Klover Research Journal of Education
Volume 1, Issue 1 October 2023
ISSN: Pending….

Enhancing College English Translation Instruction with Parallel Corpus Analysis

Dr. Xiaojun Wang
School of Applied Technology, University of Science and Technology Liaoning, Anshan, China

Abstract: Corpus linguistics hinges on the pivotal concept of "corpus," encompassing vast electronic text collections essential for linguistic research. These corpora encompass written and spoken language instances, capturing specific linguistic variations and evolving through rigorous material selection, norm establishment, and grammatical annotation processes. When appropriately scaled, corpora serve as veritable mirrors of language use within defined contexts, enabling comprehensive language observation and analysis. This linguistic discipline, a cornerstone of computational linguistics, leverages diverse corpora as indispensable primary resources for advancing linguistic theory and practical applications.

From a linguistic perspective, understanding a corpus entails recognizing it as a repository of genuine language usage, fundamentally underpinned by computational technology. However, effective utilization of corpus data mandates rigorous analysis and preprocessing. This abstract provides an overview of corpus linguistics, elucidating its significance in language research and its intricate relationship with computational linguistics.

Keywords: Corpus linguistics, Electronic corpora, Language research, Computational linguistics, Linguistic analysis

Introduction

Corpus is the core concept in corpus linguistics, which refers to large-scale language materials preserved in the form of electronic texts for the purpose of language research, including written and spoken languages that appear in language practice, representing specific languages or language variants. It has gone through the processes of scientific selection of materials, establishment of norms and annotation of grammatical structures. If the corpus has an appropriate scale, it can reflect and record the actual use of language in a certain range. We can observe and grasp the language facts and analyze and study the laws of the language system through corpus. Corpus linguistics is an important branch of computational linguistics, and various corpora have become indispensable first-hand materials for linguistic theory and application research. From the linguistic point of view, the understanding of corpus at least includes the following three points. What is stored in the corpus is the language materials that have really appeared in the actual use of language. Corpus is the basic resource for carrying language knowledge with computer as the carrier. Real corpus needs to be analyzed, processed and processed before it can become useful research data.

Since the emergence of computerized corpora, Corpus linguistics has developed rapidly. The use of corpora for language comparison research and language ontology research has achieved fruitful results. A parallel corpus is a bilingual or multilingual corpus composed of the original text and its corresponding parallel translated text. Translation is the act of cross language communication in which the thoughts and content expressed in one language are accurately and completely expressed in another language. Translation is not only a science, but also

an art. International communication cannot do without translation, and social progress cannot do without translation. At the level of translation teaching, with the rapid development of the Social market economy, translation teaching has gradually gained the attention of all sectors of society. There is a big gap between the theory and research of translation teaching and the western developed countries. Bilingual parallel corpora, as a basic research theory, can organically combine learning knowledge with theoretical practice, helping students construct knowledge and form meaning [1]. Parallel corpora have strong application value for language comparison, dictionary compilation, and translation teaching [2]. In the context of the new era, in order to effectively promote the significant development of translation teaching, universities need to follow the principles of common participation, purposefulness, and integrity, innovate traditional teaching methods, and enhance the effectiveness of translation teaching. Corpus, with its advantages of speed, comprehensiveness, simplicity, and ease of operation, lays the foundation for reforming teaching methods and improving teaching quality.

1. Functions of Parallel Corpus System for English Translation Teaching

Functional design belongs to the outline design part in the software lifecycle, which needs to follow the basic ideas of software engineering and apply basic principles such as abstraction, modularization, gradual refinement, and information concealment. On the basis of following these principles, the functions of the English translation teaching parallel corpus system are divided by user roles, which are collections of the same type of users. This system mainly includes two types of users. The first type is an administrator user, used to manage the corpus and corpus system; The second category is for teachers and student users, who use corpora for translation teaching and learning. The system functions divided by user roles are composed of a corpus management subsystem and a query statistics subsystem, each of which includes several functional modules, as shown in Figure 1.

