A Study on ChatGPT-4o's Capacity in English-Spanish Translation

Zhien Han1,* , and Leyan Zhou2

1 School of Faculty of International College, science and technology University, Macau,999078, China
2 School of Faculty of Humanities and Social Science, City University, Macau, 999078, China
*Corresponding author: H23090103623@cityu.edu.mo

Abstract:

This study assessed the ability of ChatGPT-4o to translate Gabriel Garcia Marquez's novel "Love in the Time of Cholera" from Spanish into English. Twenty participants, all of whom were either native Spanish speakers or had majored in Spanish and studied the language for at least three years, evaluated the translation of twenty excerpts. The results indicated that participants were generally satisfied with the accuracy and fluency of the translations, feeling that the translated text conveyed the meaning of the original effectively. However, there was a consensus that there was room for improvement, particularly in handling complex expressions and cultural contexts. The study concluded that while ChatGPT-4o performed well in terms of accuracy and fluency, further improvements are necessary in cultural adaptation to ensure that cultural nuances and contextual information are accurately conveyed. This highlights the ongoing need for enhancement in machine translation to fully capture the subtleties of cultural references and intricate language use.

Keywords: ChatGPT-4o's; Translation's capacity; Spanish; English.

1. Introduction

With the deepening of globalization, cross-language communication has become increasingly significant. English is the international language of today's world and plays an important role in the world [1]. Therefore, it is a crucial tool for cross-language communication. In addition, according to the data, "Spanish is the third most spoken language in the world and the second most common language. In today's increasing international cross-language communication, the increasing use of Spanish has attracted the attention of many people. As a result, the Spanish language status has increased worldwide" [2]. Therefore, these two languages are very important, and the need for cross-language communication between English and Spanish is urgent.

In the previous cross-language communications, traditional translation tools, such as the Youdao dictionary, had to be used to communicate better. However, the traditional machine translation is only a process of changing one language symbol into another language symbol, so the system has two problems [3]. First of all, traditional machine translation is just a word-for-word translation. Therefore, when it comes to the polysemy of the word, it cannot be translated. Secondly, traditional machine translation only completes the translation task but does not relate to the article's content. Nowadays, with the continuous development of science and technology, generative AI tools are gradually replacing traditional translation tools and showing unprecedented charm in translation. As a valuable and quality generative AI tool, ChatGPT-4o has strong semantic parsing ability and context understanding ability, knowledge generation ability, and multi-language support ability, so it has superior automatic translation performance [4]. Moreover, ChatGPT-4o also has gaps. First, there are still imperfections in understanding and communicating cultural context and specific shades. For example, certain idioms, slang, or expressions in specific cultural contexts may not be accurately translated by ChatGPT-4o. The second is the grasp of the emotion and tone of literary works is still lacking, and the understanding of the emotional intensity and tone of the original text is still flawed in translation. By studying the ability of ChatGPT-4o in English and Spanish translation, the accuracy, fluency and so on can be improved, and the gaps can be filled. However, few specific studies have been on this topic over the past few decades.

This research explores accuracy, fluency, cultural adaptability, context consistency, and the reader's adaptability to generative AI tools in English and Spanish translation. In this article, the author will use ChatGPT-4o to research the ability to translate English and Spanish. Through this research, the author can further provide references for the progress of generative AI and understand the greater potential of generative AI translation to facilitate cross-language communication.

2. Literature Review

2.1 Accuracy and Fluency

Comparative evaluations, such as those by Wu et al. [5], have indicated that while GPT-3 excels in generating coherent and fluent text, its translations sometimes lack precision compared to specialized models like T5 and mT5. Google's T5 (Text-To-Text Transfer Transformer) model, as detailed by Raffel et al. (2020), reframes all NLP tasks into a text-to-text format, enabling robust translation capabilities[6].

One of the most notable advancements in generative AI is OpenAI's GPT-3. Research by Brown et al. [7] demonstrates GPT-3's capability to generate human-like text, including translation tasks. Despite its versatility, GPT-3's performance in translation has been mixed, often excelling in fluency but occasionally faltering in accuracy and fidelity to the source text.

