A Corpus-based Study of Western Images in Senior High School English Textbooks

You Ma, Hui Shi*

School of Foreign Languages, Qingdao University, Qingdao 266071, China

*Corresponding email: hui_0213@yeah.net

Abstract

Against the backdrop of globalization and intercultural education, this study adopts a corpus linguistics approach to investigate the discursive construction of Western images in six sets of mainstream senior high school English textbooks in China. A specialized text corpus is built from the selected textbooks, and the Ant Conc software is employed to systematically analyze three core dimensions: high-frequency words, topic distribution, and collocation patterns. The findings reveal that the Western images constructed in these textbooks exhibit a hierarchical structure, with national spaces as the foundation, identity labels as the core, and cultural symbols as supplementary elements. The overall semantic tendency is dominated by objective, neutral descriptive knowledge, emphasizing cultural cognition while down making evaluative judgments on values. Furthermore, while reshaping Western images, the textbooks strike a balance between global vision and cultural subjectivity through comparative discussions of Chinese and Western topics. Through large-scale corpus analysis, this study provides quantitative empirical evidence for the cultural representation in English textbooks, and offers theoretical implications for textbook compilation and intercultural teaching practices.

Keywords

Corpus linguistics, Chinese senior high school, English textbooks, Western image

Introduction

Against the backdrop of advancing globalization and deepening intercultural education, English textbooks function not only as a carrier of linguistic knowledge but also as a critical discursive representation that shapes students' global perceptions and conveys cultural values. As senior high school marks a pivotal stage for the formation of students' worldviews, life outlooks, and core values, the Western images constructed in textbooks serve as a crucial lens through which learners observe the global landscape and understand heterogeneous cultures, directly influencing the development of their global vision and intercultural mindset. Such images are not transparent reflections of objective reality; rather, they are cultural symbols filtered, processed, and discursively constructed through specific ideologies, which profoundly shape learners' positioning of the relationship between the "self" and the "other" [1].

However, existing scholarship has predominantly focused on the representation of China in Western discourse, while research on the Western images actively constructed in China's basic education textbooks remains relatively underdeveloped [2]. Such studies are often limited to subjective and impressionistic descriptions and lack rigorous investigations based on large-scale textual evidence [3]. Meanwhile, relevant domestic and international studies have paid greater attention to college-level textbooks, with insufficient exploration of cultural representation and discursive construction in senior high school textbooks.

In view of this gap, the present study selects six sets of mainstream senior high school English textbooks in China as research materials and builds a dedicated textbook corpus. Leveraging the objectivity, quantifiability, and replicability of corpus linguistics, this study systematically reveals the discursive construction mechanisms of Western images in textbooks through in-depth analyses of high-frequency vocabulary distribution, core topic representation, and semantic collocation in concordance lines.

Specifically, this study addresses three core research questions:

(1) What distributional patterns characterize vocabulary

and topics related to Western images in the textbooks?

(2) What differences and emphases emerge in the topic representations of Chinese and Western images?

(3) How do specific collocational patterns construct deep-seated semantic representations of the West in the textbooks?

Using empirical data, this study seeks to uncover the priorities and tendencies of cultural representation in textbooks, so as to provide a scientific basis for the selection of cultural content in Chinese senior high school English textbooks. It also aims to offer data support and theoretical implications for balancing "global vision" and "cultural subjectivity" in textbook development, thereby optimizing the educational function of textbooks and advancing intercultural teaching practices [4,5].

Literature review

Research on cultural images in textbooks

An image is regarded as "a set of beliefs, ideas, and impressions that people hold toward an object", which is essentially a form of social cognition characterized by subjectivity and malleability. At the discursive level, an image is not a mere physical entity but a dynamic construction in the symbolic dimension through diverse linguistic resources [6,7]. To date, corpus-based research on discursive images has yielded fruitful outcomes. Researchers have extensively analyzed the discursive representation of cultural images in various contexts including media, literature, and education - from the perspectives of themes, keywords, metaphors, and collocations. Nevertheless, existing studies display a clear asymmetry: Scholarly attention has largely centered on national images from the perspective of the "other", while research on Western images actively constructed in the Chinese context remains scarce. In particular, systematic investigations using an empirical corpus-driven approach are still in the initial stage.

