Evaluating ChatGPT's Translation of Chinese Idioms in Subtitles: A Cognitive Approach Through the Case Study of The Story of Minglan

Haiou LIU[1*], Xinlei LI[2]

[1]College of Foreign Languages and Cultures, Chengdu University, Chengdu, China https://orcid.org/0000-0003-1314-8824

[2]College of Foreign Languages and Cultures, Chengdu University, Chengdu, China

***Corresponding author:** Haiou LIU

College of Foreign Languages and Cultures, Chengdu University, Chengdu, China

Abstract

This study evaluates the effectiveness of ChatGPT 4.0 in translating Chinese idioms into English subtitles from the perspective of cognitive translatology, focusing on the case study of the TV drama The Story of Minglan, which is adapted from an online novel. With the evolution of AI in translation, assessing the quality of subtitles, especially for content rich in cultural metaphors, becomes crucial. By analyzing ChatGPT's translations against traditional machine translation tools, and integrating both BLEU scores and manual evaluations, this research aims to understand ChatGPT's ability to capture metaphorical meanings and cultural nuances. Findings suggest ChatGPT outperforms other AI models in translating idioms, demonstrating a strong capability to recognize implicit metaphors and successfully achieving category conversion between languages. This study highlights ChatGPT 4.0's potential in subtitle translation, advocating for its use complemented by human editing to bridge cultural gaps and enhance the global sharing of audiovisual content.

Keywords: *subtitle translation, idioms, ChatGPT, metaphors, cognitive translatology*

1. Introduction

Audiovisual works, including TV dramas, movies, and documentaries, stand as the predominant form of entertainment, serving as pivotal conduits for the dissemination of local lifestyles, cultures, and values (Wang Yin, 2020), thereby facilitating cross-cultural comprehension. In this process, subtitles emerge as critical in transmitting information, particularly when the content is deeply intertwined with a community's culture, history, or geography, necessitating additional efforts from the subtitler (Cintas &

Remael, 2021, p. 202). Consequently, the translation of subtitles assumes significant importance.

Recently, AI translation models, with ChatGPT 4.0 leading the charge, have markedly influenced the translation industry. An increasing number of translation tasks, including the subtitling of movies and TV shows, are now being executed with the aid of, or entirely by, AI translation tools. Sun Lin (2023, p. 37) asserts, "The

interpretation of translation continues to evolve, yet the quality of translation remains a pivotal concern, often sparking considerable debate." In the specific context of subtitle translation, it becomes imperative to address not only linguistic issues but also the intricate metaphors underlying the dialogue. Lakoff and Johnson (1980) have demonstrated that metaphors permeate not just language but our thoughts and actions, asserting that metaphorical and metonymical frameworks are essential to human cognition. Given that movies and TV shows encapsulate a wide range of contextual realities, their dialogues frequently incorporate rich metaphorical thinking. Therefore, evaluating the quality of AI-generated subtitle translations necessitates a critical examination of their capacity to discern and accurately convey implicit metaphors.

2. Research Purpose and Research Questions

In an effort to explore how to actively adapt to the advancement of AI translation models and thereby offer guidance and inspiration for their application in subtitle translation, we have employed the GPT-4 version, using idioms from the dialogues in the drama *The Story of Ming Lan* as materials. Through a comparative analysis of the translations from ChatGPT with those from other machine translation tools, integrated with BLEU scores and manual evaluation scores, we aim to comprehensively assess the translation quality of ChatGPT from various perspectives to explore its potential applicability in subtitle translation. Therefore, we will also evaluate the quality of the subtitles based on the model developed from Künzli (2020)'s CIA model of subtitle quality.

This study sets out to answer the following questions:

1) From the perspective of Cognitive Translatology, how effective is ChatGPT at identifying implicit metaphors in the English translation of idioms?

2) How good is the quality of ChatGPT 4.0's idiomatic English translations (English translations of Idioms) in terms of CIA?

3) How promising is the application of AI translation models such as ChatGPT 4.0 in the field of subtitle translation?

