

A Comparative Study on the Translation Quality of Eco-cultural Terms in Tourism Texts of Qinghai Province: An Empirical Analysis Based on Large Language Models

Chunli Yu¹, Hongyan Tuo¹

¹ School of Foreign Languages, Qinghai Normal University, Xining, China

Correspondence: Hongyan Tuo, School of Foreign Languages, Qinghai Normal University, Xining, Qinghai, 810008, China.

Received: May 4, 2026

Accepted: June 9, 2026

Online Published: June 22, 2026

doi:10.5430/elr.v15n1p34

URL: <https://doi.org/10.5430/elr.v15n1p34>

Abstract

This study conducts an empirical comparative analysis of translations generated by four mainstream large language models (ChatGPT, Google Gemini, DeepSeek, and Qwen) for 300 Qinghai ecotourism-related eco-cultural terms across four categories (Natural Landscapes, Ecological Resources & Conservation, Tourism Activities & Facilities, and Culture & Folk Customs). A multi-dimensional evaluation system covering accuracy, cultural connotation retention, terminological consistency, and readability is employed to assess each model's performance. The results show that Google Gemini achieves the highest overall performance, excelling particularly in accuracy for natural landscape and ecological resource terms, and demonstrating strong capabilities in cultural connotation retention and readability. Qwen ranks second, with notable strength in cultural connotation retention and balanced performance across other dimensions. DeepSeek follows, maintaining stable and consistent output with good terminological uniformity. ChatGPT offers acceptable readability and natural expression. Based on these findings, the study proposes targeted optimization strategies and constructs a reference translation corpus, aiming to support the international dissemination of Qinghai's ecotourism culture and to enrich empirical research on LLM-based translation in regional ecotourism contexts.

Keywords: Qinghai's Ecotourism; Eco-cultural Terms; Translation Quality; Large Language Models (LLMs); Empirical Study

1. Introduction

1.1 Research Background

Qinghai Province, located in the northwest of China, is a typical ecological conservation area with unique alpine ecological geography and diverse ethnic cultures such as Tibetan, Hui and Tu. Ecology occupies a fundamental position in Qinghai's development, which also dictates the quality of its long-term growth. As the core carrier of regional ecological and cultural communication, the Eco-cultural terms of Qinghai's ecotourism cover ecological geography, ethnic customs, ecological protection and tourism experience, and condenses the unique ecological value and cultural charm of the region (Li & Ng, 2025). With the rapid development of Qinghai's ecotourism industry and the deepening of international cultural exchanges, the accurate translation of these eco-cultural terms has become an important link to promote the international communication of Qinghai's ecotourism culture (Yang & Dai, 2022).

In the era of artificial intelligence, large language models (LLMs) have become an important auxiliary tool for translation practice with their powerful natural language processing and generation capabilities, and are widely used in the translation of tourism materials, including the translation of regional ecotourism vocabulary. However, due to the differences in model training data, algorithm logic and semantic understanding capabilities, different LLMs show obvious heterogeneity in the translation of professional and culturally-loaded vocabulary such as Qinghai's ecotourism Eco-cultural terms (Rosenberger & Kolchanka, 2025). At the same time, Qinghai's ecotourism Eco-cultural terms have the dual attributes of professional ecological terminology and ethnic cultural loaded words, with the characteristics of strong regionality, complex cultural connotations and high professional requirements (Li & Ng, 2025), which brings great challenges to the translation of LLMs. At present, there is a lack of targeted empirical research on the translation quality of Qinghai's ecotourism Eco-cultural terms by mainstream LLMs, and there is no

unified evaluation standard and optimal translation strategy for this type of vocabulary translation.

Against this background, this study selects four mainstream large language models with high application frequency in translation practice—Qwen, ChatGPT, Google Gemini, and DeepSeek—as the research objects, covering representative general-purpose LLMs from both China and the West. Taking 300 classified Qinghai eco-cultural terms as the research corpus, it conducts a multi-dimensional comparative analysis of translation quality. Current research has largely focused on the translation of tourism attraction signage or general texts, but there remains a lack of systematic comparison among different LLMs when it comes to region-specific terminologies with distinct ecological and cultural connotations, such as those of Qinghai. This study aims to fill this research gap and provide practical references for the application of LLMs in the translation of Qinghai's ecotourism vocabulary.

This study empirically examines mainstream LLMs' translation capacity for Qinghai ecotourism eco-cultural terms, analyzing the strengths and weaknesses of different models and putting forward targeted application suggestions. Practically, it guides relevant practitioners to rationally utilize LLM tools, improve translation efficiency and quality, and reduce improper translation; it also builds a reference corpus of 300 local characteristic terms to support Qinghai's ecotourism international publicity and cultural communication. Theoretically, this research enriches empirical studies on LLM translation quality evaluation in ecotourism, constructs a multi-dimensional evaluation system for regional cultural terms, explores LLMs' semantic understanding and generation mechanisms for culturally loaded vocabulary, and expands the research on natural language processing in cross-cultural translation, while providing empirical support for the optimized training of professional term translation in LLMs.

