

10.18686/neet.v2i2.3922

Study on the Corpus of Bilingual Translators' Training Model during the Covid-19 Pandemic

Xinran Wang¹, Yan Wu², Zikere Bahatiyaer³

1. College of Foreign Languages, North China University of Science and Technology, Tang Shan, China

2. College of Foreign Languages, North China University of Science and Technology, Tang Shan, China

Abstract: This research examined the systematic study of corpus of bilingual translators' training model during the Covid-19 pandemic. The research utilized scientometric analysis and examined literature sourced from the Web of Science (WoS) core collection and the China National Knowledge Infrastructure (CNKI) spanning from 2019 to 2023. The main aim was to explore the present state of the bilingual (English and Chinese) translators' training model and to identify the future trends in corpus development. The findings offer valuable insights for scholars to delve deeper into this area of study.

Keywords: COVID-19; Bilingual translator; Linguistic corpus; Training model

1. Introduction

In December 2019, the World Health Organization (WHO) began receiving reports of a cluster of pneumonia cases with an unknown cause in Wuhan, China. Requests for additional information came from health authorities worldwide. By January 2020, it became evident that these pneumonia cases were linked to a new virus capable of human-to-human transmission. By the end of January 2020, the first cases of what would later be termed "coronavirus" disease emerged outside China, in the United States and Europe, signaling the potential global spread of the virus. By March 2020, the virus had spread across much of the world, with over 730,000 confirmed cases of COVID-19 and more than 36,000 reported deaths. From that point onwards until October 2021, the COVID-19 pandemic became a focal point of national and global policies, resulting in varied responses aimed at controlling virus transmission and alleviating strain on healthcare systems. These responses evolved over time and differed significantly between countries and regions. Measures included promoting physical distancing, mask-wearing, and hand hygiene, as well as implementing lockdowns that restricted public activities and mandated stay-at-home orders. Additionally, extensive testing of both symptomatic and asymptomatic individuals became commonplace, alongside secondary measures aimed at mitigating the economic and social repercussions of the primary interventions.

Language plays a crucial role in navigating the global impact of the pandemic, which has significantly altered everyday life. Its importance extends beyond simply labeling the virus and the associated disease, as seen in initiatives like the IATE database and its definitions. Language serves to conceptualize, describe, and socially negotiate the presence of the virus, its impacts, and the responses to it. With the virus's global reach and varied policy reactions at local levels, it's essential to gain a deeper understanding of how the pandemic is represented in linguistic discourse and to identify both similarities and differences across languages in how the novel coronavirus is discussed.

The pedagogical significance of corpus of bilingual translators of English vs. Chinese is conspicuous in its emerging new trends in interdisciplinary studies and technology-assisted corpus development. Chinese scholars paid substantial attention to important corpora, particularly since research in this field during the Covid-19 pandemic constantly expands and grows in depth. However, the analysis of current statuses, future trends and training models were elicited when exploring publications in WoS and CNKI. Thus, studying the corpus of bilingual translators' training model to provide researchers with holistic overviews of the field is pertinent and essential.

To date, reviews have yet to be reported on bilingual translators within the last decade, from 2019 to 2023. No literature appears to have been found to scientific employed on the summary of bilingual translators' training model publications through a scientometric analysis from the Web of Science (WoS) core collection and China National Knowledge Infrastructure (CNKI) between 2019 and 2023.

The primary aims of this research are to investigate the present state and emerging trend by generating publications found in the WoS and CNKI databases during the Covid-19 pandemic. In addition, this study aims to deliver suggestions on bilingual translators under the achievement of related corpus. This study addresses the following research inquiries:

a. What is the current status of bilingual translators' training model publications from 2019 to 2023 in WoS and CNKI databases?

b. What are the emerging trends in bilingual translators' training model under the corpus background from 2019 to 2023 in WoS and CNKI databases?

2. Literature Review

There has been a noticeable increase in research concerning language usage during the first year of the COVID-19 pandemic. Broadly, two primary areas of investigation have emerged: (1) research focused on monitoring the lexical changes brought about by the pandemic, and (2) studies that analyze linguistic data to understand the social and cultural responses and effects of the pandemic.

