

A Corpus-Driven Multi-Criteria Decision Support Framework for Complex Decision Environments: Evidence from Heterogeneous Textual Records of the Hmong Sister Festival in China

Chanjuan Wang, Jing Zhang, Guangquan Dai, Xin Xu

How to cite: Wang C, Zhang J, Dai G, Xu X. A Corpus-Driven Multi-Criteria Decision Support Framework for Complex Decision Environments: Evidence from Heterogeneous Textual Records of the Hmong Sister Festival in China. Textile & Leather Review. 2026; 9:2759-2790.

<https://doi.org/10.31881/TLR.2026.2759>

How to link: <https://doi.org/10.31881/TLR.2026.2759>

Published: 25 April 2026



A Corpus-Driven Multi-Criteria Decision Support Framework for Complex Decision Environments: Evidence from Heterogeneous Textual Records of the Hmong Sister Festival in China

Chanjuan Wang^{1,2}, Jing Zhang², Guangquan Dai^{1*}, Xin Xu¹

¹Department of Tourism Management, South China University of Technology, Guangzhou 510006, Guangdong, China

²Tourism and Air Service College, Guizhou Minzu University, Guiyang 550025, Guizhou, China

*academic686@163.com

Article

<https://doi.org/10.31881/TLR.2026.2759>

Published 25 April 2026

ABSTRACT

Decision-making in complex environments increasingly depends on the ability to transform heterogeneous textual records into structured analytical evidence. However, existing studies on festival-related texts are predominantly qualitative and rarely provide operational decision models that convert textual evidence into measurable criteria for management and policy support. To address this gap, this study proposes a corpus-driven multi-criteria decision support framework that integrates corpus linguistics, functional text analysis, and a structured multi-criteria decision-making (MCDM) procedure. Using the Hmong Sister Festival in Guizhou, China, as a case study, a structured diachronic corpus database is constructed from historical documents and multi-source textual materials. Quantitative analytical techniques, including keyword extraction, co-occurrence network analysis, topic modeling, and principal component extraction, are employed to identify dominant textual structures and temporal patterns and to transform textual evidence into measurable decision criteria. The extracted indicators are then normalized, weighted through a structured criteria-weighting process. The results reveal that the textual system exhibits strong diachronic continuity while simultaneously reflecting distinct context-dependent characteristics. By linking corpus-driven text modeling with an explicit multi-criteria decision structure, the proposed framework provides a reproducible and computationally transparent approach for supporting evidence-based decision-making in complex decision environments, including cultural heritage management, tourism planning, knowledge management, documentation-intensive governance, and text-enabled decision analytics.

KEYWORDS

multi-criteria decision-making (MCDM), TOPSIS, corpus-driven text analytics, unstructured textual data

INTRODUCTION

Traditional festivals among rural ethnic minorities provide a rich source of unstructured textual data embedded in complex managerial and institutional environments [1]. Existing research on traditional festivals has drawn on perspectives from sociology, anthropology, and cultural studies, addressing issues such as heritage protection [2], authenticity [3], tourism consumption behaviour [4,5], sense of place and meaning [6], cultural potential [7], and sustainable urban development [8]. In China, many studies have examined local traditional festivals through narrative and interpretive approaches [9,10]. However, there remains substantial scope for operational research that converts textual content, representational structure, and temporal variation into decision-relevant analytical outputs. Written texts constitute a critical source for understanding long-term change, as collective memory is inherently unstable and oral accounts are often difficult to verify consistently. The Hmong Sister Festival, held in Guizhou Province, China, has been extensively recorded across multiple documentary channels, including historical records, media reports, and academic publications. However, in the absence of systematic text-analytical approaches, these heterogeneous materials remain fragmented and difficult to transform into structured evidence for management and decision support. More importantly, existing studies rarely operationalize such textual resources through reproducible corpus-based and functional analytical procedures. In particular, there is a lack of explicit methods for translating textual features into structured indicators, deriving decision criteria from textual patterns, and supporting evidence-based ranking within a formal decision-making framework.

As a representative case of intangible cultural heritage protection, the Sister Festival has been recorded and disseminated in various forms in the modern context. Different actors and media produce heterogeneous textual records, making the case suitable for studying how unstructured textual data can be transformed into a structured decision-support framework. Rather than treating the material only as an object of cultural interpretation, this study highlights its value as a complex textual environment in which salient features, representation clusters, and temporal variation can be computationally identified and converted into decision criteria [11,12]. This study addresses the following research questions:

- Which actor groups generate the major textual signals involved in representing the Hmong Sister Festival?
- What textual elements, structural patterns, and functional signals characterize these representations across different media and historical periods, and how can they be quantified?

- How can corpus-driven text analytics be operationalized into a multi-criteria decision support model for complex management and policy contexts?

In decision-oriented contexts such as cultural heritage management, tourism planning, and knowledge management, decision-makers require structured evidence on how textual representations evolve and influence management priorities. Quantitative modeling of textual signals can inform decisions related to resource allocation, conservation strategies, interpretive programming, communication design, and institutional coordination. To meet this need, the proposed framework is designed as an operational evidence-generation tool rather than a purely linguistic inquiry. Specifically, it implements a structured analytical pipeline in which textual features are extracted through corpus-based methods, transformed into interpretable indicators via network and topic-based modeling, and subsequently organized into measurable decision criteria. These criteria are then normalized and weighted within a multi-criteria decision-making procedure, enabling the comparison and prioritization of management alternatives in a transparent and reproducible manner.

To address the above research questions, this study employs a corpus-based content analysis approach. It conducts a mixed quantitative and qualitative analysis of multi-source textual records surrounding the Hmong Sister Festival, using keyword analysis to identify signal salience, co-occurrence analysis to examine relational structures, corpus comparison to capture diachronic variation, and principal component extraction to reveal underlying representational patterns. On this basis, the extracted textual features are systematically mapped into interpretable decision criteria, normalized into an evaluation matrix, weighted based on corpus-derived indicators, and synthesized through a TOPSIS-based ranking model to support transparent comparison of alternative management priorities. This study makes the following contributions:

1. Proposes a corpus-driven multi-criteria decision support framework for analyzing unstructured textual data in complex decision environments, enabling structured and replicable evidence generation for management and policy support.
2. Reveals functional differences in textual representation across multiple actor-related discourse sources (e.g., government agencies, scholars, media, and public narratives), and translates these differences into quantifiable evaluative criteria for management and planning decisions.
3. Establishes an operational decision structure linking textual indicators, criteria weighting, and alternative ranking, thereby advancing the framework from conceptual interpretation to decision-oriented application.

4. Highlights implications for data-driven decision by providing an evidence-based analytical model that supports transparent prioritization, scalable knowledge organization, and generalizable decision support in engineering-relevant contexts.

The rest of this paper is organized as follows. Section 2 reviews the literature on temporal textual representation, multi-source text construction, and identity-related analytical dimensions, while clarifying their relevance to text analytics and decision support. Section 3 introduces the case background, research methods, and data sources. Section 4 presents the empirical results, including keyword analysis, co-occurrence analysis, and multi-source text patterns, and demonstrates how these outputs support operational ranking. Section 5 discusses the main findings, their theoretical and practical significance, and future research directions.

LITERATURE REVIEW

Temporal textual representation

Research on historical narrative has highlighted how temporal records are organized into analyzable textual structures rather than being treated only as neutral descriptions of past events [13,14]. For the purposes of this study, the key implication is methodological: historical materials can be treated as structured textual inputs that encode temporal order, value orientation, and representational emphasis. White's work [15,16] shifted attention from factual enumeration alone to the textual organization through which events are presented. In an engineering-oriented decision framework, this perspective is useful because it justifies treating historical records as heterogeneous data objects whose recurring structures can be identified, compared, and converted into measurable indicators.

Subsequent studies have extended this line of work to questions of cross-cultural applicability [17], reinterpretation [18], historical thinking [19], evaluative research [20], and comparative historiography [21,22]. Related studies on temporal analysis, textual representation, and national image construction [23-26] further demonstrate the potential of these approaches for structured text analysis. In this study, temporal textual representation is therefore not treated as a purely philosophical construct, but as an analyzable data layer that can be operationalized through corpus-based methods to examine patterns of continuity and change and to generate structured inputs for decision support.

Multi-source textual representation

Research on discourse construction is relevant here because it explains how multiple actors generate different but interrelated textual representations within the same information environment [27-30]. For engineering-oriented analysis, the important issue is not discourse theory alone but how textual signals can be extracted from heterogeneous sources and integrated into a comparable analytical structure. Prior studies show that corpus-based approaches help bridge qualitative interpretation and quantitative analysis [31-33], while work on national identity, media framing, and public communication demonstrates that textual representation often varies systematically by actor role, institutional position, and communication objective [34-36]. These studies collectively support the use of multi-source text analytics to model complex textual environments across media and historical contexts.