3. Cognitive Understanding of Chinese Idioms

3.1. Implicit Metaphorical and Metonymical Characteristic of Idioms

Based on the core principle of cognitive linguistics, namely, that reality determines cognition, which in turn determines language, the selection of existing vocabulary for conceptualization and lexicalization occurs in the human brain after processes of real-life experience and cognitive processing by a given ethnic group. Regardless of the language, its vocabulary is always finite. However, the objective world humans face, along with the objective things and phenomena to be expressed through language, is infinite. Therefore, in the process of forming cognitive significance and concept integration, people cannot avoid the influence of subjective imagination, eventually seeking similarities within the same domain or across different domains in the existing vocabulary to extend the usage of words. Idioms, as a form of semantic condensation in Chinese, naturally follow this process in their formation. They are defined as fixed phrases or expressions used over a long period in fables, becoming conventional, typically having a specific structure and components, a particular meaning,

not meant to be taken literally, and function in a sentence similarly to a single word. Idioms are conventional, derived from the conceptual integration and lexicalization of a vast amount of real-life experience by an ethnic group. Due to the strict requirements on the syntactic structure of Chinese idioms and the limited number of characters, the syntactic relationships among characters in the source are not all reflected in the condensed structure of the idiom, leading to varied interpretations of the internal syntactic relationships. It is believed that idioms can serve as representative linguistic phenomena of metaphorical thinking in Chinese. Therefore, idioms in the dialogue of the TV series "The Story of Minglan" are selected for sample translation analysis.

3.2. Cultural Difference

Building on Chomsky's theories of "Universal Grammar" and "Generative Grammar", Wang Yin (2021) introduced the concept of "Experiential Universality". This concept parallels Chomsky's belief in the innate human ability to learn languages, asserting that humanity shares a common objective world and physical structure, which naturally leads to shared thoughts and languages across different ethnicity. "We possess a fundamentally similar reality, which leads to broadly similar thought processes. This experiential cognition forms the cognitive basis for translatability between different languages." Additionally, variations in "ways of cognition" acquired later in life lead to divergent outcomes in conceptual integration and lexicalization.

Table 1 The universality of experience

If translation relies solely on a linguistic level, employing literal methods to convey cognitive meaning, differences in real-life experiences and cognitive processing will inevitably lead to variations in metaphorical thinking between the source and target languages. Such discrepancies can result in errors as the translator constructs the relationship between the "source domain" and "target domain", ultimately leading to significant ambiguity in the translation. For example, the idiom "力大如牛" is more accurately translated as "as strong as horses", since "牛" (ox) symbolizes strength in Chinese culture, while in Anglo-American contexts, the "horse" is the more typical emblem of strength. The varied perceptions of the objective world among different ethnic groups lead to distinct modes of linguistic expression, highlighting a significant aspect of cultural diversity. Idioms represent not only a unique expressive form within the Chinese language but also a vital element of Chinese culture. During the translation process, it is crucial to begin at the linguistic level and account for the variances in cultural imagery between the source text and the target language. Accounting for these considerations is indispensable in evaluating the quality of AI-assisted subtitle translation.

4. Research Design

4.1. BLEU

Introduced by Papineni et al. in 2002, BLEU (Bilingual Evaluation Understudy) serves as a metric for assessing translation quality. This metric evaluates translation accuracy through comparison between machine-generated translations and human translations. BLEU fundamentally relies on n-gram matching, assigning scores from 0 to 1, where higher scores denote superior translation

quality. However, BLEU possesses inherent limitations, notably its neglect of semantic meaning and word order. Consequently, without revealing the sources of the translations or the AI models used, this study engaged two experienced translation teachers to independently evaluate the outputs of three AI translation models.

4.2. CIA

The CIA (Correspondence-Intelligibility-Authenticity) model, developed by Künzli in 2020 through survey questionnaires, stands as a benchmark for assessing the quality of contemporary film and television subtitle translations. This study utilizes three dimensions—Denotational, Connotational, and Pragmatic Correspondence—to further assess their effectiveness in subtitles, drawing on evaluations of AI model translation quality.

