

Analysis of the Summary of the 20th National Congress of the Communist Party of China under English Vocabulary Corpus

Alatengqiqige

Inner Mongolia National Preschool Teachers College, China

Abstract: This paper uses corpus translation to study Chinese and English translations of the Summary of the 20th National Congress of the Communist Party of China (CPC) to study the translation features of the English translation of the Summary of the 20th National Congress of the CPC. Firstly, the content of the Summary of the 20th National Congress of the CPC is summarized. Secondly, a questionnaire is set up to analyze the audience of the English translation of the Summary of the 20th National Congress of the CPC. Finally, the features of the English translation of the Summary of the 20th National Congress of the CPC are analyzed using corpus translation. The analysis results show that: (1) In terms of the overall features of vocabulary, the type-token ratio and vocabulary density values of the English-translated text are relatively large, reflecting the large changes in the vocabulary of the English-translated texts of the summary and the relatively complex application of vocabulary. (2) The distribution of vocabulary occurrences in the English version of the summary is small, reflecting the diversity of vocabulary changes in the English version of the summary. (3) In terms of frequency, most of the translated high-frequency words in the English translation of the Summary of the 20th National Congress of the CPC have a frequency correspondence similar to the original high-frequency words. This paper aims to provide reflections on the countermeasures of the standard system of the English translation of documents of the Central Committee of the CPC.

Keywords: The Summary of the 20th National Congress of the CPC, Corpus, Questionnaires, Vocabulary density.

1. Introduction

Throughout the history of foreign translation studies, first, translation studies have been divided into linguistic schools, hermeneutic schools, functional schools, cultural schools, deconstruction schools, feminism, postcolonial, Soviet-Eastern schools, and other major schools [1]. Since then, translation studies have been divided into linguistic schools, literary schools, philosophical schools, functional schools, pluralistic systems and normative schools, teleological schools, cultural schools, feminist views of translation, and postcolonial translation theory [2]. Finally, translation studies such as philology, structuralist linguistics, and deconstructionism have undergone paradigm evolutions. These studies show that experts and scholars categorize and summarize translation studies from multiple dimensions [3].

New technologies have brought unprecedented opportunities and challenges to translation studies. In recent years, the application and promotion of new technologies have promoted the progress of translation practice and have greatly impacted translation research. First, corpus research, especially corpus-based translation research, has emerged in recent years as various research expands on language. It expands the volume of bilingual studies and broadens the perspective of translation studies [4]. Secondly, with the continuous application of corpus technology, corpus-based translation research has gradually developed into a means of studying translation. Based on China's previous systematic research, the paradigm of corpus translation science is summarized, and the theoretical research system is built [5]. Finally, since the 90s of the 20th century, the corpus method has been gradually used in the study of translation, which directly contributed to the birth of corpus translation [6].

Through the collection of data, it is found that the corpus

translation has made good development. Therefore, based on previous research, this paper is based on the results of corpus translation studies, which are currently at the forefront of domestic and international translation studies. It examines the language features of the English translation of the Summary of the 20th National Congress of the Communist Party of China (CPC), an integral part of the translation of documents from the Central Committee of the CPC. This paper aims to investigate the translation language features of the English version of the Summary of the 20th National Congress of the CPC by discussing the translation commonality and language-specific language features of the English version of the Summary of the 20th National Congress of the CPC.

2. Research Method

2.1. Summary of the 20th National Congress of the CPC

The summary of the 20th National Congress of the CPC points out that the next five years will be a crucial period for the beginning of the comprehensive establishment of a modern socialist country. It will also be a period of the historic transformation of the Party's central tasks. It is a period of profound change in which strategies and opportunities coexist with risks and challenges. It is also a critical period for achieving the overall 2035 goal. The next five years are to speed up the establishment of a new development landscape and effort to facilitate high-quality growth of the period. In addition, it is a period when the national economy is under the impact of various unexpected factors and unprecedented downward pressure. It is critically important to achieving the main objectives and tasks of the next five years as proposed in the summary of the 20th National Congress, especially in terms of economic development.

2.2. Questionnaire survey on the English translation of the Summary of the 20th National Congress of the CPC

According to the research design, in the process of examining the translation language features of the English translation of the Summary of the 20th National Congress of the CPC based on the corpus, the English translation of the Summary of the 20th National Congress of the CPC is used based on introspective observation of the research content in the early stage. A questionnaire survey is implemented to target the audience in the UK. The survey results are analyzed, and the linguistic features of the English translation of the summary and the problem points of audience feedback are observed. This is used as the empirical basis and preliminary reference for follow-up research. This section discusses the questionnaire survey, which echoes the corpus-based research method and is an essential element of the entire empirical research.