In festival studies, however, relatively few works have treated discourse as a source of operational evidence rather than as an interpretive endpoint. A multi-source textual perspective makes it possible to compare how government agencies, scholars, media outlets, and residents encode the same event differently, and how these differences may be transformed into knowledge structures. This is directly relevant to decision support because it allows dispersed textual records to be fused into a reproducible evidence base.

Identity-related analytical dimensions

Cultural identity is a dynamic, multi-dimensional process [37], deeply influenced by social and historical contexts [38-39]. In the present study, identity is not treated primarily as a philosophical endpoint, but as an analytical dimension that can be inferred from recurring textual signals such as actor roles, evaluative language, representational emphasis, and symbolic references. Prior studies have examined the connotation [40,41], function [42-44], structure [45,46], and development trends [47-49] of cultural identity, but quantitative operationalization remains relatively limited. This paper therefore treats identity-related signals as measurable textual features that can be extracted, structured, and incorporated into a decision model alongside other criteria relevant to management and policy analysis.

METHOD

To empirically examine the issues discussed in the previous sections, this study adopts the Hmong Sister Festival as a case study and develops a corpus-based analytical framework for decision support. The Hmong Sister Festival is a traditional annual ethnic minority festival characterized by a relatively fixed duration, a localized

geographical setting, and established ritual practices. It originated from local customs related to marriage and social interaction. During the twentieth century, local government initiatives expanded the traditional practice of “eating sister meal” into the modern form of the Sister Festival. The first Guizhou Taijiang Hmong Sister Festival was officially held in 1998, marking its institutionalization and contemporary significance. In 2006, it was included in the first batch of China’s National Intangible Cultural Heritage list. Over the past two decades, the festival has evolved from a local folk practice into a regional cultural event with increasing public visibility and institutional involvement. From an analytical perspective, this transformation provides a suitable case for examining how heterogeneous textual records generated across time, media, and actor groups can be systematically converted into structured decision-relevant evidence. The geographical location of the Hmong Sister Festival is illustrated in Figure 1.



Figure 1. The geographical location and related images of the Hmong Sister: The map is drawn based on the map of the Ministry of Natural Resources of the People’s Republic of China (Review No.: GS20161587) Source: Photographs of Mr. Gao Bing, used with permission, edited and compiled by the authors.

Research method

This study employs a quantitative content analysis approach grounded in corpus linguistics and integrates multiple text analysis tools to support a corpus-driven decision framework. Corpus-based methods enable the

systematic examination of linguistic patterns across different levels of language use and provide structured inputs for discourse-oriented analysis [50]. CorpusWordParser is used for text segmentation and part-of-speech tagging to improve lexical processing accuracy. AntConc 3.5.8 is then applied to conduct keyword frequency and collocation analysis, identifying salient vocabulary items and associative patterns within the texts. KH Coder 3.0 is subsequently employed for correspondence analysis and multivariate coding analysis to examine structural relationships, discourse variation across sources, and thematic organization. In a word, these methods support a stepwise transformation from raw textual materials to interpretable analytical indicators, which are subsequently used for criteria construction and decision modeling.

Data collection

The keyword “Sister Festival” was used as the primary search term to collect relevant textual materials. A total of 119 academic journal articles, 51 yearbooks, and 16 newspaper articles were retrieved from the China National Knowledge Infrastructure (CNKI) database. In addition, 2 monographs and 175 books were identified through an academic search of the Duxiu database, and 2 volumes of the Taijiang County Chronicle were obtained from the Guizhou Provincial Library. Web-based materials were collected using the Baidu search engine, yielding 760 initial records. Data collection was completed by 12 March 2024.

All retrieved texts were manually reviewed to ensure relevance and quality. To address potential overlap, texts were categorized by publication type and institutional origin (e.g., government, academic, media, and public sources). While some overlap remains, this classification enables comparison of dominant discourse signals across actor groups. Duplicate, redundant, overly short, incoherent, or irrelevant materials were excluded, and non-textual content such as images and videos was removed. After screening and cleaning, the final corpus consisted of 2 monographs, 2 county chronicles, 40 annals, 85 books, 67 journal articles, 14 newspaper articles, and 150 web texts. For county chronicles, annals, and books containing mixed content, only sections related to the Sister Festival were extracted, while the remaining sources were retained in full. The cleaned corpus comprised approximately 674,914 words, with a total size of about 8 MB. An overview of the data distribution is illustrated in Figure 2.

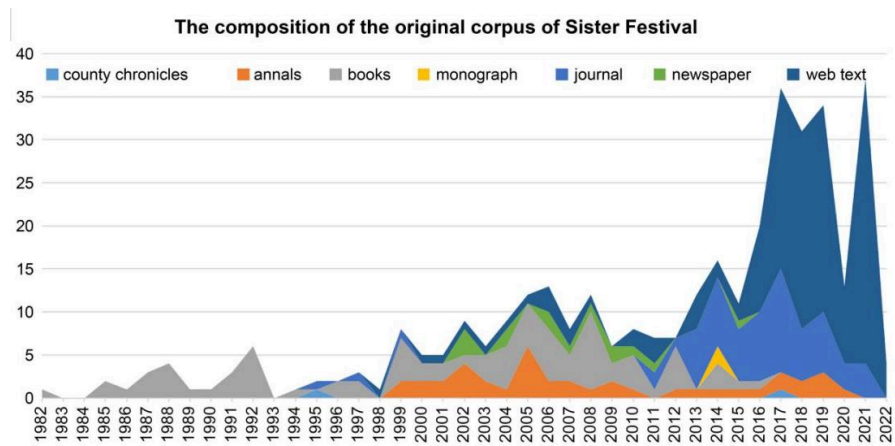


Figure 2. Distribution of the text corpus by year

Operational Multi-Criteria Decision Support Framework

Algorithm 1. Corpus-Based Multi-Criteria Decision Support Framework

Input:

$D = d_1, d_2, \dots, d_k$ // Corpus data (multi-source narrative texts)

Output:

R = ranked set of management alternatives

- 1: Construct corpus database D from multi-source texts
- 2: Perform text preprocessing: segmentation, cleaning, part-of-speech tagging
- 3: Extract discourse features F: keyword frequencies; co-occurrence patterns; topic distributions; principal components
- 4: Model diachronic patterns: analyze temporal variations in F
- 5: Construct decision criteria set: $C = c_1, c_2, \dots, c_n$
- 6: Quantify criteria values: Build decision matrix $X = [x_{ij}]$
- 7: Determine criteria weights: Compute $W = w_1, w_2, \dots, w_n$, subject to $\sum w_j = 1$
- 8: Apply the TOPSIS method:
 - Normalize $X \rightarrow R = [r_{ij}]$
 - Compute weighted matrix $V = [v_{ij}]$
 - Identify ideal solutions A^+ and A^-
 - Compute distances S^+ and S^-
 - Calculate the closeness coefficient C_i
- 9: Rank alternatives: Sort C_i in descending order $\rightarrow R$
- 10: Interpret results: Translate ranking into policy and management implications

Return R

This subsection outlines a structured analytical framework that guides the corpus-driven analysis and facilitates evidence-based decision-making for cultural heritage management. Step 1: Corpus construction -

gather and curate texts related to the festival across media and historical periods. Step 2: Feature extraction - perform text segmentation, keyword frequency analysis, and collocation analysis to identify salient terms and expressions. Step 3: Temporal and structural modeling – topic modeling, correspondence analysis, and principal component analysis are applied to capture thematic structures, discourse variation across sources, and diachronic patterns. Step 4: Criteria construction – the extracted textual features are transformed into decision criteria through explicit mapping rules. Specifically, keyword prominence is used to represent thematic salience, co-occurrence and collocation structures reflect relational density and representational stability, source differentiation indicates institutional visibility and participant diversity, and temporal persistence captures continuity of representation. Based on these mappings, five criteria are constructed: continuity of representation, identity-related salience, breadth of participant representation, visibility of institutional support, and development orientation. Step 5: Criteria weighting – criteria weights are derived from normalized corpus-based indicators to ensure computational transparency and reproducibility. Where necessary, expert interpretation is used only as a supplementary validation rather than as the primary source of weighting. Step 6: Alternative ranking – the weighted criteria are incorporated into a TOPSIS-based decision model to compare management alternatives. Step 7: Decision interpretation – the ranking outputs are translated into actionable implications for heritage governance, tourism planning, and policy decision-making. As shown in Algorithm 1, the framework integrates corpus-driven text analytics with a multi-criteria decision-making process. Heterogeneous textual records are first transformed into structured features through corpus analysis techniques. These features are then systematically mapped into decision criteria and organized into a decision matrix. A weighting scheme is subsequently applied, and the TOPSIS method is used to compute the relative closeness of each alternative to the ideal solution, producing a ranked set of management priorities in a transparent and reproducible manner.