In the field of foreign language textbook research, the construction of cultural images has developed two parallel strands. The first strand focuses on the "self-image", centered on historical narratives and model figures, aiming to construct national identity and cultural belonging. The second strand investigates the "other-image", targeting target-language countries, with a focus on the representation patterns of Western culture in textbooks [8,9]. Traditional studies mostly adopt content analysis, manually counting the national distribution and topic frequency of cultural elements to identify potential Anglo-American centrism or cultural stereotypes in textbooks [10,11]. Although such studies are insightful, they are mostly confined to frequency descriptions of surface-level elements and fail to explore the deep correlations between linguistic features (e.g., semantic collocation, concordance context) and image evaluation. Therefore, this study introduces corpus linguistics to bridge macro-scale analysis and micro-level discursive details, and to deeply examine how textbooks systematically construct the discursive meaning of Western images through the coordinated deployment of language [12].

The application of corpus linguistics in discourse and image research

The rise of corpus linguistics has brought methodological innovation to discourse research, including image studies. By collecting and analyzing large-scale authentic texts, it can uncover language use patterns that transcend researchers' intuition, such as high-frequency words, keywords, collocation networks, and semantic prosody. In the field of cultural image research, corpus methods have been successfully applied to media discourse, textbook discourse, and political discourse analysis, forming a mature research paradigm [13].

For instance, Liu et al. used a corpus approach to reveal the dominance of Anglo-American culture in Chinese college English textbooks. Some researchers combined critical discourse analysis and corpus techniques to systematically investigate ideological construction in English textbooks. In cultural representation research, Keles and Yazan each took Chinese senior high school English textbooks as materials and conducted quantitative empirical analyses of cultural presentation and image construction using corpus tools [14]. Collectively, these studies demonstrate that corpus methods can objectively and systematically uncover how cultural and national images are discursively constructed through concrete linguistic practices, with high scientific validity and replicability.

The application of corpus methods in textbook research

The application of corpus methods to textbook research serves as an effective bridge connecting the analysis of

linguistic form and cultural content. It enables the processing of full texts from complete textbook series, avoids biases caused by selective analysis, and systematically reveals patterns of cultural representation from macro-level vocabulary distribution to micro-level collocational contexts. Liu et al. adopted corpus linguistics to conduct a large-scale quantitative analysis of Anglo-American cultural representation in Chinese college English textbooks, providing a paradigmatic model for research on cultural images in textbooks. Liu et al. further integrated critical discourse analysis and corpus techniques to explore ideological construction in college English textbooks, offering important methodological references for the present study [15]. However, most of these studies focus on college English textbooks. As a critical stage of basic education and value formation, senior high school textbooks play a more fundamental and far-reaching role in cultural construction. To date, corpus-based studies that systematically and comprehensively examine Western images in complete sets of mainstream senior high school English textbooks remain limited. A systematic review of English textbook research in the Chinese context by Pan & Zhu also indicates that corpus-based empirical research on senior high school textbooks is still an underexplored area.

Against this background, this study applies a corpus-driven approach to cultural image research in senior high school English textbooks. Based on a comprehensive corpus analysis of multiple mainstream textbook versions, this study empirically addresses the research questions, with the goals of deepening theoretical understanding of discursive construction mechanisms in textbooks and providing a reliable foundation for textbook compilation and evaluation.

Research design

Corpus construction and preprocessing

This study selects six sets of current mainstream senior high school English textbooks in China as the source corpus, including those published by People's Education Press, Foreign Language Teaching and Research Press, Beijing Normal University Press, Shanghai International Studies University Press, Shanghai Education Press, and Yilin Press. The corpus covers compulsory, selective compulsory, and elective modules to ensure the comprehensiveness and representativeness of the research sample.

Corpus construction consists of three core stages. First, text digitization: Paper textbooks are converted into electronic texts using high-precision optical character recognition (OCR) technology. Second, corpus cleaning and denoising: Non-narrative content such as page numbers, exercises, pure grammatical examples, and figure captions are manually removed, retaining only core narrative texts including passages, reading articles, and cultural essays to ensure discursive integrity. Third, formatting: Cleaned texts are uniformly converted into plain text files (.txt) encoded in UTF-8, finalizing the senior high school English textbook corpus.