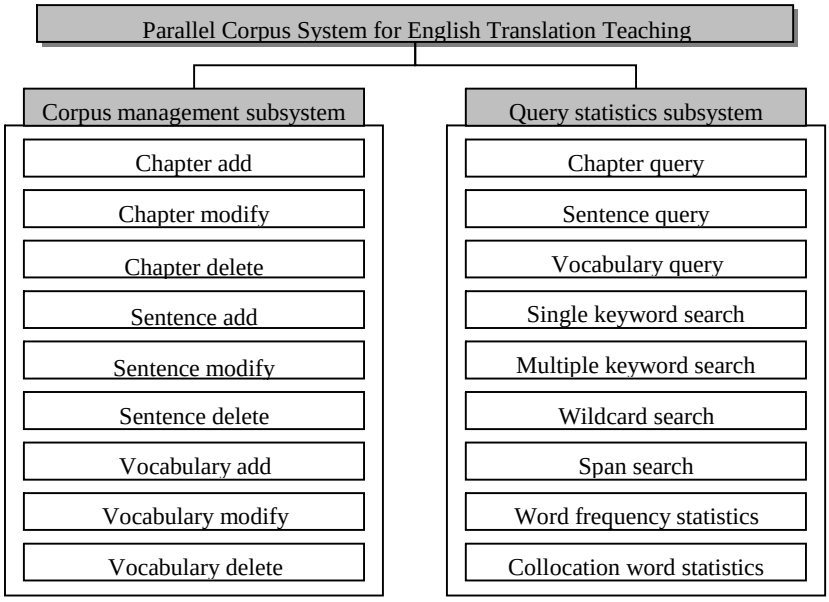

Figure 1 Functions of parallel corpus system for English translation teaching

1.1 Corpus Management Subsystem

The corpus management subsystem manages the comprehensive information of translated content, including three categories: discourse information, sentence information, and vocabulary information. Each category of information also includes three functions: adding, modifying, and deleting. Language has a certain hierarchical structure, and language representation can be divided into different granularity representations such as

vocabulary, sentences, and discourse. Indicating that language information structure includes sentence information structure and discourse information structure, and transmitting information not only relies on sentence information structure, but also relies on discourse information structure. Therefore, studying language information structure requires not only studying sentence information structure, but also studying discourse information structure. Decentralized representation has also become the most popular method in language representation, which not only enables end-to-end learning of dispersed representations of vocabulary, sentences, and texts in specific tasks, but also enables automatic learning through large-scale unlabeled texts.

1.2 Query Statistics Subsystem

The query and statistics subsystem is mainly the most commonly used function of the system, assisting teachers in completing teaching work and assisting students in completing learning tasks. Corpus query is also a type of information retrieval, playing a very important role in corpus construction, which is related to whether the corpus can meet the user's usage needs and realize the value of the corpus. It also marks the level of corpus construction [3]. According to the actual needs of users, individual queries can be performed on articles, sentences, and vocabulary, as well as multiple search methods such as single keyword, multi keyword, wildcard, and span. The statistical function includes word frequency statistics and collocations statistics. Word frequency statistics calculate the frequency of each word occurrence and the total number of all words. Collocation word statistics study the typical co-occurrence behavior between words discovered through the use of words together. The basic idea of word frequency statistics is to split the word array, define a hash map to store the word frequency statistics results, traverse the word array, and if a word is encountered for the first time, add an element to the hash map, with the word as the key and 1 as the value; If it is not the first time encountered, then find the key in the hash map and increase its value by 1.

2. English Translation Process Based on Corpus Examples

With the improvement of computer processing speed and storage capacity, and the emergence of a large number of electronic bilingual corpora, case-based translation methods have emerged in the field of translation. According to the similarity principle of Natural language processing, similar examples are imitated to overcome the weaknesses of traditional rule-based translation methods. The example based method can fully utilize parallel corpus resources, and obtain translation results by transforming and replacing the source and target languages in the corpus. The advantage is that it avoids the formulation of a large number of rules and improves translation speed. The English translation process based on corpus examples is shown in Figure 2.