2.2 Multilingual Translation Models

The advent of multilingual models like mBART and mT5 has also impacted translation studies. Liu et al. (2020) [8] discuss mBART's pre-training on a variety of languages, which enhances its ability to handle low-resource language pairs. Similarly, Xue et al. (2021) [9] highlight mT5's strengths in multilingual contexts showing notable performance improvements over single-language models.

2.3 Resource Efficiency

He, S. (2024). [10] "Prompting ChatGPT for Translation: A Comparative Analysis of Translation Brief and Persona Prompts." This study evaluates the effectiveness of using translation brief and persona prompts in ChatGPT for improving translation quality. Despite its generalist approach, offers a more balanced trade-off between computational cost and performance, making it a viable option for applications with limited resources.

2.4 Applications and Practical Implications

The practical applications of these models are vast, ranging from real-time translation services to aiding language learning and content localization. According to Torres [10], businesses leveraging these models for content localization report improved customer engagement due to more accurate and culturally relevant translations.

The comparative studies of generative AI tools for English-Spanish translation reveal a diverse landscape of capabilities and trade-offs. GPT-3 offers fluency and versatility, while T5 and mT5 provide superior accuracy and context handling. As research progresses, integrating these models into hybrid systems could potentially harness the strengths of each, paving the way for more advanced and reliable translation technologies.

3. Methods

Twenty participants evaluated ChatGPT-4o translations of twenty excerpts from Love in the Time of Cholera by García Márquez, a world-renowned Spanish writer. Love in the Time of Cholera is a literary classic containing complex syntactic structures and a rich vocabulary. It can challenge the processing capacity of translation systems and test the performance of artificial intelligence on complex sentences and polysemous vocabulary. The novel contains a great deal of specific historical context, and these elements must be accurately conveyed during translation. This helps to test the ability of the translation system to translate cross-culturally. As a literary work, Love in the Time of Cholera has a unique writing and narration style. The translation not only needs to convey the information accurately but also needs to maintain the literary style and emotion of the original text, which is an important test for the translation system. As Love in the Time of Cholera is well known and influential globally, its multilingual translations already exist to facilitate comparison and assessment of translation quality.

3.1 Accuracy

Objective: Assess whether AI-generated translations accurately convey the meaning of the original text, avoiding omissions, mistranslations, or additions.
Procedure:
Text Selection: A diverse set of texts in both English and Spanish, including technical documents, literary pieces, and everyday communications, will be selected.
Translation: Generative AI tools will be used to translate these texts.
Evaluation: Bilingual experts will compare the AI translations with human translations, identifying discrepancies as omissions, mistranslations, or additions.

3.2 Fluency

Objective: Evaluate whether the translations adhere to the grammatical and lexical norms of the target language, ensuring natural and smooth reading.
Procedure:
Review: Translations will be reviewed by native speakers of the target language.
Readability Tests: Readability tests will measure the ease with which the translated text can be read and understood.

3.3 Cultural Adaption

Objective: Determine if the translations appropriately consider cultural differences, effectively adapting or explaining culture-specific concepts.
Procedure:

Identification: Texts containing culture-specific references, idioms, and expressions will be identified.

Analysis: The AI's handling of these elements will be analyzed, noting if they are directly translated, adapted, or explained.

3.4 Contextual Consistency

Objective: Ensure the translations are coherent and consistent throughout the text, accurately reflecting the context and intent.

Procedure:

Translation: Longer texts and documents with complex contexts will be translated.

Evaluation: Translations will be evaluated for terminology, style, and intent consistency over the entire text.

3.5 Reader Appropriateness

Objective: Ascertain that translations are suitable for the target audience, considering their background knowledge and reading habits.

Procedure:

Audience Identification: The target audience for each text type (e.g., children, professionals, general public) will be identified.

User Testing: User testing will be conducted with representatives of the target audience to gather feedback on the translations' clarity, appropriateness, and effectiveness.