Decision computation and ranking procedure

Given the corpus-derived indicators and the constructed decision criteria, the framework is operationalized through a structured decision computation procedure. The indicators are organized into a decision matrix $X = [x_{ij}]$, in which each row represents a management alternative and each column represents a decision criterion derived from the text analytics process. Specifically, keyword frequencies capture thematic salience, collocation densities and co-occurrence structures reflect relational importance, topic proportions indicate thematic prominence, and principal component scores summarize higher-order variation across sources and

time. In the present context, three types of management alternatives are defined. Preservation-oriented management prioritizes cultural continuity, ritual authenticity, and heritage protection. Tourism-promotion-oriented management emphasizes visitor attraction, economic development, and external communication. Balanced management seeks to integrate both preservation and development objectives.

The indicator values are standardized to ensure comparability across heterogeneous metrics. Let r_{ij} denote the normalized value of criterion j for alternative i , and let w_j denote the weight assigned to criterion j . The weighted normalized matrix is then defined as $v_{ij} = w_j r_{ij}$. Based on this matrix, the positive ideal solution and negative ideal solution are identified, and the relative closeness of each alternative to the ideal solution is calculated using the standard TOPSIS procedure. The alternative with the highest closeness coefficient is interpreted as the management priority that best reconciles continuity, identity-related salience, institutional feasibility, and development needs. It should be noted that individual textual features (e.g., keyword frequency) do not directly determine decision priorities; rather, they function as empirical inputs within a multidimensional evaluation framework. This decision layer does not replace interpretation; instead, it translates the analytical findings into a transparent and reproducible structure for practical comparison. In this way, corpus-driven text analytics functions as an integral component of a structured decision-support system rather than remaining a purely descriptive or conceptual exercise.

RESULTS

Keyword analysis

The cleaned corpus was organized into seven text files and encoded in UTF-8 format. Text segmentation and part-of-speech tagging were conducted using the CorpusWordParser tool, and the annotated files were subsequently imported into AntConc 3.5.8 for analysis. The “Word Frequency Table of the Modern Chinese Corpus,” provided by the Computational Linguistics Laboratory of the Institute of Applied Linguistics, Ministry of Education, was adopted as the reference corpus. The self-constructed corpus contains 520,098 tokens and 26,298 types, yielding a type–token ratio (TTR) of approximately 0.05. This value provides a descriptive indication of lexical distribution within the corpus and reflects the repetitive use of core vocabulary associated with festival-related discourse. It is used here for internal characterization rather than direct comparison across corpora of different sizes. Keyword analysis was performed in AntConc using the Keyword List function, generating keyword frequency and keyness values based on log-likelihood statistics. Tool-specific statistics such as “Types Before Cut” and “Types After Cut” reflect internal segmentation and filtering processes within

the software. The resulting keyword list indicates that a relatively small set of vocabulary items accounts for a high proportion of usage, suggesting strong thematic concentration. As shown in Table 1, the top 30 notional words exhibit high frequency and positive keyness values, indicating that these terms are significantly overrepresented compared with the reference corpus. To avoid overinterpreting semantically general words, high-frequency items such as “eat,” “day,” and “use” are interpreted cautiously and are not treated as primary indicators unless supported by contextual relevance.

Table 1. Keywords, word frequency, keyness

Keywords	Word frequency	Keyness	Keywords	Word frequency	Keyness	Keywords	Word frequency	Keyness
Hmong	7586	+411.58	glutinous rice	1889	87.98	Shi Dong	895	+49.74
Sister Festival	4926	+274.81	tourism	1686	+84.1	silver ornament	892	+49.57
culture	3692	+194.48	holiday	1645	+81.86	use	1046	+49.42
festival	3571	+187.78	traditional	1578	+78.21	development	1045	+49.37
activity	2874	+149.28	day	1383	+67.62	Youfang	846	+47.01
Taijiang	2738	+152.25	eat	1351	+65.89	society	957	+44.65
ethnic	2219	+113.25	youth	1213	+58.42	costume	940	+43.74
girl	2165	+110.29	Guizhou	1150	+55.02	month	939	+43.69
sister rice	1968	+109.48	Qiannan	917	+50.96	village	698	+38.78
year	1949	+98.46	dressy	915	+50.85	tourist	738	+32.99

Keywords have seven aspects: symbolism, time, location, participants, content, mention of objects, and the deeper content of festivals. According to the researcher’s subjective interpretation, they are later used as the basic framework for content analysis [51].

Keyword diachronic analysis

Based on the composition of the original corpus, a staged diachronic analysis was conducted. As shown in Table 2, the number of word types increased across the four stages, reflecting a general expansion of lexical coverage over time. The type–token ratios (TTRs) for the four stages were 34.24%, 47.69%, 16.26%, and 6.75%, respectively. Given the substantial differences in corpus size across stages, these values are interpreted descriptively rather than as directly comparable measures of lexical richness. The second stage contains relatively fewer texts and tokens due to limited documentation and a shorter historical period. Under this condition, a higher TTR reflects a more dispersed lexical distribution within a smaller corpus, rather than

an absolute increase in vocabulary diversity. In contrast, the marked growth in word types and tokens in Stages 3 and 4 corresponds to the expansion of communication media and increased public attention. The concurrent decline in TTR in these stages suggests that discourse production relied more heavily on a stable and repeatedly used vocabulary set, with fewer newly introduced terms. Similar patterns are observed in the distribution of keyword types and tokens, indicating increasing discourse consolidation rather than lexical innovation.

Table 2. Calendar statistics for keyword types and tokens

	1. 1982–1994	2. 1995–1997	3. 1998–2010	4. 2011–2022
Word Types:	2572	1914	8883	24224
Word Tokens:	7512	4013	54641	358871
Keyword Types:	96	80	152	124
Keyword Tokens:	2262	1047	17281	94975

Keyword collocation analysis

John Firth proposed the concept of collocation [52]. Collocation research provides detailed descriptive evidence and helps reveal linguistic features that may be overlooked in frequency-based analysis. By extracting word clusters (i.e., sequences composed of two or more words) using the statistical software AntConc, this study designated “Sister Festival” as the node word to obtain collocations associated with the festival. Pronouns, prepositions, auxiliary words, and other function words were removed to retain notional collocations, which are presented in Tables 3 and 4.

Table 3. Concrete collocation before the node of “Sister Festival” as a nodal word

Before node	Nodal word	County chronicles	Annals	Books	Mono-graphs	Journals	Newspapers	Web texts	Aggregate
Hmong	Sister Festival	25	10	97	228	365	10	217	952
A certain year, a certain session, now, past, nowadays, before, ancient times		28	16	42	74	97	7	116	380
Qingshui River, Southeast Guizhou Prefecture, Gedong, Laotun Township, Pian Village, village, other surrounding counties		2	0	24	86	68	8	56	244
Taijiang County of Guizhou Province, Taijiang of Guizhou, Taijiang County		0	0	3	1	45	0	194	243
Taijiang Hmong		4	12	23	13	60	4	40	156
Shidong Hmong		5	1	19	22	76	4	28	155

Before node	Nodal word	County chronicles	Annals	Books	Mono-graphs	Journals	Newspapers	Web texts	Aggregate
Chinese Hmong		0	2	0	1	17	2	100	122
Fifteenth day of the third lunar month		4	0	19	28	25	4	40	120
Culture, tradition, folk, folk custom, intangible cultural heritage, original ecology, ethnic minority areas		2	0	5	7	57	5	38	114
Hmong of Guizhou		14	6	13	15	32	2	25	107
Grand, solemn, colorful, ancient, jubilant, romantic		0	0	15	17	21	2	42	97
The oldest oriental Valentine's Day, sugary love, hidden in stamens, hidden in the bud		0	0	14	3	17	2	43	79
Ecological environmental tour, rural revitalization, Chinese urban and rural tour, Shenzhou century tour		14	12	12	10	0	0	16	64
Hmong of Taijiang of Guizhou		2	4	0	3	22	3	7	41
Fifteenth day of the second lunar month		3	0	4	28	3	0	2	40
Be in love with, social contact, marriage, marriage customs, mate selection, young men and women, women, females		0	0	2	1	15	0	11	29

As shown in Table 3, the collocation “Hmong ethnic group” prominently represents the ethnic identity associated with the Sister Festival. Temporal expressions such as “a certain year” reflect the timeliness of historical narratives and indicate continuity and change in the chronological development of the festival. In addition, collocations combining the Sister Festival with geographic references—including “Qingshui River,” “Southeast Guizhou Prefecture,” “Gedong,” “Laotun Township,” “Pian Village,” “village,” and “other surrounding counties”—enrich the spatial scope of the discourse. The high frequency of place-related collocations suggests that regional context constitutes an important dimension of festival representation. These patterns indicate that the discourse surrounding the Sister Festival increasingly situates it within a spatial structure centered on Taijiang County, with Shidong Town as its point of origin and key villages as characteristic sites, extending to surrounding counties and, in some representations, beyond China.

