Open Access Library Journal
2025, Volume 12, e14686
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Construction and Optimization Strategies of a Multilingual Translation System for the Qilu Library in Low-Resource Language Scenarios

Wenhan Zhang[1], Yuge Zeng[2], Xiaolin Liu[3]

[1]Department of Foreign Languages, Qilu Normal University, Jinan, China
[2]Department of Information Science and Engineering (Artificial Intelligence), Qilu Normal University, Jinan, China
[3]Department of Literature and Cultural History, Qilu Normal University, Jinan, China
Email: 3307232867@qq.com, 1463794708@qq.com, 13153510947@163.com

How to cite this paper: Zhang, W.H., Zeng, Y.G. and Liu, X.L. (2025) Construction and Optimization Strategies of a Multilingual Translation System for the Qilu Library in Low-Resource Language Scenarios. *Open Access Library Journal*, **12**: e14686.
https://doi.org/10.4236/oalib.1114686

Received: November 28, 2025
Accepted: December 23, 2025
Published: December 26, 2025

Abstract

This study explores the development and enhancement of a multilingual translation system for the Qilu Library, a comprehensive cultural repository, under low-resource language conditions. It examines the challenges associated with translating culturally specific and linguistically scarce content, and proposes strategic frameworks for system construction and optimization. By analyzing current machine translation technologies, corpus development methods, and adaptive learning models, this article identifies key areas for improvement in handling low-resource languages, particularly those relevant to the cultural heritage of the Qilu (Shandong) region. The study emphasizes the integration of human-AI collaboration, data augmentation, and transfer learning as vital strategies for improving translation quality and system robustness. Furthermore, it introduces the role of knowledge graphs in cultural concept disambiguation and active learning for efficient resource allocation. The goal is to provide a scalable and culturally sensitive translation model that can be adapted to other regional cultural projects, thereby contributing to the global dissemination of local knowledge and offering practical insights for digital heritage preservation.

Subject Areas

Computational Linguistics, Digital Humanities, Low-Resource Machine Translation

Keywords

Low-Resource Languages, Multilingual Translation System, Qilu Library, Cultural Localization, Machine Translation Optimization, Knowledge Graph, Human-in-the-Loop

1. Introduction

In an increasingly globalized world, the demand for multilingual access to cultural and historical resources has never been greater [1]. However, many regional archives, such as the Qilu Library, which documents the rich cultural heritage of Shandong Province, face significant challenges in making their collections accessible to international audiences. These challenges are particularly pronounced in the context of low-resource languages, languages that lack extensive digital corpora, standardized tools, or sufficient parallel data for effective machine translation (MT) [2]. For the Qilu Library project, potential target low-resource languages could include, for example, minority languages within China that present specific challenges for machine translation due to scarce digital resources in the cultural heritage domain (e.g., Uyghur or Yi), or regionally significant languages from neighboring countries with scarce parallel data for the cultural domain.

The Qilu Library encompasses a wide range of text types, including classical literature, folk narratives, historical records, and modern academic works, many of which contain dialectal expressions, archaic terms, and culture-specific references. Traditional MT systems, such as those based on neural machine translation (NMT), often underperform when applied to such content due to data scarcity and linguistic complexity [3]. For instance, translating classical Chinese philosophical texts requires not only linguistic competence but also deep cultural understanding, which is often absent in generic NMT models trained on contemporary web data [4].

This study aims to address these gaps by proposing a structured approach to building and optimizing a multilingual translation system tailored to the needs of the Qilu Library. It seeks to answer the following research questions.

1) What are the main challenges in developing a multilingual translation system for low-resource language scenarios like the Qilu Library?

2) What strategies can be implemented to optimize translation accuracy and cultural relevance?

Through a review of existing literature and conceptual analysis, this paper offers a blueprint for constructing a sustainable and adaptive translation infrastructure that respects linguistic diversity and promotes cultural exchange [5]. It also discusses practical implementation pathways and potential impacts on similar cultural digitalization projects [6].

2. Literature Review

2.1. Low-Resource Machine Translation

Low-resource machine translation has emerged as a critical subfield within computational linguistics [7]. Early approaches relied on rule-based and statistical methods, but recent advances in neural machine translation have shifted the focus toward data-driven models [8]. However, these models require large-scale parallel corpora, which are often unavailable for low-resource languages (LRLs) [9]. Researchers have explored techniques such as transfer learning, multilingual training, and zero-shot translation to mitigate data scarcity [10]. Despite these innovations, issues such as domain mismatch, code-switching, and cultural untranslatability remain largely unaddressed [11]. For example, while multilingual models like mT5 can handle multiple languages, they often struggle with dialectal variations and culturally embedded expressions unique to specific regions like Shandong.