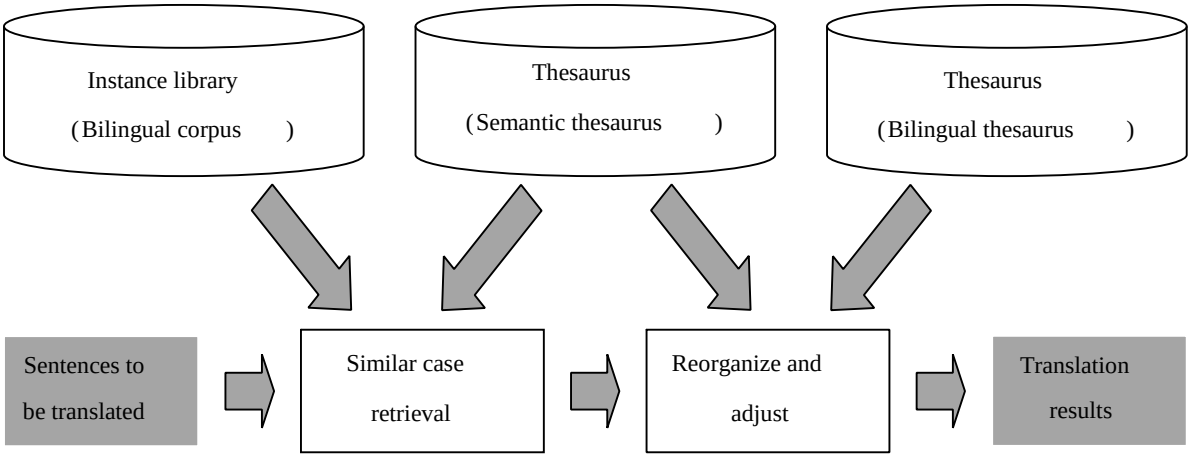

Figure 2 English translation process based on corpus examples

When translating simple sentences, people do not do a deep analysis of language. Instead, they first decompose the sentence into several phrases, use the translation of existing phrases to translate each decomposed phrase into a target phrase, and finally combine these phrases to form a long sentence. So the basic idea of instance based translation is to find the most similar translation instance to the part to be translated in the collected bilingual instance library, and then translate the instance's translation through a series of deformation operations such as replacement, deletion, and addition.

3. Parallel Corpus Alignment Tool: ABBYY Aligner

Corpus alignment is the establishment of paragraph, sentence, or even word level correspondence between the original text and the translated text, which can significantly improve translation efficiency. The current alignment techniques include sentence level, vocabulary level, multi word combination units, and alignment of clauses and paragraphs.

ABBYY Aligner is a translation alignment software, which can automatically achieve bilingual alignment according to the user's settings, and can also edit the translated content on the software to support languages of many countries. At the same time, it also supports statistical function, which can display the total number of alignment, as well as the completed and wrong contents, so as to facilitate viewing the results of local alignment processing. ABBYY's developers have specially developed a vocabulary database for ABBYY Aligner 2.0. ABBYY Aligner 2.0 can divide the text into some fragments more accurately, and find the best matching part by matching it with the translation in the text in sequence, so as to get better and more accurate translation alignment and save time for manual proofreading and alignment [4].

ABBYY Aligner 2.0 has simple text import operation, and can be edited by directly copying the text into the editor window. ABBYY Aligner2.0 saves the results in translation memory format or RTF format. TMX is an international standard supported by virtually all translation memory systems. When ABBYY Aligner 2.0 file is saved in TMX format, this result can be reused in other applications, for example, CAT tool is used by translators. Moreover, the result of alignment can be stored in RTF format. RTF format is a standard format for converting text documents and is supported by many Microsoft products. ABBYY Aligner 2.0 has a simple interface, which enables users to understand its main elements more quickly, and it is more comfortable, faster and more efficient to use.

Application Strategies of Parallel Corpus in College English Translation Teaching

In order to fully utilize parallel corpora in college English translation teaching, based on previous research results and the author's years of experience in college English teaching, specific application strategies are proposed from the following four aspects.

3.1 Fully leverage the multiple functions of parallel corpora

Based on a large amount of actual corpus in parallel corpora, design an English translation teaching outline, develop English translation textbooks, and solve the problem of "what to teach" in English translation. By using parallel corpora, students can better understand the language context of the original text, choose more flexible words for the translation, and effectively improve translation readability in the wording process [5]. Cultivate students' corpus retrieval ability, master the analysis function of corpus software, and apply corpus software to analyze the language features of texts. Guide students to make full use of the corpus to extensively collect various language materials, analyze the discourse characteristics of different language genres, discover the language patterns of English, and promote students to engage in exploratory, exploratory, and confirmatory learning

3.2 Improving openness between teaching and learning

For closed teaching, open teaching is a student-centered learning approach that encourages students to participate in learning activities. Openness is reflected in open spaces, open environments, open courses, open attitudes, open means, and open resource utilization. Teaching methods should have a certain degree of openness, mainly because teachers can expand their scope of thinking during the process of using corpora to carry out teaching activities. For texts with different themes, teachers should organize students to reasonably select corresponding corpora to minimize acquisition time. At the same time, teachers should store newly generated corpora in the teaching process into the corpus.

