a dependent variable, and the percentage of one or more parts of speech in English as an independent variable, and constantly try and study to find an equation that can explain the relationship between the two. In order to find the most suitable model, the person concerned can use the Chinese translation verb as the dependent variable and the range of parts of speech in the original English article as the potential independent variable. After some calculation, it is not difficult to find that the addition of verbs in English and analogues in English can effectively explain the changes in Chinese translated verbs. With the help of R language to build a linear model, the result correction determined that its coefficient is 0.383, and the P value of the model has a region of 2.2E-160, which means that its regression is of great significance in statistics. The prerequisite for establishing this model is that the relevant personnel do not distinguish between verbs, and in the results of part-of-speech coding, English verbs can be divided into substantive forms, situational form verbs, HAVE verbs and other categories, while Chinese verbs are generally divided into three

categories, namely general, continuous, and existential verbs [4]. In order to clarify the influence of different verbs, relevant personnel can refine the verbs in the original English article, use the corresponding functions, carry out the division work, and then obtain the most scientific model. For example, Chinese translation verb ~ English actual meaning verb + English BE verb + modal verb + type ratio. After the model is calibrated, the coefficient obtained is 0.4031. This proves that the corpus-based research on the part-of-speech relationship between English and Chinese translation, using English verbs and analogue ratios, can prove that in the process of translating English into Chinese, the corresponding verb change value is 40.31%.

6. Factors of Verb Changes in Chinese Translation

Through the regression analysis of Chinese-translated verbs, it is not difficult to find that in the analyzed corpus, the percentage change of Chinese-translated verbs in the text is more obvious, and about 40% of the change is determined by the three verbs of English and the English analogy ratio. The remaining 60% cannot simply be attributed to the influence of the translated language, but rather changes due to multiple causes. Language comparison or language translation, or discussion of translation commonality, need to clarify the concept between the system and the text, as far as the text is concerned, part of the translation commonality has nothing to do with any special language, while part of the translation commonality may be affected by the language, and even part of the translation commonality will be affected by the content type or the text. In general, it is highly likely that the translation language is affected by the content of the text [5].

7. Conclusions

Based on the self-built English-Chinese translation corpus, this paper compares the differences in the distribution of parts of speech, analyzes and thinks about the relationship between the range of related parts of speech and the differentiation of text language classes and language characteristics, and also analyzes the verb changes in the process of English-Chinese translation. It is not difficult to see that in terms of the correlation of parts of speech, there are different correlation coefficients between the original English text words and the Chinese translation words, while the original Chinese words have no obvious relationship with the two, which also reflects the transmission effect in the translation of words to a certain extent. In the regression analysis, it can also be concluded that the linguistic features of the original English text have a linear functional relationship in terms of characteristics, that is, the three types of verb features and analogues have at least 40% influence on the verb features of the Chinese translation.

References

[1] Yu Pingping. (2020). Corpus-based Equivalence of "Knowing" Translation [J]. Novel Monthly, 000(022), P.1-1.
[2] Shang Wenbo. (2020). Manifestation Characteristics of Logical Relationship of English and Chinese Academic Translation Texts Based on Corpus: A Case Study of Handbook of Social Justice in Education [J]. Contemporary Foreign Language Research, 5, 11.
[3] Liu Yang, Zhu Yubin. (2019). A Preliminary Study on the Manifestation of Causality Translation Based on Self-built Financial Discourse Corpus [J]. Journal of Hefei University: Comprehensive Edition, 36(6), 6.
[4] Wang Lisong, Zhao Han. (2018). Analysis of Lexical Characteristics of Chinese-English Translation News with Self-built Corpus [J]. Journal of Tianjin University: Social Science Edition, 20(4), 4.
[5] Wang Xiaohui, Go Juxia. (2019). An Empirical Study on Corpus-based University Translation Teaching Model [J]. Journal of Modern Linguistics, 7(4), 13.