(1) Overview of the survey

1) Survey objectives

The audience's understanding of the English translation of the Summary of the 20th National Congress of the CPC in 2022 is obtained; Audience judgment and analysis of the specific content of the English translation of the summary are captured. Overall audience ratings of the English translation of the summary are obtained.

2) Content design

An open-ended answer is adopted, and the full text of the English translation of the Summary of the 20th National Congress of the CPC is used as the survey content. Based on reading through the translation, the audience must judge the correctness of the text expression, the fluency of the writing expression, and the rationality of the logical expression. The audience explains whether they can correctly understand the content of the survey. The audience needs to give feedback on the actual problems in vocabulary, grammar, and passages of the English-translated text and give reasons and insights for judgment. The audience evaluates the overall language representation of the English translation.

3) Questionnaire implementation

Based on preliminary preparation, a questionnaire is carried out. A total of 200 questionnaires are distributed to students from a university in the UK, and 190 valid questionnaires are recovered, with an effective recovery rate of 95%, which meets the requirements of the questionnaire.

In addition, the amount of text read in the content of the summary is taken into account. The full text of the summary is divided into three parts to ensure the quality of the investigation: first, middle, and last. When surveying undergraduate students, one part of the first, middle, and last is provided in turn. This applies to other respondents as well. The content of the investigation may provoke unnecessary thinking. The purpose of the survey is indicated in the description of the questionnaire for academic research to ensure the quality of the survey. Meanwhile, the concept of the summary of the 20th National Congress of the CPC is briefly explained.

2.3. Analysis of the vocabulary features of the English translation of the Summary of the 20th National Congress of the CPC based on the corpus

(1) Corpus

The corpus consists of several electronic texts. Corpus research studies the language represented by the corpus and its variants or literary classes. Corpus translation is derived from the fusion of corpus linguistics and descriptive translation. In the construction and development of the corpus, unique research methods and principles have gradually formed. Corpus linguistics is the theoretical discussion and summary of these methods and principles. The corpus is based on the use of computers. Corpus linguistics is formed by the intersection of linguistics and computer science, using a corpus to study linguistic phenomena. Corpus linguistics has gradually established a theoretical and practical system of disciplines and has continued to develop. With the rise and development of translation studies, descriptive translation studies and corpus linguistics are integrated to form corpus translation. From the theoretical basis and research methods perspective, corpus translation and corpus linguistics are consistent. As far as research methods are concerned, it is mainly to use electronic corpus as a tool for research.

(2) Type-Token Ratio (TTR) and vocabulary density

In corpus statistics, the TTR is the ratio of class types and tokens, indicating the changing features of vocabulary application in the corpus. The larger the TTR, the greater the variation in the application of the text vocabulary. The smaller the TTR, the less variation in the application of the text vocabulary. The TTR is a common way to measure a text's vocabulary density.

(3) Distribution of vocabulary occurrences

The distribution of vocabulary occurrences in the text can help to understand the occurrence of the vocabulary in the English translation of the summary.

(4) High-frequency vocabulary

Descriptive translation research is one of the sources of corpus translation. The "required" and "non-required" in the translation transfer can be identified, the corresponding patterns of the text can be summarized, the translation process can be reconstructed, and the acceptability and importance can be observed by comparing the relationship between the original text and the paragraphs corresponding to the translated text.

(5) Design of experiments

This section will use corpus retrieval software to count the TTR and vocabulary occurrence distribution of the English translation text and reference corpus of the summary of the 20th National Congress of the CPC. Furthermore, the above features of the translated text and the reference corpus are compared and analyzed. The high-frequency vocabulary lists of the original and translated versions of the summary are searched and compared, and the overall vocabulary features of the English translation of the summary are examined. In addition, to highlight the overall vocabulary features of the English translation of the summary of the 20th National Congress of the CPC, it is also compared with the features of the English version of the summary of the 16th, 17th, 18th, and 19th National Congresses of the CPC.

3. Results

3.1. Results of the questionnaire survey on the English translation of the Summary of the 20th National Congress of the CPC

Analysis of the questionnaire shows that in the open-ended answer, most respondents indicate that they can grasp the translation of the Summary of the 20th National Congress of

the CPC as a whole. This is highly relevant to China’s sustained economic and social growth, continuous expansion of opening up to the outside world, and attaching importance to foreign propaganda work. The collation of survey feedback shows that the problems of audience feedback mainly focus on understanding the vocabulary, sentences/continuations, style, and foreign words in the translated summary. A summary of the main feedback to the questionnaire is shown in Table 1.

