Research Paper   **Open 🔓 Access**

A Study on the Construction of a Bilingual Parallel Corpus Of the Basic Terminology of Health and Wellness Industry in Panzhihua City

Yujie Tu

School of Foreign Languages and Cultures, Panzhihua University, China

ABSTRACT: *The increasing popularity of the health and wellness industry in Panzhihua has attracted a growing number of foreign visitors. However, language barriers and challenges in effectively promoting the health and wellness industry have also emerged. Consequently, the creation of a bilingual parallel corpus containing essential terms for the health and wellness industry in Panzhihua City can enhance language services and facilitate industry development.*

Keywords - *parallel corpus, the health and wellness industry, terminology*

I.   INTRODUCTION

Under the wave of ecological civilization construction in the new century, Panzhihua City has seized the opportunity for historical development and emerged as a pioneer in promoting the concept of "health and wellness" nationwide. The city has organized the first and third China Health and Wellness Industry Development Forums, with a strong focus on strategically positioning itself as a "National Health and Wellness Tourism Destination". It recognizes the importance of developing the health and wellness industry as a key driver for supply side structural reform. In line with the proposal of the Panzhihua Municipal Party Committee to formulate the 14th Five Year Plan and 2035 Long Range Goals for the National Economic and Social Development of Panzhihua City, constructing an international tourism destination stands as one of the priority objectives for the city's economic and social development during this period.

According to the "Analysis Report on the Overseas Influence of Chinese Cities (2021)", Panzhihua City was chosen as a case study for the inaugural "China City International Communication Forum". This selection greatly enhanced the city's international reputation and attracted an increasing number of foreign acquaintances. As a result, communication barriers related to language and the development and promotion of the international concept of health and wellness have become more efficiently and clearly apparent. The research on the corpus of the health and wellness industry is exceedingly limited. In fact, only a small portion of corpus research explores the health and wellness industry, as indicated by a keyword search on CNKI. Consequently, it is crucial to establish a bilingual parallel corpus of essential terms in the health and wellness industry. This corpus not only holds significant reference value for the development of the health and wellness industry in Panzhihua but also contributes to enhancing the accuracy and convenience of research in this field, as well as fostering exchanges, discussions, and collaborations within the industry.

II.   ANALYSIS OF THE CURRENT SITUATION AT HOME AND ABROAD

The history of bilingual parallel corpora is relatively short, but it has rapidly developed and holds significant value due to the exchange and integration of vast amounts of information worldwide. Research on bilingual corpora in foreign countries started early and has accumulated considerable experience in corpus technology and research methods (Wang & Huang, 2012). Several representative bilingual parallel corpora have been established. For instance, the British National Corpus currently stands as one of the most representative contemporary English corpora globally, with a word capacity exceeding 100 million and encompassing languages from diverse age groups and social classes. In comparison, the development of bilingual parallel corpora in China has been relatively brief, with a smaller scale, limited corpus representativeness, insufficient

research entry points, lagging development and application of specialized corpora, as well as an insufficient application research and technology development based on bilingual databases (Wang & Huang, 2012). Corpus linguistics has begun integrating its research with other linguistic branches, making it an internationally popular research topic. However, there still exists a certain gap between domestic corpus research and foreign research, although the development prospects are promising (Chen, 2021).

Building a bilingual parallel corpus plays a crucial role in local development as it facilitates effective language communication through the construction of a corpus. The emergence of corpora has undoubtedly revolutionized language research, as it combines theoretical frameworks with practical applications to analyze language changes and advancements through data collection (Li, 2019). Hu (2021) states that tourism corpora serve three main purposes: Firstly, they serve as valuable teaching materials for tourism translation instructors, enabling them to cultivate skilled tourism translation professionals. Secondly, they provide reliable translations for various tourism texts such as public signs, tour guides, and scenic spot descriptions. Lastly, they analyze the stylistic features of tourism texts, uncovering their textual functions and stylistic characteristics. Irrespective of the intended purpose, the standardization of translation and promotion is imperative for internationalization and cross-cultural exchange within the tourism industry. Ultimately, the construction of a bilingual parallel corpus offers fresh insights into the development of the health and wellness industry, effectively promoting its gradual internationalization.

III. CONSTRUCTION OF BILINGUAL PARALLEL CORPUS

The development of a corpus mainly involves six parts: corpus design, corpus collection, corpus processing, corpus retrieval tool design, corpus-based research, and corpus online retrieval (Wang,2012).