Analytical tools and research procedures

This study mainly employs the corpus analysis software AntConc 3.5.9, using its wordlist, collocation, and concordance functions to address the research questions hierarchically. First, at the level of word frequency and topic distribution, the Wordlist function is used to generate a full-corpus frequency list. High-frequency words related to the West are screened through manual careful reading. These words include country/region names, personal names, and culture-specific items. A topic classification system is then established based on these words, covering national space, personal identity, cultural customs, and social values. Second, at the level of comparative analysis, equivalent high-frequency words related to China are extracted and placed within the same topic framework. A quantitative comparison is conducted on the similarities and differences in the topic distribution of Chinese and Western cultures in textbooks, revealing the emphases in representing the "self" and the "other". Third, at the level of discursive construction and semantic tendency, core node words such as *America* and *Britain* are selected. A reasonable span (five words to the left and right) is set, and significant collocates are extracted based on Mutual Information (MI) scores. This study analyzes the evaluative connotations and semantic orientations of these keywords in specific contexts through an in-depth reading of their concordance lines and collocation networks. This analytical process aims to reveal how textbooks construct a multi-dimensional semantic image of the West via the coordinated deployment of linguistic resources.

Assurance of research reliability and validity

To ensure the objectivity of the research findings,

multiple quality control measures are adopted. In the corpus processing stage, a double cross-checking system is implemented to guarantee accuracy in text recognition and cleaning. In topic classification and semantic tendency judgment, independent coding and group discussion are combined; controversial items are deliberated through multiple rounds until consensus is reached. Meanwhile, triangulation of quantitative data (frequency, MI values) and qualitative textual evidence (concordance examples) is used to minimize researcher bias, ensuring the empirical reliability and interpretive validity of image analysis [16].

Results and discussion

Topics constructed by West-related high-frequency words in the textbooks

High-frequency words are words that appear relatively frequently in a text. For image construction, repeated occurrences of high-frequency words continuously project certain images into the audience's mind, thereby influencing their cognition and perception. Through the reception and psychological integration of discursive information, the audience forms impressions and understandings of specific objects, completing image construction at the discursive level.

In corpus research, frequency information serves as the starting point of analysis, and high-frequency content words reflect the core focus of the text. Examining the occurrence frequency of specific words and phrases is one of the most fundamental and core approaches to revealing discursive stance, cultural representation, and attitudinal orientation using corpus methods. Researchers' selection and preference for words directly reflect the emphasized content and focus of discursive constructors.

Therefore, a systematic analysis of recurrent linguistic expressions is crucial for uncovering discourse and ideology. In this study, a word frequency list was generated from the content of the senior high school English textbooks in the corpus, and West-related words were selected from the high-frequency words through careful manual reading, based on which a word cloud was generated (Figure 1).

Figure 1. Word cloud of West-related high-frequency words in the senior high school English textbook corpus.

To further identify the precise frequencies and the exact number of texts in which West-related high-frequency words occur in the senior high school English textbook corpus, a systematic word frequency list was used to generate Table 1 (presenting only the top 20 key high-frequency words).

Table 1. Frequencies and text distributions of West-related high-frequency words in the senior high school English textbook corpus.

Word	Frequency	Text count	Word	Frequency	Text count
British	98	36	Britain	31	17
American	93	60	France	31	22
America	78	47	German	17	12
West	65	38	Germany	17	14
England	48	35	Americans	17	11
Europe	46	34	European	17	15
French	46	22	Freedom	13	12
Australia	41	17	Canadian	13	9
Western	40	27	Australian	11	5
Canada	31	10	Europeans	10	9

Based on the word frequency statistics, the frequently occurring Western elements in Chinese senior high school English textbooks can be categorized into three main types. The distribution characteristics of these elements reflect the linguistic teaching objectives and cultural educational functions embodied in textbook compilation.

(1) Lexemes denoting Western countries and regions form the most prominent high-frequency cluster in the textbooks, including America, England, Britain, Europe, Australia, France, Germany and Canada. The high frequency and wide textual coverage of such lexemes indicate that textbooks tend to organize linguistic materials by taking specific countries as the basic unit of cultural representation. From a linguistic teaching perspective, anchoring language learning to specific countries provides a clear real-world context for students. This approach allows learners to gain geographical, historical, and social background knowledge while acquiring vocabulary and discourses, thus improving the authenticity and situationality of language input. Furthermore, as an international lingua franca, English is predominantly used in European and American countries; thus, textbook compilation naturally features a cultural source structure centered on English-speaking countries and European societies. To a certain extent, this country-centered representation constructs a visualized cognitive map of the Western world, helping learners establish a foundational framework for intercultural cognition.

