Education Reform and Development, 2025, Volume 7, Issue 4
https://ojs.bbwpublisher.com/index.php/ERD
Online ISSN: 2652-5372
Print ISSN: 2652-5364

An Empirical Analysis of ChatGPT Translation Error Types in Texts of Chinese Red Culture Based on the MQM Quality Assessment Framework

Yiming Li, Yuanpeng Huang*

Shanxi Normal University, Taiyuan 030031, China

***Corresponding author:** Yuanpeng Huang, 769015452@qq.com

Abstract: In recent years, translation quality evaluation has emerged as a major, and at times contentious, topic. The industry view on quality is highly fragmented, in part because different kinds of translation projects require very different evaluation methods. In response, the EU-funded QTLaunchPad project has developed the Multidimensional Quality Metrics (MQM) framework, an open and extensible system for declaring and describing translation quality metrics using a shared vocabulary of "issue types." As an effective approach to evaluating AI translation quality, the classification of translation errors has drawn increasing attention. This study focuses on translation errors in red texts, using the MQM quality assessment model as the analytical framework to categorize errors in translations produced by ChatGPT4.0, a leading engine among current large language models. The findings aim to provide pedagogical support for pre-editing and post-editing training in professional translator education.

Keywords: AI translation; ChatGPT; Multidimensional quality metrics; Error types

Online publication: April 28, 2025

1. Introduction

Since its proposal by Warren Weaver in 1949, machine translation has undergone multiple transformations at its core, reflecting humanity's ongoing pursuit of high-quality automated translation. In recent years, machine translation based on deep learning has achieved significant breakthroughs, prompting major companies to incorporate it into their R&D strategies and launch online translation engines, such as the widely used Google Translate, DeepL, and Baidu Translate. Although the overall quality of current online machine translation systems for English-Chinese and Chinese-English translation has reached a passable level and can generally fulfill communicative purposes, there remains a gap between the current output and higher-quality standards [1]. Even Google Translate, a leading engine, occasionally produces errors in Chinese-to-English translation that are

subtle, sometimes so much so that even intermediate English learners may struggle to detect them.

With the rapid development of artificial intelligence and natural language processing technologies, AI translation tools have become increasingly significant in cross-cultural communication. ChatGPT, developed by OpenAI, is a large-scale generative language model based on the GPT (Generative Pre-trained Transformer) architecture. Trained on massive amounts of textual data using deep learning techniques, particularly transformer-based neural networks, ChatGPT is capable of understanding and generating human-like text across a wide range of contexts. Since its release, it has been widely adopted for tasks including content creation, code generation, question answering, and language translation. As a representative of advanced large language models (LLMs), ChatGPT has significantly influenced the fields of natural language processing and AI translation, pushing the boundaries of what AI-generated language can achieve.

The emergence of generative AI tools such as ChatGPT has accelerated the evolution of neural AI translation technologies. While research on machine translation continues to flourish both in China and abroad, there remains a lack of consensus regarding evaluation methods for translation quality and standardized classification of translation errors [2]. Although certain metrics, such as the widely recognized SAE J2450 and the LISA QA Model, are available for assessing translation quality, the former is limited by its industry specificity and lacks generalizability, while the latter adopts a "one-size-fits-all" approach, sacrificing flexibility. Automatic evaluation metrics such as BLEU, METEOR, and TER can reflect the overall quality of AI translations but fail to identify specific issues in the output and tend to marginalize the role of human evaluation [3]. For a long time, the industry has lacked a method capable of classifying AI translation errors tailored to specific client needs and text types, and of proposing acceptability standards based on such error types [4]. In response, the Multidimensional Quality Metrics (MQM) framework was developed.

The MQM (Multidimensional Quality Metrics) quality assessment model is one of the outcomes of the EU-funded QTLaunchPad project. It is a dynamic, comprehensive, and customizable framework for evaluating both source and target texts. The model offers a hierarchical classification system comprising 108 error categories, each clearly defined and distinguished. This layered structure integrates mainstream quality evaluation approaches—such as the TAUS Dynamic Quality Framework, the LISA QA Model, and SAE J2450—with quality assurance tools like ApSIC Xbench, CheckMate, and XLIFF:doc, as well as theoretical models of multidimensional machine translation evaluation [5]. At the top level, the MQM framework categorizes translation issues into five dimensions: Fluency, Accuracy, Verity, Design, and Internationalization, along with a supplementary "Other" category for uncategorized issues, forming a tree-like structure. To facilitate practical application and data analysis, the model also provides a streamlined version known as the MQM Core, which includes 19 error categories—three parent categories and sixteen subcategories (see **Figure 1**). Moreover, the MQM model assigns specific weightings to each error type and classifies them by severity into Critical, Major, Minor, and None. It further offers a formula for calculating translation quality: TQ (Translation Quality Score) = 100 − TP (Translation Penalty) + SP (Source Penalty) [6]. The MQM model (**Figure 1**) truly adopts a user-centered approach, enabling users to assess translation quality across various standards, levels, and degrees of granularity [7].