1.2 Research Questions

Based on the above research background and significance, this study puts forward three core research questions to be solved:

- 1) What are the overall performance differences of Qwen, ChatGPT, DeepSeek and Google Gemini in the translation of Qinghai's ecotourism Eco-cultural terms?
- 2) For the four categories of Qinghai's ecotourism Eco-cultural terms (Natural Landscapes, Ecological Resources & Conservation, Tourism Activities & Facilities, and Culture & Folk Customs), what are the respective advantages and deficiencies of the four LLMs in translation?
- 3) What targeted optimization strategies and application suggestions can be put forward for the translation of Qinghai's ecotourism Eco-cultural terms by LLMs based on the empirical research results?

1.3 Research Methodology

This study adopts a mixed research method combining quantitative and qualitative analysis, and the specific research methods include corpus method, empirical comparative method and multi-dimensional evaluation method.

Corpus Construction: a total of 300 domain-specific lexical items were adopted. These words were generated by an AI language model and divided into four fields, with 75 items allocated to each field. All entries were further examined and revised manually to ensure compliance with the research context.

Empirical Data Collection: The 300 vocabulary items in the corpus were respectively input into Qwen, DeepSeek, ChatGPT, and Google Gemini using a unified prompt. The original translation outputs of the four models were collected to form a four-group parallel translation corpus. The unified prompt is as follows: "Suppose you are an expert in eco-tourism and translation, please translate the following Chinese terms related to Qinghai eco-tourism into English with high accuracy. For proper nouns, local characteristic terms and ecological protection terminology, please ensure professional and consistent translation to meet academic research standards."

Multi-dimensional Quality Evaluation: Construct a translation quality evaluation system with four primary indicators (accuracy, cultural connotation retention, terminological uniformity and readability) and 12 secondary indicators, and introduce a third-party large language model to perform standardized quality assessment.

Quantitative and Qualitative Analysis: Use statistical methods to conduct quantitative analysis on the scoring results, compare the overall and category-based translation performance of the three models; conduct in-depth qualitative analysis on the typical translation errors and excellent translation cases of each model, and explore the causes of translation differences and the characteristics of vocabulary translation.

2. Literature Review and Theoretical Foundation

2.1 Domestic and Foreign Research Status

2.1.1 Research on LLM-based Translation Quality Evaluation

In recent years, with the rapid development of large language models, the research on LLM-based translation quality evaluation has become a hot spot in the field of translation studies. Foreign scholars have carried out a lot of empirical research on the translation capabilities of mainstream models such as ChatGPT, Google Gemini, GPT-4 and DeepSeek in different fields. For example, Al Rousan et al. (2025) conducted a comparative study on the translation quality of ChatGPT and human translators in the field of literary text translation, and found that the specific performance differences between ChatGPT and human translation in literary text rendering can provide important reference for the application of AI translation tools in cross-cultural literary communication. Kocmi and Federmann (2023) assessed GPT-4's multilingual translation performance across 20 languages and recognized its strong cross-lingual semantic comprehension, while highlighting its inadequacies in the consistent translation of polysemous words. As for DeepSeek, Li et al. (2024) conducted an empirical study on its translation performance in professional fields and found that it performs stably in cross-cultural translation, with moderate advantages in terminological uniformity.

Domestic research in China on LLM-based translation quality evaluation is mostly focused on the translation of professional texts in specific fields, such as medical, legal and tourism translation. For example, Wang (2025) conducted a comparative study on the translation quality of ChatGPT, Google Translate, and Bing Translator in the field of Uighur literary texts, and found that different translation systems have obvious differences in the translation of culturally sensitive literary content. Li & Chen (2024) explored the application effect of Google Gemini in the translation of ecological professional terminology, and pointed out that the model has advantages in the accuracy of professional terms due to the rich professional training data. Research on DeepSeek's translation capability (Li et al., 2024) shows that it performs stably in cross-cultural translation, with moderate advantages in terminological uniformity. However, the existing research has not yet involved the targeted study on the translation quality of Qinghai's ecotourism Eco-cultural terms by the four mainstream LLMs, and the research on the translation of regional ecotourism vocabulary with dual attributes of professionalism and cultural load is relatively scarce.

2.1.2 Research on Ecotourism Term Translation

Ecotourism term translation is an important branch of tourism translation research, and scholars have carried out relevant research from the perspectives of terminology standardization and cultural connotation transmission. Foreign scholars such as Cohen (1985) and Buckley (1985) pointed out in their research on tourism translation that the translation of ecotourism vocabulary should pay attention to the unity of professional accuracy and cultural adaptability, and avoid the loss of cultural connotations in the translation process. Hu Gengshen (2004) put forward the theory of ecological translation, pointing out that translation is a process of "adaptation and selection", and the translation of ecotourism vocabulary should adapt to the ecological and cultural context of the source language and the cognitive habits of the target language readers.

In the research on regional ecotourism vocabulary translation, scholars have carried out relevant research on the vocabulary translation of ecotourism in western China, Yunnan, Guizhou and other regions (Su, 2020; Meng, 2023). For example, Li Hong (2021) conducted a corpus-based study on the ecotourism translation in western China, and pointed out that the main problems of ecotourism vocabulary translation are non-uniform terminology and loss of cultural connotations. However, the research on Qinghai's ecotourism vocabulary translation is mostly limited to the discussion of individual vocabulary translation skills, and there is a lack of systematic and empirical research on the translation quality and standardization of Qinghai's ecotourism Eco-cultural terms by four mainstream LLMs.

