

Emerging Focus in Chinese Literary Study (1927 to 2023): Latent Dirichlet Allocation (LDA) Based Topic Modeling Analysis

Ruikai Yuan¹ , Si Shen², and Yu Zheng³

Abstract

Literature has vital historical values that reveal the truth of culture, ideology of people, political focus, and human experiences like a series of no-photo films via a combination of flat words. As the carrier of history, literature plays like one decent clerk who recorded the historical presentation of modern China through a proliferation of scholarly articles. This study aims to provide a comprehensive overview of the emerging focus of modern Chinese literary research with the Latent Dirichlet Allocation (LDA) topic model to explore the research themes and trends of former researchers from 1927 to 2023. LDA is a text-mining-based approach utilized to reveal principal themes from a substantial dataset of short textual documents. A total of 14,148 articles published between 1927 and October 2023 were collected and analyzed. Findings suggest that the primary scholarly focus areas of CNKI have been significantly influenced by historical events and foreign cultures, notably the Soviet Union and America. Over the past decades, the research trend has also shifted from revolution, Chinese ambient culture, and Soviet literature to American literature, gender, digital tools, and global issues. The primary research themes and new trends identified by LDA are instrumental in aiding researchers in discerning contemporary research questions and making more informed decisions. The findings of this research could be utilized in complementarily exploring the ideology of Chinese people in different periods.

Plain Language Summary

Key trends and topics in Chinese literary studies from 1927 to 2023 using a computer-based method to analyze hidden themes in texts

This paper explores the literature in literary discipline published from 1927 to 2023 to reveal the changing ideology of China. The method applied in the study is LDA topic modelling. The result shows a great change in intellectual focus and underscores the consistency of literary research and social change.

Keywords

literary study, topic modeling, LDA, text mining, Chinese literature

Introduction

Literature has always been a focal point of study and discourse, serving as a mirror reflecting human culture and experiences across time and space (Klarer, 2013). It not only reveals cultural values and themes that transcend geographical and temporal boundaries but also reflects the diversity and complexity of human societies. The discipline of literary studies has experienced

¹National University of Malaysia, Bangi, Malaysia

²Baekseok University, Cheonan, Chungcheongnam-do, South Korea

³The Hong Kong Polytechnic University, Kowloon, China

Corresponding Authors:

Si Shen, Baekseok University, Dongnam-gu, Cheonan, Chungcheongnam-do 31065, South Korea.

Email: tinashen111@163.com

Yu Zheng, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon 999077, Hong Kong SAR, China.

Email: yuzheng@polyu.edu.hk



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of

the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

substantial changes in recent decades, particularly in the volume and quality of scholarly articles. With the evolution of research methodologies and the adoption of new analytical tools, literary research is entering a new phase, including the use of text-mining techniques for a deeper understanding of literary works and themes (Chu et al., 2022; Moretti, 2013). However, scholars from both within and without China have not systematically analyzed the evolving changes of literary concerns caused by the ideological transformation of modern China.

Research has been conducted to review specific themes and methods in certain areas of the literary field using qualitative or quantitative methods. For example, Yaacoba and Newberry (2019) provide a historical overview of the development of comparative literature, discussing its origins, key schools of thought, and the theoretical debates surrounding its scope. Durán (2020) critically examines the transnational turn in American literary studies, analyzing its impact on comparative and international approaches. H. Z. Yuan (2018) reviews a decade of research on trauma narratives in CSCI journals, highlighting key themes such as family, social, war, and racial trauma in both Chinese and foreign literary works. Guy et al. (2018) explore the intersection of literary stylistics and authorial intention, critically assessing the claims of cognitive poetics in integrating empirical approaches. Despite providing insights into evolving literary discourses in shaping contemporary literary scholarship, these studies are often limited by their qualitative nature, small sample sizes, or selection bias, making it difficult to capture large-scale shifts in research trends comprehensively (Munthe-Kaas et al., 2019; Queirós et al., 2017). As for the quantitative method, Sun (2023) and Wei & Li (2022) conducted visual analysis of literature related to English and Chinese literature in the CNKI database in a literature review, discussing the research hotspots of different genres and topics. D. Wang (2022) explored the research horizon in literature studies is tend to merge with other disciplines, with a clear trend toward interdisciplinary research within the Chinese literary community. However, their research only focuses on a specific literary scope within a certain period.

This study aims to address these limitations by employing Latent Dirichlet Allocation (LDA), a machine learning-based topic modeling technique, to analyze a large corpus of 14,148 scholarly articles published between 1927 and 2023. Unlike traditional literature reviews that rely on manual categorization and subjective interpretation, LDA enables the identification of latent thematic structures in an objective and scalable manner (Chauhan & Shah, 2021; Gurcan & Cagiltay, 2019; Vulić et al., 2015). This approach allows for a data-driven exploration of how historical, political, and cultural

changes have shaped Chinese literary scholarship over nearly a century. Additionally, this study differs from prior research by providing a longitudinal analysis of the shifts in literary focus over multiple distinct historical periods. By systematically tracing changes from the revolutionary and socialist realism-dominated early 20th century to the present-day emphasis on digital humanities, globalization, and gender studies, we offer an insight into how literary research has evolved in response to external influences. To be specific, this research contributes to the field by: *a.* providing a comprehensive, data-driven review of Chinese literary research trends over nearly a century using computational methods; *b.* identifying previously unexamined macro-level shifts in Chinese literary studies by analyzing a large-scale dataset; *c.* highlighting key historical and ideological influences that have shaped literary scholarship in China; *d.* offering insights into emerging trends that may guide future research directions and funding priorities. To achieve these goals, we break down our overarching purpose-level questions into the following four research questions (RQs):

RQ1: What have been the bibliometric characteristics of Chinese literary research during the period between 1927 and 2023?

RQ2: What have been the emerging topics in the Chinese literary field in the period between 1927 and 2023?

RQ3: How have the topics of interest in Chinese literary studies changed from 1927 and 2023?

RQ4: What are the future trends in the Chinese literary field?

This study seeks to bridge the gap between traditional literary analysis and computational approaches, demonstrating the potential of machine learning techniques in advancing literary scholarship. The findings of this study uncover the historical, cultural, and political influences on Chinese literature and guide future research directions based on these insights. The findings are instrumental in aiding researchers to discern contemporary research questions and make more informed decisions, complementarily exploring the ideology of Chinese people in different periods.

Literature Review

Topic Model

Topic modeling is a machine learning technique that has gained significant traction in the field of natural language processing and text mining (Abdelrazek et al., 2023; Churchill & Singh, 2022; Vayansky & Kumar,

2020; Blei & Lafferty, 2006). It is an unsupervised method that aims to discover the latent themes or “topics” present within a collection of documents (D. M. Blei et al., 2003; Suominen & Toivanen, 2016). Topic modeling can analyze the co-occurrence patterns of words and their distributions across the corpus, then automatically identify and extract the underlying topics that permeate the text data. The aims of topic modeling are twofold: to categorize documents into specific topics and to uncover the blend of topics that each document comprises (L. Liu et al., 2016; Mohr & Bogdanov, 2013).