(2) Lexemes marking identities and ethnic groups (e.g., American, British, European, Australians, Canadians) also occupy a prominent position in the word frequency list. This finding reveals that the Western images portrayed in the textbooks are materialized not only through geographical space but also through the images of "people". Such lexemes typically appear in contexts of character stories, campus life, social interaction or intercultural communication, rendering culture no longer an abstract backdrop but a set of lifestyles and behavioral patterns embodied by specific social members. In terms of discursive function, identity labels contribute to constructing communication subjects and help students develop a clear awareness of interlocutors during reading and dialogue. This practice aligns with the core principle of communicative language teaching, which places a strong emphasis on authentic interactive contexts. Meanwhile, this person-centered mode of cultural representation reduces the difficulty of cultural comprehension and reinforces students' cognition of the behavioral norms and social customs of different cultural groups through narratives of daily life, thus integrating language learning with the cultivation of intercultural awareness.

(3) Lexemes signifying cultural and value symbols (e.g., freedom, church), though relatively low in overall frequency, carry strong symbolic significance in cultural representation. Such lexemes usually appear in texts introducing history and culture or social systems, serving to present the religious traditions, social concepts and value ideals of Western societies. However, the quantity of these lexemes is significantly lower than that of country and identity lexemes. This gap indicates that senior high school English textbooks employ a generally prudent and neutral strategy in cultural representation,

prioritizing informative cultural introduction while avoiding direct reinforcement or systematic interpretation of values. This distribution characteristic reflects the educational orientation of language curricula in textbook compilation: emphasizing intercultural understanding and information cognition rather than the transmission of value positions, thereby striking a balance between cultural input and educational guidance.

Comparative analysis of high-frequency words for Western and Chinese images in the textbooks

To gain a clearer understanding of the attention paid to Chinese and Western cultures in senior high school English textbooks, this study, following an analysis of West-related high-frequency words and a sorting of common topics related to the West in the textbooks, further calculated the frequencies of references to China and Chinese cities in the textbook corpus. The aim of this comparison was to explore whether senior high school English textbooks exhibit a tendency to prioritize Western elements. Using the wildcard + (representing one or more characters), the corpus was searched for the root Chin+, revealing the occurrence frequencies of China (410 times), Chinese (412 times) and Chinas. The results showed that these frequencies are far higher than those of France* (31 times), England* (48 times) and America* (78 times), where the wildcard * denotes zero or more characters. Further retrieval and analysis also found that *Beijing* and *Shanghai* appear 61 and 18 times respectively, with frequencies fairly close to those of Western cities such as Paris (16 times) and London (61 times). In other words, in the senior high school English textbooks examined in this study, references to China are far more frequent than those to Western European and American countries, while the mention frequencies of Chinese and Western cities are roughly comparable. Simply put, this discrepancy reflects the balance that textbook compilers strive to achieve between two goals: telling Chinese stories well and cultivating a global vision.

In addition to China and Chinese cities, the corpus was also searched for Chinese lexemes analogous to Western elements, including representative figures (Confucius), traditional festivals (Spring Festival) and iconic enterprises (Huawei). The retrieval results indicated that Confucius or Confucianism is mentioned 26 times in the textbooks, a frequency close to that of the Western given name Jack (43 times). In contrast, the Chinese traditional festival Spring Festival appears 33 times, while the Western festival Christmas occurs 80 times; iconic Chinese enterprises such as Huawei receive no mention at all. Topic selection reflects the focal points of image construction, and the comparison of Chinese and Western high-frequency words shows that senior high school English textbooks address Chinese traditional festivals and iconic enterprises less frequently than their Western counterparts. The authors argue that textbook compilation in the future may appropriately consider increasing the coverage of such content [17].

Collocation analysis of core high-frequency words for Western images in the textbooks

Image construction is multi-dimensional, and the definition of the "West" in this study is correspondingly broad. To explore the specific representation of Western images in senior high school English textbooks in greater depth, this study selected representative high-frequency words and analyzed the specific content of the topics they involve. Based on the preceding statistical analysis, there are significant differences in the mention frequencies of Western countries, cities and people in the textbooks, with the United States/Americans and Britain/British ranking the highest. This indicates that when depicting the West, the textbooks tend to take these two representative developed countries as the core entry points. Collocation is defined as the co-occurrence of two or more lexemes within a specific textual span (Sinclair, 1991: 170). Starting the analysis with high-frequency content words and examining their concordance lines and collocational patterns can effectively reveal meaningful semantic patterns in discourses. Therefore, this study selected high-frequency words associated with the United States and Britain respectively to conduct a more in-depth investigation. It systematically analyzed the American and British images constructed in senior high school English textbooks by examining their collocational structures and concordance line content.