Research examining the impact of SARS-CoV-2 on language stocks within various languages can be categorized into lexicographical and lexicological studies, as well as natural language processing approaches. Initially, efforts to monitor coronavirus-related neologisms have been undertaken by linguists, lexicographical institutions, news agencies, language standardization bodies, and lexicologists. Notable examples include glossaries of coronavirus terminology compiled by Time, the Oxford English Dictionary, and the Institut für Deutsche Sprache. These glossaries document new words and are subject to analysis by lexicologists, who explore word formation processes and types of neology, often through qualitative methods. Various studies have been conducted on this topic, such as those by Balnat (2020) for French and German, Pietrini (2021) and Sgroi (2020) for Italian, Ladilova (2020) and Rodríguez Abella (2021) for Spanish, Belhaj (2020) for French terms in Moroccan newspapers, Bowker (2020) comparing COVID-19-related terms in Canadian and European French, and Roig-Marín (2021) for English.

In addition, researchers in the field of natural language processing have begun compiling lists of coronavirus-related terms to aid in the analysis of coronavirus discourse. The Coronavirus Corpus, introduced in May 2020, offers a sizable and regularly updated collection of coronavirus-related data sourced from online newspapers and magazines across 20 English-speaking countries (Davies, 2021). Leaman and Lu (2020) have developed an extensive dictionary cataloging unique terms found in scientific literature referring to SARS-CoV-2 and COVID-19. Taking a broader approach, Ma et al. (2021) not only document names for the virus and associated disease but also encompass COVID-19 terminology across ten different categories, including clinical manifestations and epidemic prevention, resulting in a compilation of 464 concepts corresponding to 724 Chinese and 887 English terms. Lew and Kosem's contribution to Tan et al. (2020) introduces a tool enabling researchers to identify COVID neologisms within a timestamped corpus. Additionally, many linguistic tools developed by and for researchers focusing on COVID-19 language rely on Twitter data and often adopt a broad cross-linguistic perspective. Abuld-Mageed et al. (2020) released MegaCov, a billion-scale geo-located Twitter dataset encompassing over 65 languages (see also Lopez et al. 2020, Chen et al. 2020a).

The databases comprising tweets and keywords, provided by natural language processing, serve as crucial inputs for the second main avenue of research: analyzing the linguistic expressions reflecting the social impact and reception of the COVID-19 pandemic (e.g., as undertaken by Chen et al. 2020b utilizing the dataset from Chen et al. 2020a). Research within this strand is characterized by a methodological division between those adopting more quantitative and qualitative perspectives. However, both approaches typically aim to conduct content analyses of COVID-related messages. They seek to identify potential stigmatizing social stereotypes present in COVID discourse and to assess the emotional weight of COVID language, often paying attention to small-scale diachronic trends observed during the initial months of the pandemic.

In the quantitative domain, content analysis is predominantly conducted automatically, with researchers utilizing various methods such as raw frequency counts of tweets (Singh et al. 2020), the frequency of specific strings in tweets (Abuld-Mageed et al. 2020 focusing on the most common hashtags), topic modeling (e.g., Kurten and Beullens 2021), and keyword identification (Makhachashvili and Bilyk 2020, utilizing SketchEngine on a corpus of French news media, also explored by Spina 2020). Wicke and Bolognesi (2020) investigate the prevalence of WAR metaphors among automatically identified COVID topics, contrasting them with other source domains for COVID (Semino, 2021).

Furthermore, natural language processing techniques have been employed to track changes in sentiment and emotions within COVID discourse. While Kurten and Beullens (2021) integrate sentiment analysis into their broader content analysis of a Twitter corpus, Aslam et al. (2020) utilize automatic emotion classification for a dataset comprising over 140,000 COVID-related newspaper headlines, revealing a predominantly negative sentiment in the headlines.

Various qualitative approaches provide further insights into language usage as indicative of the social impact during the initial year of the COVID-19 pandemic. Nossem (2020) examines how virus names contribute to linguistic rebordering, framing the disease as foreign and thus "a threat to the nation from the outside" (Nossem, 2020). Black (2020), Chen (2017), and Du (2020) identify markers of stigma, stereotyping, and racial profiling in COVID discourse. Parvin et al. (2020) and Kasperè and Motiejūnienė, J (2020) employ a qualitative keyword approach to analyze COVID discourse in mass media corpora.