Topic modeling offers several advantages over traditional methods of text analysis (Debortoli et al., 2016). First, they enable the automatic discovery of topics without the need for predefined categories or manual labeling (Ball & Lewis, 2020; X. Chen et al., 2024; Park et al., 2024). This unsupervised approach allows for the exploration of previously unknown or unanticipated themes within the data (Bansal et al., 2020; Hannigan et al., 2019; Koo, 2022). Second, topic modeling provides a quantitative representation of the topics, allowing for further analysis and comparisons across different subsets of the corpus or over time (Isoaho et al., 2021; Sbalchiero & Eder, 2020). Third, it offers a powerful dimensionality reduction technique, transforming high-dimensional text data into a lower-dimensional representation of topics, which can facilitate various downstream tasks such as information retrieval, document clustering, and trend analysis (George & Sumathy, 2023; J. W. Lee & Han, 2023).

Beyond its applications in natural language processing, topic modeling is widely used in various domains, including digital humanities (Callaway et al., 2020; C. M. Chen et al., 2023), social media (Boon-Itt & Skunkan, 2020; Negara et al., 2019), marketing (Amado et al., 2018; Reisenbichler & Reutterer, 2019), and psychology (S. Liu et al., 2021).

LDA Topic Modeling

One of the most widely used topic modeling techniques is Latent Dirichlet Allocation (LDA). Introduced by Blei and Jordan in 2003, LDA has become a cornerstone methodology for extracting hidden topics from vast amounts of unstructured text data. At its core, LDA is a generative probabilistic model that posits documents as mixtures of latent topics, where each topic is characterized by a distribution over words. This model allows for a deeper understanding of the content by identifying patterns of word clusters that frequently co-occur across documents, thereby revealing the predominant themes or topics within the corpus.

Implementing LDA in literary analysis involves several steps: pre-processing the text data (Sbalchiero &

Eder, 2020; Shahbazi & Byun, 2020), selecting an appropriate number of topics (Hasan et al., 2021; Li & Lei, 2021), running the LDA algorithm, and then interpreting the resulting topic distributions. The model can efficiently handle large datasets which makes it an invaluable tool in digital humanities, where the analysis of extensive text collections is usual. Especially in literary study, the application of LDA marks a paradigm shift from traditional close reading methods to a more data-driven approach (Underwood, 2019). It facilitates the uncovering of topics and structures that might not be immediately apparent through conventional reading methods. This technique is particularly valuable in the exploration of large bodies of work, like an author’s complete oeuvre, or across different literary periods and genres.

The elegance of LDA lies in its ability to decompose a document corpus into two key components: the mixture of topics that each document comprises, and the distribution of words that defines each topic. This dual focus not only facilitates an analysis of individual documents in the context of their dominant themes but also enables the examination of words’ significance and their contribution to the topics they belong to. LDA improves its understanding of the structure of the corpus by going through each document over and over again and changing the topics based on how common words are within topics and how they are spread out across documents. The algorithm proceeds until it reaches a stable state where the topic distribution throughout the documents no longer changes significantly, culminating in a comprehensive portrayal of the corpus’s topics.

Scholars such as Jockers and Mimno (2013) have demonstrated LDA’s utility in identifying themes across literary works. It helps reveal patterns that transcend individual authors, historical periods, or genres (Underwood et al., 2018). For instance, Jockers and Talken (2020) used topic modeling to analyze 19th-century Anglo-American novels, uncovering thematic trends and relationships among authors. In an eminent work by Goldstone and Underwood (2014), they utilize the topic model to analyze 21,367 literary articles from 1889 to 2013, revealing gradual shifts in themes and scholarship rationales that challenge established historical narratives. Their findings spotlight the emergence of subjects like violence and the innovation within core theoretical concepts, extending beyond traditional frameworks like “New Criticism.” Later, Piper (2015) uses computational models as tools to explore how narrative structure and emotional engagement reshape the reader’s critical perceptions by emphasizing the affective responses elicited by literary engagement. In analyzing historical documents, LDA can bring out previously unknown thematic structures, helping historians interpret shifts in discourse

over time. Researchers like Newman and Block (2006) have utilized LDA to track ideological shifts in the American press throughout the 19th century. LDA has also contributed to cultural studies by providing a means to identify the thematic structures that underpin various media forms. For instance, DiMaggio et al (2004) used it to explore topic patterns in social science research articles.

Important Changes in Modern China from 1927 to the Present

The feature of Chinese literature shifts with the power shifts, cultural norms, and social structures of the cultures. The modern history of China from 1927 to 2000 witnessed tumultuous events and profound ideological shifts that shaped the nation's course (Cheek, 2016; Hutchings, 2001). The period began with the conflict of civil war. The First Chinese Civil War (1927–1937) saw a direct confrontation between capitalist ideals and the communist vision (Esherick, 2000). Literature of this era delved into the socio-economic implications of these opposing systems and the ideological underpinnings that fueled the conflict. In the next period, China experienced the Second Sino-Japanese War (1937–1945), and drama emerged as the most prevailing genre (Esherick, 2000).

The Second Chinese Civil War (1946–1949) reignited the ideological struggle between the two parties, with the Communist Party of China ultimately emerging victorious (Esherick, 2000). From 1950 to 1978, China underwent significant transformations (Hutchings, 2001). During this period, China gradually got rid of the influence of the Soviet Union and integrated with the world. Transmitting to the 1970s, China initiated the Reform and Opening-up policy. This marked a pivotal turning point for China, steering it away from the isolationist policies of the past and toward a path of economic liberalization and global integration (Hutchings, 2001). From 2001 to 2023, China experienced rapid economic growth, technological advancements, and a rise in global influence. This recent period, which will be the focus of further analysis, witnessed the continuation of market reforms, the emergence of a growing middle class, and the challenges of balancing economic development with environmental sustainability and social equity (Huang & Bernhardt, 2024).

Methodology

This study employs a mixed-methods computational text analysis research model based on unsupervised machine learning techniques. The model follows a sequential exploratory design where quantitative text mining approaches are used to identify patterns, followed by

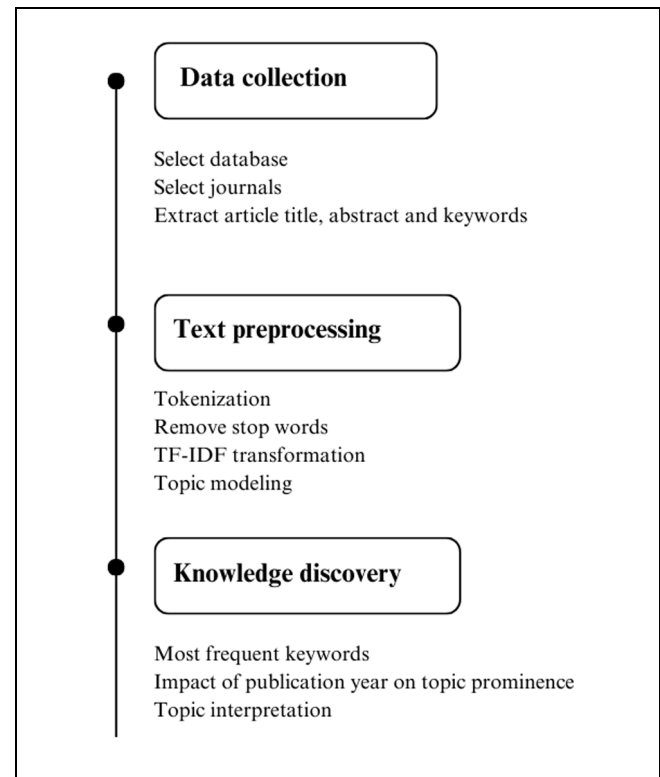


Figure 1. Research process flowchart.

qualitative interpretation of the emergent topics. The research model consists of three primary components: (1) a comprehensive bibliometric analysis of literary research publications to address RQ1, (2) an LDA topic modeling component used to identify emergent themes and address RQ2, and (3) a temporal analysis component that examines the evolution of topics over defined historical periods to address RQ3 and RQ4.