4. Results

4.1 English-Spanish

The experimental results indicate that participants rated ChatGPT-4o's translation of excerpts from "Love in the Time of Cholera" highly, showcasing its proficiency in both accuracy and fluency. This study involved ten participants who were either native Spanish speakers or had majored in Spanish with at least three years of study. They evaluated twenty selected excerpts translated by ChatGPT-4o from the novel written by Gabriel García Márquez. Participants expressed high satisfaction with the translations, noting that they accurately conveyed the original meaning and were natural and coherent when read in English. Most felt the translations accurately reflected the original text, with only minor errors or interpretive deviations noted. Fluency was another strong point, with participants widely recognizing that the translated text read naturally in English, almost free from grammatical errors or discontinuities. Participants unanimously agreed that the translations retained the original text's style and tone, capturing García Márquez's distinctive narrative style and ensuring the translated text resonated with the same emotional and artistic impact. While participants pointed out some challenges in translating complex expressions or cultural contexts, they generally felt these did not significantly impact the overall comprehension or quality. "The original Spanish word 'los amores contrariados' means lost or broken love, whereas the English translation here is' unrequited love', which is close in meaning but slightly different." More appropriate words like 'broken love' should be used." So commented one participant who grew up in Spain and read the English version of Love in the Time of Cholera. Suggestions for improvement included handling cultural expressions more detail and retaining the original text's rhythm and beauty in poetic expressions. Overall, the results demonstrate that ChatGPT-4o is proficient in translating "Love in the Time of Cholera," especially regarding accuracy and fluency. While improvements can be made in cultural and poetic details, the high quality of the translations indicates ChatGPT-4o's reliability as a tool for literary translation, highlighting the potential of AI in bridging language barriers and bringing literary works to a broader audience.

In summary, the experimental results indicate that ChatGPT-4o demonstrated a high level of proficiency in translating "Love in the Time of Cholera," especially in terms of accuracy and fluency, which the participants highly rated. However, there is room for improvement, particularly in handling cultural details and poetic expressions.

4.2 Spanish-English

After analyzing ten net friends' comments on ChatGPT-4o's translation capacity, I reached two conclusions. Firstly, for general Spanish learners, the translation ability of ChatGPT-4o is relatively good; general Spanish learners, the translation ability of ChatGPT-4o is relatively good; it can carry out effective and accurate translation of the text. For example, "The translation can convey the meaning of the original text with 95% accuracy, basically faithful to the original text, and there is no missing content." is written by an English and Spanish learner who passed English PETS4, CET6 and Spanish DELE C1; A reviewer who got seven scores on his IELTS and passed Spanish DELE C1 reviewed, "There is almost no problem of accuracy, the translation content is very accurate, and because the structure of the clause is very similar, I only feel a little stiff in the processing of a few long sentences and complex sentences, but the meaning is OK." In addition, it can also be done to match the background culture given in the prompt and ensure that the context is smooth and consistent. For instance, a reviewer who passed English CET6 and Spanish CET4 and obtained a Spanish tour guide certificate reviewed, "Cultural adaptability is generally good, and context is more coherent, and cohesion is better, logical, and detailed." ; A reviewer who learned

Spanish for two years and learned English as a child reviewed, "The cultural adaptation is quite good, and the translation has properly explained the concepts specific to the culture. Moreover, the translated content conforms to the coherence of the context, and there is no perceived inconsistency." More importantly, its translation results can make it easier for readers to adapt. As a reviewer who learned English and Spanish when she was a child reviewed, "Yes, it is adapted to target readers."; A reviewer who learned Spanish three years ago and learned English when she was a child reviewed, "Readers can adapt it." However, I came to a second conclusion after analyzing the comments of native Spanish speakers. ChatGPT-4o's translation ability is strong, and the content it translates is more human than traditional translation software, such as the Youdao dictionary. However, for native speakers, its translation is still lacking. ChatGPT-4o's translation prioritizes accuracy over consistency and lacks authentic language. For example, "It's an accurate translation, but it feels off at times. We're still lacking idiomatic expressions that would be used by a native speaker. Accuracy was placed before consistency." written by a reviewer who is a native Spanish speaker, and he obtained Advanced (CAE). The original hypothesis is that ChatGPT-4o's translation skill is excellent in accuracy, fluency, cultural adaptability, context consistency, and reader adaptability. However, this hypothesis is not established.