AI	1	2	3	4	5	6	7	8	9	10	11	Average	
Starkdesk	11.36	≈0	≈0	≈0	≈0	14.95	33.34	18.58	5.37	3.8	≈0	8.0	115.8
DeepL	≈0	≈0	≈0	3.3	≈0	4.08	20.21	9.55	5.37	≈0	≈0	3.9	35.6
ChatGPT4.0	16.99	3.3	3.93	≈0	7.73	≈0	3.8	2.28	5.37	≈0	17.78	5.7	23.4

Table 2 The scores of BLEU

AI	1	2	3	4	5	6	7	8	9	10	11	Average
Starkdesk	8	3	3	9	3	5	3	6	5	6	5	5.09
DeepL	4	3	3	3	3	5	5	6	5	7	5	4.45
ChatGPT4.0	9	7	8	9	8	8	8	8	8	6	9	8.00

Table 3 The Scores from the First Teacher

AI	1	2	3	4	5	6	7	8	9	10	11	Average
Starkdesk	8	6	6	6	4	6	6	4	8	6	4	5.82
DeepL	6	6	6	6	6	8	4	6	8	6	4	6.00
ChatGPT4.0	10	8	8	8	8	8	8	8	8	8	6	8.00

Table 4 The Scores from the Second Teacher

5. Result and Analysis

5.1. Stability

According to the BLEU translation quality assessment criteria, the results suggest that Starkdesk's translations are comparatively superior, followed by DeepL, while ChatGPT4.0 ranks lowest in quality. However, these findings stand in direct contradiction to the outcomes of human evaluations. The paper further scrutinizes the limitations of the BLEU assessment method, noting its heavy dependence on the quality of reference translations and its neglect of sentence structure and specific meanings. Specifically, culturally dependent expressions such as idioms necessitate flexible translation approaches to encapsulate their cognitive meanings effectively. Consequently, the study calculates variance data for the three AI translation models as a measure of stability. ChatGPT 4.0 exhibited the least variance, in contrast to Starkdesk, which recorded the highest, suggesting that ChatGPT 4.0's translation quality is comparatively more stable. Furthermore, in the assessment of 11 idiomatic translations, while all three AI models have scored nearly 0 for at least once, ChatGPT 4.0 received the least number of such low scores, with only three occurrences. Considering the outcomes of human evaluations, this paper proposes that regarding idiom translation, ChatGPT4.0's stability and quality exceed those of the other two models.

5.2. Capability of Identifying Metaphors

Human evaluation scores clearly demonstrate that ChatGPT 4.0 has superior metaphor recognition capabilities compared to the

Ultimately, the mean of the human evaluations, when aggregated with the BLEU scores, facilitates a comprehensive comparison of the translation capabilities of the three AI models, thereby allowing for a multidimensional assessment of translation quality from both machine and human evaluative lenses.

other two AI models. The term "青眼相加" has been literally translated to "eyes looking straight at each other", metaphorically conveying affection and respect. ChatGPT 4.0 interprets this as "not despising me, showing favor and kindness", while DeepL and Starkdesk translate it respectively as "they don't dislike me" and "don't mind me, showing favor". Idiom translation encompasses three dimensions: linguistic, cognitive, and contextual. ChatGPT 4.0 not only exhibits a profound understanding of the original text but also conveys the fundamental cognitive meaning with accuracy and clarity. This underscores ChatGPT 4.0's robust capability to ensure the accurate rendition of cognitive meanings, a critical aspect for AI models in the translation of idioms.

5.3. Category Conversion

Wen Xu (2020) distinguishes between linguistic and non-linguistic category conversions. Observations indicated that ChatGPT4.0 demonstrated proficiency in applying both types of category conversion patterns. An instance involves "信口雌黄", which ChatGPT4.0 rendered as "based on groundless accusations". Words with high cultural dependency, like "雌黄" are highly reliant on cultural and social contexts, anchoring themselves within a specific culture. Once removed from their specific cultural setting, these words lose their meaning. This phenomenon is most commonly seen in the concept of category vacancies in the context of high cultural dependency. Realgar, or orpiment, is a yellow mineral

used as a pigment. Ancient people wrote on yellow paper, and when mistakes were made, they used realgar to cover them up for corrections. In the idiom "信口雌黄", realgar, through the mechanism of metonymy, represents baseless statements, suggesting they are subject to arbitrary alteration.