This paper aims to conduct an empirical analysis of translation error types produced by ChatGPT when translating texts of Chinese Red Culture, using the MQM quality assessment framework as the analytical tool. By classifying and analyzing the nature and severity of translation errors, the study seeks to identify systematic weaknesses in ChatGPT's handling of such texts and to provide insights that may inform both MT improvement and translator training, particularly in pre-editing and post-editing practices.

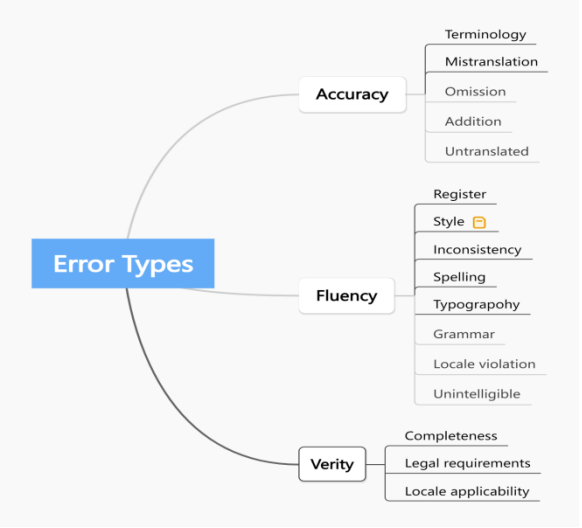

Figure 1. MQM core with error types.

2. Literature review

In recent years, neural machine translation (NMT) based on deep learning has gradually become mainstream, achieving breakthroughs in semantic understanding and contextual awareness through training on massive corpora. Online translation engines such as Google Translate, DeepL, and Baidu Translate have demonstrated considerable advancements in English-Chinese and Chinese-English translation, though issues with nuanced or domain-specific content persist [8]. For instance, Baidu's Chinese-English translation system significantly improved its handling of Chinese-specific vocabulary (such as idioms and colloquialisms) from 2017 to 2019 by optimizing its models. Similarly, domestic large models like DeepSeek have enhanced their ability to recognize technical terms and disciplinary associations through the construction of knowledge graphs, and have already been applied in the automatic summarization of academic literature. These technologies provide a technical foundation for translating complex historical terms in red culture texts. Such texts often contain unique expressions related to revolutionary historical events, political slogans, and material cultural terms (e.g., "Hongchuan Jingshen," "Jinggangshan Huishi"), which pose challenges due to semantic gaps and difficulties in conveying intended meaning in the target language. Research shows that while ChatGPT can produce literal translations of material culture terms (e.g., "Tu Su Jiu," "He Huan Tang"), it often fails to convey the deeper cultural connotations, necessitating human annotation or interpretive translation strategies to compensate. Moreover, red classic texts frequently involve ideological expressions, and AI translation may misrepresent politically sensitive content due to a lack of contextual understanding, highlighting the need for dynamic translation adjustment mechanisms.

Many scholars have already begun to explore the opportunities and challenges that ChatGPT brings to language teaching and academic writing. Salvagno et al. argued that ChatGPT can assist writers in organizing materials, generating drafts, and proofreading, while also acknowledging risks such as plagiarism and inaccuracy [9]. Thorp has explicitly expressed serious concerns about the use of ChatGPT in academic writing, asserting that ChatGPT cannot replace the role of the author [10]. Wu provided an in-depth analysis of the challenges and methodologies in detecting large language model (LLM)-generated texts. While its focus is on detection rather than translation, the article underscores the increasing indistinguishability between LLM outputs