Because traditional Sister Festival celebrations take place in multiple locations within the county, the timing of celebrations varies. As a result, the gendered social nature of the festival is not always foregrounded in the discourse, while greater attention is given to its contemporary social and cultural significance. The development of the Sister Festival has also been closely associated with national tourism initiatives, and its external publicity discourse has correspondingly aligned with broader tourism development trends. With the increasing involvement of media, representations of the Sister Festival have become more diversified and

rhetorically elaborated. Expressions such as “the oldest oriental Valentine’s Day,” “sugary love,” and “hidden in the stamens” appear frequently in everyday discourse and media narratives, reflecting popularized and metaphorical modes of description.

Table 4. Concrete collocation after the node of “Sister Festival” as a node word

Nodal word	After node	County chronicles	Annals	Books	Mono-graphs	Journals	News-papers	Web texts	Aggregate
Sister Festival	1. Place name or scope: Guizhou, Guiyang, Southeast Guizhou, Kaili, Taijiang, Shidong, Gedong, Fangnan, community, folk, village, square, countryside	2	8	29	82	47	2	48	218
	2. Time frame: period, occasion, day, time, schedule, time, lunar calendar, March, February, May Day Golden Week	21	25	119	120	144	14	78	521
	3a. Characters: young women, Hmong youth, protagonists, girls, outsiders, rookies, tourists, runners	1	1	28	36	33	2	49	150
	3b. Appellations: modern free translation, name, also known as, Hmong language, also called, Hmong dialect, the ancient name	7	1	53	22	34	1	76	194
	4. Nicknames: honorary title, selected intangible cultural heritage, listed as, praised as, rated as, named	1	5	8	3	39	1	51	108
	5. Publicity: advertisement, feature film, promotional film, brand effects, market positioning, promotion, logo, image, announcements, live broadcast, mascot	2	3	2	9	24	1	39	80
	6. Origin: birthplace, hometown, origin, spread in, legend, predecessor, derived from, source	6	0	52	40	71	4	86	259
	7. Definition or nature: ethnic festivals, marriage, ancestral commemoration, traditional festivals, folk customs, cultural carriers, blind dates, tourist activities, cultural banquets	4	1	36	44	80	3	98	266
8. Description: long history, simple, world-famous, rich and colorful, unique, the most romantic, ancient, crowded, spectacular	4	2	47	27	23	5	184	292	

Nodal word	After node	County chronicles	Annals	Books	Mono-graphs	Journals	News-papers	Web texts	Aggregate
	9. Iconic events: opening ceremonies, celebrations, jubilees, press conferences, media seminars, symposiums, economic and trade fairs, tourism development conferences, photography exhibitions	20	3	3	12	1	2	75	116
	10. Scene introduction: grand occasion, condition, situation, scenario, overview, beauty, atmosphere	15	1	6	12	3	1	30	68
	11. Process introduction: activities, tourism activities, contents, themes, materials, projects, customs, etiquette, procedures, folk customs, cloud tours, folklore	25	5	59	239	116	4	73	521
	12. Concrete activity projects: Youfang, parade, stepping on drums, sacrifice, singing and dancing, eating, catching fish, meeting uphill, returning gifts, killing pigs, getting married, betrothal, long table	6	1	29	86	85	2	67	276
	13. Items involved: food, bacon, road blocking wine, tokens, sisters' rice, Hmong costumes, Lusheng, schedule, activity guide	5	0	13	71	50	1	41	181
	14. Organization and management: organizing committee, county government, state government, local government, village elders, women's director, lead, association organization, launch, host	3	1	5	44	20	0	22	95
	15. Introduction to a certain aspect: culture, music, dance, songs, history, tourism, ancient songs, transportation	16	1	17	43	198	3	69	347
	16. Evolution: following, preservation, protection, inheritance, become, formation, change, variation, follow the old habits	2	0	12	5	47	2	3	71
	17. Research: evaluation research, investigation research, thinking	0	0	1	81	195	2	101	380
	18. Other festivals: Dragon Boat Festival, Eating New Festival, Women's Festival, Dong Samar Festival, Bridge Worship Festival, Chabai Song Festival, Drum Tibet Festival	5	0	11	9	84	3	9	121
	19. Related attractions: wooden drum dance of Fanpai, dragon dance with fireworks, Xijiang	12	0	3	14	11	3	11	54

Nodal word	After node	County chronicles	Annals	Books	Mono-graphs	Journals	News-papers	Web texts	Aggregate
	thousand household Hmong village, Jiabang terraces, Fan-jing Mountain								
	20. Results: successfully held, realized, attracted investment, completed successfully, attracted, continuously held, grandly held, function, effect	4	7	4	4	78	4	19	120

The collocations following the node “Sister Festival” in Table 4 demonstrate a rich and relatively comprehensive coverage of informational content. These collocations encompass temporal, spatial, participatory, and activity-related aspects, indicating that discourse representations of the festival have become increasingly concrete and detailed.

The Sister Festival is consistently described as being centered in Taijiang County, while its sphere of influence extends both outward (e.g., Kaili, Southeast Guizhou, Guiyang, and Guizhou) and inward to nearby towns and villages (e.g., Shidong, Gedong, and Fangnan). This distribution of place-related collocations suggests an expansion of the festival’s perceived influence and growing community participation. Festival descriptions tend to emphasize differences in historical background and temporal context, and the alignment of traditional festival dates with modern public holidays is particularly noticeable. Discourse representations also reflect diversity among festival participants in terms of age, ethnicity, region, occupation, and gender.

The name “Sister Festival” originates from the Hmong language, and an indigenous term is commonly used within the community, whereas the modern translated name “Sister Festival” is more prevalent in external discourse. The presence of additional nicknames in the corpus further reflects the popularity and symbolic resonance of the festival. Influenced by the communicative functions of different text genres, promotional discourse is more prominent in journal articles and web texts than in other sources. Given the festival’s long history, references to its origin and birthplace are frequently included. Overall, discourse surrounding the Sister Festival is richest in descriptions of scenes, processes, specific activities, and organizational arrangements. Local cultural practitioners, government agencies, and local organizations tend to employ more standardized and tradition-oriented expressions, whereas external media sources favor more elaborate and diversified rhetorical strategies. Collocations centered on iconic events, organization, management, and outcomes further indicate that traditional festivals also exhibit characteristics associated with modern festival practices. Taken together, the collocational patterns show that discourse on the Sister Festival extends beyond

the description of customary practices to include reflections on its development, transformation, and evaluative interpretations by different discourse subjects.

Multiple discourse expressions

KH CODER 3.0 software was used to analyse the correspondence of feature words in multiple discourse expressions and to generate Figure 3. This correspondence analysis aims to identify how different discourse sources foreground distinct thematic and functional emphases in representing the Sister Festival.

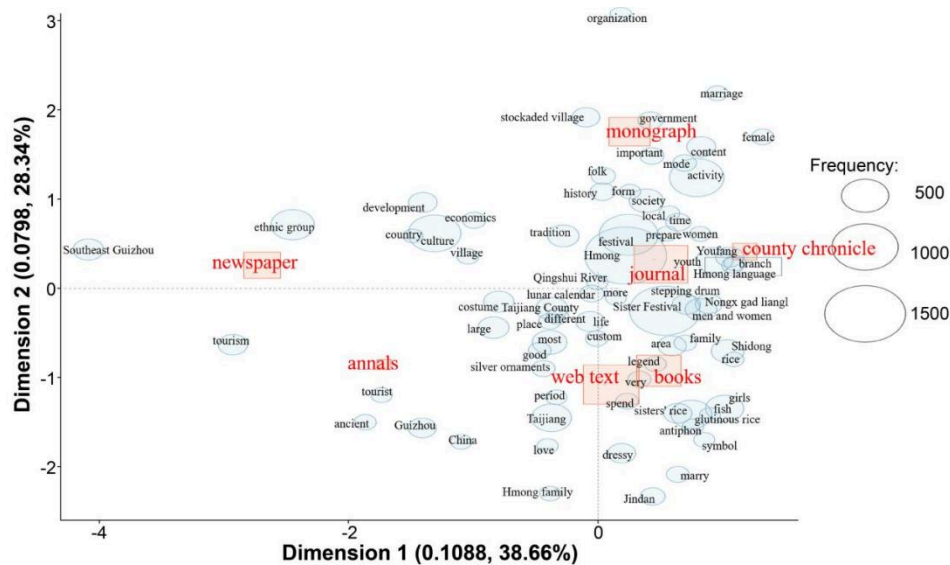


Figure 3. Correspondence analysis of feature words

The discourse features words from the county chronicles, including ‘expression’, ‘activities’, ‘youth’, ‘female’, ‘Shidong’, and others, introducing and presenting an interpretation of the Sister Festival. Words from the annals, including ‘held’, ‘opening ceremonies’, ‘TV station’, ‘news’, and ‘festival’, present the introduction to the representative events and a summary of the process, achievements, and influence. Words from the books, including ‘girls’, ‘sisterhood’, ‘men and women’, ‘glutinous rice’, ‘young man’, and others, provide a descriptive introduction to the concrete content of the Sister Festival. Words in the monographs, including ‘activities’, ‘festival’, ‘Hmong ethnic’, ‘Sister Festival’, ‘content’ and others, present a comprehensive introduction and discussion of the issue. Words from the journals, including ‘Hmong ethnic’, ‘Sister Festival’, ‘culture’, ‘activities’, ‘Shidong’ and others, present concrete aspects of the issue to be explored. Words from the newspapers, including ‘Southeast Guizhou’, ‘culture’, ‘ethnic group’, ‘protection’, ‘journalist’ and others present an expansion of the scope of attention and the understanding of new issues. Taken together, institutional sources such

as county chronicles and annals primarily function to introduce and standardize the basic profile of the Sister Festival, reflecting an institutional and descriptive discourse orientation.