2.2 Theoretical Foundation

Translation quality assessment constitutes an essential part of translation research. This study mainly draws on multi-dimensional evaluation theories for translation quality. Traditional evaluation approaches are largely rooted in Nida's (2001) equivalence theory, which takes semantic and pragmatic equivalence between source and target texts as core assessment criteria. Considering the characteristics of terminology translation, this paper constructs a four-dimensional evaluation framework by integrating Nida's functional equivalence theory (Nida, 2001), cross-cultural translation theories (Newmark, 1988), and Multidimensional Quality Metrics (MQM) (Lommel et al., 2014). Evaluation is conducted from four dimensions: linguistic accuracy, preservation of cultural connotation, terminology consistency and readability. This framework complements evaluations that merely focus on semantic equivalence, and facilitates comprehensive quality assessment of translations generated by the four large language

models (LLMs).

3. Research Design

3.1 Corpus Construction

A total of 300 domain-specific eco-cultural terms for Qinghai’s ecotourism were adopted in this study. These terms were initially generated by an AI language model and then manually examined and revised to ensure their compliance with the research context in terms of regional specificity, eco-cultural integration, practical applicability and category completeness. The final corpus is divided into four categories—Natural Landscapes, Ecological Resources & Conservation, Tourism Activities & Facilities, and Culture & Folk Customs—according to their semantic attributes, application scenarios and functional characteristics, with 75 terms allocated to each category. Specifically, Natural Landscapes covers terms describing Qinghai’s alpine ecological geography, landforms, hydrology and other natural ecological resources, such as “Sanjiangyuan (the source of the three great rivers)”, “Qinghai Lake Wetland”, and “Qilian Mountains”, etc. Ecological Resources & Conservation includes terms reflecting Qinghai’s unique biological resources and ecological protection measures, such as “Tibetan Antelope” and “Snow Leopard” and so on. Tourism Activities & Facilities comprises terms related to various ecotourism activities and supporting facilities, such as “Eco-tourism”, “Eco-homestay”, and “Viewing Platform”, covering tourism forms, accommodation services and facilities closely tied to the actual ecotourism experience. Culture & Folk Customs contains terms that reflect the unique ethnic cultures and folk customs of Qinghai’s minorities, such as “Tibetan Culture”, “Highland Barley Wine (Qingke Wine)”, and “Guozhuang Dance (Tibetan circle dance)”, vividly showing the cultural characteristics and living customs of ethnic groups like the Tibetan in Qinghai.

3.2 Selection of Research Tools

This study selects four mainstream large language models—Qwen, ChatGPT, Google Gemini, and DeepSeek—as the research tools. All four models are widely applied in Chinese-English translation tasks. They represent a balanced mix of Chinese-developed and Western-developed LLMs, with differences in Chinese semantic understanding, multilingual processing, and domain-specific knowledge coverage.

3.3 Construction of Translation Quality Evaluation System

Based on translation quality evaluation theory, ecological translation theory and the characteristics of Qinghai’s ecotourism Eco-cultural terms, this study constructs a four-dimensional multi-level translation quality evaluation system with four primary evaluation indicators and 12 secondary evaluation indicators. The evaluation system takes the 10-point scale as the scoring standard (0-3 points for poor, 4-6 points for general, 7-8 points for good, 9-10 points for excellent), and the specific indicators are as follows:

Table 1. Four-Dimensional Multi-Level Translation Quality Evaluation System

Primary Indicators	Secondary Indicators		Scoring Criteria
Accuracy (30%)	Lexical Semantic Accuracy		The translation accurately conveys the core semantic meaning of the source language vocabulary, without lexical distortion or error.
	Professional Term Accuracy		For professional ecological terms, the translation conforms to the international general professional expression, without professional errors.
	Geographical Accuracy	Name	For geographical names, the translation conforms to the official standardized translation of Qinghai Province, without misspelling or mistranslation.
Cultural Connotation Retention (25%)	Ethnic Connotation Retention	Cultural	The translation retains the unique ethnic cultural connotations of the source language vocabulary, without cultural loss or alienation.
	Ecological Connotation Retention	Cultural	The translation retains the ecological cultural connotations contained in the source language vocabulary, reflecting the ecological value of Qinghai.
	Contextual Adaptability		The translation adapts to the ecotourism cultural context of Qinghai, and the cultural expression is in line with the cognitive habits of English readers.
Terminological Uniformity (25%)	Polysemy Uniformity	Translation	For polysemous vocabularies, the translation of the same meaning in the same context is unified, without random change of translation expressions.
	Synonym Uniformity	Translation	For synonymous ecotourism vocabularies, the translation follows the same terminology standard, without non-uniform translation.
	Proper Noun Uniformity	Translation	For proper nouns such as geographical names and ethnic cultural terms, the translation conforms to the unified standard, without multiple translations.
Readability (20%)	Target Language Fluency		The translation conforms to the grammatical rules and expression habits of

	English, with smooth and natural language.
Vocabulary Appropriateness	The translated vocabulary is appropriate for the ecotourism context, without the use of overly obscure or colloquial words.
Communicative Effectiveness	The translation is easy for English readers to understand and accept, and can achieve the expected cross-cultural communication effect.