2.2. Cultural Localization in Translation

Translation is not merely a linguistic task but also a cultural one [12]. Scholars such as Nida (1964) and Venuti (1995) have long emphasized the importance of cultural equivalence in translation [13]. In digital contexts, this requires systems that can recognize and adapt culture-specific terms, historical context, and rhetorical styles [14]. For projects like the Qilu Library, where texts are deeply embedded in local traditions, a culturally-aware translation system is essential. Recent studies have begun exploring the integration of external knowledge sources, such as ontologies and cultural databases, to enhance MT systems' ability to handle culture-specific items, though practical implementations remain limited [15].

2.3. Existing Models and Gaps

While models like mBART, mT5, and OPUS-MT have shown promise in multilingual settings, their performance on low-resource languages with dialectal variations remains limited [16]. There is a notable lack of frameworks that integrate cultural experts into the MT pipeline, or that use active learning to continuously improve system output based on user feedback [17]. Furthermore, most existing systems prioritize linguistic accuracy over cultural appropriateness, leading to translations that may be grammatically correct but culturally insensitive or misleading [18]. This gap is particularly critical for cultural heritage projects where contextual fidelity is as important as textual accuracy.

3. System Construction Framework

3.1. Data Collection and Corpus Building

The first step in constructing the translation system is the development of a specialized parallel corpus. This includes not only digitizing and aligning existing Qilu Library materials in Chinese and target languages but also implementing a multi-layered annotation strategy [19]. For instance, classical texts require special markup for archaic terms and historical references, while folk narratives need dialectal annotations and cultural notes. To make the dialect processing more concrete, consider a Shandong dialectal term such as "夜来" (yè lái), which means "yesterday" in standard Mandarin. In the corpus, this term would be annotated with its dialectal provenance (e.g., <term dialect = "Shandong" std_mandarin = "昨天">夜来</term>), along with a contextual note explaining its usage. This structured annotation enables the model to learn region-specific mappings and supports the generation of culturally-glossed translations (e.g., "yesterday (Shandong dialect: yè lái)”). The corpus building process should also incorporate oral histories and expert-validated translations of key cultural terms to ensure contextual accuracy. Collaboration with cultural institutions is crucial for enriching the corpus with metadata and contextual notes, which can later serve as training data for culture-aware translation models. Moreover, techniques such as web scraping of relevant contemporary materials and controlled data augmentation using back-translation can help expand the corpus size while maintaining domain relevance [20].

3.2. Model Selection and Adaptation

A hybrid model combining pre-trained multilingual NMT systems with fine-tuning on domain-specific data is recommended. The system should leverage transfer learning from high-resource languages while incorporating domain adaptation techniques specifically designed for cultural content. This includes implementing a cultural knowledge graph that maps key concepts, entities, and their relationships within the Qilu cultural domain. The structure of this knowledge graph would use entities such as historical figures (e.g., Confucius, Mencius), philosophical concepts (e.g., “仁” Benevolence, “礼” Ritual), cultural artifacts (e.g., “泰山石刻” Mount Tai Stone Inscriptions), literary works, and geographic locations as nodes. The edges would define relationships like “is a student of,” “is a core concept of,” “is created in,” or “is located in.” For example, when translating terms like “儒家思想” (Confucianism), the system can reference connected concepts in the knowledge graph to generate contextually appropriate translations [21]. This allows the system to generate not only lexically accurate but also contextually and culturally coherent translations for compound terms or phrases. Additionally, incorporating specialized modules for handling proper nouns and technical terms through transliteration and contextual disambiguation can significantly improve named entity recognition and consistency across different text types and historical periods.

3.3. Human-in-the-Loop Workflow

To ensure cultural and linguistic accuracy, a human-in-the-loop mechanism is essential throughout the translation pipeline. This involves not only post-editing by bilingual cultural experts but also their active participation in the pre-translation phase through glossary development and translation memory curation. The distinct roles within this workflow should be clarified: Bilingual cultural experts are primarily responsible for high-level tasks requiring deep domain knowledge, such as defining core cultural terminology, developing and validating the initial cultural knowledge graph, setting translation guidelines for complex texts (e.g., classical philosophy), and performing quality assurance on critical outputs. In contrast, the community engaged via crowdsourcing is better suited for scalable tasks that benefit from collective local knowledge, such as validating dialectal expressions, collecting regional variants of folk narratives, flagging potentially ambiguous translations for expert review, and expanding the corpus with contemporary usage examples. The workflow should be designed as an iterative process where human feedback is systematically incorporated into model retraining. For instance, a dedicated interface can allow experts to flag culturally problematic translations, which are then used to update the knowledge graph and fine-tune the model. Community engagement through crowdsourcing platforms can further enhance dialectal and regional validation, particularly for texts with strong local characteristics. This continuous collaboration between human experts and the AI system creates a virtuous cycle of improvement, gradually enhancing the system's cultural competence while maintaining efficiency.