A Critical Discourse Analysis approach is evident in MIRCo (2020), where a research group reflects on entries made in a quarantine discourse diary. Finally, Zhang and Li (2020)'s special issue on Multilingua features several papers on multilingual and translanguaging practices amidst COVID-19 (Abdul-Mageed, et al., 2020), illustrating how global pandemic trends and local interpretations intersect in social media discourse (Lin, 2021), as well as in specific linguistic practices such as Mongolian fiddle stories poetry (Chen 2017).

In conclusion, the research presented here can be categorized based on three main criteria. Firstly, concerning data sources, many researchers predominantly focus on social media platforms, especially Twitter, although some also examine mass media corpora. There appears to be a need for further investigation into large newspaper corpora reflecting international news coverage.

3. Methodology

As previously mentioned, two databases were utilized for the search. Firstly, in the initial phase of WoS, the key search terms “bilingual translator” combined with either “English” and “learning” were employed. Refinements were made by adding the keyword “Covid-19 pandemic” and selecting document types “Article” and “Review”. This search yielded a total of 107 records between 2019 and 2023. Secondly, in CNKI, the terms “战疫语料库 (corpus of translator during Covid-19 pandemic)” and “中英双语翻译人才培养模式” (English-Chinese translator training model)” were separately searched under the advanced index and then combined with "OR." Refinements were made by selecting document type "Articles" and subject classification "language and education." After deduplication and sorting, the final dataset in CNKI contained 48 records. The total number of publications retrieved from both databases is summarized in Table 1. Upon preliminary examination, it was observed that papers published in CNKI primarily originate from China, whereas those in WoS predominantly represent international studies.

Table1. Retrieval results of the WoS and CNKI database

Database	Year	Retrieving Results
WoS collection	2019-2023	107
CNKI	2019-2023	48

Next, co-word analyses, including co-word citation, co-word clustering, burst detection, and analysis of major citing articles of detected keywords, were conducted on CiteSpace (6.1.R2) to examine the themes and emerging trends in TBTL studies.

4. Findings & Discussion

The COVID-19 pandemic indeed underscored the critical role of translation in overcoming barriers to communication, combatting misinformation, and ensuring access to vital information across languages. Numerous initiatives and measures have been implemented to address the urgent need for multilingual communication during this crisis. Some of the key strategies and developments include:

4.1 Increased Translation Efforts

Governments, international organizations, and NGOs have ramped up translation efforts to disseminate accurate information about COVID-19 in multiple languages. This includes translating public health guidelines, educational materials, and official communications into diverse languages to reach broader populations.

4.2 Technology-Driven Solutions

Translation technologies such as machine translation and AI-powered language tools have played a significant role in facilitating rapid translation of COVID-19-related content. These tools have enabled quick and scalable translation of information across various languages, helping to bridge linguistic gaps.

4.3 Collaborative Translation Platforms

Collaborative platforms and crowdsourcing initiatives have emerged to facilitate community-driven translation efforts. Volunteers, language professionals, and bilingual individuals contribute to translating COVID-19 resources, ensuring a wider reach and cultural relevance.

4.4 Accessible Formats

Efforts have been made to ensure that translated materials are available in accessible formats, including audio recordings, videos with subtitles, and easy-to-read formats. This helps to cater to diverse audiences, including those with visual or literacy challenges.

4.5 Cultural Sensitivity

Translation efforts have emphasized cultural sensitivity and linguistic nuance to effectively convey public health messages across different cultural contexts. Adapting translations to local customs, beliefs, and communication preferences enhances their effectiveness and resonance.

4.6 Coordination and Information Sharing

Collaboration among translation professionals, healthcare authorities, and communication specialists has been vital in coordinating trans-

lation efforts and sharing best practices. This collaborative approach ensures consistency, accuracy, and relevance in translated materials.

4.7 Training and Capacity Building

Efforts to strengthen translation capacities, particularly in languages that are less commonly spoken or under-resourced, have been prioritized. Training programs, workshops, and initiatives to empower local translators and interpreters have been instrumental in building sustainable language support systems.