5. Discussion

5.1 English-Spanish

The feedback on retaining the original text's rhythm and beauty, particularly in poetic expressions, suggests another area for improvement. Literary texts often employ a specific rhythm and aesthetic that contribute significantly to their overall impact. While ChatGPT-4o managed to retain the general style and tone, further refinement is needed to capture the prose's more delicate and artistic elements. This might involve developing specialized training models focused on literary and poetic texts to better handle the subtleties of such language. While ChatGPT-4o has demonstrated a high level of proficiency in translating "Love in the Time of Cholera," the discussion highlights important areas for enhancement. Focusing on better handling of cultural nuances and improving the translation of poetic and artistic elements will further elevate the quality of AI-generated translations. This will not only improve the accuracy and fluency of translations but also ensure that a global audience fully preserves and appreciates the rich cultural and artistic value of literary works. The potential of AI in literary translation is significant, and with continued development, tools like ChatGPT-4o could play

a crucial role in making literature more accessible across language barriers.

5.2 Spanish-English

From the result analysis, the hypothesis is invalid because the author did not consider the idiomatic expression of a native speaker. In the review of this native Spanish speaker, "It's an accurate translation, but it feels off at times. We're still lacking idiomatic expressions that would be used by a native speaker. Accuracy was placed before consistency." shows that the translation of ChatGPT-4o is not adapt for a native speaker. It still lacks the human touch; it is only a mechanical translation. It puts accuracy before consistency, which is still different from the habitual speech of the native language. Therefore, the suggestions give AI more knowledge about local idioms and increase the diversity and quality of training data through continuous improvement of algorithms.

Due to limitations, this study only asked ten net friends to review the context translated from Spanish to English. Therefore, the study result can not universally reflect the capacity of ChatGPT-4o's translation. However, this imperfection does not influence the validity of the result because the Spanish native speaker at least reveals that ChatGPT-4o lacks local usage habits. In future studies, more Spanish-English bilinguals will be found.

6. Conclusion

In conclusion, through this study, the authors found two areas where ChatGPT-4o can do better. The first is its ability to translate literary texts. There will be certain rhythms and aesthetics in literary texts, and while ChatGPT-4o translation retains the style and feel of literature, it needs further refinement. Capturing the more artistic elements of prose may require the development of specialized training models. And focusing on literary and poetic texts can better translate the style of the article. Therefore, focusing on dealing with cultural differences and improving the translation of poetry and artistic elements will further improve the quality of ChatGPT-4o's translations. The other is whether the translation content conforms to the local language habits. From the evaluation results, it can be seen that the translation of chatgpt-40 does not conform to the language habits of native speakers, and its translation is still a little blunt and mechanical, so it does not feel like ordinary speech and lacks human touch. It puts accuracy before consistency in translation, which is still different from the customary speech of the mother tongue. Therefore, the authors' suggestion is that Chat-GPT-4o can learn more about local idioms and increase the diversity and quality of training data by continuously improving the algorithm.

Through this study, the authors identified two problems in ChatGPT-4o that could help it optimize its translation capabilities in the future and better facilitate cross-language communication.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Rao, P. S. (2019). The role of English as a global language. Research journal of English, 4(1), 65-79.

[2] Shan guoyang. (2012). Spanish: The Language of Conversation with God. College Enrollment (College Entrance Examination Guidance), (3), 32-32.

[3] Qin Meng, & Wang Xian, (2023). Artificial Intelligence, Linguistic features of machine and Human translation: A Case study of the novel To Live. Modern Linguistics, 11, 5169.

[4] Siu, S. C. (2023). ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. Available at SSRN 4448091.

[5] Wu, Z., et al. (2022). Comparative Analysis of Generative Models in Translation Tasks. *Computational Linguistics Journal*, 23-45.

[6] Raffel, C., et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 45-78.

[7] Brown, T., et al. (2020). Language Models are Few-Shot Learners. *arXiv preprint arXiv:2005.14165*, 12-34.

[8] Liu, Y., et al. (2020). Multilingual Denoising Pre-training for Neural Machine Translation. *Proceedings of the Transactions of the Association for Computational Linguistics*, 78-102.

[9] Xue, L., et al. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. *Proceedings of the Conference on Computational Linguistics*, 102-134.

[10] He, S. (2024). Prompting ChatGPT for Translation: A Comparative Analysis of Translation Brief and Persona Prompts. In *Proceedings of the Conference on Computational Linguistics*, 34-56.