

A Study of Leshan's International Image Based on a Self-Built Web Corpus (2022-2024)

Xian,Xinyu

Sichuan University Jinjiang College, Meishan, Sichuan, 620860, China

Abstract: This paper constructs a self-built web corpus and employs a corpus-based critical discourse analysis approach to explore the discourse features and constructed international image of Leshan in overseas social media reports from 2022 to 2024. The study finds that Leshan, centered on the Leshan Giant Buddha, is portrayed as an international destination that blends natural wonders, cultural heritage, and convenient tourism experiences. Analysis of high-frequency words and collocations highlights the "Giant Buddha" attributes of the Leshan Giant Buddha, while its geographic context, tourism activities, and cultural value enhance its appeal.

Keywords: International image; Leshan; Corpus-based discourse analysis

DOI: 10.62639/sspjess12.20250203

1. Introduction

Leshan, recognized as an outstanding tourism city and a historical-cultural city in China, is endowed with abundant natural resources, a rich cultural heritage, and a prominent tourism industry. Home to the "Emei Mountain-Leshan Giant Buddha Scenic Area," a UNESCO World Cultural Heritage site, Leshan serves as a vital platform for telling China's story and promoting its global image. This paper conducts a discourse analysis of the linguistic features in overseas social media reports about Leshan using a self-built web corpus, aiming to explore the international image of Leshan constructed by these media and provide academic references for optimizing Leshan's international communication strategies.

2. Research Background

Corpus-based Critical Discourse Analysis (CBCDA) is an emerging research method that integrates the strengths of corpus linguistics and critical discourse analysis (CDA). It seeks to reveal power relations, ideologies, and social structures embedded in language use through the statistical analysis and contextual interpretation of large-scale corpora. Unlike traditional CDA, which relies heavily on subjective interpretation, CBCDA incorporates quantitative linguistic data while retaining a critical exploration of social significance, making it a key approach at the intersection of linguistics and sociology.

The core of CBCDA lies in combining the quantitative methods of corpus linguistics with the qualitative perspective of CDA. Corpus linguistics leverages computer tools (e.g., AntConc, WordSmith) to analyze texts, extracting patterns such as word frequency, collocations, and keywords to provide objective data support (Baker, 2006). CDA, on the other hand, focuses on how language reflects and constructs social power, identity, and inequality, emphasizing the sociocultural meanings behind texts (Fairclough, 1992). The integration of these approaches overcomes their respective limitations: corpus methods enhance the systematicity and verifiability of

(Manuscript NO.: JIESS-25-3-62002)

Funding

This study is supported by the 2024 project of the Leshan Science and Technology and Industrial Internationalization Service Collaborative Innovation Base, titled "A Study of Leshan's Urban International Image from the Perspective of Critical Discourse Analysis (2021-2023)" (Project No.: LSKJ24G06).

analysis, while CDA adds depth to the data by providing a socially critical lens, avoiding mere linguistic description (Partington et al., 2013).

CBCDA typically involves the following steps: first, constructing a corpus around a specific theme, such as news reports, social media posts, or political speeches; second, using corpus tools for quantitative analysis to identify significant linguistic features; and third, conducting qualitative interpretation within a social context to uncover the ideological implications behind linguistic patterns (Baker, 2006).

In recent years, with the rapid development of the internet, big data, and other new-generation information technologies, an increasing number of Chinese scholars have utilized large-scale corpora, treating “discourse as a core issue” (Xu Jiagin, 2019: 37) to study the construction of China’s international image. For instance, Qian Yufang (2010: 198) systematically outlined a methodology combining corpus linguistics with critical discourse analysis, using the British tabloid *The Sun*’s portrayal of certain celebrities as a case study to explore how newspapers embed political and social factors—such as national interests, business objectives, and audience targeting—into discourse. Shao Bin and Hui Zhiming (2014: 28) extracted English-language reports on the “Chinese Dream” from the LexisNexis news database, creating a dedicated “Chinese Dream” corpus to empirically investigate Western media’s interpretations, evaluations, and discourse construction using concordance lines and collocation networks. Hu Kaibao and Tian Xujun (2018: 79) analyzed China’s diplomatic image in English translations of its diplomatic discourse, focusing on high-frequency words, keywords, modal verbs, and collocations with “we.” Qian Yufang and Dong Yingying (2020: 28) utilized the NOW Corpus, developed by Brigham Young University, to examine the discursive representations and image construction of Yiwu in overseas media over the past decade.