This computational text analysis model was selected for its ability to systematically process large volumes of textual data while minimizing researcher bias in the initial identification of themes. The model allows for both the discovery of latent patterns in the corpus and the interpretation of these patterns within historical and cultural contexts, making it particularly suited for longitudinal analysis of literary scholarship trends. The overall workflow of the study is presented in Figure 1. The first step is data collection. In the next step, we preprocess the data for analysis. The third step involves discovering the knowledge to obtain comprehension from the corpus using text-mining techniques.

Data Collection

The research aim of this study is to identify the academic trend in Chinese literature academy, National Knowledge Infrastructure (CNKI) and Chinese Social

Table 1. Number of Articles from Each Period.

Period	Database	Number of papers
1927–1937	CNKI	28
1938–1945	CNKI	57
1946–1949	CNKI	10
1950–1978	CNKI	772
1979–2000	CNKI	2,000
2001–2024	CSSCI	2,000

Sciences Citation Index (CSSCI) databases were selected. CSSCI is the core database in CNKI, it was developed by the China Social Science Research Evaluation Centre of Nanjing University. The database is used to search for papers included and literature cited in Chinese social sciences and has been a landmark project in the field of humanities and social sciences evaluation in China since 1998. The search keywords in the title or abstract include “literature,” “novel,” or “poet.”

The review period began in 1927, as abstracts from the aforementioned databases have typically been included in CNKI since 1927. Before 1979, there were 860 papers searched. After that, the number of papers increased significantly. For 1979 to 2000 and 2001 to 2024, we chose the 2,000 most highly indexed papers respectively. The period from 1927 to December 2023 is 96 years. In total, 4,095 results were selected, the result of each period is presented in Table 1.

Data Processing

Text preprocessing involves several steps to refine the data. The first step in text preprocessing is tokenization. This involves breaking down the text into smaller units called tokens, which are typically words or terms. Tokenization is done using delimiters such as spaces, hyphens, and other punctuation marks. The next step is removing stopwords which add little to no semantic value to the text. In Chinese, common stop words include particles and auxiliary words such as “的” (de), “了” (le), and “是” (shì). For Chinese text, the process can be more complex due to the nature of the written language, which does not use spaces to separate words. Jieba is a popular Python library used for processing Chinese text, it can effectively handle the segmentation of Chinese text into tokens, considering the context and the structure of sentences.

In the next phase, the method of Term Frequency-Inverse Document Frequency (TF-IDF) is applied to evaluate the significance of each term within the body of text. Commonly occurring terms are sometimes not

indicative of valuable content. Certain terms, often found in scientific literature such as “study,” “meaning,” “presentation,” and the like, may not carry significant meaning and are therefore omitted in subsequent topic modeling processes. The TF-IDF method adjusts the word corpus by diminishing the weight of frequently appearing terms across multiple documents and amplifying the significance of rarer terms that may be more telling of a document’s content, as noted by Kaiser & Ali (2018). After computing the TF-IDF scores for all terms and arranging them accordingly, a threshold is set after a manual review of terms with low TF-IDF scores. Those falling below the threshold are discarded and the count of the remaining terms is updated. The computation of TF-IDF encompasses two elements: TF and Inverse Document (IDF. TF measures how often a term appears, normalized by the document’s length to account for variability in document size. Assuming a collection with D documents and V distinct terms, the TF for a term i in document j (TF_{ij}) is ascertained by its frequency relative to the total term count within the document.

$$TF_{ij} = \frac{C_{ij}}{N_j}$$

The variables C_{ij} and N_j represent the frequency of occurrence of the i th term and the overall amount of terms in the j th text, accordingly.

IDF serves as the second component, quantifies the significance of a phrase. A phrase that appears frequently across several documents, such as “model” and “study” in scientific articles, may be seen as less significant. The IDF value is employed to impose a penalty on frequently occurring phrases. It is computed as follows:

$$IDF = \log_2 \left(\frac{D}{D_i} \right)$$

The IDF value of the i th phrase is denoted as IDF_i , while D_i represents the count of documents containing the i th term. The TF value of the i th term in the whole corpus will be obtained by scaling the TF value with the IDF value in order to standardize frequency values, as demonstrated below:

$$TF_i = \sum_{j=1}^D TF_{ij} * IDF_i$$

After applying the preprocessing steps (tokenization, stopword removal, and TF-IDF weighting), the same text excerpt is transformed as follows:

“康拉德”，“英国”，“晚唐”，“余华”，“心理学”，“翻译”，“白话”，“自由主义”，“革命”，“田园”，“斗争”，“农村”，“变革”……]

Topic Modeling

In this study, the LDA package was used. The number of topics K shall be defined when specifying the LDA. LDA topic clustering was obtained as a group of keywords, and we assigned labels to each topic cluster after summarizing them. for the purpose of word cloud visualization, we translated each Chinese word with the computation result was translated. Due to the complexity of the literary field, we artificially translated each word to preserve the original cultural meaning as much as possible.

The reliability and validity of the LDA topic model were insured by implementing multiple validation approaches. First, coherence scores were utilized to evaluate the quality of the generated topics. Specifically, we calculated the C_V coherence measure, which assesses the semantic similarity between high-scoring words in each topic (Röder et al., 2015). We tested models with different numbers of topics (K ranging from 5 to 30) and selected $K=10$ as it produced the optimal coherence score of 0.57, indicating strong semantic relatedness within topics (Newman et al., 2010). The ten topics were characterized by a list of high-probability words. Below are examples of two topics generated by the LDA model, illustrating the types of themes identified in the corpus:

Topic1:

(1, ‘0.0381*“革命” + 0.0209*“阶级” + 0.0132*“政治” + 0.0117*“城市” + 0.0074*“知识分子” + 0.0058*“意识形态” + 0.0050*“讽刺” + 0.0037*“幽默” + 0.0014*“左翼” + 0.0013*“郁达夫”’)

Topic2:

(2, ‘0.0353*“艺术” + 0.0083*“科学” + 0.0082*“苏联” + 0.0067*“诗歌” + 0.0057*“戏剧” + 0.0051*“战争” + 0.0042*“上海” + 0.0010*“台湾” + 0.0008*“英勇” + 0.0007*“报告文学”’)

Then, we employed perplexity measures to assess how well the probabilistic model predicts unseen data. Lower perplexity values indicate better generalization performance (Wallach et al., 2009). We conducted 5-fold cross-validation, training the model on 80% of the corpus and evaluating it on the held-out 20%. The perplexity score for our selected model ($K=10$) was 437.2, which outperformed models with alternative K values. Third, we ensured the stability of our topic model by performing multiple runs with different random initializations and measuring the similarity between resulting topic distributions using the Jensen-Shannon divergence. The mean pairwise divergence