Table 1. Summary table of audience feedback from questionnaires

| Category | Feedback |
|-------------------------|--|
| Vocabulary | There are many expressions of meditation combined. However, from the point of view of “summary,” there should be more specific content; It is difficult to understand some Chinese words that are used many times. It is more natural for some vocabularies not to be used together to conform to English expression habits. |
| Sentences/continuations | Single sentences are very long, and it is necessary to pay attention to juxtaposition when reading. It is difficult to interpret the causal relationship between the contents, and it is necessary to divide the sentence and add the continuation. |
| Style | The “summary” genre is used, and the reader and the listener are reminded. However, many expressions seem to ask others and show a strong will. |
| Foreign words | There is much katakana that cannot understand its meaning perhaps due to insufficient personal knowledge. Foreign words that are not necessary to use are used. |
| Other | Overall, there are no grammatical errors, but many sentences express unnaturalness. |

From Table 1, in terms of semantic understanding of translated words, the audience will pay attention to whether the translated words apply to the type and subject matter of the text. The audience wants to use translations that directly reflect objective facts, which also shows that the audience does not fully understand the semantics of the translated words. From the perspective of structure, the audience pays very much attention to the logical semantic understanding in the sentence. Audiences are more aware of understanding the relationships between clauses. They believe it is necessary to include continuations in the translation to make it fit the English convention. Studies have shown that audiences not only are concerned with understanding basic semantic

messaging but also pay attention to the functional effects of translations.

3.2. Corpus-based analysis of the vocabulary features of the English translation of the Summary of the 20th National Congress of the CPC

(1) TTR and vocabulary density analysis

The TTR of each summary translated from English is retrieved, the TTR and vocabulary density are calculated, and Figure 1 presents the results.

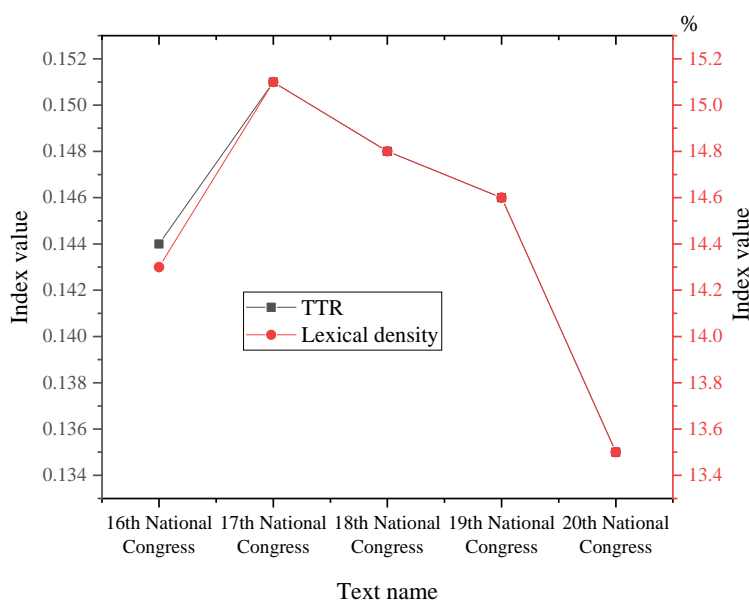


Figure 1. Analysis of TTR and vocabulary density of English-translated texts in each summary

From Figure 1, the TTR for the English-translated versions of the summary is relatively stable, ranging from 0.135 to

0.151, with a mean of about 0.145. Vocabulary density and TTR are homogeneous. Therefore, the vocabulary density

value is correspondingly between 13.5% and 15.1%, and the mean vocabulary density is about 14.5%.

(2) Distribution of vocabulary occurrences

Based on the English corpus, descriptive statistics and comparative investigations are carried out on the features of

the “occurrences” of each summary text and reference corpus. The characteristic data of the word “occurrences” of each summary translated into English are retrieved using the retrieval software. Figure 2 depicts the results.

