Characteristics of Post-pandemic International Chinese Language Learning: Insights from X-Data

Yanyan Li[1, *], Wenhui Sun[2]

[1]*School of Literature and Journalism, Yantai University, Yantai, Shandong, China*
[2]*School of Mathematical Sciences, Soochow University, Suzhou, Jiangsu, China*
*Corresponding Author.

Abstract: This study uses sentiment analysis and topic modeling to analyze X-data on Chinese language learning both before and after the pandemic. Our analysis reveals shifts in sentiment and topics, as well as their underlying causes. We observe a notable increase in online learning as the primary mode of international Chinese language learning, accompanied by an escalated demand for high quality online learning resources. There is also a discernible polarization of learning sentiments in the post-pandemic era. In addition, the former emphasis on cultural exchange in language learning has shifted to economic and trade interactions, highlighting the commercial facets of international Chinese education.

Keywords: X-data; International Chinese Education; Online Learning; Sentiment Analysis; Topic Modeling

1. Introduction

The field of international Chinese language learning is experiencing unprecedented opportunities and extensive room for development [2,5]. Understanding the needs and characteristics of learners in this context is crucial to improving the quality of international Chinese education [10]. Existing research on the development of international education in China focuses mainly on institutional construction, talent development, curriculum development and evaluation systems [9]. However, most of the research uses traditional methods such as small-scale surveys and interviews.

Due to the COVID-19 pandemic and shifts in the international landscape, traditional offline survey and interview research methods are facing obstacles due to disruptions in offline teaching modes [4,7]. With the widespread adoption and development of social networking platforms, major international social media platforms such as X (formerly Twitter), Quero and Reddit communities have attracted large numbers of Chinese language learners and educators. These platforms serve as important data sources where users share learning experiences, methods, and resources, making them invaluable for understanding and capturing the views and emotions among Chinese language learners.

This study explores Chinese language learners' discussions on social media, specifically on the X platform. By analyzing the content shared by learners, we gain valuable insights into their needs, confusions, and emotional fluctuations in learning. In addition, topic modelling allows us to uncover deeper discourse themes and learning patterns. Our research provides valuable insights into the field of international Chinese education based on massive online data.

2. Methodology

We collected X-data from January 1, 2019, to April 30, 2023, using Python web scraping scripts, resulting in a total of 28,800 tweets related to Chinese language learning. To ensure data accuracy, we conducted preprocessing, including removing low relevance texts generated by fuzzy indexing and retaining only text data due to the complexity of processing image and audio texts. After preprocessing, we retained a total of 5,737 valid tweets, including information about the publisher, publication region, publication time, and text content.

We divided the collected data into three stages based on time. Data from January 2019 to December 2019 is designated as the pre-pandemic stage. Given that major countries announced relaxation of restrictions and implemented a policy of "living with the

pandemic" in early 2022, we designated the period from January 2020 to February 2022 as the pandemic stage. Finally, data from March 2022 to the end of April 2023 is designated as the post-pandemic stage.

We intend to address the following three questions. First, during and after the pandemic, what are people's main concerns in Chinese language learning? Second, what are their attitudes towards learning? How do emotions vary among learners over time? What might be the reasons for these variations? Third, in the post-pandemic era, how do people's preferences and motivations differ from those of the past?

For the first question, we use word frequency analysis [1] to uncover the focus of Chinese language learning. For the second question, we use sentiment analysis tools [8] to analyze X-data, with a particular focus on the changes of learners' emotional attitude before and after the pandemic. For the third question, we use the Latent Dirichlet Allocation (LDA) model [3] to extract themes and uncover underlying characteristics and trends.

3. Research Results

3.1 Word Analysis

TF-IDF stands for "Term Frequency-Inverse Document Frequency". It is a widely used algorithm in text mining that assesses the significance of a term in a document compared to a collection of documents [1]. It achieves this by calculating the frequency of a term within a document and then adjusting it based on the inverse of its frequency across all documents.