and human-written texts, raising concerns about authenticity and bias [11]. David explored the utility of ChatGPT in providing accurate, actionable, and understandable generative medical translations in English, Spanish, and Mandarin about Otolaryngology [12]. Alm et al. explored ChatGPT's role in language education through Freire's critical pedagogy, emphasizing the dual potential of AI tools to either empower or constrain learners, and the study critiques ChatGPT's tendency to replicate dominant cultural norms due to its Anglo-centric training data, a point highly pertinent to this paper's concern with red culture translation [13]. Athanassopoulos investigated the effectiveness of ChatGPT in enhancing L2 writing among socially vulnerable students, such as refugees and migrants. The findings demonstrate measurable improvements in vocabulary richness, grammatical accuracy, and sentence length after ChatGPT-assisted revisions [14]. While these studies have analyzed the significant impact of ChatGPT on the field of translation, few have paid attention to its translation quality when applied to Chinese-specific discourse. In the context of China's entry into a new era, the theoretical framework and macro perspective for exploring and reflecting on new technologies still require further development. In this context, the Multidimensional Quality Metrics (MQM) framework has gained traction as a flexible and comprehensive quality evaluation model. Developed as part of the EU-funded QTLaunchPad project, MQM enables evaluators to classify translation errors into over 100 fine-grained categories across multiple dimensions, including Fluency, Accuracy, Verity, and Design [15]. Compared to earlier evaluation standards such as SAE J2450 and the LISA QA Model, MQM offers greater adaptability and has been integrated with professional translation tools and quality assurance systems. Lommel introduced the hierarchical taxonomy of error types and allowing tailored granularity, and highlighted that MQM enables comparative, systematic translation quality assessments [16]. Freitag focused on comparing the professional translator annotations for MT systems in the WMT2020 task. This article demonstrates that MQM error analysis significantly alters system rankings compared to crowd-sourced evaluations and reveals that mistranslation remains the dominant error type in modern NMT outputs [17]. Laurer explored the integration of multilingual BERT-based models and machine translation for cross-lingual political text classification. Their empirical findings confirm that transformer-based models can produce valid and substantively meaningful outputs across languages [18].

Nevertheless, these studies have provided valuable inspiration for this paper in terms of corpus selection, error type classification, and error avoidance strategies in research design. Despite this growing body of research, few studies have systematically applied MQM to evaluate the output of LLMs like ChatGPT, particularly in the context of culturally and politically embedded texts such as those from Chinese Red Culture. This study aims to fill that gap by combining the strengths of the MQM framework with the unique linguistic characteristics of red-themed discourse, providing new insights into the capabilities and limitations of generative AI in cross-cultural translation.

3. Research design

3.1. Research questions

To avoid the limitations of the "Mentalist View" in translation evaluation and to ensure the standardization and representativeness of translation error classification, this study selects ChatGPT, currently regarded as one of the most advanced large language models (LLM)-based translation systems [19], as the subject of evaluation. Using the MQM Core framework, the study systematically assesses the translations generated by ChatGPT. The research aims to address the following three questions:

(1) What is the overall translation quality of ChatGPT when translating texts related to Chinese Red Culture?

(2) What are the typical translation error types that occur in ChatGPT's output?

(3) How can these errors be effectively avoided or mitigated in practice?

3.2. Corpus selection

The selection of appropriate corpora is critical to the reliability and validity of this study's analysis of ChatGPT's translation errors. Given the focus on Chinese Red Culture texts, the corpus used in this study consists of texts that embody key elements of China's revolutionary history, socialist ideologies, and political rhetoric. These texts include, but are not limited to, excerpts from historical documents, speeches by prominent political figures, and literary works that reflect the values and principles associated with the Chinese Red Culture movement.

To ensure the diversity and richness of the corpus, a balanced selection of texts was made, covering various genres and linguistic styles. This includes formal political documents, propaganda literature, and narratives associated with China's revolutionary struggle. The content of these texts is ideologically dense, with terminology and expressions specific to Chinese socialist culture, making them particularly challenging for AI translation systems, especially those trained on general corpora [20].

The corpus used in this study comprises both short passages and longer texts to assess the performance of ChatGPT in different translation contexts. The selected texts are also carefully annotated to reflect the unique linguistic and cultural features of Chinese Red Culture, which will aid in the subsequent error categorization and analysis.

By focusing on this specific domain, this study aims to provide an in-depth understanding of how ChatGPT handles translations of culturally and politically charged texts and to explore the specific challenges faced by large language models in such contexts.

This study utilizes a self-compiled mini-corpus consisting of 100 carefully selected examples drawn from texts representative of Chinese Red Culture. The selected corpus contains a total of approximately 44,000 Chinese characters, ensuring sufficient textual volume for a reliable and comprehensive evaluation, while remaining manageable for detailed manual analysis. Each text was chosen to reflect diversity in genre, linguistic complexity, and contextual richness, highlighting cultural references, political terminology, and discourse features typical of Chinese Red Culture.