In summary, Chinese scholars’ analyses of China’s image using mainstream corpora have primarily focused on macro-level studies of domestic or diplomatic discourse, employing methods such as word indexing, collocation analysis, semantic tendencies, and semantic prosody to examine discourse features. However, research on the international image of specific Chinese cities based on self-built corpora remains limited, with studies on Leshan’s international image using corpora being particularly scarce.

3. Research Design

(1) Research questions

To investigate the image of Leshan constructed in overseas social media, this study conducts a discourse analysis based on a self-built web corpus around the following three questions:

- 1) What are the high-frequency words related to the node word “Leshan” in overseas social media, and what topics do these words cover?
- 2) What are the high-frequency collocations of the node word “Leshan” in overseas social media?
- 3) What international image of Leshan is constructed by these high-frequency words and collocations?

(2) Corpus construction and research steps

The data for this study is sourced entirely from a self-built Leshan International Image Corpus. To construct this corpus, the study used “Leshan” as a keyword to retrieve all English-language social media web pages containing this term from 2022 to 2024, including but not limited to blogs, microblogs, and comments. Web crawler software was employed to extract the text from these pages, with each page saved as a text document in UTF-8 encoding. The final corpus contains 22,000 tokens.

Using the constructed corpus, this study first employed AntConc to retrieve high-frequency words, generating a high-frequency word list and related data to understand the distribution of different vocabulary items. Second,

it retrieved collocation information for the node word “Leshan” in the “collocates” section to identify its primary collocating words, using visualization techniques to analyze the overall impressions and attitudinal evaluations of Leshan by overseas social media. Finally, based on the data from the corpus analysis, the study summarizes Leshan’s international image and proposes suggestions for further enhancing its global image.

4. Results and Discussion

(1) High-Frequency word retrieval and analysis

The top ten high-frequency words mentioned in overseas social media reports about Leshan from 2022 to 2024 are shown in Table 1.

[Table 1: Top Ten High-Frequency Words in Overseas Social Media Reports on Leshan, 2022-2024]

Rank	Word	Freq	Range	NormFreq	NormRange	Semantic Category
1	Buddha	561	37	24677.781	0.974	Religion/Culture
2	Leshan	478	38	21026.701	1	Place Name
3	giant	298	34	13108.697	0.895	Adjective/Characteristic
4	statue	134	24	5894.515	0.632	Artifact/Attraction
5	Chengdu	101	27	4442.88	0.711	Place Name
6	boat	91	19	4002.991	0.5	Transportation/Tourism
7	city	77	25	3387.146	0.658	Geography/City
8	river	72	23	3167.202	0.605	Natural Landscape
9	Sichuan	72	27	3167.202	0.711	Place Name
10	China	69	25	3035.235	0.658	Country/Background

These words clearly outline the core themes of Leshan’s coverage: the Leshan Giant Buddha and its geographic and tourism context. The exceptionally high frequency and near-complete coverage of “Buddha” and “Leshan” (with NormRange values of 1.0 for Leshan and 0.974 for Buddha) indicate that the Leshan Giant Buddha is not only a defining symbol of the city but also the absolute focus of English-language social media attention, with a NormFreq of 24,677.781 far exceeding other words, further emphasizing its dominant role. Geographically related terms like “Chengdu,” “Sichuan,” and “China” appear frequently, suggesting that reports often situate Leshan within Sichuan and China’s tourism landscape, particularly linking it to Chengdu (101 mentions, NormRange 0.711), likely due to its role as a transportation hub or tourism gateway. Terms related to natural landscapes and tourism activities, such as “river” and “boat” (the latter with a NormFreq of 4,002.991), point to the confluence of three rivers in Leshan and the unique experience of viewing the Giant Buddha by boat.