In a nutshell, The adept mastery of these category conversion patterns allows ChatGPT4.0 to thoroughly analyze the linguistic structure of the source category and elaborate translations based on conceptual and experiential frameworks.

5.4. Cultural Sensitivity

Wen Xu (2020) proposes categorizing words into low and high cultural dependency based on their extent of reliance on cultural context and social environment. Words with low cultural dependency often exhibit cross-linguistic commonalities, effectively managed by these three AI models at linguistic or cognitive levels, as illustrated by "turning black and white upside down" and "adding fuel to the fire". Taking "火上浇油" for example, when translating low culture-dependency words like "fire" and "oil"", translators only need to convert these words from the source language to the corresponding words in the target language, ensuring conceptual consistency without worrying about the impact of cultural differences on the accuracy of the translation.

In contrast, words with high cultural dependency demonstrate a greater reliance on cultural context and social environment. For example, when translating "铁面菩萨", Starkdesk and DeepL primarily adopt a linguistic perspective, rendering it as "an iron-faced Bodhisattva". This method may create comprehension barriers for target language readers, particularly those unfamiliar with Buddhist culture, thereby complicating understanding. In Chinese, "Bodhisattva" symbolizes kindness and compassion, exemplified by "She helped many impoverished children, truly a living Bodhisattva". Readers not acquainted with this cultural imagery might find a literal translation confusing or misleading, requiring additional explanation. ChatGPT4.0, from a cognitive significance standpoint, interprets "铁面" as "impartia", reflecting an understanding that in Chinese culture, "铁面" conveys not only the literal meaning of having an iron-like, emotionless face but also the metaphorical meaning of being strict and impartial. Thus, by translating "铁面菩萨"" as "an impartial Bodhisattva", it captures both the literal and cultural nuances, facilitating a deeper comprehension among readers not familiar with the original cultural context. This approach underscores the importance of cognitive translation studies, where understanding the interplay between language, cognition, and culture is crucial for achieving more accurate and culturally sensitive translations. Cognitive linguistics supports this by emphasizing the role of human cognition in the process of language understanding and translation, advocating for translations that consider not just the literal meanings of words, but their cultural and cognitive dimensions as well.

6. The Analysis for the Translations of Subtitles

In summary, we utilized BLEU scores and human evaluations to assess the quality of idiom translations by three AI models. Overall, we regard ChatGPT 4.0's performance as superior. However, our selected reference texts, sourced from movies and TV shows, contain a wealth of contextual reality. The efficacy of subtitle translation cannot be solely judged by the linguistic and cognitive processing of individual idioms. As Beaugrande (1980:

13) suggests, "The basic unit of translation is not the word or the sentence, but the text." I posit that each scene in a movie or TV show should be considered a collection of statements, namely, a text. In translating a dialogue, one must begin with the reality level, considering the entire text, namely, the contextual reality in which the dialogue appears. For this reason, we conducted our evaluation based on the CIA model as proposed by Künzli (2020). Furthermore, we adopted three dimensions (Denotational, Connotational, Pragmatic) to comprehensively evaluate how well ChatGPT4.0's idiom translations function as subtitles:

1) Using the idiom "树大招风" as an example, the dialogue containing this idiom occurs in a scene where the character Zhao Zongquan is compelled to enter the competition for the throne. Although Zhao Zongquan had no intention of participating, other royal families, as a precaution, dispatched assassins to eliminate him. Consequently, Gu Tingye describes the other royal houses as "tall trees catching the wind", implying that their persistent malicious actions would lead to a loss of the current emperor's trust and favor. This ultimately results in the throne being handed over to Zhao Zongquan. ChatGPT4.0, with its advanced metaphor recognition capabilities, interprets this as "those who attract trouble due to their prominence". From a semantic standpoint, this interpretation reveals the implicit meaning behind the characters' words: other royal houses invite disaster through their impatience and aggressiveness. Pragmatically, it also communicates the characters' intended message of maintaining optimism about their situation, foreshadowing plot developments.