4. Empirical Research Results and Analysis

4.1 Overall Translation Quality Evaluation Results

Based on the evaluation data of four large language models translating 300 eco-cultural terms related to Qinghai ecotourism, this study calculated the average score and weighted comprehensive score of each model in four primary indicators (accuracy, cultural connotation retention, terminology consistency, and readability). The overall translation quality results are shown in Table 2.

Table 2 Overall Translation Quality Evaluation Results of Four Large Language Models

Large Language Models	Accuracy (30%)	Cultural Connotation Retention (25%)	Terminology Consistency (25%)	Readability (20%)	Comprehensive Average Score
ChatGPT	7.5	6.5	7.5	7.0	7.2
Google Gemini	9.0	8.5	8.0	8.0	8.4
DeepSeek	8.0	7.0	7.5	7.0	7.4
Qwen	8.5	7.5	8.0	7.0	7.8

Note: The rankings are descriptive only; no claims of statistical significance are made.

The four models exhibited different performance characteristics and advantages in translating eco-cultural terms related to Qinghai ecotourism, with no absolute superiority or inferiority, but rather obvious differences in functional focus. Each model has its own unique functional orientation and translation advantages, which are closely related to its training corpus characteristics and technical positioning, and there is no model that can achieve leading performance in all evaluation dimensions. As shown in Table 2, Google Gemini ranks first with a comprehensive score of 8.4, achieving the highest scores in all four evaluation indicators, particularly excelling in accuracy (9.0) and cultural connotation retention (8.5); Qwen ranks second with a comprehensive score of 7.8, demonstrating relatively balanced performance with a notable accuracy score of 8.5; DeepSeek ranks third with 7.4 points, featuring stable performance across all indicators and good terminology consistency (7.5); ChatGPT ranks fourth with 7.2 points, with its scores slightly lower than the others, though its readability (7.0) is on par with DeepSeek and Qwen.

Table 2 shows that Google Gemini leads across all four evaluation indicators, while the other three models each have their own relative strengths and weaknesses. The specific performance of each model in each indicator is as follows:

ChatGPT performs modestly across all indicators, with its highest scores being 7.5 in accuracy and terminology consistency. Its English translation is generally acceptable, but it lacks the fluency and naturalness of the top-performing model. Its terminology consistency is moderate (7.5), though it occasionally shows inconsistent translations of the same term in different contexts, which slightly affects the overall standardization. Its readability score of 7.0 is the same as DeepSeek and Qwen, but lower than Gemini.

Google Gemini excels in all four indicators, with the highest accuracy score of 9.0, the highest cultural connotation retention score of 8.5, the highest readability score of 8.0, and a top-tier terminology consistency score of 8.0 (tied with Qwen). It performs particularly well in the standardization of professional ecological and geographical terms, strictly following the official translation standards and international academic norms for terms such as geographical names, ecological concepts and professional nouns. Its cultural connotation retention is also strong, effectively conveying regional cultural nuances and ethnic characteristics.

DeepSeek achieves a terminology consistency score of 7.5, which is tied with ChatGPT but lower than Gemini and Qwen (both 8.0). However, it maintains a relatively stable performance across all four evaluation dimensions, with no extreme weaknesses. Its accuracy (8.0) is higher than ChatGPT (7.5), and its readability (7.0) is on par with ChatGPT and Qwen. It shows good stability and reliability in standardized translation scenarios.

Qwen ranks second overall with a comprehensive score of 7.8. Its accuracy (8.5) is the second highest after Gemini, and its terminology consistency (8.0) is tied with Gemini for the highest. Its cultural connotation retention (7.5) and readability (7.0) are moderate, neither leading nor lagging significantly. While it does not have an absolute advantage in any single indicator, its balanced and relatively high performance across the board secures its second-place

ranking.

4.2 Translation Quality Evaluation Results by Category

To further explore the performance of the four models in different categories of terms, this study compared their average scores in four self-defined categories of terms (consistent with the term classification of this study): Natural Landscapes, Ecological Resources & Conservation, Tourism Activities & Facilities, and Culture & Folk Customs. The results are shown in Table 3.

Table 3. Translation Quality Results of Different Categories of Eco-Cultural Terms

Large Language Models	Natural Landscapes	Ecological Resources & Conservation	Tourism Activities & Facilities	Culture & Folk Customs
ChatGPT	7.5	7.5	7.5	6.5
Google Gemini	9.0	9.0	8.5	8.5
DeepSeek	7.5	7.5	7.5	7.0
Qwen	8.5	8.5	8.0	8.0

Note: The rankings are descriptive only; no claims of statistical significance are made.

Based on the data in Table 3, the four models show obvious category-specific characteristics in translating different categories of eco-cultural terms, with each having its own applicable scenarios.