In addition, each of the Chinese texts is accompanied by an official English translation, which serves as a reference version in the comparative analysis between the source text and the translations generated by ChatGPT. Due to the rapid growth in demand for the translation of specialized texts, an in-depth exploration of such texts can enhance the efficiency of both pre-editing and post-editing processes [21]. It helps quickly identify error types, optimize translation quality, and provide valuable insights for improving AI translation algorithms and training models.

3.3. Statistical procedures

(1) The selected 100 source texts were individually input into ChatGPT, asked it to translate into English, and the generated translations were collected and archived with proper annotation.

(2) Following the definitions provided in the MQM Core framework, translation errors were manually identified and annotated after the annotators had gained a thorough understanding of all error types and their definitions. In cases where disagreements arose during the annotation process, decisions were made through group discussion and consensus. Since the MQM Core does not cover all possible error types, any errors not listed in **Figure 1** were categorized under the corresponding parent category or

assigned to the "Other" dimension, concerning the full MQM model. To emphasize the classification of error types—rather than the frequency or severity of repeated instances—each type of error was counted only once per sentence, even if it occurred multiple times. Distinct error types within the same sentence, however, were separately identified and recorded.

(3) Error Type Statistics and Comparative Analysis. The final statistical results were manually compiled into an Excel spreadsheet. The dataset consisted of two parts: errors manually identified by annotators and errors automatically detected by the platform itself. The percentage of a single error type was calculated by dividing the frequency of that specific error by the total number of errors.

4. Research findings

The statistical results indicate that ChatGPT exhibits several issues in translating Chinese Red Culture texts into English, with the majority of errors concentrated in the dimensions of accuracy and fluency, showing both typified and repetitive patterns (**Table 1**). Although ChatGPT is generally capable of achieving a high degree of equivalence between source and target texts, limitations inherent to its model architecture become apparent when dealing with complex sentence structures, semantic ambiguity, and multiple layers of modifiers. These issues are most evident in the accuracy dimension, where errors such as terminological inappropriateness and misinterpretation of information, often caused by over-adherence to source text syntax, are prevalent, leading to a total error rate of 73.69%.

Admittedly, AI translation systems, including ChatGPT, operate primarily at the sentence level, with translation rules and matching mechanisms designed around sentence units. This results in limited consideration of broader contextual factors such as paragraph structure or discourse cohesion. Consequently, translation strategies such as antithesis, summarization, and omission—commonly used by human translators—are rarely employed [22], negatively impacting fluency. For example, in the parallel texts examined in this study, ChatGPT often neglected or misrepresented structural consistency in handling parallel elements and subheadings, leading to decreased readability across sentences. The total error rate for fluency-related issues reached 28.94%. Moreover, the use of functional words such as conjunctions, participles, and pronouns was frequently problematic, contributing an additional 21.05% to the fluency error rate.

According to the MQM framework, the Verity dimension refers to cases in which a factually true statement in the source language becomes false in the target language, covering issues such as procedural completeness, regulatory compliance, and regional applicability. However, no such errors were identified in this study. Therefore, the analysis focuses on the accuracy and fluency dimensions. In the following sections, representative examples of frequent errors in these two areas will be presented and discussed in detail.

Table 1. Statistics of error types of ChatGPT

Error Types	Frequency	Percentage (%)
Accuracy	28	73.69
Terminology	9	23.68
Mistranslation	19	50
Omission	-	-
Addition	-	-

Table 1 (Continued)

Error Types	Frequency	Percentage (%)
Untranslated	-	-
Fluency	11	28.94
Register	-	-
Style	-	-
Inconsistency	3	7.89
Spelling	-	-
Typography	-	-
Grammar	8	21.05
Locale violation	-	-
Unintelligible	-	-
Verity	0	0
Completeness	-	-
Legal requirements	-	-
Locale applicability	-	-
Total	38	100

4.1. Accuracy dimension

According to the MQM quality assessment framework, the accuracy dimension refers to errors in which the target text fails to accurately convey the meaning of the source text, excluding any authorized or intentional deviations. This dimension includes issues such as terminology errors, mistranslations, omissions, additions, and untranslated segments. It is important to note that within the MQM framework, "terminology" is defined as the use of a term in the target language that differs from the expected or conventional usage in the relevant field. In this translation practice, no issues of omission, addition, or untranslated content were found. Therefore, the following discussion will focus solely on terminology errors and mistranslations.