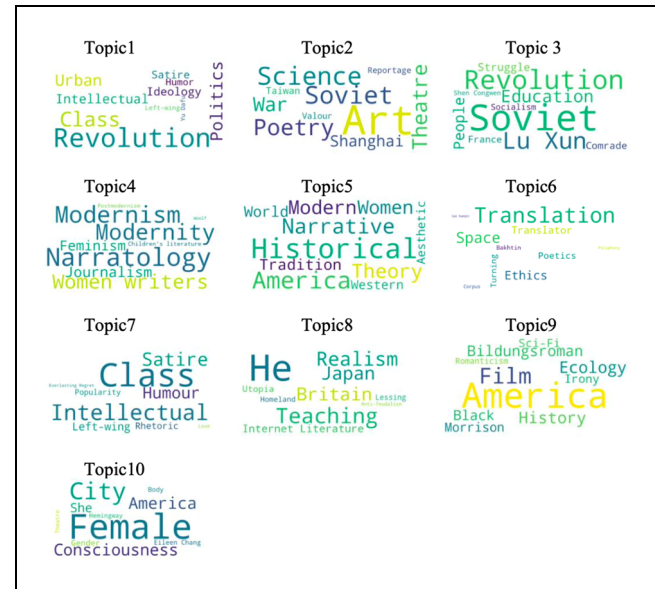


Figure 2. Word cloud.

across ten runs was 0.08, indicating high reproducibility of the topic structure (Maier et al., 2021).

Results and Discussions

To provide a comprehensive understanding of the evolving focus in Chinese literary studies from 1927 to 2023, our analysis integrates both quantitative and qualitative approaches. Through LDA topic modeling, we identified key themes and their distribution across the entire dataset. This section presents the detailed findings, highlighting how these themes reflect the influence of historical events, cultural exchanges, and ideological shifts over nearly a century. Additionally, we explore the prominence of these topics across different historical periods, providing insights into the evolution of Chinese literary research.

Topic Modeling Analysis

The current study employed LDA to analyze a century-long dataset of Chinese literary research, revealing thematic shifts influenced by historical events and cultural exchanges. Figure 2 presents the ten topics along with word clouds of keywords for each topic. The displayed keywords are those with the highest likelihood of occurring within each topic, where a larger font size indicates a higher probability of occurrence. To analyze the prominence of each topic across different historical periods, we categorized the approximately 100-year time span into six distinct eras based on historical stages (refer to literature review). We then conducted a topic

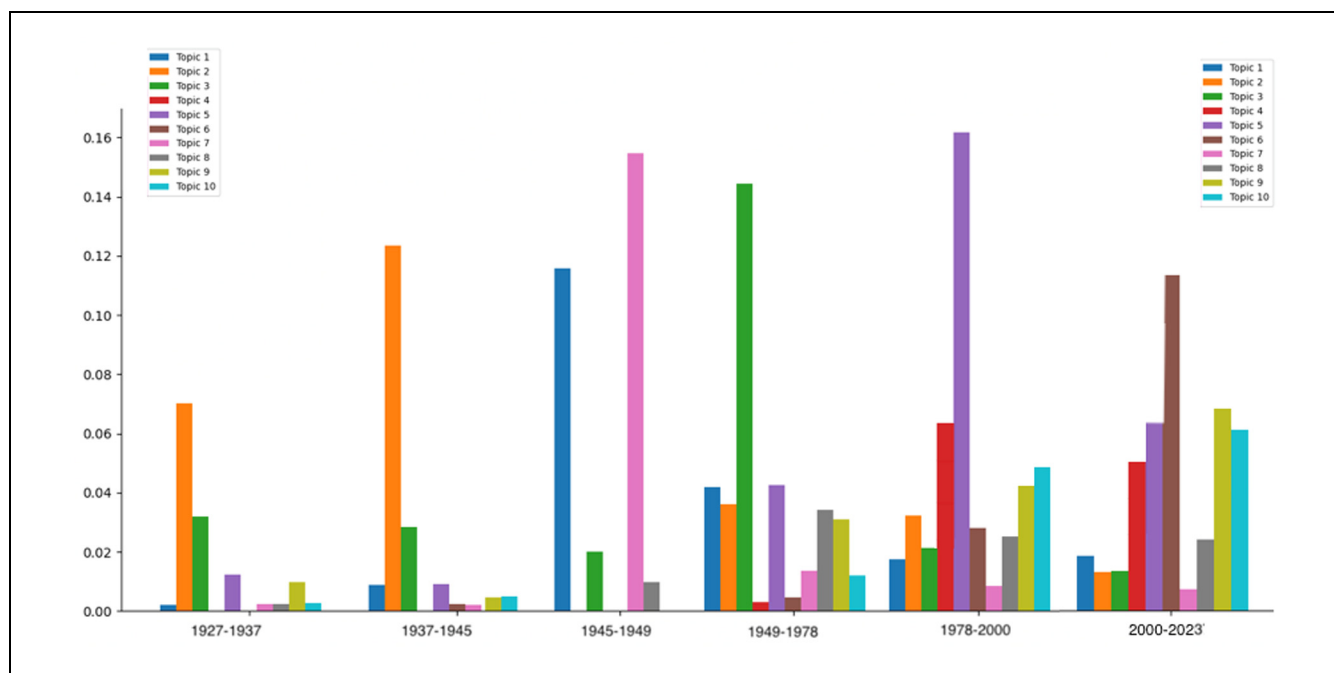


Figure 3. Publication period on topic prominence.

prominence visualization, as shown in Figure 3, to illustrate how each topic emerged and evolved over these periods.

During 1927 to 1937, some left-wingers in China were influenced by the Soviet Russian Revolution and attempted to draw on the revolutionary experiences of the Soviet Union after the October Revolution (X. Liu, 2004). As a result, “Soviet Union” became a prominent literary theme and the most important keyword for Topics 2 and 3 during this period (see Figure 3). Meanwhile, America experienced the prosperity era, becoming one of the world’s largest economies and dominating in fields such as technology and culture. However, the period witnessed the bud of cultural exchange between China and America (Yeh, 2000; Zhao, 2021), the study of American literature (Topic 9) in the Chinese literary community was still minimal.

This period saw China face cultural influences and challenges from the West, since in the mid-19th century, China had been subjected to a series of unequal treaties with Western powers, leading to significant territorial and economic concessions (Guo, 2022; Le Moli, 2021). This created a direct Western presence and influence in China, which exposed Chinese society to Western ideas and lifestyles. Intellectuals and cultural figures began actively exploring modern Western culture and thought, attempting to absorb and adapt its excellent achievements to the Chinese context (Li, 1991; Yang, 2022). Meanwhile, China was in the midst of the Chinese Civil

War, experiencing political turmoil and frequent social changes. This turmoil and change spurred reflection and reconsideration of social realities and political systems, motivating cultural figures to actively participate in the cultural movement, thereby promoting cultural prosperity and innovation (Ding, 2018; Furth, 1983). This process is also known as the New Cultural Movement. Keywords during this period include “science” and “art” from Topic 2, and “revolution” from Topic 3. A core aspect of the New Cultural Movement was the advocacy of science and the democratization of art education. Some intellectuals and scientists began focusing on modern scientific theories and technologies, and this interest in science was also reflected in literature. One of the key figures, “Lu Xun,” emerged as a leading figure in Topic 3. His works, such as *A Madman’s Diary* and *The True Story of Ah Q*, reflect themes of revolution, struggle, and the plight of the common people (L. O. F. Lee, 2023). During this time, there was also some research into literary theory (Topic 5), although its proportion was still very small.