(1) Collocational analysis of high-frequency words for the American image

Using AntConc 3.5.9, the wildcard symbol * (denoting zero or more characters) was first employed to retrieve the collocates of America* (including America, American and Americans) in the senior high school

English textbook corpus. This analysis aimed to explore the American image constructed in the textbooks by examining the high-frequency content words co-occurring with the target lexemes. Table 2 presents the frequencies, MI values and other relevant information of the high-frequency content words collocating with America* in the senior high school English textbook corpus.

Table 2. High-frequency content words collocating with America* in the senior high school English textbook corpus.

Collocate	Frequency	Left frequency	Right frequency	MI value
people	14	8	6	5.11
culture	8	0	8	7.51
African	8	8	0	10.29
South	7	7	0	8.40
North	7	7	0	8.67
Food	7	2	5	5.98
Famous	6	5	1	7.29
Example	6	4	2	6.51
Europe	6	5	1	8.76
English	6	1	5	7.05
Teenagers	4	0	4	7.75
Jazz	4	2	2	9.65

As shown in Table 2, among the high-frequency content words co-occurring with America*, lexemes related to geographical locations and ethnic identities account for a remarkably high proportion and exhibit prominent MI values, forming the foundational framework of the American image in senior high school English textbooks. America* primarily appears in expressions such as North America, South America and African American, and frequently cooccurs with region- or language-related lexemes including American English and Europe.

In terms of collocational positions, the right frequency of African, south and north are zero, with these lexemes exclusively forming left collocations in the structure of Noun + America as fixed designations for geographical divisions and ethnic identities. By contrast, culture and teenagers have a left frequency of zero and only appear as right collocations in the structure of America + Noun, manifesting as categorical expressions such as American culture and American teenagers. Further verification through concordance lines reveals that lexemes such as people, culture and jazz all frequently occur in objective identity and cultural labels like American people, American culture and American jazz, without any explicit affective evaluation, as illustrated in Examples (1), (2) and (3):

(1) American people are known for their diverse cultural backgrounds.

(2) Thanksgiving is an important tradition in American culture.

(3) American jazz originated in the late 19th and early 20th centuries.

Famous is the only lexeme with a positive semantic prosody in the table. A review of its concordance lines shows that its co-occurrence with America* mostly takes the structure of famous American + Noun (e.g., famous American writers). The word serves merely as an attributive modifier for the objective introduction of people or objects, rather than a positive commendation of the American image as a whole, as seen in Example (4):

(4) Hemingway is a famous American writer known for his novels.

The magnitude of MI values reflects the exclusivity of the co-occurrence between a given lexeme and America*. In Table 2, African (MI=10.29) with an MI value exceeding 10 and jazz (MI=9.65) with an MI value close to 10 emerge as the iconic collocates of the American image in senior high school English textbooks.

An analysis of the concordance lines of African indicates that its co-occurrence with America* almost exclusively refers to the core expression African American, which centers on the ethnic composition of the United States and acts as a key carrier for presenting America's diverse social structure in the textbooks, as shown in Examples

(5) and (6):

(5) African American musicians have made great contributions to jazz music.

(6) Martin Luther King Jr. was a key figure in the African American civil rights movement.

As a native cultural symbol of the United States, jazz cooccurs with America* predominantly in culture-themed contexts (e.g., American jazz music), serving as a core example for showcasing the characteristics of American popular culture, as illustrated in Example (7):

(7) Louis Armstrong is one of the most famous American jazz musicians.

In addition, the high MI values of Europe (8.76), north (8.67) and south (8.40) also reflect that the textbooks tend to situate the United States within the geographical framework of the Americas and the cultural context of Europe and America when introducing it, highlighting its geographical and cultural relevance, as seen in Examples (8), (9) and (10):

(8) North America includes countries like Canada, the United States, and Mexico.

(9) South America is home to the Amazon rainforest, the world's largest tropical rainforest.

(10) The relationship between Europe and America has a long and complex history.