In terms of Natural Landscapes, Google Gemini achieves the highest score (9.0), demonstrating superior performance in accuracy and adherence to official geographical naming standards. Qwen ranks second (8.5), showing reliable handling of landscape terminology. DeepSeek and ChatGPT both score 7.5, indicating moderate performance, with DeepSeek being stable but occasionally less accurate, and ChatGPT generally fluent yet sometimes deviating from standard renderings.

In terms of Ecological Resources & Conservation, Google Gemini again leads with 9.0 points, reflecting its strong compliance with international ecological and biological terminology norms. Qwen follows with 8.5 points, showing good accuracy. DeepSeek and ChatGPT both score 7.5; DeepSeek is generally reliable but may lack scientific precision in some cases, while ChatGPT occasionally exhibits inconsistency or verbosity in policy-related ecological terms.

In terms of Tourism Activities & Facilities, Google Gemini achieves the highest score (8.5), demonstrating strong performance in handling terms related to tourist facilities and activities. Qwen ranks second with 8.0 points, showing good accuracy and fluency. DeepSeek and ChatGPT both score 7.5, with DeepSeek providing consistent and standardized expressions, while ChatGPT is fluent but may occasionally lack terminological precision.

In terms of Culture & Folk Customs, Google Gemini holds the highest score (8.5), excelling in providing clear and context-aware translations. Qwen follows closely with 8.0 points, actively offering useful cultural annotations to enhance cross-cultural understanding. DeepSeek scores 7.0, delivering basic translations but lacking deeper cultural context. ChatGPT scores 6.5, conveying only essential information without sufficient depth or explanatory notes.

4.3 Qualitative Analysis of Typical Translation Cases and Performance Characteristics

Based on the quantitative evaluation results, this section selects representative excellent cases and cases with performance differences from the actual translation outputs of the four models for qualitative analysis, and summarizes the translation characteristics and potential influencing factors of each model.

4.3.1 Accuracy

Accuracy in translating Qinghai's eco - tourism terminology is the linchpin for effective communication. Given the region's unique ecological and cultural landscapes, precise translations are essential for conveying information accurately across different fields. Inaccuracies can distort the understanding of Qinghai's natural wonders, ecological systems, and cultural heritage. Whether it's for international research collaborations, tourism promotion, or educational purposes, an accurate translation serves as the foundation, ensuring that the essence of the original terms is maintained without any misrepresentation.

Example 1

ST: 青海湖裸鲤 (Qīng hǎi Hú luǒ lǐ)

TT 1 (Google Gemini): Qinghai Lake Naked Carp (Gymnocypris przewalskii)

TT 2 (Qwen): Qinghai Lake Naked Carp (*Gymnocypris przewalskii*)

TT 3 (DeepSeek): Qinghai Lake Naked Carp (*Gymnocypris przewalskii*)

TT 4 (ChatGPT): Qinghai Lake Naked Carp

Analysis: Google Gemini, Qwen, and DeepSeek include the Latin scientific name “*Gymnocypris przewalskii*”, adhering to international biological classification standards. ChatGPT, by only providing the common name, lacks this academically standardized information, thus scoring lower in accuracy.

Example 2

ST: 年宝玉则 (Niánbǎo Yùzé)

TT 1 (Google Gemini): Nianbaoyuze (Mt. Nyenpo Yurtse)

TT 2 (Qwen): Nianbaoyuze

TT 3 (DeepSeek): Nyainqêntanglha (Nianbaoyuze) Scenic Area

TT 4 (ChatGPT): Nianbaoyuze

Analysis: Google Gemini accurately represents the mountain’s origin by supplementing the phonetic transcription with the Tibetan name “Mt. Nyenpo Yurtse”. Qwen and ChatGPT only offer the phonetic form. DeepSeek makes a significant error by confusing it with “Nyainqêntanglha”. Google Gemini clearly demonstrates higher accuracy in geographical name translation.

4.3.2 Cultural Connotation Retention

Qinghai’s eco-tourism terms are deeply rooted in its rich cultural tapestry. Retaining cultural connotations in translation is crucial for presenting a holistic view of the region. These terms are not just words but carriers of local traditions, beliefs, and historical legacies. A translation that fails to capture these subtleties risks reducing the cultural experience to a mere surface-level description. By preserving cultural connotations, we enable international audiences to engage with Qinghai’s culture on a deeper level, fostering cross-cultural understanding and appreciation.

Example 3

ST: 甜醅 (tián pēi)

TT 1 (Google Gemini): Tianpei (Sweet Fermented Oats/Barley)

TT 2 (Qwen): Tianpei (Fermented Grain Dessert)

TT 3 (DeepSeek): Tianpei (Fermented Grain Dessert)

TT 4 (ChatGPT): Sweet Fermented Grain (Tianpei)

Analysis: Google Gemini precisely captures the regional characteristics by specifying the ingredients “Oats/Barley”. Qwen and DeepSeek use a more general “Fermented Grain Dessert”, and ChatGPT’s translation with a reversed order and a general “Grain” misses key cultural details. Google Gemini is superior in retaining the cultural connotation of this local cuisine.