From 1937 to 1945, China remained in a state of internal turmoil (Ma, 2014), and the literary world largely continued the basic characteristics of the previous period. However, more importantly, this period saw the outbreak of the Second Sino-Japanese War. During the war, several literary works emerged, reflecting themes of resistance (Zhang, 2023), such as *Langya Shan Five Heroic Men*. However, before the founding (1949) of the PRC,

the illiteracy rate in China was over 80% (Jowett, 1989), making “drama” (Topic 2) the most accessible form of art for the public.

After China finished the war in 1945, Chinese literature reflected on the experiences of the war, exploring the impact of war on China and beginning to pay attention to the changes in post-war society. The study of “Japan” (Topic 8) helped the Chinese understand and reflect on the traumas of the past war. After the war, internal conflicts in China resurfaced, and there was conflict between two opposite classes: one was the landlord and bourgeois classes, which wanted to be conservative and traditional in cultural matters, the other was the peasant and working classes, which leaned toward revolution and progress (S. S. Hu & Wang, 1997). The struggle between classes had a significant impact on literature, with writers actively expressing their support for or opposition to different political forces, responding to the political challenges of the time (Wen, 2023; Zheng, 2008). According to the results of LDA topic modeling, prominent themes can be seen, such as the keywords revolution, class, politics, and ideology (Topic 1), and left-wing, intellectual (Topic 7). The literary techniques of this period often used humor and satire to convey their messages (Topic 7).

The end of this period was marked by the establishment of the PRC in 1949. Literature was regarded as an important educational tool for cultivating people in terms of values (Da, 2019). Therefore, socialist realism became the dominant literary theory and practice, encouraging writers to create works that reflected the positive aspects of socialist construction (W. S. Wang, 2007). Keywords like “education” and “socialism” in Topic 3 which were prominent in this period better demonstrate this situation. The decline of Topic 1 (class, intellectual, urban) and Topic 7 (humor, satire) indicates that in a political environment emphasizing class struggle, urban intellectuals and artistic forms like humor and satire were restricted. The decline of Topic 2 (art, poetry, science) points to limitations on broader artistic and scientific discussions during this period, reflecting the direct intervention of political movements in literary and artistic fields.

At the same time, China gradually began cultural and political engagements with the West (Ye, 2003; Zhu & Webber, 2016), which had an indirect impact on the field of literature. Keywords such as “Britain” and “Lessing” in Topic 8, and “America,” “film,” “history,” “modern,” and “narrative,” in Topics 5 and 9 reflect how Chinese literary works of this period began to more openly explore relationships with the world. Additionally, “translation” (Topic 6) became an important task in the literary community, highlighting the increasing openness and exchange with global cultures.

This phenomenon became more pronounced in 1978 when China officially launched its policy of reform and opening up (Lu et al., 2019). Following the formal establishment of diplomatic relations between China and America in 1979, political, economic, and cultural exchanges between the two countries significantly increased. America became one of China’s main trade partners and sources of foreign investment after the reform and opening up (Morrison, 2019). This close economic relationship facilitated broader cultural exchanges and mutual influence (Zhang & Musa, 2023). Prominent topics during this period included Topic 4 (western literary theories) and Topics 5, 9, and 10, which centered on “America,” indicating that literary works began to extensively explore relationships with America and the wider world.

At the end of the 20th century, the Chinese literary community actively absorbed and adapted Western literary theories and aesthetic concepts, such as modernism, postmodernism, feminism, and cultural studies (Ren, 2004). During this period, there was a significant increase in the exploration of female roles and gender issues (Ho et al., 2018). This shift reflects a growing focus on the contributions of female writers, the representation of gender in their works, and societal attention to gender equality and changes in women’s status (see topics 5 and 10). Two writers most relevant to this topic are “Hemingway” and “Eileen Chang.” Their novels have been widely analyzed under the theme of female characters and gender issues (Kadhim, 2018; R. Yuan, Vengadasamy, & Zheng, 2024; R. Yuan, Vengadasamy, et al., 2024). During this period, Chinese literary interest in America surpassed that in Britain.

After entering the 21st century, the interactions between China and the world have increased (Brooks & Wohlforth, 2015), marked by China’s accession to the World Trade Organization (WTO) in 2001. Chinese literature increasingly focuses on exploring the integration of global and local elements. Topic 6 (translation, translator) merged as the most studied topic. As China’s interactions with the world increased, the mutual communication between Chinese and global culture (S. Liu et al., 2023). As a consequence, translation played a critical role in this cultural exchange (Topic 6). The rise of digital tools (“corpus”) has made the collection and analysis of large text corpora much easier, thus advancing research in this area.

This period also saw the movement of the third wave of feminism, which emerged in the 1990s and continues today (Wu & Dong, 2019). Themes such as women’s literature and ethnic literature are even more widely explored, with works giving more attention to the voices of marginalized groups and issues of social justice, as seen in topics like Topic 9 (black, Morrison) and Topic

Table 2. Top Frequent Keywords.

No.	Frequent	Keywords
1	2,138	Culture
2	1,558	Translation
3	1,209	Him
4	1,071	History
5	830	Art
6	766	Tradition
7	614	America
8	594	Female
9	590	World
10	495	Lu Xun
11	492	Western
12	483	Revolution
13	366	Her
14	360	Political
15	349	Education
16	307	Film

10 (female, gender, body). It emphasizes diversity and intersectionality, encouraging the exploration of how race, class, and other identities intersect with gender. Notably, many of these keywords are related to America. As a focal point of global culture, politics, and economics, America's literature and culture are crucial areas of study.

It is also worth noting that with the development of the production and consumption patterns of Chinese literature have shifted, "film" adaptations of literary works have rapidly emerged. The global popularity of science fiction has increased, influenced by technological advancements and concerns about climate change, making science fiction a significant literary genre recognized internationally. For example, Liu Cixin's *The Three-Body Problem* has achieved great success worldwide, marking the international influence of Chinese science fiction literature.

Keywords Analysis

We identify the most frequently occurring keywords to discern major themes and areas of interest in literary study as a complement to the result of topic modeling, the result is presented in Table 2. The result highlights diverse topics. "Culture" emerged as the most prominent keyword, reflecting how early 20th-century China was undergoing a cultural clash between new Western enlightenment ideas and traditional Chinese feudal culture. In the 1920s, China experienced a movement that rethought traditional Chinese thought and advocated for a new form of Chinese culture based on modern ideals like electoral politics and the scientific method. The prominence of "tradition" also indicates the transformation of literary studies in Chinese academia. Closely following

this, "Translation" was identified as the second most common keyword. Translation is crucial for mutual understanding between Eastern and Western cultures, especially since China's participation in international affairs after the 21st century. Numerous Western novels were introduced to China, while over 150 contemporary Chinese writers' works have been translated into Western languages. For instance, translations of Nobel laureate Mo Yan's works have found international acclaim, introducing Western audiences to contemporary Chinese literature.

"Art" is also a prominent word. This not only indicates a close relationship between literature and art but also underscores how Chinese artists were significantly influenced by Western culture in the 20th century. For instance, themes of romance and love reflect the impact of Western Romanticism.