3.1 Collection and processing of corpus

Numerous sources of language materials exist, and their content is inherently intricate. When selecting these materials, it is crucial to assess their persuasiveness and accuracy. First, the sources of the corpus must undergo rigorous screening. The collected language materials originate from influential sources such as official documents, newspapers, brochures, and books. Additionally, basic terms relevant to the health and wellness industry are annotated based on the winning bidder. These materials undergo secondary screening to compile a comprehensive Excel vocabulary.

3.2 Corpus alignment

Corpus alignment involves aligning the source language corpus and the corresponding translated corpus, paragraph by paragraph or sentence by sentence (Gale & Church, 1994). Import the filtered and organized Excel documents into Language X software. The AI alignment function of Language X can segment sentences more precisely, thus enhancing the accuracy of automatic alignment. Utilize the alignment processing capabilities of Language X software using its AI alignment function. To ensure ultimate accuracy, Language X processing precedes manual proofreading.

3.3 Construction process and results

This study conducted extensive research on corpus construction and successfully outlined the construction process as shown in Figure 1. The final construction results can be seen in Figure 2.
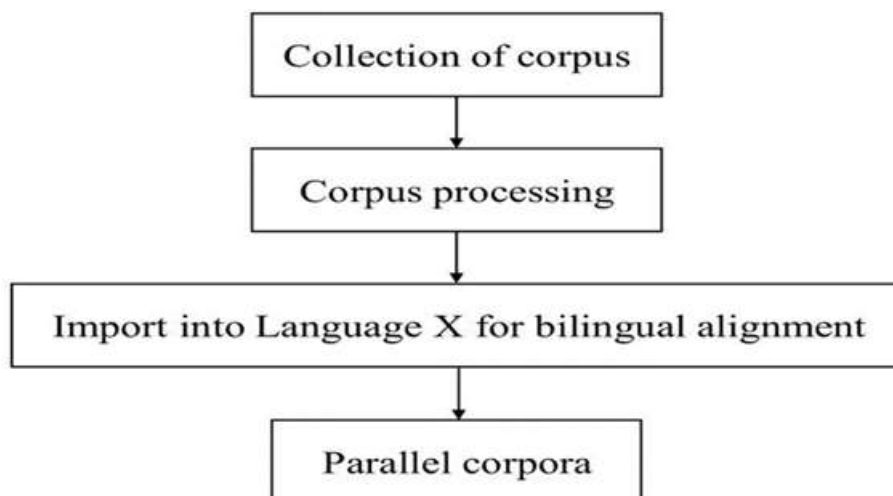

Figure1. Construction process of corpus

Source text	Target text
健康	health
养生	health-preservation
康复	rehabilitation
康养	health-preservation and rehabilitation
健康素养	health literacy
健康预期寿命	health expectancy life expectancy
亚健康	sub-health
治未病	cure is not ill
康养产业	health-preservation and rehabilitation industry
康养全域化	Holistic regional health-preservation and rehabilitation
康养全员化	all-age health-preservation and rehabilitation
康养全时化	all-time health-preservation and rehabilitation
海拔高度	altitude
温度	temperature
湿度	humidity
空气洁净度	air cleanliness
优产度	superiority
和谐度	harmony
医疗机构	medical institutions
康复机构	rehabilitation agencies
养老机构	senior care organization

Figure2. Samples of the bilingual parallel corpus

3.4 Typical examples
3.4.1 Case 1
Source text: 养生
Target text: health-preservation
Health preservation refers to a range of activities aimed at promoting life, enhancing physical well-being, and preventing diseases in order to achieve longevity. The term "preservation" carries the notion of maintenance. In comparison to the phrase "keep in good health," it is more aesthetically pleasing and conveys greater depth of meaning.
3.4.2 Case 2
Source text: 健康素养
Target text: health literacy
Health literacy refers to an individual's capacity to acquire and comprehend fundamental health information and services, and to effectively utilize this knowledge to make informed decisions in order to maintain and enhance personal well-being. Literacy is defined as the proficiency in reading and writing, and this terminology aptly encompasses the process of acquiring and comprehending health-related information.
3.4.3 Case 3
Source text: 优产度
Target text: superiority
Superiority refers to the diversity and quality of agricultural and other products across various seasons. The original meaning of superiority is more skillful, powerful, or greater than others. In contrast to high yield, "superiority" places a greater emphasis on evaluating quality in a formal manner.

IV. PROBLEMS AND SOLUTIONS IN CORPUS CONSTRUCTION

4.1 Incomplete corpus data
The health industry encompasses a broad array of fields, such as medical treatment, rehabilitation, and wellness. In order to effectively meet user needs and facilitate in-depth language analysis and research, it is crucial for the corpus data to be comprehensive. Collection of ample corpus data pertaining to the health and

wellness industry is necessary to enhance both the comprehensiveness and practicality of the corpus. Additionally, continuous updating and expansion of the corpus data is essential to cater to user requirements and facilitate thorough language research. Furthermore, researchers must possess a significant level of literary knowledge, literacy, and accurate judgment.