Thematically, the corpus can be clustered into five categories:

- 1) Leshan Giant Buddha (“Buddha,” “giant,” “statue,” “carved,” “cliff,” “stone”), emphasizing its scale and craftsmanship.
- 2) Geography and Natural Landscapes (“Leshan,” “Sichuan,” “Chengdu,” “river,” “mountain,” “Lingyun,” “rivers”), highlighting its location and riverine features.
- 3) Tourism Experiences (“boat,” “tour,” “visit,” “scenic,” “park,” “bus,” “train,” “station”), focusing on transportation and activities.
- 4) Culture and Heritage (“temple,” “ancient,” “heritage,” “dynasty”), reflecting Buddhist and historical value.

5) Daily Life and Consumption ("food," "local," "yuan," "people"), mirroring the tourist lifestyle perspective.

From a sociocultural perspective, these high-frequency words reflect the image construction of Leshan in English-language social media. The Leshan Giant Buddha, as a cultural ambassador, attracts attention not only for its natural wonder ("giant," "cliff") but also for its religious culture ("temple") and heritage value ("heritage"), catering to international tourists' interest in Eastern culture. The use of "meters" and "feet" (42 mentions, Range 16) may reflect Western audiences' amazement at its scale, while the high frequency of "Chengdu" and "Sichuan" reinforces Leshan's position within Sichuan's tourism network. Practical tourism information ("boat," "bus," "train") and lifestyle elements ("food," "local") indicate that reports balance utility, meeting tourists' needs. Overall, from 2022 to 2024, English-language social media has shaped Leshan as an international destination that integrates natural landscapes, historical culture, and tourism experiences, with the prominent and multifaceted portrayal of the Leshan Giant Buddha underscoring its global recognition and appeal.

(2) Collocation retrieval and analysis based on "Leshan"

As shown in Table 2, the node word "Leshan" in overseas social media reports on Leshan from 2022 to 2024 includes 11 significant collocations, covering metrics such as frequency (Freq(Scaled) ranging from 4,488 to 160), left-right co-occurrence frequency (FreqLR from 319 to 13, FreqL from 51 to 3, FreqR from 268 to 1), distribution range (Range from 36 to 4), likelihood ratio (Likelihood from 462.029 to 17.05), and effect size (Effect from 2.387 to 1.163). These metrics provide a rich basis for revealing linguistic patterns, thematic tendencies, and underlying meanings in Leshan-related reports.

Top-ranking collocations include "giant" (Freq(Scaled) 2,384, FreqLR 255), "Buddha" (4,488, 319), "city" (616, 35), "Sichuan" (576, 31), and "Chengdu" (808, 37). "Buddha" boasts the highest frequency at 4,488, with a left-right co-occurrence of 319 (FreqL 51, FreqR 268), a Range of 36 (near full coverage), a Likelihood of 367.053, and an Effect of 1.797, indicating a strong and widespread association with "Leshan," likely in expressions like "Leshan Buddha" or "Buddha of Leshan," underscoring the Giant Buddha's central role. "Giant" follows with 2,384 mentions, FreqLR 255 (FreqL 20, FreqR 235), a Likelihood of 462.029, and an Effect of 2.387, with right co-occurrence far exceeding left, suggesting "Leshan giant" as a fixed phrase emphasizing the Giant Buddha's grandeur. "Chengdu" (808 mentions, FreqLR 37, FreqL 25, FreqR 12) has a high frequency but a lower Effect of 1.163, with left co-occurrence dominating, possibly appearing in "from Chengdu to Leshan," reflecting Chengdu's role as a transportation or tourism starting point. "Sichuan" (576 mentions, FreqLR 31) and "city" (616 mentions, FreqLR 35) serve as geographic modifiers, with Effects of 1.396 and 1.474, respectively, and moderate significance, often used in "Leshan in Sichuan" or "Leshan city," providing regional and urban context.

