Global Journal of Advanced
Research and Reviews
Journal homepage: https://gsjournals.com/gjarr/
ISSN: 2980-423X (Online)

(REVIEW ARTICLE)

GENERATIVE-AI-ORIENTED MULTIMODAL LINGUISTIC LANDSCAPE: EMPIRICAL EVIDENCE FROM YANCHENG CULTURAL-TOURISM SITES

Qin Liu *

School of Foreign Languages, Yancheng Teachers University.
* Corresponding Author

Global Journal of Advanced Research and Reviews, 2026, 04(04), 043–047

Publication history: Received on 22 September 2026; revised on 30 October 2026; accepted on 02 November 2026

Article DOI: https://doi.org/10.58175/gjarr.2026.4.4.0088

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Traditional linguistic landscape research largely concentrates on the semiotic analysis of physical signage, while dynamic digital linguistic symbols generated by large language models remain understudied. Taking three types of cultural-tourism spaces in Yancheng as field sites, this study collects three corpora: physical signs, digital navigation interfaces and AI-generated texts, and constructs a local multilingual cultural-tourism corpus. Controlled experiments are conducted to compare the outputs of a general large language model and a local-knowledge RAG model on sea-salt cultural proper terms and intangible heritage narratives. Field data show that the physical linguistic landscape of Yancheng tourism follows a pattern of "dominant Chinese, limited English coverage and fragmented Korean distribution", and translations of historical cultural concepts commonly suffer from semantic simplification and erosion of cultural information. Model results reveal that general large models exhibit high hallucination rates when processing low-resource local cultural concepts such as Zaohu and Tuanjian; after connecting to the local corpus knowledge base, term translation accuracy and cultural fidelity improve. This study identifies a prominent semiotic disjuncture between digital and physical linguistic landscapes in tourism settings: generative AI brings instrumental empowerment alongside risks of cultural drift. By incorporating algorithmically generated texts into the multimodal linguistic landscape framework, this paper provides an empirical case for research on local cultural-tourism linguistic signs.

Keywords: Multimodal linguistic landscape; Generative artificial intelligence; Cultural-tourism space; Cultural proper terms; Retrieval-augmented generation; Yancheng

1 INTRODUCTION

Since Landry & Bourhis established the linguistic landscape framework, scholarship in this field has long centered on static physical signs in public space. With the advancement of smart tourism, AR guides, AI customer service and dynamic interpretive texts have entered cultural-tourism spaces continuously, forming an evolving virtual linguistic landscape. These algorithm-generated signs are free from material constraints and long production cycles of physical signage yet inherit knowledge biases and cultural stereotypes embedded in pre-trained models.

Existing "AI + cultural tourism" research prioritizes technical development and system deployment. Few sociolinguistic studies empirically describe and evaluate AI-produced linguistic landscape texts. Scholars including Shohamy and Gorter have argued that linguistic landscape research should move beyond static signs to embrace fluid, shifting

linguistic resources; nevertheless, algorithmically generated virtual landscapes have not yet received sufficient empirical scrutiny. Yancheng boasts coastal World Natural Heritage, sea-salt culture and Huai opera intangible heritage, with a steadily growing inbound tourist market. The city's ongoing development of a multilingual smart tourism infrastructure offers an ideal field site for this investigation.

Drawing on fieldwork and controlled large-model experiments, this paper addresses two research questions:

- RQ1: What differences exist in code distribution and textual features between physical and digital virtual linguistic landscapes in Yancheng's cultural-tourism spaces?
- RQ2: What linguistic limitations emerge when general large language models generate linguistic landscape texts for low-resource regional cultural concepts such as sea-salt and intangible heritage items?

This study focuses on quantitative assessment of code distribution patterns and model performance across physical and digital linguistic landscapes in Yancheng.

2 DATA AND EXPERIMENTAL DESIGN

2.1 Field Corpus Collection

Fieldwork covers three categories of tourism sites: (1) coastal World Heritage corridors (Tiaozini, Chinese Milu Park, Red-crowned Crane Wetland); (2) urban cultural core zones (Chuanchang River Sea-Salt Cultural Belt, KK-Park, European-style Flower Street); (3) integrated tourism complexes (Holland Flower Sea, Dazong Lake, Huai Opera Town). Following Backhaus's individual sign sampling method, 842 valid physical signs are collected. Meanwhile, 317 publicly available digital navigation texts from scenic mini-programs and smart guide platforms are gathered. A test set of 62 cultural proper terms is built, covering sea-salt historical terminology and Huai opera intangible heritage concepts.

A dedicated sea-salt subcorpus, built by filtering the 842 signs for those containing at least one sea-salt or Huai opera cultural proper term, is retained for a finer-grained semiotic analysis; the present paper, however, reports the overall code distribution and the model-experiment results.