On the basis of analyzing the concordance lines of high-frequency content collocates, a further review of all concordance lines containing America* reveals that the core content of references to the United States in senior high school English textbooks revolves around four dimensions: ethnic composition, geographical division, cultural symbols and social life. These references include not only the objective definition of ethnic groups and geography (e.g., African American, North America) and the presentation of cultural symbols (e.g., jazz, culture), but also descriptions of daily life and social groups (e.g.,

food, teenagers), as shown in Examples (11), (12) and (13):

(11) American food has been influenced by many different cuisines.

(12) American teenagers often spend a lot of time on social media.

(13) American English and British English have some differences in spelling and vocabulary.

Concordance line analysis shows that all expressions mentioning America* in senior high school English textbooks are neutral factual descriptions. Even positive lexemes such as famous are only used to describe the attributes of specific objects and do not extend to an overall evaluation of the national image of the United States.

In summary, through the high-frequency collocates of America*, senior high school English textbooks construct an American image grounded in geographical identity, centered on cultural diversity and oriented toward objective facts. This mode of representation focuses on knowledge transmission and avoids explicit value judgment, which aligns with the core objective of cultural cognition in English teaching at the senior high school level [18].

(2) Collocational analysis of high-frequency words for the British image

Besides America*, lexemes related to Britain, namely British and Britain, also appear frequently in the word list and word cloud. Accordingly, this study further analyzes collocates and concordance lines of these two lexemes to examine the British image constructed in senior high school English textbooks. Using the wildcard symbol | (denoting "or"), AntConc 3.5.9 was employed to retrieve collocates of British|Britain. Table 3 presents information of high-frequency content words with high MI values co-occurring with the target lexemes.

Table 3. High-frequency content words collocating with British|Britain in the senior high school English textbook corpus.

Collocate	Frequency	Left frequency	Right frequency	MI value
People	20	5	15	7.29
Museum	17	0	17	11.76
Comedy	12	3	9	12.48
Columbia	12	1	11	13.48
Great	10	8	2	8.02
British	8	4	4	9.87
Film	6	1	5	8.98

Collocate	Frequency	Left frequency	Right frequency	MI value
English	6	1	5	8.71
Kingdom	5	5	0	11.41
Culture	5	2	3	8.50
United	4	4	0	9.68
School	4	1	3	7.01

As illustrated in Table 3, among the high-frequency content words co-occurring with British/Britain, lexemes related to cultural scenarios (museum, comedy, film, culture) and those denoting identity/national designations (people, kingdom, united, British, English) account for a disproportionately high proportion and exhibit prominent MI values, forming the core framework of the British image in senior high school English textbooks.

In terms of collocational positions, museum has a left frequency of zero and exclusively functions as a right collocate in the structure of British/Britain + Noun (i.e., British Museum), serving as a fixed expression for a cultural scenario. By contrast, kingdom and united have a right frequency of zero and only appear as left collocates in the structure of Noun + British/Britain, manifesting as fixed designations in proper nouns such as the United Kingdom of Great Britain and Northern Ireland. Other lexemes including comedy, film and English predominantly act as right collocates, focusing on specific cultural or linguistic scenarios such as British comedy, British film and British English.

Further verification through concordance lines reveals that lexemes such as people, museum and comedy frequently occur in expressions referring to objective cultural labels or scenarios (e.g., British people, British Museum, British comedy), without any explicit affective evaluation, as shown in Examples (14), (15) and (16):

(14) British people are known for their love of tea and football.

(15) The British Museum houses artifacts from all over the world.

(16) British comedy often uses dry wit and self-deprecating humour.

Great is the sole lexeme with a positive semantic prosody in the table. An analysis of its concordance lines shows that its co-occurrence with British/Britain mostly takes the form of the proper noun Great Britain or the structure great British + Noun (e.g., great British writers). The word serves merely as an attributive modifier for objective description or in specific contextual usage, rather than a positive commendation of the British national image as a whole, as illustrated in Examples (17) and (18):

(17) Great Britain is an island that includes England, Scotland, and Wales.

(18) William Shakespeare is a great British playwright.