Example 4

ST: 撒拉族篱笆楼 (Sālāzú líbā lóu)

TT 1 (Google Gemini): Salar Wattle House (Liba Building)

TT 2 (Qwen): Salar Fence House (Traditional architecture)

TT 3 (DeepSeek): Salar Hedgerow Houses (Li Ba Lou)

TT 4 (ChatGPT): Salar Fence House

Analysis: Google Gemini provides the most culturally informative translation by using “Wattle” (a technical term for a traditional building method involving woven branches) and supplementing with the phonetic transcription “Liba Building”, which helps preserve the original term’s ethnic and regional identity. DeepSeek offers a slightly less precise rendering with “Hedgerow Houses” and retains the Chinese Pinyin “Li Ba Lou”, also making an effort to preserve cultural specificity. Qwen adds a general explanatory note “Traditional architecture”, which provides basic context but lacks the architectural detail of “wattle” or the phonetic specificity of “Liba”. ChatGPT gives the simplest translation, “Salar Fence House”, which conveys only the most literal meaning and loses most of the

cultural and architectural connotations. Therefore, in terms of cultural connotation retention, Google Gemini performs best, followed by DeepSeek, while Qwen and ChatGPT show progressively less cultural depth.

4.3.3 Terminology Consistency

Terminology consistency is of paramount importance in the translation of Qinghai's eco-tourism vocabulary. In a context where multiple stakeholders, including researchers, policymakers, and tourism operators, rely on these translations, inconsistent usage can lead to confusion and misinterpretation. Standardized translations across various documents and platforms ensure seamless communication. It also aids in building a common knowledge base, facilitating the exchange of ideas, and promoting coordinated efforts in the development and conservation of Qinghai's eco-tourism resources.

Example 5

ST: 生态修复 (shēngtài xiūfù)

TT 1 (Google Gemini): Ecological Restoration

TT 2 (Qwen): Ecological Restoration

TT 3 (DeepSeek): Ecological Restoration

TT 4 (ChatGPT): Ecological Rehabilitation (occasionally)

Analysis: Google Gemini, Qwen, and DeepSeek consistently use "Ecological Restoration". ChatGPT, however, occasionally uses "Ecological Rehabilitation", showing inconsistent terminology. This lack of consistency can lead to confusion, especially in technical or specialized texts.

4.3.4 Readability

Readability plays a vital role in the translation of Qinghai's eco-tourism terminology. Given that these translations are often targeted at a diverse audience, including tourists, students, and the general public, a high level of readability is necessary for effective dissemination of information. Complex or convoluted translations can act as barriers, preventing the audience from fully grasping the significance of the terms. Readable translations, on the other hand, make it easier for people to engage with the content, encouraging them to explore and appreciate Qinghai's unique ecological and cultural offerings.

Example 6

ST: 草畜平衡 (cǎo xù pínghéng)

TT 1 (Google Gemini): Forage-Livestock Balance

TT 2 (Qwen): Grass-Livestock Balance

TT 3 (DeepSeek): Balancing Livestock with Grassland Capacity

TT 4 (ChatGPT): Grass-Livestock Balance

Analysis: Google Gemini's "Forage-Livestock Balance" accurately captures the ecological function of "草" as livestock feed and is concise and natural, fitting English policy-term conventions. Qwen and ChatGPT's "Grass-Livestock Balance" is a more literal translation and less natural. DeepSeek's "Balancing Livestock with Grassland Capacity" is overly verbose, reducing readability.

Example 7

ST: 生态体验游 (shēngtài tīyàn yóu)

TT 1 (Google Gemini): Eco-experience Tourism

TT 2 (Qwen): Eco-experience Tour

TT 3 (DeepSeek): Eco-experience Tourism

TT 4 (ChatGPT): Eco-experience Tourism

Analysis: Google Gemini, DeepSeek, and ChatGPT use "Eco-experience Tourism", which accurately reflects the nature of the activity and aligns with English tourism terminology, offering good readability. Qwen's use of "Eco-experience Tour" with "Tour" emphasizing a specific trip slightly misaligns with the "type of tourism" concept, making it less readable.

Example 8

ST: 旅游攻略 (lǚyóu gōnglǜ)

TT 1 (Google Gemini): Travel Guide

TT 2 (Qwen): Travel Guide / Strategy

TT 3 (DeepSeek): Travel Guides / Itineraries

TT 4 (ChatGPT): Travel Guide

Analysis: Google Gemini provides the most direct and idiomatic translation, “Travel Guide”, which accurately captures the meaning of “旅游攻略” as a comprehensive guide for travelers. ChatGPT offers the same translation, equally clear and readable. Qwen adds a less common alternative “Strategy”, which may confuse readers because “strategy” is not typically used for travel planning in English, slightly reducing readability. DeepSeek uses “Travel Guides/Itineraries”: “Travel Guides” is acceptable, but “Itineraries” refers more to a detailed day-by-day plan rather than a general guide, which narrows the meaning of the original term. Therefore, Google Gemini and ChatGPT perform best in terms of accuracy and readability, while Qwen and DeepSeek introduce less appropriate synonyms that may affect clarity.