3.3 Strengthen the cultivation of students' cross-cultural awareness

Cultural awareness refers to the initiative generated in the learning process of cultural cognition, which involves absorbing language knowledge, comprehending the connotation of culture, receiving cultural influence, and gradually transforming it into an emotional pursuit of culture. Translation not only involves language issues, but also cultural issues. Translators not only need to understand foreign cultures, but also have a deep understanding of their own ethnic culture [6]. Translation, as an act of language to language communication, will inevitably integrate into cultural and cultural exchanges. To carry out English translation teaching based corpus, it is necessary to strengthen cross-cultural awareness cultivation, which is conducive to deepening the understanding of the original text, optimizing the expression of translated content, correctly accepting the information conveyed by the other party, and equivalently expressing one's own attitude and viewpoints [7].

3.4 Constructing a Translation Teaching Model Based on Parallel Corpus

The traditional English translation teaching model has many drawbacks in terms of teaching mode and translation teaching philosophy. Teachers are the leaders of the entire classroom, and students receive limited information and knowledge, making it difficult to truly master translation skills. Bilingual parallel corpus translation teaching provides convenience and has also led to the emergence of new translation teaching models [8]. Teachers are no longer the absolute authority of the classroom, but the organizers and participants of teaching activities. Teachers strive to create a relaxed heuristic learning environment, transforming "Require me to learn" into "I want to learn", so that students can give full play to their independent learning ability. While learning translation theory, students can fully utilize the corpus as a platform to search for authoritative corpus of interest. Compared to the limited and isolated texts in translation textbooks, the corpus provides comprehensive and rich information. Corpus can provide timely and effective assistance in sentence structures, common expressions, and vocabulary, reducing students' burden on language form, investing more energy in content, and ultimately achieving the goal of fundamentally improving students' translation skills.

4. Conclusions

Traditional English translation teaching has certain limitations, such as insufficient teaching time, monotonous teaching by teachers, incomplete textbook content, insufficient translation resources, passive acceptance by students, and unsuitable teaching according to their aptitude. These problems have long plagued English translation teaching, affecting students' enthusiasm and teaching effectiveness. With the development of information technology and the Internet, students' learning methods have undergone fundamental changes, and they can choose learning content and suitable learning methods according to their needs. Corpus based English translation teaching can not only serve as a supplement to English education, broaden students' English learning channels, but also promote changes in learning methods. By utilizing corpus resources and presenting teaching

scenarios based on teaching content, information such as sound and images can stimulate students' hearing and vision, stimulate their enthusiasm for thinking, promote rapid perception and firm memory, thereby mobilizing students' observation and imagination, deepening their learning impression, and improving learning efficiency.

References

R. R. Liang. Construction of EnglishTranslation Teaching Model Driven by Bilingual Parallel Corpus [J]. Journal of Xinzhou Teachers University, 2022, 38(02): 98-101.

X. Yu, Y. G. Cai. The Application of English Monolingual Corpora in Translation Teaching [J]. Journal of Guizhou Education University, 2019, 35(10): 52-57.

B. L. Zhang, "A Perspective on the Retrieval System of Chinese Interlanguage Corpus," Journal of Tianjin Normal University(Social Sciences), 2021, 48(06): 29-37.

J. Q. Wan, F. Xiao. The main functions of the ABBYY Aligner and Tmxmall Aligner tools are explored in comparison[J]. English Square, 2020, 10(25): 27-29.

S. L. Yang. The Application of Bilingual Parallel Corpus in Chinese English Translation Teaching [J]. Modern Communication, 2019, 33(19): 17-18.

L. Y. Yuan. An Empirical Study on the Correlation Between English Majors' Translation Competence and Intercultural Awareness[J]. Journal of Wuyi University, 2021, 40(04): 70-73.

H. Wang, W. H. Qi. The relationship between translation ability and cross-cultural awareness [J].

English Square, 2023, 13(05): 28-31.

T. Dai, Q. Zhang. Construction of Translation Teaching Model under Bilingual Parallel Corpus [J]. Comparative Study of Cultural Innovation, 2017, 1(33): 70+72.