MI values reflect the exclusivity of co-occurrence between a given lexeme and British/Britain. In Table 3, Columbia (MI=13.48), comedy (MI=12.48), museum (MI=11.76) and kingdom (MI=11.41) - all with an MI value exceeding 10 - emerge as the iconic collocates of the British image in senior high school English textbooks. The co-occurrence of Columbia and British/Britain mostly refers to proper geographical nouns such as British Columbia, acting as a key carrier for presenting the geographical affiliations of the Commonwealth of Nations in the textbooks. As an important symbol of British popular culture, comedy cooccurs with British/Britain primarily in film, television and literary contexts (e.g., British comedy series), serving as a core example for showcasing the characteristics of British entertainment culture. Museum frequently appears in the expression British Museum, referring to Britain's rich historical and cultural heritage. Kingdom mainly occurs in the proper noun United Kingdom, acting as the core designation of Britain's national identity. These collocational patterns are exemplified in Examples (19), (20), (21) and (22):

(19) British Columbia is a province on Canada's Pacific coast.

(20) British comedy series like Mr. Bean are popular worldwide.

(21) The British Museum is one of the oldest public museums in the world.

(22) The United Kingdom is a constitutional monarchy in northwestern Europe.

In addition, the high MI values of British (9.87), united

(9.68) and film (8.98) also reflect that the textbooks tend to situate Britain within the framework of national identity and the global cultural industry when introducing it, highlighting its historical and cultural influence, as shown in Examples (23), (24) and (25):

(23) British and American English have many differences in vocabulary and pronunciation.

(24) The United Kingdom of Great Britain and Northern Ireland are a member of the Commonwealth.

(25) British film directors have made significant contributions to world cinema.

On the basis of analyzing the concordance lines of high-frequency content collocates, a further review of all concordance lines containing British/Britain reveals that the core content of references to Britain in senior high school English textbooks revolves around four dimensions: cultural scenarios, proper nouns, linguistic identity and social life. These references include not only the presentation of cultural scenarios (e.g., British Museum, British comedy) and the designation of proper nouns (e.g., United Kingdom, British Columbia), but also descriptions of education and culture (e.g., British school, British culture).

Concordance line analysis demonstrates that all expressions mentioning British/Britain in senior high school English textbooks are neutral factual descriptions. Even positive lexemes such as great are only used to describe the attributes of specific objects and do not extend to an overall evaluation of the British national image, as illustrated in Examples (26) and (27):

(26) British school systems have a long history of academic tradition.

(27) British culture is a rich tapestry of literature, music, and art.

In summary, through the high-frequency collocates of British/Britain, senior high school English textbooks construct a British image grounded in historical heritage, centered on cultural influence and oriented toward objective facts. This mode of representation focuses on cultural cognition and language learning, which aligns with the core objective of intercultural understanding in English teaching at the senior high school level [19].

Conclusion

Drawing on corpus linguistics approaches, this study systematically deconstructs the discursive construction characteristics of the "Western image" in six sets of mainstream senior high school English textbooks in China. The findings reveal that the representation of the West in these textbooks exhibits a hierarchical structure featuring national space as the foundation, identity labels as the core, and cultural symbols as supplementary elements. This pattern profoundly reflects that in the functional screening of cultural materials, Chinese senior high school English textbooks prioritize constructing a relatively neutral, knowledge-based portrait of the "Other", thereby guiding students to establish objective intercultural cognition while serving the goals of language teaching.

In terms of academic contributions, this study breaks through the limitations of traditional textbook analysis that relies on subjective qualitative interpretation, and realizes a paradigm shift from empirical observation to quantitative analysis through large-scale corpus-based empirical research, which significantly enhances the generalizability and verifiability of the research conclusions. Meanwhile, by operationalizing abstract cultural images into measurable linguistic units, the three-dimensional structural analytical framework proposed in this study not only provides an empirically grounded linguistic approach for textbook evaluation, but also offers data-driven support for balancing global vision and cultural subjectivity in future textbook compilation.

This study is still limited by the analytical depth of corpus linguistics. The current research focuses on explicit lexical frequency and static collocational distribution, with insufficient exploration of the implicit cultural positions embedded in the deep narratives of discourses. Additionally, the actual impact of the research conclusions on learner reception remains to be verified. Based on these limitations, future research can further integrate Critical Discourse Analysis (CDA) with close reading of concordance lines to explore the dynamic construction process of the Western image in complex narrative contexts. Meanwhile, diachronic comparative studies of textbooks across different eras can be conducted to examine the driving mechanisms of social changes on cultural representation in textbooks. Furthermore, integrating corpus analysis with empirical research on learner psychology will enable an in-depth investigation of the empirical effects of textbook input on the formation of students' cultural identity and values.