According to the frequency analysis of the X-data, we can see from Figure 1 that the focus of Chinese learning during and after the pandemic mainly include disciplines and fields related to Chinese learning, methods and content, and language and culture.

Figure 1. Word Clouds of X-data

In the field of Chinese language and its disciplines, the most frequently mentioned

words are such as methods (use), experiences (way), resources (app), and examinations (HSK).

In terms of learning methods and content, the word "online learning" appears with a high frequency, indicating that social media users pay attention to online learning resources [4]. Specifically, online language courses (class), fee models (free), and applications (app) have become core issues for users.

Note that language serves not only as a means of communication, but also as a channel for cultural exchange. In X-platform discussions, the highly frequent use of the term "Culture" suggests that learners are actively engaging with both linguistic and cultural dimensions. This underlines their desire for a deep understanding of China and its cultural intricacies [9].

3.2 Sentiment Distribution

To explore the emotional variation of Chinese language learners, we employed sentiment analysis using ROST-CM, which is social computing platform developed and coded by Wuhan University, to compute sentiment scores for each text data. By aggregating sentiment scores across all texts, we gained insights into the distribution of positive, negative, and neutral sentiments.

Overall, attitudes towards learning Chinese show a positive trend in sentiment, with most learners expressing a favorable attitude towards learning. Even during the pandemic, learners' positive emotions surpassed neutral and negative sentiments. Combining the insights from Figure 1, it seems that most learners find learning Chinese "fun", "interesting", or even "enjoyable".

However, noteworthy is the polarization of learners' emotions before and after the pandemic. On one hand, some learners exhibited relatively higher levels of positive emotions before the COVID-19 outbreak (56.48%), which declined during the pandemic to 51.59%. Yet, post-pandemic, learners' emotions swiftly rebounded, with interest and willingness to learn Chinese returning to normal or even heightened levels, reaching 58.80% in positive sentiment. On the other hand, some learners' negative emotions did not alleviate post-pandemic; in fact, they increased. As observed in the figure, during the pandemic,

there was a noticeable rise in negative emotions among some learners, from 13.95% pre-pandemic to 21.94% during the pandemic, and even post-pandemic, negative sentiment reached 15.59%, exceeding pre-pandemic levels. This demonstrates the impact of the pandemic on learners' emotional fluctuations, with some rediscovering joy, while others experienced heightened apprehension.

The reasons behind negative emotions are complex. Through further analysis, we identified several sources contributing to negative emotions, including the difficulty of learning leading to frustration and psychological pressure, anxiety due to constraints in accessing necessary learning resources caused by technical factors such as network connectivity and learning platforms, challenges in adapting to online teaching mode resulting from the shift from traditional face-to-face instruction, and feelings of loneliness and helplessness stemming from the lack of face-to-face interaction and communication.

3.3 Topic Themes in X-Data

This section employs Latent Dirichlet Allocation (LDA) [3], a probabilistic graphical model for topic modeling, to explore the thematic evolution within X-data. By extracting the top four ranked topics (see Table 1) in each phase and analyzing the text-to-topic and word-to-topic distributions, we unveil the dynamics of thematic content.

Prior to the pandemic, discussions primarily revolved around Chinese language learning, encompassing language acquisition, cultural exchange, and travel experiences.

However, during the pandemic, shifts in teaching methods propelled discussions on online learning and the utilization of new media technologies, alongside a notable prominence of discussions intertwining the pandemic with political discourse.

Post-pandemic, topics related to cultural exchange and travel experiences significantly declined, reflecting a reduction in international interactions. Conversely, discussions among Chinese language learners increasingly centered on topics related to economic and trade activities, including discussions on business opportunities, trade dealings, abilities, and careers, highlighting the enduring appeal of the vast Chinese market and the pivotal role of deepening international economic cooperation in the dissemination of international Chinese education.