4.1.1. Terminology errors

AI translation systems, including ChatGPT, benefit from extensive training on large-scale corpora, enabling them to correctly render most domain-specific terms. However, due to the flexible and context-dependent use of terminology in technical and political discourse, translation errors still occur. These errors often involve unexpected or non-standard terms that deviate from the conventional usage within the domain, thereby reducing the accuracy and acceptability of the translation.

Example 1

ChatGPT: Beiyang Warlords

Reference: the Northern Warlords

Example 2

ChatGPT: strategic pivot

Reference: strategic fulcrum

Example 3

ChatGPT: War of Attrition

Reference: protracted war

In Example 1, the term “Beiyang Junfa” is a historical and cultural term in China. ChatGPT’s translation is a direct one, focusing too much on literal correspondence, which fails to capture the deeper meaning behind the term. As a result, Western readers may not fully understand the connotations of the term. To avoid such situations, translators can either refine the translation in the post-translation phase by providing an indirect interpretation of the implied meaning or pre-edit the original text during the pre-translation phase to supplement the core information.

Furthermore, although ChatGPT has strong logical reasoning and internet search capabilities, this large language model was developed by OpenAI, an American company. As such, Chinese content makes up a smaller portion of its pre-training data, and its search ability for Chinese domestic websites is relatively weak. Consequently, it may not accurately search for and translate certain war-related terms in Chinese history and culture, such as “strategic fulcrum” in Example 2 or “protracted war” in Example 3.

It is evident that, in addition to effectively utilizing search engines to enhance information retrieval capabilities, providing a dedicated terminology database during the AI translation process plays a crucial role in improving translation quality. For instance, during the pre-editing phase, specific translation rules can be established for key terms; alternatively, in the post-editing phase, AI-translated terms can be further refined by explaining their underlying meanings or selecting more appropriate term variants to enhance the readability of specialized texts.

4.1.2. Mistranslations

Due to limitations in training models, AI translation engines are prone to mistranslations that lead to information mismatches between the source text and the translation at lexical, syntactic, and discourse levels. In the corpus used for this study, such errors account for approximately 50% of the total.

Example 4

ChatGPT: Wuxiang Drum Opera

Reference: Wuxiang Dagu Storytelling

Example 5

ChatGPT: the Chinese Civil War

Reference: the War of Liberation

Example 6

ChatGPT: Japanese Imperialism

Reference: the Japanese imperialists

Example 7

ChatGPT: Local Resistance

Reference: Regional War of Resistance against Japanese Aggression

Example 8

ChatGPT: against Japanese resistance

Reference: against Japanese aggression

Example 9

ChatGPT: Commemorating the revolutionary martyrs and inheriting the revolutionary spirit

Reference: Pay tribute to the revolutionary martyrs and passing on the traditions of revolution

In Example 4, “Wuxiang Gushu” is officially translated as “Wuxiang Dagu Storytelling.” It is listed as a second batch of provincial intangible cultural heritage in Shanxi Province and is a unique form of local art in the region, where performers drum while singing a play, with the script telling local stories. Therefore, it has been translated as “Wuxiang Dagu Storytelling.” However, ChatGPT, due to its limited understanding of Chinese culture, translated it simply as “opera,” which is a typical mistranslation. In Example 5, “Jiefang Zhanzheng” is officially translated as “the War of Liberation,” referring to the great war in which the Chinese people fought for their liberation and national independence. However, due to its political bias and misunderstanding of Chinese history, ChatGPT translated it as the “Chinese Civil War,” which is also a typical mistranslation. In Example 6, “Riben Diguozhuyi” refers to Japan’s imperialist aggression, but ChatGPT, lacking understanding of the term’s deeper meaning, translated it as “imperialism.” In Example 7, “Jubu Kangzhan” refers to part of the war in which China resisted Japanese invaders. ChatGPT translated it literally, which could lead to misunderstandings among readers.