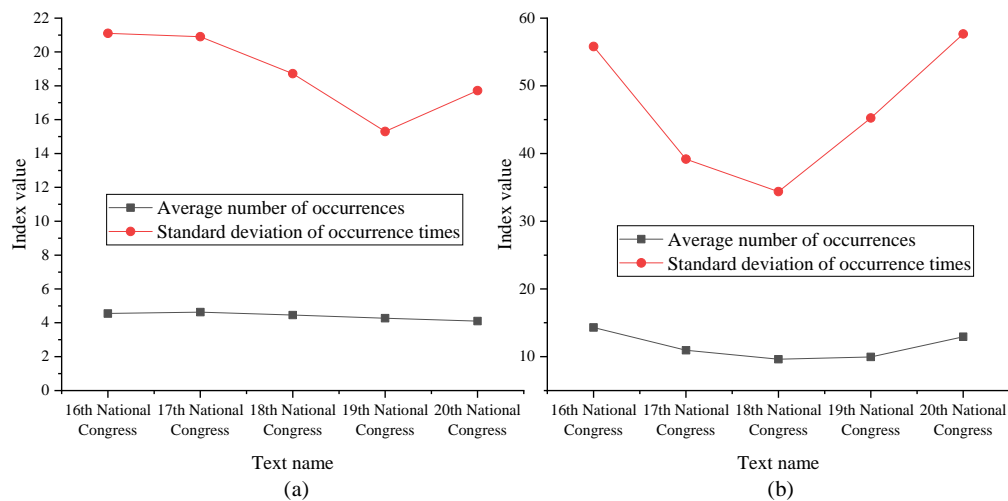


Figure 2. Distribution of the number of occurrences of vocabulary in the English version of each summary

(a) Statistics on the “number of occurrences” features of the English-translated vocabulary of each summary; (b) Characteristic statistics of the “number of occurrences” of the vocabulary reference corpus

Figure 2 shows that the “mean occurrences” value of the English translation vocabulary in each summary ranges from 4.10 to 4.63, and the “standard deviation of occurrences” value ranges from 15.31 to 21.2. Moreover, the reference corpus vocabulary has a “mean occurrence” value between

9.62 and 14.3 and a “standard deviation of occurrences” value between 34.38 and 57.65.

(3) Distribution of high-frequency vocabulary

Here, the top ten high-frequency vocabularies each year of the original Chinese and English translations of the Summary of the 20th National Congress of the CPC are examined. A review of texts from more recent years can reflect the latest features of the text. Table 2 illustrates the results of the examination.

Table 2. The top ten high-frequency vocabularies list of the original Chinese and English translations of the Summary of the 20th National Congress of the CPC

| Number | Original vocabulary | The vocabulary of speech use | Frequency | Translated vocabularies | Frequency |
|--------|---------------------|------------------------------|-----------|-------------------------|-----------|
| 1 | reform | Noun | 100 | reform | 142 |
| 2 | develop | Noun | 78 | develop | 98 |
| 3 | advance | Verb | 75 | advance | 84 |
| 4 | economics | Noun | 75 | economics | 68 |
| 5 | development | Verb | 70 | development | 66 |
| 6 | do | Verb | 67 | do | 64 |
| 7 | strengthen | Verb | 58 | strengthen | 57 |
| 8 | new | JJ | 54 | new | 49 |
| 9 | society | Noun | 53 | society | 48 |
| 10 | comprehensively | Adv | 52 | comprehensively | 48 |

From Table 2, in the Chinese version and the English version of the Summary of the 20th National Congress of the CPC, although the order of the top ten words is different, the words contained are the same. Regarding frequency, most translated high-frequency vocabularies have a frequency correspondence similar to the original high-frequency vocabularies.

4. Conclusion

Based on the theory of corpus translation, this paper focuses on the translation language features of the Summary of the 20th National Congress of the CPC. Also, it conducts a questionnaire survey on the British audience. The translation

language features of the English translation of the Summary of the 20th National Congress of the CPC are examined by analyzing the vocabulary features of the English translation of the summary. The findings indicate that: (1) Most audiences can understand the translated vocabulary through form and meaning. A certain percentage of audiences disagree on the judgment of semantic logical relations. (2) The vocabulary of the English translation of each summary is highly variable, and the application of vocabulary is relatively complex. The downside is that some of the language features of translation presented through data or descriptive corpus are not sufficiently well explained. Follow-up research will continue to improve and strengthen the construction of the

English discourse system.

References

- [1] Sayogie F, Supardi M. Equivalence Levels of Literary Corpus Translation Using a Freeware Analysis Toolkit[J]. Buletin Al-Turas, 2021, 27(1):55-70.
- [2] Qi D, Haladin N B. Corpus translation system based on Web Services and its application[J]. Journal of Intelligent and Fuzzy Systems, 2021(6):1-12.
- [3] Grabowski U, Groom N. Functionally-defined recurrent multi-word units in English-to-Polish translation:A corpus-based study[J]. Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics, 2022, 35(1):1-29.
- [4] Matvieieva S. Terminological Synonymy: Corpus-Driven Translation[J]. International Journal of Science and Research (IJSR), 2020, 9(3):971-973.
- [5] Ramos M. Teaching English for Medical Translation: A Corpus- based Approach[J]. Language Teaching Research Quarterly, 2020, 8(2):25-40.
- [6] Wang X. Building a Parallel Corpus for English Translation Teaching Based on Computer-aided Translation Software[J]. Computer-Aided Design and Applications, 2021, 18(S4):175-185.