Overall, the county chronicles and books are mainly descriptive introductions to the Sister Festival. Monographs and journals are evaluative research based on a descriptive introduction, which is either concrete or comprehensive. They are based on book content and county chronicles, combined with in-depth analysis and interpretation of the actual investigation. The newspapers reflect the descriptive characteristics of the news media’s contemporary discourse. The annals focus on the achievements and influence of public affairs, while the web-text features are not distinctive. This distribution indicates a clear hierarchy of discourse functions, ranging from foundational description to evaluative interpretation, shaped by the institutional roles and communicative purposes of different media types.

The generated data are coded manually to further understand the intentions of individuals and organisations expressed through different discourses. According to the three elements of language function [53], there is also an authorial evaluation, which is distinct from the author’s expressive function. It is the author’s thinking on certain issues. Using the coding cross-analysis of KH CODER 3.0 software, with description, expression, evocation, evaluation, and thinking as the theme coding and multiple discourse sources as the cross-analysis selection items, the data running results are displayed in Figure 4.

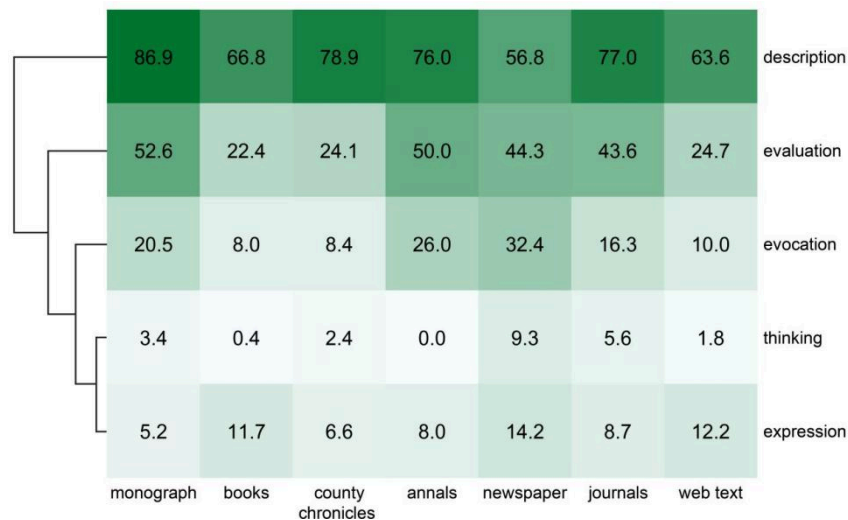


Figure 4. Cross-analysis of text type groups and coded themes

The chi-square formula is used to measure whether there is a correlation between two categorical variables.

$$\chi^2 = \sum [(O_i - E_i)^2 / E_i] \quad (1)$$

Where χ^2 represents the chi-square value, O_i represents the observed value, and E_i represents the expected value. In general, the larger the chi-square value, the smaller the p-value, indicating that the greater the difference between the observed value and the expected value, the stronger the correlation between the two variables. The chi-square values are 99.732**, 32.324**, 203.745**, 163817** and 244.213**, indicating that the chi-square values are large, the p-value is less than 0.05, and the variables are significantly correlated. The relevance of the function of description is the strongest, indicating that the multi-discourse is based on the description as the basic prerequisite. This dominance of descriptive discourse provides the narrative foundation upon which shared cultural identity is stabilized and transmitted across discourse communities. This result suggests that discourse function distribution is not random across text types but is systematically shaped by discourse source indicating structurally differentiated patterns of cultural representation. Although effect size measures (e.g., Cramer's V) are not explicitly reported, the consistent significance patterns across discourse types provide sufficient evidence for systematic functional differentiation.

Operational decision-support implications

The preceding corpus results can be interpreted not only as descriptive findings but also as inputs to an operational decision model. In particular, the identified textual patterns provide measurable evidence for five decision criteria: continuity of representation, salience of identity-related signals, breadth of participant representation, visibility of institutional support, and intensity of development-oriented communication. These criteria define the evaluative dimensions through which managers can compare alternative priorities in a transparent and reproducible manner.

From the Hmong Sister Festival case, the corpus evidence indicates that continuity-related and identity-related signals remain the most stable dimensions across historical periods, while media-oriented publicity and development-oriented communication are more sensitive to institutional context and contemporary communication demands. This pattern suggests that decision systems using unstructured textual data should prioritize criteria associated with continuity and semantic stability, while evaluating promotion-oriented measures in balance with, rather than in place of, core contextual values. Accordingly, the proposed framework produces a decision path of "text signal extraction – criteria construction – weighting – ranking – management interpretation". This path operationalizes the framework by specifying how unstructured textual evidence can

be systematically transformed into comparable decision variables and used to support transparent prioritization and reproducible decision-making.

The descriptions of the Sister Festival in county chronicles, monographs, and journals tend to follow a structured narrative pattern, including origin, overview, legend, scene description, and cultural meaning. This consistency contributes to the stability of representational continuity, which serves as one of the core decision criteria in the framework. Discourse evaluation varies across sources, with monographs, journals, annals, and newspapers exhibiting different evaluative tendencies shaped by their institutional roles. Individual discourse sources tend to emphasize descriptive evaluation, while organizational sources are more likely to present evaluative descriptions. These differences contribute to variations in participant representation and institutional visibility, which are incorporated into the corresponding decision criteria. The discourse evocation is less relevant in books, county chronicles, and web texts, which have obvious calling functions that are generally short and appear at the end. The author's intention is clearly expressed and plays a role in promoting and propagating the Sister Festival. Discourse thinking is the least relevant of the themes used because multiple discourses are more about narrative communication and expression.

In terms of emotional expression, monographs and journals tend to focus on investigative analysis with limited emotional content, while books, newspapers, and web texts often include stronger subjective expressions reflecting cultural identification and local attachment. Although predominantly positive, such expressions are unidirectional and do not capture audience reception. Within the decision framework, these patterns are interpreted as part of identity-related salience and communication intensity, contributing to the evaluation of cultural representation and public engagement. Taken together, these patterns demonstrate that cultural identity is not expressed through a single dominant discourse mode, but is collectively constructed through the interaction of descriptive stability, evaluative framing, and emotionally resonant narratives across discourse sources. From a systemic functional perspective, cultural identity is operationalized through recurring discourse functions, evaluative positioning, and narrative roles distributed across discourse sources rather than through explicit ideological statements. These dimensions are explicitly incorporated into the decision framework as indicators of identity-related salience and representational stability, enabling the translation of qualitative discourse patterns into structured inputs for decision support.

DISCUSSION

The textual representation of the Hmong Sister Festival involves a diversified configuration of actor groups, including both individuals and organizations. From a decision-system perspective, this multi-participant structure can be interpreted as a heterogeneous information environment in which different actors contribute distinct but complementary textual signals. Local cultural practitioners function as core signal generators, while administrative bodies, industry associations, residents, researchers, news media, tourists, and external scholars introduce additional layers of standardized representation, external evaluation, and institutional framing. Such a configuration is characteristic of complex decision environments, where structured outputs must be derived from multi-source and partially inconsistent textual inputs.

The specific elements of textual representation in the Hmong Sister Festival are examined through the analysis of key keywords. The clustering of keywords indicates that the textual system follows a stable cognitive framework of iconic name, time, place, participants, activity content, the items involved and deeper content. Historical analysis of the keywords shows that the representation in each historical development stage is unevenly distributed. The overall textual capacity has increased substantially, and the richness of vocabulary has increased marginally, but repetition of core vocabulary is still the main feature. Overall, the self-built corpus indicates that the textual representation of the research object is highly consistent across actor groups. For decision analytics, this consistency is important because it supports the construction of stable criteria from repeated textual patterns while retaining sensitivity to temporal variation, particularly in relation to continuity of representation and identity-related salience.