Table 1. Topic Themes in X-Data

Period	Topic Theme	Key Words
Before pandemic	Education and family activities	write hard practice free garden teach kids video back festival
	Travel and culture	tour city complete walkthrough amazing reading pavilion sun writing temple
	Daily life and history	confine shantang street museum places started liveliest daily keju minutes
	Language learning and art	bilingual pages western painting describe classes week back American free
During pandemic	Epidemic and politics	science ban cotton coronavirus steal vaccine states story united teaching
	Online education and home learning	online teacher family video working interesting week medical calligraphy educational
	Technology application and family education	American talking company find back stay level kids app teacher
	Arts and learning	art hard life back machine beginner schools quarantine Arabic online books
After pandemic	Online education and skills development	progress app lessons duolingo kids teaching graph skills schools lesson
	Business and Work	business ability symbol career astrology integrity dealings standards highest free
	Teaching and learning methods	grammar lesson easy common tongue effort hear native traditional stop
	Learning emotion	fun back literally native daily thinking hard future advance happy pleased

4. Discussion

Based on the previous findings, we recommend that educational administrators and instructors proactively plan and prioritize the current hot topics and themes of learners, understand their learning motivations and needs, and help mitigate external barriers affecting learners. We propose the following suggestions for planning international Chinese education.

First, emphasizing the integration of online and offline teaching. Given the prominence of online learning and the utilization of new media technologies observed during the pandemic, emphasizing the integration of online and offline teaching is crucial. This approach can enhance accessibility and flexibility for learners while leveraging the advantages of digital platforms [7] to provide high-quality educational resources.

Second, leveraging the communicative and economic value of Chinese learning. Recognizing the increasing interest in topics related to economic and trade activities post-pandemic [2], aligning Chinese language education with professions and economic sectors can enhance its practical relevance and appeal to learners. This strategy can also contribute to the broader goals of economic cooperation and cultural exchange between countries.

Finally, addressing external factors affecting learning emotions and international communication risks. Acknowledging the impact of external factors, such as international communication risks and barriers to cultural exchange, underscores the importance of proactive measures to mitigate these challenges. Strengthening international cooperation and introducing supportive policies can foster a conducive environment for Chinese language learning and promote cross-cultural understanding [6].

References

[1] F. Heimerl, S. Lohmann, S. Lange and T. Ertl. Word Cloud Explorer: Text Analytics Based on Word Clouds[C]. 47th Hawaii International Conference on System Sciences, 2014.

[2] F. Hu, H. Zhang and J. Chen. Challenges, Opportunities and Countermeasures of International Chinese Education in the Post-epidemic Era[J]. TCSOL Studies, 2022, 2:49-56.

[3] H. Jelodar, Y. Wang, C. Yuan, et al. Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, A Survey[J]. Multimed. Tools Appl., 2019, 78:15169-15211.

[4] Ivana. The Effectiveness of Online Learning in Learning Chinese Language[J]. International Journal of Research and Review, 2021, 7:116-120.

[5] B. Li and X. Shang. New Opportunities, Challenges and Accomplishment of Chinese International Communication Against the Background of the Belt and Road Initiatives[J]. Journal of Liaoning University (Philosophy and Social Sciences Edition), 2018, 46(02):121-130.

[6] J. Li and X. Eryong. The Rising Soft Power: An Educational Foreign Exchange and Cooperation Policy Conceptual Framework in China[J]. Educational Philosophy and Theory, 2022,55:1329 - 1338.

[7] R. Ma and J. Xu, Construction of Smart Education Evaluation System: A Case Study of International Chinese Language Education[J]. International Journal of Information and Education Technology, 2023, 13:279-288.

[8] P. Nandwani and R. Verma. A Review on Sentiment Analysis and Emotion Detection from Text[J]. Soc. Netw. Anal. Min., 2021, 11:81.

[9] J. Ruan, X. Duan and X. Duan. Tasks and Learner Motivation in Learning Chinese As a Foreign Language[J]. Language, Culture and Curriculum, 2015, 28:170-190.