The skills and competences that need to be acquired by way of contact studying in students' opinion are simultaneous interpreting, practical translation skills and skills related to the use of computer-assisted translation tools and translation and localization software. Language competences were also mentioned by the students among those requiring contact studying. Overall, students felt that translation studies became more monotonous, some of them lost motivation and struggled keeping track of the tasks and deadlines. On the other hand, such weariness and fatigue with the remote mode of studying may be well expected in any kind of training.

Overall, the COVID-19 pandemic has highlighted the indispensable role of translation in facilitating global communication, fostering understanding, and combating the spread of misinformation during crises. Continued investment and innovation in translation infrastructure and resources will be essential to address future challenges and ensure equitable access to information for all communities.

This paper employed the scientometric analysis method using CiteSpace to compare TBTL studies retrieved from WoS and CNKI from 2013 to 2022. The current status, major themes, and emerging trends of TBTL studies were compared, and the major findings include the following:

Unlike the exponential growth in WoS bilingual translator studies, the related studies in CNKI showed a downward trend. Two main reasons were identified for the disparities: a) the research which are related to Covid-19 is decline because of the cure of this disease, in China is generally confined to theoretical discussions, while international research on this area combines theoretical discussions with empirical studies; b) the research in China has a high similarity in research content and design, with a lack of innovation due to the limitation of medicine science.

Both similarities and differences were identified in bilingual translators' training model themes in WoS and CNKI. English-Chinese translator, self-built corpus, and Covid-19 are major themes in both databases. While in WoS, in addition to the fight for Covid-19, studies are also found to be conducted on other aspect of pandemic disease. The relevant themes include pneumonia and infectious disease, and the influence of language development in both English and Chinese. Nonetheless, major related themes in CNKI include college English teaching during the Covid-19 pandemic, training model of bilingual translator etc., which are generally based on different school-running and student enrollment modes.

Through keyword burstness and major citing articles analysis, online teaching and speaking and listening teaching are prominent trends of bilingual translator in CNKI. By contrast, emerging trends in WoS studies involve technology-assisted training model, especially mobile-assisted training model. Oral proficiency development and corrective feedback were also detected as major emerging trends in WoS related studies.

This study is not without limitations. First, scientometrics analysis was conducted by using CiteSpace to extract the themes and emerging trends of training model. Different results may be generated if other tools, such as VOSviewer and Science of Science (Sci2) Tool, were employed. Second, in predicting potential trends, inevitable deviations may exist. Despite these limitations, there is no denying that this study has shed light on this research field and future researchers are recommended to address these limitations to verify the findings of the study.

In addition, new directions were provided for future researchers interested in bilingual translator training model and scholars who focus on the China context. Through the comparison of related studies in WoS and CNKI, it is found that the research in China needs to be further developed, and future researchers could conduct more empirical studies instead of merely focusing on theoretical discussions. In addition, as related research in China is mainly confined to language education, interdisciplinary research should be explored more in the future, such as investigating the combination of the rest with the subjects of psychologies and information technologies.

References

- [1] Abdul-Mageed, M., A. R. Elmadany, E. M. B. Nagoudi, D. Pabbi, K. Verma, and R. Lin. 2021. Mega-COV: A Billion-Scale Dataset of 100+ Languages for COVID-19. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, 3402–20, Online. Association for Computational Linguistics. 10.18653/v1/2021.eacl-main.298. Search in Google Scholar

- [2] Chen, C. (2017). Science mapping: a systematic review of the literature. *Journal of Data and Information Science*, 2(2), 1–40.
- [3] Chen, Y. (2018). The Application of Task-based Teaching Method in Higher Vocational English Teaching. *Journal of Liaoning Higher Vocational*, 20 (8), 44-46.
- [4] Chen, C., Hu, Z., Liu, S., & Tseng, H. (2012). Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. *Expert Opinion on Biological Therapy*, 12(5), 593–608.
- [5] Ellis, R. (2016). Focus on form: a critical review. *Language Teaching Research*, 20(3), 405–428.
- [6] Kasperė, R. and Motiejūnienė, J. (2021). Impacts of global pandemic on translator’s career and translator training. *Current Trends in Translation Teaching and Learning E*, 8, 154 – 195. <https://doi.org/10.51287/cttle 20216>

Funding: The partially achievement of the research “ A study on the training model of Chinese-English bilingual translators against COVID-19 based on self-built corpus”. (SD2021073)