Mid-frequency collocations (Freq(Scaled) 200–500) include "located" (328, FreqLR 23), "visit" (440, 26), "Emeishan" (200, 16), "station" (304, 20), and "province" (280, 18), expanding the thematic dimensions. "Visit" (440 mentions, FreqLR 26, with left co-occurrence at 25 nearly covering all, FreqR 1) has a Likelihood of 21.924 and an Effect of 1.53, often appearing in "visit Leshan," reflecting a tourism activity theme with a clear syntactic preference for left-placed verbs. "Station" (304 mentions, FreqLR 20, FreqL 7, FreqR 13) shows balanced distribution, with a Likelihood of 19.866 and an Effect of 1.685, possibly in "bus station in Leshan" or "Leshan station," highlighting the importance of transportation facilities. "Emeishan" (200 mentions, FreqLR 16, FreqL 3, FreqR 13) has more right co-occurrences, a Likelihood of 20.602, an Effect of 1.968, and a Range of 4, suggesting concentration in specific documents, likely in "Leshan and Emeishan," reflecting the tourism linkage with Mount Emei. "Located" (328 mentions, FreqLR 23, FreqL 13, FreqR 10) as a locative descriptor has a Likelihood of 25.005 and an Effect of 1.777, often in "located in Leshan," reinforcing geographic positioning. "Province" (280 mentions, FreqLR 18) typically appears in "Leshan in Sichuan Province," with an Effect of 1.652, as a standard collocation. The low-frequency collocation "railway" (160 mentions, FreqLR 13, FreqL 5, FreqR 8) has a Range of 10, a Likelihood of 17.05, and an Effect of 1.99, indicating a high effect despite low frequency, possibly in "Leshan railway" or "railway to Leshan," emphasizing rail transport focus.

[Table 2: Significant Collocations of the Node Word “Leshan” in Overseas Social Media Reports on Leshan, 2022-2024]

Collocate	Rank	Freq(Scaled)	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
giant	1	2384	255	20	235	31	462.029	2.387
Buddha	2	4488	319	51	268	36	367.053	1.797
city	3	616	35	9	26	16	27.695	1.474
located	5	328	23	13	10	14	25.005	1.777
Sichuan	7	576	31	12	19	17	22.308	1.396
visit	8	440	26	25	1	11	21.924	1.53
Emeishan	9	200	16	3	13	4	20.602	1.968
station	10	304	20	7	13	10	19.866	1.685
Chengdu	11	808	37	25	12	15	19.341	1.163
province	12	280	18	7	11	15	17.278	1.652
railway	13	160	13	5	8	10	17.05	1.99

From a sociocultural perspective, the strong collocations of “giant” and “Buddha” position the Leshan Giant Buddha as a majestic cultural landmark in English-language social media, with “Leshan Giant Buddha” as a core expression appealing to international audiences’ interest in Eastern wonders, reinforced by its high Likelihood and Effect. The high frequency of “visit” and transportation-related terms (“station,” “railway”) reflects a tourism orientation, serving the practical needs of potential visitors. The collocations with “Chengdu,” “Sichuan,” and “Emeishan” situate Leshan within Sichuan’s tourism network, with “Chengdu”’s left co-occurrence bias suggesting its role as a transportation hub, and “Emeishan” pointing to bundled promotion with Mount Emei, potentially attracting a broader tourist base. Linguistically, the disparity in left-right co-occurrences (e.g., “giant” and “Buddha” leaning right, “visit” leaning left) reflects English syntactic habits and the fixed phrasing of proper nouns, with “Leshan Giant Buddha” as a branded phrase recurring in social media, enhancing Leshan’s international recognizability.

In summary, this collocation table reveals the linguistic features of English-language social media reports on Leshan from 2022 to 2024: centered on “Leshan Giant Buddha,” supplemented by geographic context, tourism activities, and transportation information. The high frequency, significance, and wide distribution of “giant” and “Buddha” underscore the Buddha’s iconic status, while “Chengdu” and “Emeishan” expand its tourism network, and terms like “visit” and “station” address practical needs. This pattern positions Leshan as a destination blending natural wonders, cultural heritage, and tourism convenience, with English-language social media reinforcing its image as an international tourism landmark through its practical value and cultural appeal.