The collocation before the node is simple and clear, enabling an intuitive preliminary understanding through textual combination. The collocation after the node has many combinations and richer content, and the basic function of serving different media presents distinctive features. The overall word cluster develops through a basic understanding of ethnic cultural inheritance. The reduction of some combinations and the emergence of new combinations reflect historical changes in communicative demand. Taken together, these collocational patterns indicate the emergence of a structured knowledge system encompassing multiple interrelated dimensions, including identity signals, spatiotemporal framing, cultural attributes, evaluative positioning, and modern festival management. This structure shows how unstructured textual data can be reorganized into engineering-relevant knowledge components for decision support and mapped onto interpretable decision criteria within the proposed framework.

The quantitative feature analysis of multi-discourse vocabulary revealed that the characteristic words of monographs and journals, books, and county chronicles were similar, which reflects the authentic reproduction of local cultural expressions and the secondary interpretation and evaluation based on such expressions. Greater similarity and more repetition indicate a consistent expression of the same phenomenon and reflect the effectiveness of inheritance of the original culture, but also provide a practical advantage for computational modeling because recurring lexical templates improve the stability of feature extraction. These patterns contribute to the robustness of criteria construction, particularly in relation to representational stability and institutional visibility within the decision model.

From another perspective, the similarity of multiple discourses is very strong, with a relatively consistent cognition of the Hmong Sister Festival. In short, the media characteristics of multi-subject discourse are prominent, with discourse communication exhibiting co-occurrence relationships. This reflects the consistency and shared nature of cultural inheritance and transmission, while also showing that multi-source textual data can be fused into a coherent analytical basis for decision support. With the participation of multiple discourses, the overall discourse is dominated by a descriptive function, and descriptive evaluations or evaluative descriptions are derived. There are a few functions of emotion, inspiration, and thinking. Overall, this functional distribution demonstrates that stable descriptive structure is the primary mechanism through which heterogeneous textual sources can be transformed into comparable decision criteria. This supports the operationalization of the framework by ensuring that discourse-derived features can be consistently translated into structured inputs for multi-criteria decision analysis.

CONCLUSION

In contrast to macro-level historical narratives of Hmong history, the Sister Festival functions as a micro-level textual environment characterized by fragmented and dispersed records across different text types. Such micro-level textual systems provide valuable insight into everyday cultural practices, local meanings, and socially embedded forms of representation, but they also constitute a challenging data source for decision support because the information is heterogeneous, distributed, and only partially structured. Rather than pursuing the objectivity of historical events in a traditional sense, this study treats historical records as analyzable textual evidence that can be converted into structured decision inputs. By examining how the Sister Festival is described, framed, and evaluated across multiple textual sources, the analysis reveals how micro-level textual patterns contribute to the articulation of cultural value and identity within broader historical

contexts. These patterns are further operationalized within the proposed framework as measurable indicators and decision criteria, enabling the transformation of qualitative textual evidence into structured inputs for decision-making. [19]. At the same time, the study demonstrates that these patterns can be computationally organized into a transparent and reproducible decision-support framework, enabling the systematic transformation of qualitative textual insights into structured analytical inputs.

Hardt-Mautner.G [32] pointed out that corpus analysis functions can effectively break the boundary between quantitative and qualitative research. The self-built corpus enables the integration of large volumes of Sister Festival data and allows structured analysis across pictures and texts; it presents the basic situation of the textual system in a simplified way and addresses otherwise fragmented research problems. The structural analysis reveals core representational features and provides explanatory support for detailed textual interpretation. More importantly, the interdisciplinary use of linguistic analysis technology demonstrates that unstructured textual data can be transformed into measurable, comparable, and rankable indicators. This provides a new attempt not only for festival research, but also for data-driven decision modeling in knowledge-intensive contexts.

Individuals and organisations of different identities, regions, and professions participate in the textual construction of traditional ethnic minority festivals diachronically and synchronically, which is a manifestation of multiple actors participating in social activities and social practice. The difference between individual and organisational sources is that the former often reflects a single communication identity, while the latter may combine multiple institutional identities, making the analysis more complex. Therefore, the core status identities of indigenous cultural practitioners were identified in the research. Their textual expressions, perceived as authentic, often influence others or are reused by other subjects. The media, tourists, experts, and scholars from outside the community compare and contrast what they see, thereby producing different representations. Such expressions fully reflect the fact that textual representation is constructive and shapes social subjects, social relations, knowledge, and belief systems [54-57]. From an operational perspective, this complexity is precisely what makes the case relevant to complex decision environments: multiple actors generate heterogeneous inputs, and the proposed framework provides a way to structure them into a transparent analytical process (Figure 5).

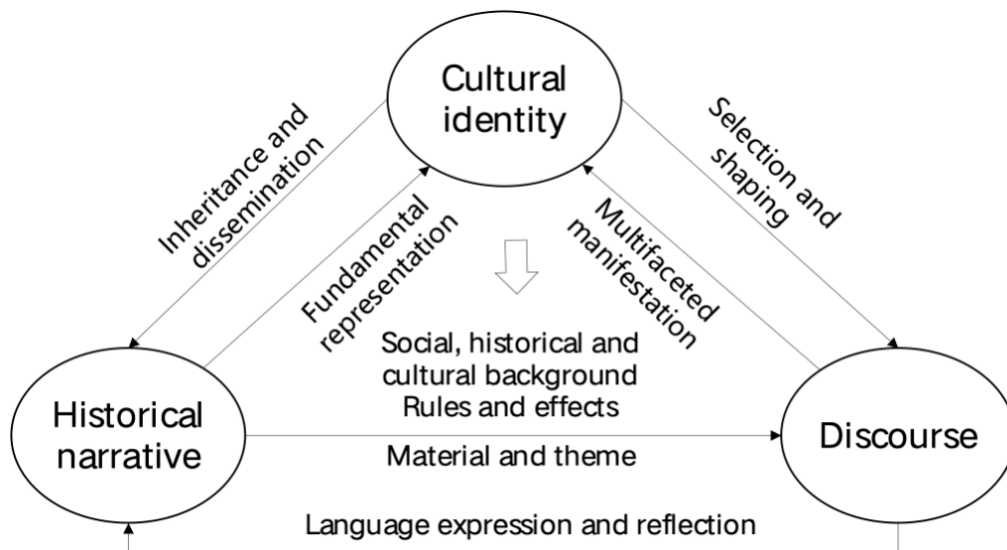


Figure 5. The relationship between historical narrative, discourse and cultural identity

Historical narration serves as a foundational source of material and thematic content for textual representation. Textual representation is the mechanism through which historical materials are articulated and conveyed. Cultural identity is a crucial component of this process because it underpins both historical narration and textual representation and influences their selection and shaping. In this study, these relationships are not only interpreted conceptually but also operationalized analytically: historical narration provides textual input, text analytics extracts structured features, and identity-related signals become one dimension within a broader decision-support model. In this sense, the framework connects historical narration, textual representation, and cultural identity to a reproducible process of evidence generation, through which qualitative historical materials are transformed into structured indicators and decision-relevant analytical inputs for management and policy analysis.

Study limitations and future research

The present study adopts a corpus-based discourse analytical approach that focuses on the production and representation of discourse rather than on interactive reception processes. While this design enables systematic comparison of discourse patterns across text types and historical periods, it does not capture how audiences interpret, negotiate, or rearticulate these discourses in practice. From a decision-support perspective, this limitation implies that the current framework primarily reflects production-side textual signals, while user perception and feedback are not yet incorporated into the decision criteria or evaluation process. Future research should aim to address this gap by explicitly investigating audience reception, for example, by eliciting

reader responses to representative texts, conducting surveys on discourse perception, or collecting audience comments from online platforms. Integrating such data would allow the incorporation of reception-based indicators into the decision model, thereby improving the robustness and completeness of decision criteria, particularly in relation to communication effectiveness and public engagement. At the methodological level, clarifying and modeling discourse functions, communication pathways, and interaction dynamics should be a priority for future work. As interactive discourse analysis remains an emerging research area, its practical feasibility and analytical validity require continued empirical exploration.