In Example 9, ChatGPT’s translation, using “commemorating,” is highly appropriate. It accurately conveys the reverent and memorial tone of “Mianhuai.” Compared with “remembering” or even “cherish the memory,” the word “commemorating” evokes a more solemn, formal tone, making it particularly fitting for contexts like “Mianhuai Geming Xianlie.” It captures both the honor and ceremony intended in the original, which is why this version received the highest evaluation. Compared to the official translation “Pay tribute to the revolutionary martyrs and pass on the traditions of revolution”, the phrase “pay tribute to” is used to show respect, admiration, or gratitude for someone or something, often publicly. It is in an emotional, respectful, and sometimes personal tone. The word “commemorate” is used to remember and honor a person or event, especially with a ceremony, monument, or official observance. It’s often used with events, holidays, or historical remembrance. Therefore, the word is inappropriate here. ChatGPT cannot understand the deeper meaning of the phrase “Mianhuai,” and mistranslates it.

This type of error highlights the need for translators to thoroughly understand the cultural and historical context of China in the pre-translation phase, carefully considering the deeper meaning of the original text. This approach helps improve the quality of AI-generated translations and reduces the workload in post-translation editing.

4.2. Fluency dimension

In the MQM quality assessment framework, “fluency” refers to aspects of the translation that are closely related to the form and presentation of the text, rather than its meaning. It encompasses three major categories: content, conventions, and readability. The content-related dimension includes factors such as register, adherence to style guides, consistency, and ambiguity. The conventions category covers elements such as spelling, typography, grammar, and regional linguistic variations.

4.2.1. Fluency of translation

Although the translations of ChatGPT are generally capable of rearranging clause structures to highlight implicit logical relationships and avoid rigid, literal renderings caused by “linear translation,” the paratactic nature of Chinese, particularly its use of flowing sentence structures composed of layered coordinate clauses, often poses challenges to fluency. As a result, the translated output may suffer from reduced naturalness and coherence in the target language.

Example 10

ChatGPT: Wangjiayu Village was the location of the headquarters of the Eighth Route Army during the War of Resistance Against Japan. Over 70 years ago, the older generation of Chinese Communist Party revolutionaries

lived and fought here for an extended period, leading the guerrilla warfare in various anti-Japanese bases across North China. Today, it has become an important red tourism destination in Shanxi Province.

Reference: Home to the former headquarters of the Eighth Route Army, a major military force led by the CPC who fought the Japanese invaders more than 70 years ago, Wangjiayu is an important red tourism destination in Shanxi province.

Ellipsis of redundant elements is a common technique in Chinese-to-English translation to avoid Chinglish and enhance the readability of the translated text. However, due to the limitations of AI translation engines, the issue of repetition in the English translation is particularly prominent. In Example 10, in the translation provided by ChatGPT, 'Wangjiayu Village' is repeated twice, while the official translation only mentions it once. This does not align with the English language's preference for conciseness. Repetition in translations remains a significant issue that current AI translation engines struggle to avoid. Translating 'Suozaidi' as 'the location of' is also redundant. The official translation is concise and clear, whereas the ChatGPT translation is more verbose and complex.

Overall, one significant error type in current AI translation engines, such as ChatGPT, is their inability to effectively use lexical cohesion beyond pronouns to link discourse and enhance readability. To address this issue, it is necessary to employ superordinates, synonyms, near-synonyms, and general terms to connect the text, avoiding repetition and achieving a more elegant choice of words.

4.2.2. Grammar

Due to the differences in sentence structures between Chinese and English, when there are clear connectors between clauses in the Chinese text, AI translation can most effectively preserve the internal logic (although occasional issues such as the misuse or confusion of connectors may still occur). However, when there are no clear connectors in the original text, the error rate in the translation increases significantly.

Participles are widely used in specialized texts due to their more concise nature compared to clauses. However, when multiple participles are used together, the sentence becomes loose, with unclear information hierarchy and a stacking effect, which goes against English language expression habits. This kind of error is often influenced by the flowing sentence structure of Chinese. The corresponding preventive measures mainly include pre-editing the original text, adding appropriate Chinese conjunctions, merging short clauses, or restructuring sentences. These strategies can help reduce the error rate in AI translation and improve translation output efficiency.

The usage of tense is also a major issue in AI translation. Since ChatGPT sometimes fails to accurately grasp the underlying meaning of the text, it may misuse verb tenses. For example, in Example 11, the official translation employs the present continuous tense to indicate an ongoing action, while ChatGPT uses the present perfect tense instead — a clear instance of tense misuse. AI translation cannot truly interpret context, and it often makes errors in sentences involving mixed tenses or requiring inference of the subject's intent. The present continuous tense emphasizes actions that are currently in progress, while the present perfect tense (has done) highlights actions that have been completed but are still relevant to the present. The semantic distinction between the two is clear, and misuse can lead to a deviation in meaning. This sentence emphasizes that the Communist Party of China is currently leading the Chinese people toward new goals, rather than having already achieved them. Therefore, ChatGPT's translation contains a grammatical error in its use of tense.