5. Conclusion

Leshan’s international image can be summarized as a global tourism destination centered on the Leshan Giant Buddha, integrating natural landscapes, cultural heritage, and convenient tourism experiences. As a focal point, the Buddha’s “giant” scale and “Buddhist” identity are widely emphasized, showcasing its grandeur and unique Eastern charm while catering to international tourists’ interest in Eastern culture through its Buddhist heritage and historical value. Additionally, Leshan’s significance within Sichuan’s tourism network is highlighted, with its close ties to Chengdu, Mount Emei, and the confluence of three rivers strengthening its geographic and regional appeal. In overseas social media reports, the high frequency of transportation-related terms (e.g., “boat,” “bus,” “station,” “railway”) and tourism activities (e.g., “visit,” “tour”) reflects Leshan’s practicality and accessibility as a destination. Furthermore, the inclusion of daily life elements (e.g., “food,” “local”) enriches the tourist experience, making Leshan’s

image more multidimensional, relatable, and imbued with local character. Overall, English-language social media has successfully crafted Leshan as a multifaceted international tourism landmark combining cultural depth, natural allure, and tourism convenience.

Based on this international image, future global promotion strategies for Leshan could be optimized as follows: First, the brand core of the Leshan Giant Buddha should be continuously reinforced, expanding the reach of the international phrase "Leshan Giant Buddha" through video, virtual reality, and other technologies to showcase its scale and craftsmanship, appealing to visually oriented overseas audiences. Second, tourism linkages with Chengdu and Mount Emei should be deepened, promoting the concept of a "Sichuan Golden Tourism Circle" to highlight Leshan's strategic value as a key node in the regional network, enhancing its visibility. Third, to meet tourists' practical needs, multilingual travel guides could be developed, integrating transportation, culinary, and cultural experience information, with targeted social media campaigns reaching potential visitors. Additionally, hosting international cultural festivals or Buddhist-themed events and inviting overseas social media influencers to participate could boost Leshan's global interactivity and cultural appeal. Finally, promoting sustainable tourism principles by balancing natural landscape conservation with visitor experiences could position Leshan as a modern, responsible tourism destination. Through these strategies, Leshan can not only solidify its current international image but also further elevate its global recognition and attractiveness.

References

- [1] Baker, P. *Using Corpora in Discourse Analysis*. London: Continuum, 2006.
- [2] Fairclough, N. *Discourse and Social Change*. Cambridge: Polity Press, 1992.
- [3] Partington, A., Duguid, A., & Taylor, C. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins, 2013.
- [4] Hu Kaibao, Tian Xujun. "A Corpus-Based Study of China's Diplomatic Image in English Translations of Chinese Diplomatic Discourse." *Foreign Languages in China*, 2018, 15(06): 79–88.
- [5] Nie Yanmei. "Research Progress and Perspectives in the Field of Urban Image Dissemination in Europe and America." *Urban Culture Studies*, 2016(01): 95–109.
- [6] Qian Yufang. "Corpus Linguistics and Critical Discourse Analysis." *Foreign Language Teaching and Research*, 2010, 42(03): 198–202+241.
- [7] Qian Yufang, Dong Yingying. "A Study of Yiwu's Image Construction in Overseas Media." *Corpus Linguistics*, 2020, 7(02): 28–38.
- [8] Shao Bin, Hui Zhiming. "The 'Chinese Dream' in Western Media: A Critical Discourse Analysis Based on Corpus Linguistics." *Foreign Language Studies*, 2014(06): 28–33.
- [9] Tian Hailong. "Revisiting the Essence of Critical Discourse Analysis: Starting from Three Related Issues." *Foreign Languages and Their Teaching*, 2016(02): 1–9+144.
- [10] Wang Qi. "The Image of Chinese Enterprises in English Media of the 'Belt and Road': A Critical Discourse Analysis Supported by Corpus Linguistics." *Contemporary Foreign Language Studies*, 2019(03): 99–113.