4.2 High cost of corpus construction

The construction of a corpus incurs significant costs, as it demands substantial manpower and material investment for its collection, organization, screening, and processing. These excessive costs can potentially undermine the corpus's sustainability and maintainability. To mitigate these expenses, researchers have opted for more efficient data processing techniques and tools, thereby reducing the required manpower and resources. Additionally, a collaborative approach with other institutions can be adopted to foster the development of bilingual parallel corpora in the health and wellness industry.

4.3 Poor quality of corpus data

Bilingual data in the health and wellness industry may have issues such as inaccurate translation and unclear context. Therefore, in addition to using Language X software, manual checks and corrections should also be carried out.

4.4 Data alignment error

Bilingual parallel corpora need to ensure that the sentences in the source and target languages correspond semantically, but alignment errors may occur in practical operations. The accuracy of alignment directly affects the subsequent use effect. So automatic alignment technology can be used. Utilize advanced natural language processing techniques to achieve automatic alignment of sentences in the source and target languages.

V. PRACTICAL IMPLICATIONS

The construction of a bilingual parallel corpus provides new ideas for actively exploring the development of the health and wellness industry, improves the language service level of the health and wellness industry, and promotes the gradual internationalization of the health and wellness industry. The benefits of constructing a bilingual parallel corpus for the health and wellness industry include:

1. The Chinese and English bilingual texts collected during the establishment of a bilingual parallel corpus for the health and wellness industry can be used to train machine translation systems and improve translation quality.
2. Train professional translation talents. Utilize bilingual parallel corpora for translation training to improve the English proficiency of the health and wellness industry practitioners, enabling them to better communicate and cooperate with international peers, and enhance the language service level of the healthcare industry.
3. Develop a bilingual information platform for the health and wellness industry. Utilizing a bilingual parallel corpus to develop a bilingual information platform for the health and wellness industry, providing real-time and accurate health and wellness industry information to domestic and foreign customers, is conducive to the internationalization process of the health and wellness industry.
4. Strengthen the brand construction of the health and wellness industry. Utilize bilingual parallel corpora to strengthen brand building in the health and wellness industry and enhance international awareness and reputation.
5. Promote technological innovation in the health and wellness industry. Utilize bilingual parallel corpora to promote technological innovation in the health and wellness industry and increase the added value of products and services.
6. Improve the international competitiveness of the health and wellness industry. By utilizing bilingual parallel corpora, we aim to enhance the international competitiveness of the healthcare industry and attract more international customers and investments.
7. Develop an international development strategy for the health and wellness industry. Based on the research results of the bilingual parallel corpus, formulate an international development strategy for the health and wellness industry, and clarify development goals and paths.
8. Organize international exchange activities in the health and wellness industry: Utilize bilingual parallel corpora to organize international exchange activities in the health and wellness industry, such as seminars, training courses, etc., to promote international technological exchange and cooperation.

VI. CONCLUSION

The construction of a corpus has greatly enriched the vocabulary of terms related to the health and wellness industry, and in the process of collecting and screening relevant terms. We can also ensure the accuracy of core terms. Secondly, consistency in terminology has been achieved, providing a reliable corpus for the health and wellness industry, effectively avoiding translation errors, and greatly improving translation

quality. At the same time, the convenience of language greatly improves work efficiency and better serves tourists who come to visit.

VII. FUNDING

ACKNOWLEDGEMENTS

REFERENCE

[1].   Wang Kefei & Huang Libo, 2012.Construction and Application of Parallel Corpora: Issues and Comments. Foreign language audio-visual teaching (06),3-10.
[2].   Chen Zhao. (2021). An Overview of Development of Corpus Linguistics at Home and Abroad. Journal of Liaoning Institute of Education Administration (03),83-87.
[3].   Li Rong. (2014). On the Role of Corpus in Language Research New course: Teachers, (11), 81-81.
[4].   Hu Saier. (2021). Design and Construction of Bilingual Parallel Corpus for Tourist Scenic Spots: Taking Zhoushan as an Example. Western Academic Journal (08),66-68.
[5].   Wang Kefei. (2012). On the Design and Construction of the Super-large-scale China English-Chinese Parallel Corpus. Foreign Languages in China (06),23-27.
[6].   Gale, W. A., & Church, K. W. (1994). A program for aligning sentences in bilingual corpora. Computational linguistics, 19(1), 75-102.

Yujie Tu
School of Foreign Languages and Cultures, Panzhihua University, China