Example 11

ChatGPT: In the past century, the Communist Party of China has delivered an outstanding answer to the

people and to history. Now, the Communist Party of China, united with the Chinese people, has embarked on a new journey toward achieving the second centenary goal.

Reference: Over the past century, the Communist Party of China has secured extraordinary historical achievements on behalf of the people. Today, it is rallying and leading the Chinese people on a new journey toward realizing the second centenary goal.

Example 12

ChatGPT: However, just as the vanguard of the Anti-Japanese Army was about to occupy the Tongpu Railway and actively prepare to move eastward into Hebei to engage directly with Japanese imperialism, Chiang Kai-shek sent more than ten divisions into Shanxi, cooperating with Yan Xishan to block the Red Army's route to resist Japan. He also ordered Zhang Xueliang, Yang Hucheng, and the Shaanxi-Ningxia forces to advance into the Shaanxi-Gansu Soviet area, thereby disrupting our anti-Japanese rear.

Reference: But when it occupied the Tatung-Puchow Railway and was energetically preparing to drive eastward into Hebei to engage the Japanese imperialists directly, Jiang Jieshi sent more than ten divisions into Shanxi and co-operated with Yan Xishan in barring its advance against the Japanese. He also ordered the troops under Zhang Xueliang and Yang Hucheng, as well as the troops in northern Shensi, to march on the Shanxi-Gansu Red area to harass our anti-Japanese rear.

Conjunctions, which function as function words linking words, phrases, or clauses, can generally be categorized into coordinating and subordinating conjunctions. Among these, coordinating conjunctions account for the highest frequency of errors in AI translation. In the data examined in this study, translations of ChatGPT most frequently made mistakes with conjunctions such as "and", "together with," and "as well as." As seen in Example 12, the use of 'and' in ChatGPT's translation is a direct translation of the Chinese word "Bing." However, from the perspective of English grammar, when listing more than two coordinated elements, the first and second items should be separated by commas rather than connected by conjunctions. Moreover, the conjunction 'as well as' places emphasis on the preceding element, which distorts the meaning of the original text. In Example 8, the translation error is directly related to the lack of precision in the Chinese source text. It is recommended that translators pre-edit the source text for clarity and also develop a solid understanding of function words to improve post-editing efficiency and produce high-quality translations.

5. Conclusion

This study conducts an empirical analysis of the performance of the Large Language Model ChatGPT in Chinese-to-English translation, using a self-compiled mini-corpus of specialized texts. The results indicate that ChatGPT can meet the basic translation needs for informational texts, but it still struggles with complex sentences. Translation errors mainly occur in two areas: "accuracy" and "fluency." Typical errors are manifested in three aspects: inappropriate terminology, inaccurate information, and lack of fluency in the translation.

Indeed, while the proliferation of artificial intelligence continues to drive the emergence of new translation technologies in the language services industry, "no machine translation system is currently capable of simultaneously reproducing the conceptual, discourse, and interpersonal meanings of a source text." Therefore, mastering translation technologies—especially the "AI translation + post-editing" model—and understanding how AI translation works can help translators quickly identify common errors. Moreover, by combining AI translation with the MQM quality assessment framework, translators can utilize the model's definitions and classifications of translation errors as a reference tool for targeted editing and refined translation output.

For example, post-editing can be used to clarify terminological concepts and enhance appropriateness through the use of contextually suitable terms; to eliminate ambiguity and improve the accuracy of the source content; to simplify the source text by removing redundancies and repeated phrases to improve readability; and to refine the use of cohesive devices and tighten sentence and paragraph structure. In addition, the MQM framework can also assist translation learners and even developers of translation technologies in identifying the typical weaknesses of AI translation systems, enabling them to proactively avoid common pitfalls and improve output precision. Ultimately, this allows AI translation to better support human translation practices.

Furthermore, to fully harness the capabilities of large language models in culturally and politically sensitive translation tasks, future research may incorporate larger, domain-specific corpora for model fine-tuning. Such corpora should include comprehensive annotations for key terms, historical references, and nuanced rhetorical devices frequently found in red culture texts. By integrating contrastive analyses across multiple AI translation engines, researchers could also investigate whether domain adaptation methods or hybrid human-AI workflows best mitigate error severity. As the industry evolves, establishing targeted translator training programs and dynamic evaluation standards will remain vital for ensuring that AI-assisted translations uphold both linguistic fidelity and cultural authenticity.