Notably, "him" and "her" both appear in the list and the topic modeling results. "Him" was mentioned 1,209 times and was assembled in Topic 8, while "her" appeared only 360 times and was in Topic 10. Before the introduction of modern Western culture, there was no distinction between genders in third-person pronouns in Chinese culture, as the pronoun for "她" (she) did not exist. With Western cultural influence, gender awareness became more prominent, leading to the coexistence of the pronouns "他" (he) and "她" (she), marking a significant shift in gender identity recognition (Chan, 2021).

According to the topic prominence (Figure 3), this influence and discrepancy emerged around 1949. Before 1949, the prominence of Topic 10 was zero. However, after the founding of the PRC in 1949, Topic 10 began to appear and, following the reform and opening up in 1978, its prominence surpassed that of Topic 5. Early in its establishment, China enacted laws to protect women's rights, encouraged female participation in social labor, and allocated a quota of government positions to women (Tiefenbrun, 2016), significantly elevating women's status.

As previously mentioned, the Chinese literary focus on America began in 1927. However, prior to 1978, the main areas of research were the Soviet Union, Britain, and Japan. Despite this, the United States was the only country to appear frequently in high-frequency word lists, indicating scholars' shifting attention to America. This largely contributes to America's rising international role after the two world wars, garnering considerable international attention on a cultural level as well (Brooks & Wohlforth, 2015). In the 1990s, Bercovitch (1994) pointed out in the preface of *The Cambridge History of American Literature* that, over the past 30 years, American literary criticism had evolved from a marginal field to a central aspect of humanistic studies. Over the course of more than half a century, American literature

has become one of the fastest-growing disciplines in the humanities (Bercovitch (1994). Its literary works have also garnered significant attention from the Chinese academic community, especially concerning topics like race issues, exemplified by Toni Morrison's (Topic 9) exploration of racial identity and history in her works.

The following highly frequent word is "world" in Chinese literary academia. After 1978, globalization has increasingly integrated Chinese literature into the world literary framework, prompting scholars to engage in cross-cultural research and comparative analysis. The translation of foreign literary works into Chinese and the dissemination of Chinese literature globally further contributed to this trend (see Topic 6). Additionally, the rise of modernity and global awareness led to a more prominent global perspective in literary discussions, highlighting the unique and universal aspects of Chinese literature. Increased international academic exchanges and collaborative projects also resulted in the frequent appearance of "world" in scholarly literature.

The presence of Lu Xun (1881–1936) was the only novelist mentioned by name. As mentioned before, he was one of the leading figures in the New Cultural Movement, and is considered one of the most important and influential Chinese writers of the 20th century (Xianfei, 2024; Liu, 2022), often regarded as the founder of modern Chinese literature (Pusey, 1998).

Conclusion

This study has provided a comprehensive and data-driven analysis of the evolving research themes in Chinese literary studies from 1927 to 2023, using Latent Dirichlet Allocation (LDA) topic modeling to uncover the shifting focus areas influenced by historical, political, and cultural contexts. The findings reveal a clear trajectory of thematic evolution, reflecting the interplay between domestic events and global influences over nearly a century. The study provides a comprehensive, data-driven review of nearly a century of Chinese literary research trends using computational methods, filling a gap in the literature by offering a macro-level perspective that has not been systematically explored in previous studies.

The analysis shows that early research themes (1927–1937) were heavily influenced by revolutionary ideas and the Soviet Union, with a focus on class struggle and political ideologies. The subsequent period (1937–1945) was dominated by wartime literature, highlighting themes of resistance and national identity. The post-war era (1945–1949) and the early years of the People's Republic of China (1949–1978) saw a rise in socialist realism and political themes, with restricted discussions on broader artistic and scientific issues. The reform and opening-up policy in 1978 marked a significant turning

point, leading to an influx of Western literary theories, gender studies, and a growing emphasis on globalization and multiculturalism. In the 21st century, Chinese literary research has increasingly integrated global and local elements, with a notable rise in digital literature, cross-cultural exchanges, and the use of digital tools for analysis.

The high-frequency keyword analysis corroborates these findings, demonstrating the cultural negotiation between traditional Chinese and Western influences. The emergence of gendered pronouns in Chinese scholarly discourse reflects growing gender awareness, supported by Topic 10's increasing prominence after 1949 and particularly after 1978. Meanwhile, America's presence as the only country appearing in the high-frequency word list confirms its outsized cultural influence compared to other nations.

The evolution of Chinese literary research themes over the past century is a testament to the profound impact of historical events, political movements, and cultural exchanges. The study offers a longitudinal analysis of the shifts in literary focus over multiple distinct historical periods, tracing the evolution of research themes from the revolutionary and socialist realism-dominated early 20th century to the present-day emphasis on digital humanities, globalization, and gender studies. Future trends in the Chinese literary field are likely to further integrate global and local literary elements, with increased cross-cultural research, comparative analysis, and the use of digital tools. The findings of this study not only provide valuable insights into the historical and ideological influences on Chinese literature but also offer guidance for future research directions, emphasizing the need to explore emerging themes such as film adaptations, digital literature, and the interplay between literature and other media forms.

Limitations and Future Implications

This study has several limitations that should be acknowledged. First, the analysis relies primarily on LDA for topic modeling, which, while effective for identifying latent themes in large corpora, may not capture all nuances present in the data. Comparing LDA results with other topic modeling techniques or machine learning models could provide additional insights but was beyond the scope of this study. Second, the dataset was sourced mainly from CNKI and CSSCI databases, which, although comprehensive, may not fully represent the entirety of Chinese literary research, especially independent publications or those outside mainstream academic databases. Third, the study focuses on thematic evolution over time, but qualitative analysis of specific topics could reveal deeper contextual understanding

that quantitative methods alone cannot provide. Finally, the study's reliance on keyword and topic prominence may overlook emerging themes that have not yet gained significant traction in the literature.

Future research can address these limitations and build on our findings in several ways. First, employing multiple topic modeling techniques (e.g., Non-negative Matrix Factorization, Structural Topic Model) could validate and enrich the thematic patterns identified in this study. Second, expanding the dataset to include a broader range of sources, such as independent journals, online publications, and international databases, would provide a more comprehensive view of Chinese literary research. Additionally, future work could focus on emerging topics that are currently underrepresented but show potential for growth, such as digital literature, environmental humanities, and cross-media adaptations.

ORCID iD

Ruikai Yuan  <https://orcid.org/0000-0002-9041-2900>

Ethical Considerations

This article does not involve any studies on human participants conducted by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

- Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A. (2023). Topic modeling algorithms and applications: A survey. *Information Systems*, 112, Article 102131.
- Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics*, 24(1), 1–7.
- Ball, Z., & Lewis, K. (2020). Predicting design performance utilizing automated topic discovery. *Journal of Mechanical Design*, 142(12), Article 121703.
- Bansal, P., Gualandris, J., & Kim, N. (2020). Theorizing supply chains with qualitative big data and topic modeling. *Journal of Supply Chain Management*, 56(2), 7–18.
- Bercovitch, S. & Patell, C. R. (Eds.). (1994). *The Cambridge history of American literature: Volume 7, prose writing, 1940–1990* (Vol. 7). Cambridge University Press.
- Blei, D., & Lafferty, J. (2006). Correlated topic models. *Advances in Neural Information Processing Systems*, 18, 147.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Boon-Itt, S., & Skunkan, Y. (2020). Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4), Article e21978.
- Brooks, S. G., & Wohlforth, W. C. (2015). The rise and fall of the great powers in the twenty-first century: China's rise and the fate of America's global position. *International Security*, 40(3), 7–53.
- Callaway, E., Turner, J., Stone, H., & Halstrom, A. (2020). The push and pull of digital humanities: Topic modeling the “what is digital humanities?” genre. *Digital Humanities Quarterly*, 14(1).
- Chan, C. H. Y. (2021). OF 她 (FEMININE TA) AND 它/牠 (NEUTER TA). In N. Ramière & R. Varshney (Eds.), *Rhizomes: Connecting Languages, Cultures and Literatures* pp.178–194.
- Chauhan, U., & Shah, A. (2021). Topic modeling using latent Dirichlet allocation: A survey. *ACM Computing Surveys (CSUR)*, 54(7), 1–35.
- Cheek, T. (2016). *The intellectual in modern Chinese history*. Cambridge University Press.
- Chen, C. M., Ho, S. Y., & Chang, C. (2023). A hierarchical topic analysis tool to facilitate digital humanities research. *Aslib Journal of Information Management*, 75(1), 1–19.
- Chen, X., Xie, H., Tao, X., Xu, L., Wang, J., Dai, H. N., & Wang, F. L. (2024). A topic modeling-based bibliometric exploration of automatic summarization research. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(5), Article e1540.
- Chu, K. E., Keikhosrokiani, P., & Asl, M. P. (2022). A topic modeling and sentiment analysis model for detection and visualization of themes in literary texts. *Pertanika Journal of Science & Technology*, 30(4), 2535–2561.
- Churchill, R., & Singh, L. (2022). The evolution of topic modeling. *ACM Computing Surveys*, 54(10s), 1–35.
- Da, N. Z. (2019). The computational case against computational literary studies. *Critical Inquiry*, 45(3), 601–639.
- Debortoli, S., Müller, O., Junglas, I., & Vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems (CAIS)*, 39(1), 7.
- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). Digital inequality: From unequal access to differentiated use. In D. B. Grusky & S. Szelenyi (Eds.), *Social inequality* (pp. 355–400). Westview Press.
- Ding, J. Z. (2018). The spread and impact of Cartesian philosophy in China: Historical and comparative perspectives. *Asian Philosophy*, 28(2), 117–134.
- Durán, I. (2020). What is the transnational turn in American literary studies? A critical overview. *Atlantis*, 42(2), 138–159.

- Esherick, J. W. (Ed.). (2000). *Remaking the Chinese city: Modernity and national identity, 1900-1950*. University of Hawaii Press.
- Furth, C. (1983). *Intellectual change: From the reform movement to the May Fourth movement, 1895-1920*. na.
- George, L., & Sumathy, P. (2023). An integrated clustering and BERT framework for improved topic modeling. *International Journal of Information Technology*, 15(4), 2187-2195.
- Goldstone, A., & Underwood, T. (2014). The quiet transformations of literary studies: What thirteen thousand scholars could tell us. *New Literary History*, 45(3), 359-384.
- Guo, Z. (2022). *Translation of treaties signed between China and Great Britain during 1842-1911: Translation, ideological reconstruction, and discursive power* [Doctoral dissertation, University of Leeds].
- Gurcan, F., & Cagiltay, N. E. (2019). Big data software engineering: Analysis of knowledge domains and skill sets using LDA-based topic modeling. *IEEE Access*, 7, 82541-82552.
- Guy, J. M., Conklin, K., & Sanchez-Davies, J. (2018). Literary stylistics, authorial intention and the scientific study of literature: A critical overview. *Language and Literature*, 27(3), 196-217.
- Hannigan, T. R., Haans, R. F., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals*, 13(2), 586-632.
- Hasan, M., Rahman, A., Karim, M. R., Khan, M. S. I., & Islam, M. J. (2021). Normalized approach to find optimal number of topics in Latent Dirichlet Allocation (LDA). In M. S. Kaiser, A. Bandyopadhyay, M. Mahmud, & K. Ray (Eds.), *Proceedings of international conference on trends in computational and cognitive engineering: Proceedings of TCCE 2020* (pp. 341-354). Springer.
- Ho, P. S. Y., Jackson, S., Cao, S., & Kwok, C. (2018). Sex with Chinese characteristics: Sexuality research in/on 21st-century China. *The Journal of Sex Research*, 55(4-5), 486-521.
- Hu, S. S., & Wang, H. L. (1997). *The Chinese civil war: Political struggles, 1945-1949*. China Youth Press. (Chinese)
- Huang, P. C., & Bernhardt, K. (2024). Fifty years of Modern China: An International Journal of History and Social Science. *Modern China*, 50(4), 399-412.
- Hutchings, G. (2001). *Modern China: A guide to a century of change*. Harvard University Press.
- Isoaho, K., Gritsenko, D., & Mäkelä, E. (2021). Topic modeling and text analysis for qualitative policy research. *Policy Studies Journal*, 49(1), 300-324.
- Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics*, 41(6), 750-769.
- Jockers, M. L., & Thalken, R. (2020). *Text analysis with R*. Springer International Publishing.
- Jowett, A. J. (1989). Patterns of literacy in the People's Republic of China. *GeoJournal*, 18, 417-414.
- Kadhim, R. J. (2018). Negative treatment of women in selected short stories by Katherine Mansfield and Ernest Hemingway. *African Educational Research Journal*, 6(4), 228-235.
- Klarer, M. (2013). *An introduction to literary studies*. Routledge.
- Koo, J. W. (2022). Global perceptions of South Korea's COVID-19 policy responses: Topic modeling with tweets. *Journal of Human Rights*, 21(3), 334-353.
- Le Moli, G. (2021). 'Parity with all nations': The 'coolie' trade and the quest for recognition by China and Japan. *Leiden Journal of International Law*, 34(4), 879-897.
- Lee, J. W., & Han, D. H. (2023). Data analysis of psychological approaches to soccer research: Using LDA topic modeling. *Behavioral Sciences*, 13(10), 787.
- Lee, L. O. F. (Ed.). (2023). *Lu Xun and his legacy*. University of California Press.
- Li, X. (1991). *A comprehensive history of the Chinese new democratic revolution: The rise of the nationalist revolution, 1923-1926*. Shanghai People's Publishing House. (Chinese)
- Li, X., & Lei, L. (2021). A bibliometric analysis of topic modeling studies (2000-2017). *Journal of Information Science*, 47(2), 161-175.
- Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, 5, 1-22.
- Liu, P. (2022). Lu Xun's literary revolution in Chinese Marxism. In C. Lye & C. Nealon (Eds.), *In After Marx: Literature, theory, and value in the twenty-first century* (p. 161). Cambridge University Press.
- Liu, S., Gallois, C., & Volcic, Z. (2023). *Introducing intercultural communication: Global cultures and contexts*. Sage.
- Liu, S., Zhang, R. Y., & Kishimoto, T. (2021). Analysis and prospect of clinical psychology based on topic models: hot research topics and scientific trends in the latest decades. *Psychology, Health & Medicine*, 26(4), 395-407.
- Liu, X. (2004). *Frontier passages: Ethnopolitics and the rise of Chinese communism, 1921-1945*. Stanford University Press.
- Lu, Y., Zhang, Y., Cao, X., Wang, C., Wang, Y., Zhang, M., Ferrier, R. C., Jenkins, A., Yuan, J., Bailey, M. J., Chen, D., Tian, H., Li, H., von Weizsäcker, E. U., & Zhang, Z. (2019). Forty years of reform and opening up: China's progress toward a sustainable path. *Science Advances*, 5(8), Article eaau9413.
- Ma, L. (2014). The Forgotten Ally: China's Second World War (1937-1945). W. Shiwei & L. Anfang (Eds.), *Frontiers of Foreign Social Sciences* (pp. 549-550). Shanghai People's Publishing House.
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2021). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In W. van Atteveldt & T.-Q. Peng (Eds.), *Computational methods for communication science* (pp. 13-38). Routledge.
- Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Poetics*, 41(6), 545-569.
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Morrison, W. M. (2019). China's economic rise: History, trends, challenges, and implications for the United States. *Current Politics and Economics of Northern and Western Asia*, 28(2/3), 189-242.
- Munthe-Kaas, H. M., Glenton, C., Booth, A., Noyes, J., & Lewin, S. (2019). Systematic mapping of existing tools to