Through the integration of multi-dimensional methods, research on foreign language textbooks is expected to provide more targeted theoretical guidance and pedagogical implications for cultivating intercultural talents with firm cultural confidence and international competitiveness, on the basis of revealing the logic of cultural representation in textbooks.

Funding

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

References

[1]   Risager, K. (2021) Language textbooks: windows to the world. *Language, Culture and Curriculum*, 34(2), 119-132.

[2]   Liu, Y., Zhang, L. J., May, S. (2022) Dominance of Anglo-American cultural representations in university English textbooks in China: a corpus linguistics analysis. *Language, Culture and Curriculum*, 35(1), 83-101.

[3]   Pan, M. X., Zhu, Y. (2022) Researching English language textbooks: a systematic review in the Chinese context (1964-2021). *Asian-Pacific Journal of Second and Foreign Language Education*, 7(1), 30.

[4]   Liu, H., Fang, F. (2022) Towards a global Englishes-aware national English curriculum of China. *ELT Journal*, 76(1), 88-98.

[5]   Lu, J., Liu, Y., An, L., Zhang, Y. (2022) The cultural sustainability in English as foreign language textbooks: Investigating the cultural representations in English language textbooks in China for senior middle school students. *Frontiers in Psychology*, 13, 944381.

[6]   Elmiana, D. S. (2019) Pedagogical representation of visual images in EFL textbooks: a multimodal perspective. *Pedagogy, Culture & Society*, 27(4), 613-628.

[7]   Suwandani, F., Hidayati, S. (2025) Conceptual Representation of culture in EFL textbook images: a multimodal analysis. *Dinamika Ilmu*, 25(1), 99-122.

[8]   Gonja, C., Musaev, T. (2025) "We" in English language textbooks for Japanese high school students. *World Englishes*, 44(3), 360-377.

[9]   Lee, J. F., Li, X. (2020) Cultural representation in English language textbooks: a comparison of textbooks used in mainland China and Hong Kong. *Pedagogy, Culture & Society*, 28(4), 605-623.

[10] Gheisari, M., Akbari, O. (2024) A comparison of cultural representation and ideologies in the multimodal discourses of textbooks used in public and private Iranian contexts: a cross-textual study. *Social Semiotics*, 34(2), 289-309.

[11] Nguyen, T. T. M., Marlina, R., Cao, T. H. P. (2021) How well do ELT textbooks prepare students to use English in global contexts? An evaluation of the Vietnamese English textbooks from an English as an international language (EIL) perspective. *Asian Englishes*, 23(2), 184-200.

[12] Lavrenteva, E., Orland-Barak, L. (2023) Conceptual-analytical framework for exploring culture in EFL coursebooks: analysis of teaching materials from a multimodal perspective. *Social Sciences & Humanities Open*, 7(1), 100441.

[13] Jang, G., Tinker Sachs, G., Park, J. H. (2024) Conflicting understandings of multicultural society, global world, and English: multimodal content analysis of 5 Korean elementary EFL textbooks. *Critical Inquiry in Language Studies*, 21(2), 153-176.

[14] Keles, U., Yazan, B. (2023) Representation of cultures and communities in a global ELT textbook: a diachronic content analysis. *Language Teaching Research*, 27(5), 1325-1346.

[15] Liu, Y., May, S., Zhang, L. J. (2024) The ideological underpinnings in university English textbooks in China: a critical discourse analysis and corpus linguistics approach. *Journal of Multilingual and Multicultural Development*, 45(10), 4448-4467.

[16] Chen, F. W. (2025) An Examination of English language textbooks used in China's senior high schools, focusing on character, culture, linguistic norms, and communication strategies. *English Teaching & Learning*, 49(4), 783-807.

[17] Li, Z., Zeng, J., Nam, B. H. (2023) A comparative analysis of multimodal native cultural content in

English-language textbooks in China and Mongolia. *SAGE Open*, 13(2), 1-15.

[18] Wang, J., Hiew, W., Swanto, S. (2022) Problems and countermeasures on applying the transformative learning theory to China's college English teaching. *International Journal of Education, Psychology and Counselling*, 7(45), 376-385.

[19] Wu, L., Wu, L. (2025) Values Education in China: a critical discourse analysis of the People's Education Press's Seventh Grade English Textbook. *JOLLT Journal of Languages and Language Teaching*, 13(2), 632-647.