Coding dimensions include sign type (top-down official signage / bottom-up commercial signage), language combination, translation strategy for cultural proper terms and modality (text-only / image-text multimodal).

2.2 Controlled AI Experiment Setup

The two experimental groups adopt identical prompts for cultural-tourism interpretive text generation. The control group uses an open-source general large model without access to the local knowledge base. The experimental group adopts a retrieval-augmented generation (RAG) architecture, with the retrieval repository built from the field-collected Yancheng tourism corpus. Evaluation metrics include linguistic accuracy, cultural fidelity and term adaptation; hallucination frequency and definition correctness of key concepts are recorded. Two linguistically trained annotators conduct independent coding, with disagreements resolved by a third coder.

3 KEY FINDINGS FROM FIELDWORK

3.1 Code Distribution of the Physical Linguistic Landscape

Table 1: Code distribution of physical signage

Code type	Frequency	Proportion
Monolingual Chinese	517	61.4%
Chinese–English bilingual	272	32.3%
Multilingual with Korean	40	4.8%
Other / unclassified	13	1.5%
Total	842	100%

Among the 842 physical signs, 517 (61.4%) are monolingual Chinese, 272 (32.3%) are Chinese–English bilingual, 40 (4.8%) contain Korean, and the remaining 13 (1.5%) belong to other mixed-language or unclassified combinations.

Korean signs cluster exclusively in the Korean-style block of KK-Park; few Korean markers appear at other sites, and no standardized trilingual signage at scale is observed.

Top-down official signs are dominated by Chinese–English bilingualism, with translation difficulties concentrated on historical cultural proper nouns. Foreign languages on commercial signs mostly perform decorative rather than informational functions. Ecological wetland labels achieve more consistent translation quality compared with sea-salt history and Huai opera heritage signs.

3.2 Linguistic Features of Translations for Cultural Proper Terms

Across the 62 sea-salt and heritage concepts, three recurrent translation problems emerge on physical signs. First, conceptual simplification: Zaohu and Tuanjian are rendered directly as salt-worker and salt-making, erasing institutional meanings of ancient salt household registration and collective production. Second, conceptual alienation: some Huai opera terms are transcribed purely in pinyin without glosses, preventing overseas visitors from establishing semantic connections. Third, inconsistent renderings: a single sea-salt concept receives 2–4 incompatible English variants across different sites.

Zaohu serves as a representative example: one exhibition label at the Sea-Salt Museum writes salt worker, reducing the institutionally registered salt household to a generic labourer; another panel simply uses the transliteration Zaohu with no supplementary explanation. Parallel divergent renderings hinder external readers from recognizing it as a distinctive salt administration concept.

Table 2: Examples of translations for cultural proper terms

Chinese concept	Common renderings on physical signs	Core issue
Zaohu	salt worker / Zaohu	Institutional identity reduced to labourer; transliteration without semantic compensation
Tuanjian	salt-boiling	Collective production organization meaning lost
Yanke	salt tax / salt tribute / salt duty	Three competing renderings with divergent fiscal connotations
Huaiju	Huai opera / Huaiiju	Some signs use pinyin only, no introduction to the opera genre

3.3 Semiotic Disjuncture between Physical and Digital Linguistic Landscapes

Eighty-seven of the 317 digital navigation texts (27.4%) demonstrate narrative inconsistency with matching physical signs: interpretive wording and proper-term translations on online platforms diverge from on-site signage. Physical signs have long renewal cycles and high revision costs, while online digital texts can be adjusted casually with low barriers. The two semiotic systems lack a unified benchmark, creating misaligned physical and virtual linguistic landscapes.

4 EXPERIMENTAL RESULTS OF MODEL TESTS

On the test set of 62 cultural proper terms, the general model (control) achieves correct definitions for 22 terms (35.5%) and produces hallucinations for 29 terms (46.8%). Faced with low-resource local concepts such as Zaohu and Yanke, the model defaults to generic salt-production knowledge and overlooks Yancheng-specific institutional histories; for Huai opera, it frequently conflates it with other southern Chinese opera traditions. The RAG experimental group reaches 48 correct definitions (77.4%) and only 8 hallucinatory outputs (12.9%). Concept generalization still occurs when retrieved context fragments are insufficient.

Table 3: Model performance on 62 cultural proper terms

Metric	Control group (general LLM)	Experimental group (local RAG)
Correct definition	22 (35.5%)	48 (77.4%)
Hallucination observed	29 (46.8%)	8 (12.9%)

A stable pattern emerges: even factually accurate AI outputs tend toward de-localized rewriting, reframing Yancheng-specific cases as generic tourism narratives and diluting regional cultural uniqueness. This tendency is invisible from accuracy metrics and requires close textual reading.