This study's approach is innovative within the limitations of the research on traditional festivals of ethnic minorities. The use of historical narratives based on Western historiography and philosophy to understand Chinese issues is debatable because of cross-cultural adaptability. Kragh [17] pointed out that White's narrative method of topology belongs to the refraction of mythic consciousness. Therefore, using Western theories to analyse China's issues may not fully capture China's unique historical events and cultural background, potentially overlooking distinct theories and perspectives in Chinese history and philosophy. This approach could result in an overly Westernised interpretation of Chinese historical phenomena. Future studies should explore China's specific historical events and cultural contexts, presenting local perspectives through comprehensive literature research and analysis of historical data. It is crucial to emphasise the unique theoretical viewpoints in Chinese historiography and philosophy, such as the 'historiography system' and the 'historiography of righteousness', integrating them into the framework of historical interpretation. This research can also determine which factors apply to Western thought to enhance the possibility of White's historical narration from the outside to the inside. White's historical narrative theory has a relatively mature analytical framework. Future research based on existing data can go deeper into the typical text to conduct concrete content analysis, and the deduction process will simultaneously test the practical applicability of historical narrative thought. Collaboration and dialogue with Chinese scholars are essential to ensure that research methods and interpretations remain true to the core of Chinese culture and concepts. These efforts will enhance the cultural adaptability and depth of understanding in historical research.

Corpus analysis has inherent limitations, and the temporal scope of the data included largely influences its results. Subject-related data evolve significantly over time, and narrative data from specific time points may only constitute a relatively limited proportion of the overall dataset, potentially leading to these aspects being overlooked or underestimated in the analysis. For example, during the COVID-19 pandemic, the case study

site in mainland China was less affected by the outbreak. Between 2019 and 2021, only the 2020 festival was suspended, resulting in fewer records of festivals impacted by the epidemic. In addition, due to the lag in recording cultural events and the influence of cultural psychology, people tend to focus more on the past and present happy moments. At the same time, unpleasant memories of the epidemic are easily diluted. Thus, specific points in time may appear insignificant in the broader historical narrative.

During the research, it was found that the absence of the festival in the year it was not held remained a significant topic of discussion among participants. Everyone showed great enthusiasm when it was held normally in the second year. However, this information was obscured by a large volume of other data and was not highlighted prominently. Furthermore, if future research enables corpus analysis followed by question exploration, a corpus can be built for more quantitative research. For example, keyword co-occurrence analysis helps researchers better understand research trends and hot topics in a particular field. Part-of-speech analysis reveals the functional and semantic changes of words in different contexts, enhancing our understanding of the structure and content of the text. Part-of-speech sentiment analysis combines part-of-speech tagging with sentiment analysis techniques to grasp emotional expressions better in text. These in-depth analyses, based on the descriptive analysis in this paper, serve the research problems being addressed and suggest future research directions. Furthermore, different software can be used for reference in the corpus analysis. The present version has a limited effect on Chinese analysis, and the software analysis is only a quantitative analysis process. 'Quantitative findings need to be interpreted according to existing theories'[58-59], and we must select representative texts for in-depth discourse analysis to form new theories.

As a representative traditional ethnic minority festival in the region, the Hmong Sister Festival case provides an empirically grounded example for examining how cultural heritage data can be structured for decision support. While the analysis highlights issues such as cultural preservation, diversity, and identity formation, the study primarily demonstrates how these dimensions can be operationalized within a corpus-driven analytical framework. Nevertheless, the inherent limitations of a single-case study remain. These include potential sampling bias, as the selected corpus may not fully capture the diversity of broader Miao cultural practices, and the possibility of interpretation bias influenced by researchers' perspectives. From a decision-support perspective, these limitations may affect the generalizability and robustness of the constructed criteria and ranking results. Future research should therefore extend the framework to multiple cases and incorporate cross-context validation to enhance model reliability. In addition, as a demonstration case for text-

based decision modeling, the study shows how localized narrative data can be systematically converted into structured analytical evidence, forming measurable indicators and decision-relevant inputs. Although broader industrial validation remains a direction for future research, the proposed approach establishes a replicable analytical pathway for transforming unstructured textual data into decision-support outputs.

Beyond its theoretical contributions, this study provides practical insights for decision-making in cultural heritage management, tourism planning, and policy design. By transforming unstructured textual records into measurable analytical structures, the proposed corpus-driven framework enables heritage managers, tourism planners, and policymakers to derive structured evidence for resource allocation, risk mitigation, and strategic planning. The analytical workflow is applicable to other knowledge-intensive domains that rely on heterogeneous textual data, such as smart manufacturing knowledge management, documentation-intensive governance, and decision analytics platforms. In this sense, the study demonstrates how text analytics can be integrated into applied decision systems, thereby strengthening its alignment with the interdisciplinary aims of DMAME.

Author Contributions

Conceptualization – Chanjuan Wang; methodology – Chanjuan Wang; formal analysis – Jing Zhang; investigation – Jing Zhang; resources – Chanjuan Wang; writing-original draft preparation – Chanjuan Wang, Guangquan Dai, Xin Xu; writing-review and editing – Chanjuan Wang, Jing Zhang; visualization – Guangquan Dai; supervision – Xin Xu. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This work was supported by the National Natural Science Foundation of China under Grant number 42171228.

Acknowledgements

Not applicable.

REFERENCES

- [1] Getz D, Page SJ. *Event Studies: Theory, Research and Policy for Planned Events* (4th ed.). Routledge; 2019. doi: 10.4324/9780429023002.

- [2] Chan CS. Folklore without a folk: questions in preserving the Marinduque Moriones heritage. *Int. J. Herit. Stud.* 23 (1)(2017) 29-40. doi: 10.1080/13527258.2016.1232746.
- [3] Knox D. Spectacular tradition - Scottish folksong and authenticity. *Ann. Touris. Res.* 35 (1)(2008) 255-273. doi: 10.1016/j.annals.2007.07.008.
- [4] Meretse AR, Mykletun RJ, Einarsen K. Participants' benefits from visiting a food festival - the case of the Stavanger food festival (Gladmatfestivalen). *Scand. J. Hosp. Tour.* 16 (2)(2016) 208-224. doi: 10.1080/15022250.2015.1108865.
- [5] Shen SY. Intention to revisit Traditional Folk Events: A case study of Qinhuai Lantern Festival, China. *Int. J. Tour. Res.* 16 (5)(2014) 513-520. doi: 10.1002/jtr.1949.
- [6] Cobb MC. "For a While They Live a Few Feet Off the Ground": Place and cultural performance at the Walnut Valley Festival. *J. Contemp. Ethnogr.* 45 (4)(2016) 367-395. doi: 10.1177/0891241614568192.
- [7] Rossetti G, Quinn B. Understanding the cultural potential of rural festivals: A conceptual framework of cultural capital development. *Journal of Rural Studies.* 86 (2021) 46-53. doi: 10.1016/j.jrurstud.2021.05.009.
- [8] Perry B, Ager L, Sitas R. Cultural heritage entanglements: festivals as integrative sites for sustainable urban development. *Int. J. Herit. Stud.* 26 (6)(2020) 603-618. doi: 10.1080/13527258.2019.1578987.
- [9] Liao Z, Dai G. Inheritance and dissemination of cultural collective memory: An analysis of a traditional festival. *SAGE Open.* 10 (1)(2020). doi: 10.1177/2158244020901601.
- [10] Xu X, Wang SY. The evolution of intangible CH digital resources: The case of the Qingming Festival. *Libr. Trends.* 69 (1)(2020) 201-222. doi: 10.1353/lib.2020.0020.
- [11] Clair RP, Carlo S, Lam C, Nussman J, Phillips C, Sánchez V, Schnabel E, Yakova L. Narrative theory and criticism: An overview toward clusters and empathy. *Review of Communication.* 14 (1)(2014) 1-18. doi: 10.1080/15358593.2014.925960.
- [12] Zavadskas, E. K., J. Antucheviciene, and S. H. Zolfani. 2017. "Multi-Criteria Decision-Making (MCDM) for the Assessment of Sustainability: A Review of the State of the Art and Future Directions." In *Springer Series in Reliability Engineering*, 1-26. Springer. doi:10.1007/978-3-319-48454-3_1z
- [13] Roth PA. Hayden White in philosophical perspective: Review essay of Herman Paul's *Hayden White: The historical imagination*. *Philos. Soc. Sci.* 44 (1)(2014) 102-111. doi: 10.1177/0048393113480609.
- [14] Elliott J. *Using narrative in social research: Qualitative and quantitative approaches*. Sage; 2005.
- [15] Rusen J. A turning point in theory of history: The place of Hayden White in the history of Metahistory. *Hist. Theory.*