4. Optimization Strategies

4.1. Active Learning and Adaptive Training

Active learning strategies can significantly enhance resource efficiency by prioritizing the translation of uncertain or complex segments for human review. The system can be designed to calculate confidence scores for each translation output, automatically flagging low-confidence segments, such as those containing rare cultural terms or complex syntactic structures, for expert verification. This approach ensures that human effort is focused where it adds most value, thereby improving model performance over time with minimal expert input. Adaptive training cycles can then leverage this human feedback to incrementally update the model, allowing it to learn from corrections and incorporate new linguistic patterns. For example, the system can be retrained weekly with newly verified translations, gradually expanding its coverage of cultural concepts and dialectal variations while reducing error rates in problematic areas.

4.2. Evaluation Metrics beyond BLEU

Traditional metrics like BLEU often fail to capture cultural adequacy, as they primarily measure surface-level similarity to reference translations. This study advocates for a comprehensive evaluation framework that combines multiple assessment methods. Human evaluation rubrics should be developed to systematically assess fluency, accuracy, and cultural appropriateness, with special attention to the handling of culture-specific items. Semantic similarity measures using embeddings can provide additional insights into meaning preservation beyond n-gram matching. Furthermore, implementing a detailed error classification system based on cultural and linguistic typologies enables targeted improvements, for instance, identifying whether errors stem from linguistic misunderstanding, cultural misinterpretation, or contextual insensitivity. This multi-faceted approach ensures that translation quality is assessed holistically, with equal emphasis on technical accuracy and cultural resonance.

4.3. Long-Term Sustainability

To ensure the system's longevity, it is crucial to establish clear governance and maintenance protocols from the outset. Key implementation challenges must be acknowledged and planned for, including the ongoing cost and availability of bilingual cultural experts for annotation and validation, and the need for dedicated technical expertise to maintain, update, and troubleshoot the system's complex components (e.g., the knowledge graph, active learning module, and model retraining pipelines). This includes version control for both models and corpora, regular quality audits, and systematic updates based on user feedback and evolving linguistic standards. Developing open-access tools for community contributions can foster a sustainable ecosystem around the translation system, encouraging ongoing improvement and adaptation. Integration with broader digital heritage platforms, such as national digital library initiatives and international cultural databases, enhances the system's utility and visibility while facilitating resource sharing. Additionally, implementing a modular architecture allows for component-level upgrades and adaptations, enabling the system to evolve alongside technological advancements and expanding requirements. These measures collectively ensure that the translation system remains relevant, accurate, and accessible over the long term, maximizing its impact on cultural preservation and dissemination.

5. Conclusion

The development of a multilingual translation system for the Qilu Library represents a significant step toward preserving and promoting regional culture in a global context [1]. While low-resource languages pose distinct challenges [2], strategies such as corpus enrichment, model fine-tuning [10], and human-AI collaboration can substantially enhance system performance. The integration of knowledge graphs for cultural concept management and active learning for efficient resource allocation offers promising pathways for maintaining both linguistic accuracy and cultural appropriateness. Future work should focus on real-world implementation, user experience studies, and the expansion of the system to other cultural archives. Particular attention should be paid to developing more sophisticated methods for handling dialectal variations and measuring cultural adequacy in translation outputs. It should be noted that the framework proposed herein is conceptual, and its practical efficacy requires validation through implementation in the Qilu Library context, particularly when dealing with ultra-low-resource oral historical materials. By bridging the gap between technology and tradition, this project aims to set a precedent for the digital preservation of linguistic diversity while providing practical solutions for cultural institutions worldwide facing similar challenges in multilingual accessibility.

Funding Statement

2025 National Undergraduate Innovation and Entrepreneurship Training Program Project "Transformer-Based Multilingual Intelligent Translation System for the Qilu

Canon and Cross-Cultural Communication Design" (Project No. 202514276003).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Khoboko, P.W., Marivate, V. and Sefara, J. (2025) Optimizing Translation for Low-Resource Languages: Efficient Fine-Tuning with Custom Prompt Engineering in Large Language Models. *Machine Learning with Applications*, **20**, Article ID: 100649. https://doi.org/10.1016/j.mlwa.2025.100649

[2] Ngo, T., Nguyen, P., Nguyen, V.V., Ha, T. and Nguyen, L. (2022) An Efficient Method for Generating Synthetic Data for Low-Resource Machine Translation. *Applied Artificial Intelligence*, **36**, Article ID: 2101755. https://doi.org/10.1080/08839514.2022.2101755

[3] Ul Qumar, S.M., Azim, M. and Quadri, S.M.K. (2025) Enhancing Low-Resource Neural Machine Translation with Decoding-Based Data Augmentation. *International Journal of Information Technology*, 1-13. https://doi.org/10.1007/S41870-025-02710-X

[4] Pakray, P., Dadure, P. and Bandyopadhyay, S. (2024) Empowering Low-Resource Languages with NLP Solutions.