- appraise methodological strengths and limitations of qualitative research: first stage in the development of the CAMELOT tool. *BMC Medical Research Methodology*, 19, 1–13.
- Negara, E. S., Triadi, D., & Andryani, R. (2019, October). *Topic modelling twitter data with latent dirichlet allocation method* [Conference session]. International Conference on Electrical Engineering and Computer Science (ICECOS), Batam, Indonesia (pp. 386–390). IEEE.
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010, June). *Automatic evaluation of topic coherence* [Conference session]. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Omnipress Inc. (pp. 100–108).
- Newman, D. J., & Block, S. (2006). Probabilistic topic decomposition of an eighteenth-century American newspaper. *Journal of the American Society for Information Science and Technology*, 57(6), 753–767.
- Park, G., Cho, M., & Lee, J. (2024). Leveraging machine learning for automatic topic discovery and forecasting of process mining research: A literature review. *Expert Systems with Applications*, 239, Article 122435.
- Piper, A. (2015). Novel devotions: Conversional reading, computational modeling, and the modern novel. *New Literary History*, 46(1), 63–98.
- Pusey, J. R. (1998). *Lu Xun and evolution*. SUNY Press.
- Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25–29.
- Queirós, A., Faria, D., & Almeida, F. (2017). Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*, 3(9), 369–387.
- Reisenbichler, M., & Reutterer, T. (2019). Topic modeling in marketing: Recent advances and research opportunities. *Journal of Business Economics*, 89(3), 327–356.
- Ren, X. Y. (2004). An overview of the study of Western literary theory in China in the 20th century. *Journal of Linguistics: Higher Education Edition*, 6, 3. (Chinese)
- Röder, M., Both, A., & Hinneburg, A. (2015, February). *Exploring the space of topic coherence measures* [Conference session]. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Shanghai, China (pp. 399–408). Association for Computing Machinery.
- Sbalchiero, S., & Eder, M. (2020). Topic modeling, long texts and the best number of topics. Some Problems and solutions. *Quality & Quantity*, 54, 1095–1108.
- Shahbazi, Z., & Byun, Y. C. (2020). Analysis of domain-independent unsupervised text segmentation using LDA topic modeling over social media contents. *International Journal of Advanced Science and Technology*, 29(6), 5993–6014.
- Sun, Y. (2023). Visual analysis of domestic English literature research. *Jingu Wenchuang*, 31, 55–58. <https://doi.org/10.20024/j.cnki.CN42-1911/I.2023.31.017>
- Suominen, A., & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology*, 67(10), 2464–2476.
- Tiefenbrun, S. (2016). China's employment laws and their impact on women working in china. *UC Davis Journal of International Law and Policy*, 23, 253.
- Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. University of Chicago Press.
- Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in English-language fiction. *Journal of Cultural Analytics*, 3(2), 1–25.
- Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, 94, Article 101582.
- Vulić, I., De Smet, W., Tang, J., & Moens, M. F. (2015). Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management*, 51(1), 111–147.
- Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009, June). *Evaluation methods for topic models* [Conference session]. *Proceedings of the 26th Annual International Conference on Machine Learning*, Quebec, Canada (pp. 1105–1112). Association for Computing Machinery.
- Wang, D. (2022). Research on the hotspots and trends of foreign literature studies in China from 2016 to 2020: Based on China's five major journals of foreign literature studies. *Journal of Kunming University*, 2, 120–126.
- Wang, W. S. (2007). The choice and loss of modernity—An interpretation of the realist literary trend during the “seventeen years.” *Journal of Nanjing Normal University: Social Science Edition*, 1, 5. (Chinese)
- Wei, J. H., & Li, J. (2022). A bibliometric analysis of Chinese literature going global (2011–2020) based on CiteSpace. *Shandong Foreign Language Teaching Journal*, 43(1), 95–106.
- Wen, Y. H. (2023). The situation and fate of the youth in wartime—Centered on the suicide of female literary youth Xue Ruijuan. *Modern Chinese Culture and Literature*, 4, 143–157. (Chinese)
- Wu, A. X., & Dong, Y. (2019). What is made-in-China feminism(s)? Gender discontent and class friction in post-socialist China. *Critical Asian Studies*, 51(4), 471–492.
- Xianfei, Z. (2024). Research on Lu Xun's concept of dialogue in the revolutionary literature movement. *Theoretical Studies in Literature and Art*, 44(4), 29–39.
- Yaacoba, A., & Newberryb, T. S. (2019). The scope of comparative literary studies: Review of schools of study. *International Journal of Innovation, Creativity and Change*, 7(3), 1–19.
- Yang, R. (2022). *The Chinese idea of a university: Phoenix reborn*. Hong Kong University Press.
- Ye, S. X. (2003). The cross-century evolution of the Western cultural root-seeking trend—An analysis of the “New Age Movement.” *Journal of Literary and Historical Studies*, 1 (01), 110–121. (Chinese)
- Yeh, W. H. (2000). *The alienated academy: Culture and politics in Republican China, 1919-1937* (Vol. 148). Harvard University Asia Center.
- Yuan, H. Z. (2018). A decade review of “trauma narrative” research—Based on papers published in CSSCI journals. *Journal of North China University of Water Resources and Electric Power (Social Science Edition)*, 3, 126–129. <https://doi.org/10.13790/j.cnki.issn1008-4444.2018.03.031> (Chinese)

- Yuan, R., Vengadasamy, R., Perry, M. S., & Hashim, R. S. (2024). Exploring feminist consciousness through music in Eileen Chang's postcolonial narratives. *GEMA Online Journal of Language Studies*, 24(1), 173–183.
- Yuan, R., Vengadasamy, R., & Zheng, Y. (2024). A bibliometric analysis of study on Eileen Chang using CiteSpace. *SAGE Open*, 14(2), Article 21582440241254892.
- Zhang, S. (2023). *The long shadow: Literary and cinematic representation and re-imagination of Chinese female traumas in the second Sino-Japanese War* [Doctoral dissertation, Purdue University].
- Zhao, X. (2021). *Displacement and anachronism: Art education and exhibitions in meta-institutionalizing China, 1912-1937*. University of California, San Diego.
- Zheng, Z. F. (2008). *Rebuilding the social center: A study on the group of free intellectuals after the war (1945–1949)* [Doctoral dissertation, East China Normal University]. (Chinese)
- Zhu, Y., & Webber, M. (2016). The impact of Western Economics on China's reforms from the late 1970s to the present: an overview. In M. Warner (Ed.), *The diffusion of Western economic ideas in East Asia* (pp. 364–375). Routledge.