- 59 (1)(2020) 92-102. doi: 10.1111/hith.12147.
- [16] White H. *Metahistory: The historical imagination in nineteenth-century Europe*. jhu Press; 2014.
- [17] Kragh UT. The Refraction of White: The Primary Colors of Hayden White's Tropological Theory Of Discourse. *Hist. Theory*. 60 (4)(2021) 82-99. doi: 10.1111/hith.12238.
- [18] Tsytsarau, M., and T. Palpanas. 2012. "Survey on Mining Subjective Data on the Web." *Data Mining and Knowledge Discovery* 24 (3): 513-552. doi:10.1007/s10618-012-0263-9
- [19] Domanska E. Hayden White and liberation historiography. *Rethink. Hist.* 19 (4)(2015) 640-650. doi: 10.1080/13642529.2014.959361.
- [20] Dean CJ. *Metahistory: The historical imagination in nineteenth-century Europe*, by Hayden White. *Am. Hist. Rev.* 124 (4)(2019) 1337-1350. doi: 10.1093/ahr/rhz863.
- [21] Ahlskog J. Michael Oakeshott and Hayden White on the practical and the historical past. *Rethink. Hist.* 20 (3)(2016) 375-394. doi: 10.1080/13642529.2016.1153320.
- [22] Vizan I. Americans versus Japanese Americans: a new historicist perspective of the World War ii period. in: *International Conference on Literature, Discourse and Multicultural Dialogue*. Arhipelag Xxi Press; 2015. p. 250-256.
- [23] Fonfarova V. A star-shaped crossroad: From (counterfactual) Historiography to Historiographic Metafiction. in: *5th International Conference on Anglophone Studies*. Tomas Bata Univ Zlin; 2013. p. 201-212.
- [24] Ali MS. Marking time and writing histories. *Hist. Theory*. 60 (2)(2021) 271-291. doi: 10.1111/hith.12206.
- [25] Sastrawan WJ. Temporalities in Southeast Asian Historiography. *Hist. Theory*. 59 (2)(2020) 210-226. doi: 10.1111/hith.12155.
- [26] Dowling ME. Romancing Russia: The effect of emplotment on the representation of a nation. *Rethink. Hist.* 24 (3-4) (2020) 523-542. doi: 10.1080/13642529.2020.1832290.
- [27] Gee JP. *An Introduction to Discourse Analysis: Theory and Method*. Foreign Language Teaching and Research Press, Routledge Press; 2000. doi: 10.4324/9781315819679.
- [28] Widdowson HG. Discourse analysis: a critical view. *Language and literature*. 4 (3)(1995) 157-172. doi: 10.1177/096394709500400301.
- [29] Huang G. *Summary of Discourse Analysis*. Hunan Education Press; 1988.
- [30] De Beaugrande R. The story of discourse analysis. *Discourse as structure and process*. (1997) 35-62. doi:

10.4135/9781446221884.n2.

- [31] Carvalho A. Media(ted) discourse and society: Rethinking the framework of Critical Discourse Analysis. *Journal. Stud.* 9 (2)(2008) 161-177. doi: 10.1080/14616700701848162.
- [32] Hardt-Mautner G. 'Only connect': Critical discourse analysis and corpus Linguistics UCREL Technical Paper 6. University of Lancaster; 1995.
- [33] Mu JF, Ma R. A CiteSpace-based analysis of the application of Critical Discourse Analysis in news discourse. *Discourse Commun.* (2022) 23. doi: 10.1177/17504813211070670.
- [34] Yuan ZM, Leng TY, Wang H. Understanding National Identity Construction in China-ASEAN Business Discourse. *SAGE Open.* 12 (1)(2022) 9. doi: 10.1177/21582440211069385.
- [35] Zhang WM. Analysis of 'proactive discourse' in the construction of local community by the media during the 2013 Tasmanian bushfire in Australia. *Journalism.* 22 (6)(2021) 1467-1483. doi: 10.1177/1464884918816249.
- [36] Ye, B. H., H. Q. Zhang, and J. J. Y. Ye. 2015. "Understanding the Impact of Festival Attributes on Perceived Value, Satisfaction and Revisit Intention: A Study of the Macau Food Festival." *International Journal of Contemporary Hospitality Management* 27 (7): 1618-1640. doi:10.1108/IJCHM-03-2014-0133.
- [37] Rutherford J. *Identity: community, culture, difference.* Lawrence & Wishart; 1990.
- [38] Hall S. Culture and power. *Radical Philosophy* 86.27 (1997): 24-41.
- [39] Katz E, Lazarsfeld PF, Roper E. *Personal influence: The part played by people in the flow of mass communications.* Routledge; 2017. doi: 10.4324/9781315126234.
- [40] Hamkins S. Narrative psychiatry and family therapy: Promoting resilience and collaboration with persons with intense mental and emotional experiences at risk of causing violence. *Australian and New Zealand Journal of Family Therapy.* 40 (3)(2019) 295-307. doi: 10.1002/anzf.1380.
- [41] Bourassa KJ, Allen JJ, Mehl MR, Sbarra DA. The impact of narrative expressive writing on heart rate, heart rate variability, and blood pressure following marital separation. *Psychosomatic Medicine.* 79 (6)(2017) 697-705. doi: 10.1097/PSY.0000000000000475.
- [42] Meca A, Sabet RF, Farrelly CM, Benitez CG, Schwartz SJ, Gonzales-Backen M, Lorenzo-Blanco EI, Unger JB,

- Zamboanga BL, Baezconde-Garbanati L. Personal and cultural identity development in recently immigrated Hispanic adolescents: Links with psychosocial functioning. *Cultural diversity and ethnic minority psychology*. 23 (3)(2017) 348. doi: 10.1037/cdp0000129.
- [43] Liu Y, Vrontis D, Visser M, Stokes P, Smith S, Moore N, Thrassou A, Ashta A. Talent management and the HR function in cross-cultural mergers and acquisitions: The role and impact of bi-cultural identity. *Human Resource Management Review*. 31 (3)(2021) 100744. doi: 10.1016/j.hrmr.2020.100744.
- [44] Usborne E, De La Sablonnière R. Understanding my culture means understanding myself: The function of cultural identity clarity for personal identity clarity and personal psychological well-being. *Journal for the Theory of Social Behaviour*. 44 (4)(2014) 436-458. doi: 10.1111/jtsb.12061.
- [45] Schwartz SJ, Zamboanga BL, Rodriguez L, Wang SC. The structure of cultural identity in an ethnically diverse sample of emerging adults. *Basic and Applied Social Psychology*. 29 (2)(2007) 159-173. doi: 10.1080/01973530701332229.
- [46] Hu FW, Wang P, Li LJ. Psychometric structure of the Chinese multiethnic adolescent cultural identity questionnaire. *Psychological assessment*. 26 (4)(2014) 1356. doi: 10.1037/a0037690.
- [47] Chao GT, Moon H. The cultural mosaic: A metatheory for understanding the complexity of culture. *Journal of applied psychology*. 90 (6)(2005) 1128-1140. doi: 10.1037/0021-9010.90.6.1128.
- [48] Sui M. Ethnic Selective Exposure: A Test of Cultural-Identity Based Media Selectivity Theory. *Mass Communication and Society*. 26 (2)(2023) 227-251. doi: 10.1080/15205436.2021.2007266.
- [49] Avila-Saavedra G. Ethnic Otherness Versus Cultural Assimilation: U.S. Latino Comedians and the Politics of Identity. *Mass Communication and Society*. 14 (3)(2011) 271-291. doi: 10.1080/15205431003650478.
- [50] Teubert W. My version of corpus linguistics. *Int. J. Corpus Linguist*. 10 (1)(2005) 1-13. doi: 10.1075/ijcl.10.1.01teu.
- [51] Ping Y. Representations of the 2014 Hong Kong protests in journalistic translation: A corpus-based critical framing analysis of Chinese and English news coverage. *Journalism*. 23 (7)(2022) 1509-1529. doi: 10.1177/14648849221074556.
- [52] Hu X. *Corpus Stylostatistics: Methods and Applications*. Foreign Language Teaching and Research Press; 2021.
- [53] Halliday MAK, Matthiessen CM, Halliday M, Matthiessen C. *An introduction to functional grammar*. Routledge;

2014. doi: 10.4324/9780203783771.

- [54] Liu B, Zhang G, Peng G, Zhao L, Guo Y. Postmodernism and history. *Historiography Quarterly*. (02)(2004) 4-17.
- [55] Taylor C. The affordances of metaphor for diachronic corpora & discourse analysis: water metaphors and migration. *Int. J. Corpus Linguist*. 27 (4)(2022) 451-479. doi: 10.1075/ijcl.22004.tay.
- [56] Mercan Küçükakın P, Engin-Demir C. A critical analysis of discourses on gender equality in education: the case of Turkey. *Gend. Educ*. 34 (8)(2022) 940-956. doi: 10.1080/09540253.2022.2050679.
- [57] Fairclough N. *Discourse and Social Change* Blackwell Publishers; 1993.
- [58] Fitzgerald S. The discursive representation of the International Baccalaureate in the global press: a computer-assisted discourse analysis. *Discourse: Studies in the Cultural Politics of Education*. (2022) 1-18. doi: 10.1080/01596306.2022.2104811.
- [59] Baker P, Gabrielatos C, Khosravini M, Krzyzanowski M, McEnery T, Wodak R. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Int. J. Soc. Res*. 19 (3)(2008) 273-306. doi: 10.1177/09579265080888962.