Example 9

ST: 仙女湾 (Xiānnǚ Wān)

TT 1 (Google Gemini): Xiannü Bay (Fairy Bay)

TT 2 (Qwen): Fairy Bay (Xiannü Wan)

TT 3 (DeepSeek): Xiannü Bay (Fairy Bay)

TT 4 (ChatGPT): Fairy Bay

Analysis: ChatGPT’s “Fairy Bay” is the most readable for English-speaking tourists, as it uses a pure English semantic translation without any Pinyin, making it immediately understandable. However, it completely loses the original Chinese place name. Google Gemini and DeepSeek adopt a balanced approach: they place the Pinyin “Xiannü Bay” first (which may be slightly less readable for non-Chinese speakers) but provide “Fairy Bay” in parentheses, ensuring clarity. Qwen reverses the order, putting the English meaning first for immediate readability, followed by the Pinyin “Xiannü Wan”--this prioritizes readability while still offering the original name. In terms of pure readability for an English-speaking general audience, ChatGPT ranks highest, followed by Qwen (with English first), then Google Gemini and DeepSeek (with Pinyin first). However, if the goal is to balance readability with authenticity, Qwen’s structure is most reader-friendly.

4.4 Discussion

Our results reveal functional divergence among four LLMs in translating Qinghai ecotourism terminology, extending previous domain-specific evaluations. Zhao, Zhang and Yang (2024) found that Ernie Bot and iFlytek Spark outperformed ChatGPT in literary translation, suggesting that culturally oriented training data benefit literary and culturally loaded tasks. Huang and Zhao (2026) reported that different LLMs excel on different metrics (e.g., lexical diversity, readability, BLEU), with no single model dominating all dimensions. Xu and Shen (2025) showed that DeepSeek’s semantic accuracy in metaphor translation surpassed ChatGPT, attributing this to cultural bias in training data. Collectively, these findings align with our observation: Google Gemini leads in accuracy for ecological terms, Qwen excels in cultural connotation retention, DeepSeek achieves the highest terminology consistency, and ChatGPT stands out in readability. Thus, model selection should be task-dependent rather than assuming universal superiority.

A second pattern concerns the tension between cultural sensitivity and terminological consistency. Xu and Shen (2025) demonstrated that culturally biased training enhances metaphor handling – consistent with Qwen’s advantage in folk customs through contextual annotations. However, Zhao, Zhang and Yang (2024) noted output inconsistency in culturally specific items. In our study, accuracy-focused models (e.g., Gemini) occasionally miss subtle cultural meanings, while culturally strong models (e.g., Qwen) show moderate consistency fluctuations – reflecting a known trade-off. Limitations include the limited term set (300 items), subjectivity in manual scoring, and focus on a single language pair. Future research should expand corpora, include more models (e.g., Claude, Wenxin Yiyan), and incorporate automated metrics alongside human evaluation, potentially through fine-tuning on Qinghai-specific corpora or multi-model pipelines.

5. Conclusion

Through the comparative analysis of four large language models translating 300 eco-cultural terms related to Qinghai ecotourism, this study verifies the good applicability of large language models in the translation of Qinghai ecotourism eco-cultural terms, and clarifies the functional characteristics, advantages and applicable scenarios of each model:

Based on the overall translation quality evaluation (Table 2) and the category-specific results (Table 3), Google Gemini demonstrates the strongest overall performance, ranking first in comprehensive average score and achieving the highest scores across all four term categories. It excels particularly in accuracy and cultural connotation retention, making it the most suitable choice for professional academic texts and scenarios requiring strict standardization of geographical, ecological, and cultural terminology. Qwen ranks second in overall comprehensive score and shows particularly strong performance in the Culture & Folk Customs category, significantly outperforming both ChatGPT and DeepSeek in that domain; its balanced strengths across accuracy and terminology consistency make it well suited for cross-cultural popularization scenarios that require effective transmission of cultural information. DeepSeek achieves a moderate comprehensive score with stable and balanced performance across all four evaluation indicators; it ties with ChatGPT in most categories while slightly surpassing it in cultural and folk terms, making it a reliable option for general translation scenarios that demand unified terminology and standardized expression. ChatGPT ranks fourth in overall comprehensive score, with its strength lying in readability and natural expression; however, it scores relatively lower in cultural connotation retention and performs weakest in the Culture & Folk Customs category, making it more suitable for daily communication and general cross-cultural scenarios, though caution is needed when handling culturally specific terms.

To improve the translation quality of eco-cultural terms related to Qinghai ecotourism, this study proposes a classification-based optimization strategy: for academic and professional terms such as Qinghai's ecological protection zones and cultural heritage terminology, Google Gemini with high academic rigor should be prioritized; for daily professional documents and work reports, DeepSeek with stable and consistent performance is more suitable; for frontline work scenarios that require quick access to key information, Qwen which balances professionalism and conciseness is preferred. At the same time, targeted manual proofreading should be carried out on proper nouns unique to Qinghai, such as place names, ethnic cultural terms, and characteristic ecological terms, to eliminate translation deviations. Additionally, it is necessary to unify the terminology standards of eco-cultural terms related to Qinghai ecotourism and optimize prompt instructions according to the functional characteristics of each model, so as to give full play to the advantages of each model and improve overall translation accuracy and standardization.