[5] Liu, C., Karakanta, A., Tong, A.N., Aulov, O., Soboroff, I.M., Washington, J., *et al.* (2020) Introduction to the Special Issue on Machine Translation for Low-Resource Languages. *Machine Translation*, **34**, 247-249. https://doi.org/10.1007/s10590-020-09256-8

[6] Wang, H., Wu, H., He, Z., Huang, L. and Church, K.W. (2022) Progress in Machine Translation. *Engineering*, **18**, 143-153. https://doi.org/10.1016/j.eng.2021.03.023

[7] Zhu, J., Sun, H. and Kong, B. (2025) Improving Multilingual English Translation Performance through T5 and MAML Integration. *Systems and Soft Computing*, **7**, Article ID: 200394. https://doi.org/10.1016/j.sasc.2025.200394

[8] Shivakumar, B., *et al.* (2025) Multi-Head Attention Transformer for Text2Text Translation. *Procedia Computer Science*, **258**, 1962-1971. https://doi.org/10.1016/j.procs.2025.04.447

[9] Li, Z.J., Yan, X.R., Yang, W.Z. and Pan, S.H. (2025) Corpus Machine Translation System Based on Compaction Algorithm and Self-Attention Mechanism Model. *Neural Computing and Applications*, **37**, 28479-28494. https://doi.org/10.1007/S00521-025-11351-X

[10] Wang, Y. and Liu, S. (2025) A Cognitive Study on the Strongly Condensed and Implicit Features of Chinese-Characteristic Chunked Discourse and Human-Machine Multilingual Translation among English, French and Spanish. *Lecture Notes in Education*, *Arts*, *Management and Social Science*, **3**, 171-177. https://doi.org/10.18063/lne.v3i6.1147

[11] Zou, L. (2025) Evaluating the Effectiveness of Machine Translation Tools in Translating China-Specific Discourse. *Scientific Journal of Humanities and Social Sciences*, **7**, 52-59. https://doi.org/10.54691/30qy6287

[12] Moorkens, J., Way, A. and Lankford, S. (2024) Automating Translation. Routledge. https://doi.org/10.4324/9781003381280

[13] Cui, Z.H. and Xu, J. (2025) A Corpus-Based Study on the Characteristics of Machine

Translation in Historical Texts: Taking an Excerpt from the First Biography in Volume 101 of the Book of Jin. *International Journal of Social Science and Education Research*, **8**, 57-63.

[14] Luo, S.H. (2025) Machine Translation Post-Editing Studies in China: A Research Review (1995-2025). *Education Journal*, **8**, 144-154.

[15] Xu, R. and Bai, Q. (2025) Research on the Development Status and Difficulties of Machine Translation in the Era of Artificial Intelligence. *Journal of Electronic Research and Application*, **9**, 39-43. https://doi.org/10.26689/jera.v9i3.10789

[16] Guo, J. and Zhou, Z. (2025) Personalised Foreign Language Learning Path Recommendation Strategy Based on Disciplinary Knowledge Graph. *International Journal of Information and Communication Technology*, **26**, 49-69. https://doi.org/10.1504/ijict.2025.145719

[17] Lu, H. and Li, X. (2025) Chinese Language Curriculum Resources Pushing Method Based on Knowledge Graph and Similarity Algorithm. *International Journal of Continuing Engineering Education and Life-Long Learning*, **35**, 77-94. https://doi.org/10.1504/ijceell.2025.143804

[18] Lv, F. (2024) Knowledge Graph Analysis of International Chinese Language Textbooks Based on Citespace. *Journal of Contemporary Educational Research*, **8**, 163-175. https://doi.org/10.26689/jcer.v8i4.6481

[19] Li, J. (2024) The Establishment of Chinese Culture Teaching Resources by Knowledge Graph Applied to Chinese International Education. *Applied Mathematics and Nonlinear Sciences*, **9**, 1-17. https://doi.org/10.2478/amns-2024-0555

[20] Wang, Y. and Hou, Y. (2024) The Construction of a Shared Resource Base for Teaching Chinese Culture under the Architecture of Disciplinary Knowledge Mapping. *Applied Mathematics and Nonlinear Sciences*, **9**, 1-20. https://doi.org/10.2478/amns-2024-1314

[21] Geng, Y. (2024) Citespace-Based Knowledge Mapping to Analyze the Hot Spots and Frontier Research on the Integration of Excellent Traditional Chinese Culture into English Teaching in Colleges and Universities. *Applied Mathematics and Nonlinear Sciences*, **9**, 1-11. https://doi.org/10.2478/amns-2024-1963