Funding

Graduate Innovation Project of Shanxi Normal University (Project No.: 2024XSY31); Philosophy and Social Sciences Planning Project of Shanxi Province (Project No.: 2024YB046)

Disclosure statement

The author declares no conflict of interest.

References

[1] Okpor M, 2014, Machine Translation Approaches: Issues and Challenges. International Journal of Computer Science Issues (IJCSI), 11(5): 159.

[2] Moorkens J, Castilho S, Gaspari F, et al., 2018, Translation Quality Assessment. Machine Translation: Technologies and Applications, 1: 299.

[3] Babych B, 2014, Automated MT Evaluation Metrics and Their Limitations. Tradumàtica Tecnologies de la Traducció, 2014(12): 464–470.

[4] Putri T, 2019, An Analysis of Types and Causes of Translation Errors. Etnolingual Journal, 3(2): 93–103.

[5] Lommel A, Burchardt A, Uszkoreit H, 2014, Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. German Research Center for Artificial Intelligence (DFKI), 2014: 455–463.

[6] Burchardt A, 2013, Multidimensional Quality Metrics: A Flexible System for Assessing Translation Quality. Proceedings of Translating and the Computer 35, Aslib, London.

[7] Li Y, Suzuki J, Morishita M, et al., 2024, MQM-Chat: Multidimensional Quality Metrics for Chat Translation. arXiv preprint: 2408.16390.

[8] Peng K, Ding L, Zhong Q, et al., 2023, Towards Making the Most of ChatGPT for Machine Translation. arXiv

preprint: 2303.13780.

[9] Salvagno M, Taccone F, Gerli A, 2023, Can Artificial Intelligence Help for Scientific Writing? Critical Care, 27(1): 75.

[10] Thorp H, 2023, ChatGPT Is Fun, but Not an Author. Science, 379(6630): 313–313.

[11] Wu J, Yang S, Zhan R, et al., 2025, A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions. Computational Linguistics, 2025: 1–66.

[12] Grimm D, Lee Y, Hu K, et al., 2024, The Utility of ChatGPT as a Generative Medical Translator. European Archives of Oto-Rhino-Laryngology, 2024: 1–5.

[13] Alm A, Watanabe Y, 2023, Integrating ChatGPT in Language Education: A Freirean Perspective. Iranian Journal of Language Teaching Research, 11(3 Special Issue): 19–30.

[14] Athanassopoulos S, Manoli P, Gouvi M, et al., 2023, The Use of ChatGPT as a Learning Tool to Improve Foreign Language Writing in a Multilingual and Multicultural Classroom. Advances in Mobile Learning Educational Research, 3(2): 818–824.

[15] Lommel A, Burchardt A, Popović M, et al., 2014, Using a New Analytic Measure for the Annotation and Analysis of MT Errors on Real Data. Proceedings of the 17th Annual Conference of the European Association for Machine Translation, 2014: 165–172.

[16] Lommel A, Gladkoff S, Melby A, et al., 2024, The Multi-Range Theory of Translation Quality Measurement: MQM Scoring Models and Statistical Quality Control. arXiv preprint: 2405.16969.

[17] Freitag M, Foster G, Grangier D, et al., 2021, Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation. Transactions of the Association for Computational Linguistics, 9: 1460–147.

[18] Laurer M, Atteveldt W, Casas A, et al., 2023, Lowering the Language Barrier: Investigating Deep Transfer Learning and Machine Translation for Multilingual Analyses of Political Texts. Computational Communication Research, 5(2): 1.

[19] Jiao H, Peng B, Zong L, et al., 2024, Gradable ChatGPT Translation Evaluation. arXiv preprint: 2401.09984.

[20] Guofu X, Jieli X, 2025, Theoretical Interpretation and Practical Path of Red Culture Empowering the Consolidation of the Awareness of the Chinese National Community. Journal of Sociology and Education (JSE), 1(1): 44–53.

[21] Gao Y, Wang R, Hou F, 2023, Unleashing the Power of ChatGPT for Translation: An Empirical Study. arXiv preprint: 2304.02182.

[22] Jiao W, Wang W, Huang J, et al., 2023, Is ChatGPT a Good Translator? Yes with GPT-4 as the Engine. arXiv preprint: 2301.08745.