This study makes certain contributions to the translation of eco-cultural terms related to Qinghai ecotourism. Firstly, it systematically analyzes the differences in professional semantic expression of mainstream large language models in this specific field, clarifying the advantages and applicable scenarios of each model, which provides a useful reference for the selection of translation tools. Secondly, it proposes a targeted classification optimization strategy aiming at the translation quality of Qinghai's eco-cultural terms, including model selection based on scenarios, manual proofreading of proper nouns, terminology standardization and prompt optimization, helping to address the key problems of inaccuracy and inconsistency in the translation of regional characteristic terms and providing a preliminary foundation for improving the standardization and accuracy of related translations. In future research, the core focus will be on expanding the corpus scale of Qinghai eco-cultural terms, incorporating more large language models, and conducting targeted model fine-tuning to further enhance the translation performance of regional characteristic terms.

Acknowledgments

Not applicable.

Author contributions

Dr. Chunli Yu drafted the manuscript and revised it. Assoc. Prof. Hongyan Tuo were responsible for study design and data collection. All authors read and approved the final manuscript.

Funding

This work was supported by the Research Center for Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era, Qinghai Normal University [project number Qsd25szh20]. This work was also funded by the 2026 Research Startup Fund (1st batch), Qinghai Normal University [project number QHSF-BS1-2618].

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Al Rousan, R., Jaradat, R., & Malkawi, M. (2025). ChatGPT translation vs. human translation: an examination of a literary text. *Cogent Social Sciences*, 11(1), 2472916. <https://doi.org/10.1080/23311886.2025.2472916>
- Buckley, R. (1985). Ecotourism: A New Direction for Tourism in Natural Areas. *Tourism Management*, 6(3), 183-190.
- Cohen, E. (1985). *The Sociology of Tourism: Approaches, Issues, and Findings*. Sage Publications.
- Hu, G. S. (2004). Eco-translatology: A New Perspective on Translation Studies. *Chinese Translators Journal*, 1, 10-15.
- Huang, X., & Zhao, S. (2026). An Empirical Comparative Study on Chinese-English Translation Quality of Large Language Models (LLMs). *Language and Translation*, 1, 50-59.
- Kocmi, T., & Federmann, C. (2023). GPT-4 as a Translator: An Evaluation of Translation Quality and Style. *ACL Anthology*, 5678-5690.
- Li, H. (2021). A Corpus-Based Study on Ecotourism Translation in Western China. *Journal of Regional Tourism Research*, 10(4), 56-72.
- Li, J., Zhang, H., & Wang, L. (2024). An Empirical Evaluation of DeepSeek's Translation Performance in Cross-Cultural and Professional Scenarios. *Journal of Artificial Intelligence in Translation*, 6(1), 78-92.
- Li, Q. J., & Ng, Y. (2025). Effective Chinese-to-English biotic interpretation in ecotourism destinations: A corpus-based interdisciplinary study. *Journal of Sustainable Tourism*, 33(1), 207-242. <https://doi.org/10.1080/09669582.2024.2322129>
- Li, Y., & Chen, W. (2024). The Application Effect of Google Gemini in Ecological Professional Terminology Translation. *Journal of Ecological Linguistics*, 12(2), 103-119.
- Meng, C. P. (2023). A Study on the English Translation of Guizhou's Tourist Attraction Names—Based on the Sociosymbolic Translation Approach [J]. *Journal of Guizhou Normal University*, 39(10), 78-84.
- Nida, E. A. (2001). *Language, Culture and Translating*. Shanghai Foreign Language Education Press.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2019). *Improving Language Understanding by*

Generative Pre-Training. OpenAI Blog.

- Rosenberger, P., & Kolchanka, N. (2025). Assessing translation quality in large language models and machine translation systems: The BLEU metric and language pair effects. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics*, 33, 130-155. <https://doi.org/10.55549/epstem.1280>
- Su, Y. Y. (2020). English Translation of the Exhibition Signs:a Case of the Broze Exhibition at the Yunnan Provincial Museum. *Relics and Museology*, 1, 97-103.
- Wang, Q (2025) Evaluating Uighur literary translation: A comparative study of ChatGPT, Google Translate, and Bing Translator. *PLoS One*, 20(10), e0335261. <https://doi.org/10.1371/journal.pone.0335261>
- Xu, Y., & Shen, W. (2026). An Inquiry into the Localized Advantage of Large Language Models in Translating Metaphors: A Quantitative Comparative Study of ChatGPT and DeepSeek. *Translation Research and Teaching*, 2, 88-96.
- Yang, Y., & Dai, Y. (2022). Construction and application of Chinese-English parallel corpus for Sanjiangyuan National Park. *Modern Linguistics*, 10(1), 1-8. <https://doi.org/10.12677/ML.2022.101001>
- Zhao, Y., Zhang, H., & Yang, Y. (2024). Comparative Study on the Translation Quality of Large Language Models——Taking the Translation of “Fan Hua” as an Example. *Technology Enhanced Foreign Language Education*, 4, 60-66+109.