5 DISCUSSION

Linguistic landscape should not be reduced to visible physical signs. Algorithmically generated AR narration, AI dialogue and dynamic recommendations together constitute a consequential virtual linguistic landscape in tourism space. These virtual signs carry semiotic power equivalent to physical signage yet are harder to regulate and standardize.

The central tension of Yancheng's physical linguistic landscape lies not in insufficient linguistic diversity, but in weak capacity for cross-linguistic transfer of local cultural proper concepts. Adding foreign-language labels does not automatically achieve internationalization; preserving cultural semantics constitutes the core linguistic challenge of linguistic landscape translation.

The model experiments warn that general large models cannot be deployed directly to produce linguistic landscape texts for cultural tourism. Even grammatically fluent outputs may contain conceptual generalization, historical hallucination and erasure of local identity. Local RAG knowledge bases mitigate these risks, yet performance heavily depends on corpus coverage and quality. Importantly, RAG is only a technical remedy: it cannot fundamentally resolve the untranslatability of embedded cultural concepts, and the knowledge base itself relies on human semantic annotation of cultural proper terms.

This study has limitations. The test set is restricted to sea-salt and Huai opera terminology; AI transfer of wetland heritage concepts is not examined. Tourist subjective perceptions are not quantified. Future linguistic landscape research needs dedicated coding frameworks and evaluation criteria for algorithm-generated texts.

6 CONCLUSION

Fieldwork and controlled modelling jointly demonstrate that the linguistic landscape in the generative-AI era comprises static physical signs and dynamic algorithmic signs. Physical signage suffers from erasure of cultural proper-term meanings. Texts generated by general LLMs carry overlapping risks of hallucination, de-localization and cultural drift. Narrative rupture frequently separates physical signage from algorithmic virtual landscapes. Integrating AI-generated texts into linguistic landscape analysis enables a renewed understanding of the semiotic ecology of smart tourism spaces.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Funding Source

This research was supported by the Social Science Foundation of Yancheng City (No. 26skC192).

REFERENCES

[1] Al-Aziz, R., & Hassan, L. (2026). Linguistic landscape in the age of generative AI: Algorithmic signs and semiotic risk in smart heritage sites. Journal of Sociolinguistics, 30(2), 214–237.

[2] Chen, Y., & Wang, H. (2026). RAG for low-resource cultural terminology: Reducing LLM hallucination in regional heritage interpretation. Computational Linguistics and Chinese Language Processing, 31(1), 45–72.

[3] Gao, S., & Liu, M. (2025). Multimodal linguistic landscape and cultural untranslatability: Evidence from Chinese coastal heritage tourism. Language, Culture and Curriculum, 38(3), 289–306.

[4] Gong, J., & Zhang, L. (2025). De-localization in LLM-generated tourism texts: A corpus-based semiotic analysis. Discourse, Context & Media, 64, Article 100921.

[5] Huang, X., & Li, T. (2025). Ontological annotation of cultural proper terms: A remedy for translation loss in heritage signage. Journal of Multilingual and Multicultural Development, 46(7), 612–628.

[6] Jiang, R. (2026). Virtual linguistic landscape: AI interpretive texts as emergent semiotic resources in smart tourism. Social Semiotics, 36(4), 441–458.

[7] Li, S., & Zhao, Y. (2025). Hallucination and cultural bias in LLMs: Testing regional historical terminology. Journal of Computer-Assisted Language Learning, 38(5), 501–522.

[8] Lin, Q., & Wu, B. (2026). Complementary corpus sampling in paired linguistic landscape studies: Boundaries between primary dataset and thematic subsets. Linguistic Landscape, 12(1), 78–96.

[9] Liu, Z., & Fang, H. (2025). Code distribution and asymmetric multilingualism in Chinese second-tier tourism cities. International Journal of the Sociology of Language, 236, 117–140.

[10] Mendes, A., & Costa, P. (2026). Sociosemiotic loss in automated heritage translation: Beyond accuracy metrics. Sign Systems Studies, 54(1), 89–112.

[11] Landry, R., & Bourhis, R. Y. (1997). Linguistic landscape and ethnolinguistic vitality: An empirical study. Journal of Language and Social Psychology, 16(1), 23–49.

[12] Backhaus, P. (2007). Linguistic landscapes: A comparative study of urban multilingualism in Tokyo. Multilingual Matters.

[13] Shohamy, E., & Gorter, D. (Eds.). (2009). Linguistic landscape: Expanding the scenery. Routledge.

[14] Kress, G., & van Leeuwen, T. (2006). Reading images: The grammar of visual design (2nd ed.). Routledge.

[15] Lewis, P., Perez, E., Piktus, A., Petroni, F., Goyal, N., Blum, A., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.