However, from a denotational perspective, coordination between subtitle translation and visuals must be considered. The writer suggests the translated version for "树大招风" as "Tall trees catches the wind, as those who attract trouble due to their prominence". Adding "Tall trees catches the wind" can evoke a tense and serious atmosphere through its literal meaning, making the imagery accessible to readers of the target language. Retaining the literal meaning of "catching the wind" partially enhances the visual experience, engages the audience's senses more fully, and immerses them in the atmosphere of the film or series. Alternatively, beginning with a realistic perspective and blending the metaphorical meaning of "wind" with the film or series' contextual reality can clarify the situation, achieving a similar effect.

2) Taking "班门弄斧" as an example, three AI models interpreted its cognitive meaning respectively as "displaying our incompetence", "acting as mere fiddlers" and "being amateurs in comparison". This approach, demonstrating the AIs' strong cultural sensitivity, captured the culturally specific imagery of "Lu Ban" as "master" and effectively conveyed its metaphorical meaning. However, considering the intrinsic function of subtitles, they also serve the pragmatic role of conveying culture. Therefore, when such highly culture-specific terms appear in subtitles, translating them literally at the linguistic level and providing annotations for the unique cultural imagery they contain is a more effective approach. Although this method may somewhat increase the audience's reading difficulty, it can stimulate thought and deepen understanding of the source culture, enabling

subtitles to fulfill a more diversified role and align with the practical significance of films and series going international.

7. Conclusion

Following a comprehensive analysis of three AI translation models, ChatGPT4.0 excels in translating idioms within dialogues into English. Its prospects for film and television subtitle translation appear promising. Given its superior ability to recognize implied metaphors, ChatGPT4.0 can effectively translate most dialogues from linguistic, cognitive, and practical perspectives. Furthermore, ChatGPT4.0 can adeptly capture diverse cultural imagery in dialogues, performing suitable category conversions to prevent ambiguity and misunderstanding among the audience. However, ChatGPT 4.0 exhibits limitations in dialogue translation effectiveness, requiring manual editing and optimization by translators, particularly in integrating the contextual nuances of the film or series, ensuring alignment with visuals and plot, and enhancing coherence. When handling dialogues with unique cultural imagery, choosing to directly translate culturally specific terms first and add annotations permits the language to fully convey the essence of the national culture, underscoring the cultural transmission role of the film or series. The international dissemination of films and television series has long been a crucial channel for projecting a country's culture abroad, with subtitles playing an indispensable role in this process. The quality of their translation can significantly influence the effectiveness of cultural dissemination. Currently, ChatGPT 4.0 has demonstrated significant potential in film and television subtitle translation, with the capacity to markedly improve translation efficiency and reduce labor costs. Nevertheless, faced with challenges such as cultural differences and the comprehension of complex contexts, ChatGPT4.0 requires further optimization and enhancement. Looking ahead, subtitle translators should leverage ChatGPT4.0's powerful capabilities for effective post-translation editing, brightening the prospects for its use in film and television subtitle translation.

References

1. Beaugrande. (1978). *Factors in a Theory of Poetic Translating*. Assen: Van Gorcam.
2. Cintas, J. D., & Remael, A. (2021). *Subtitling: Concepts and Practices*. London and New York: Routledge.
3. Künzli, A. (2020). From inconspicuousness to flow: The CIA model of subtitle quality. *Perspectives*, 28 (3), 1-13.
4. Kishore, P. Salim, R. Todd, W. and Wei Jing, Z. (2002) BLEU: a Method for Automatic Evaluation of Machine Translation. InIsabelle,P.et al. (eds.). Proceedings of the 40th Annual Meeting on Association for Computational Linguistics[C]. Philadelphia, USA: Association for Computational Linguistics: 311-318.
5. Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. Chicago: University of Chicago Press.
6. Liu Jiexiu. (1985). *Chinese Idioms*. Beijing: The Commercial Press.
7. Ming，H. (2014). The reconstruction of syntactic relations in idioms from the
8. perspective of syntactic iconicity. *Dong Yue Tribune*.
9. Shi Shi. (1976). *The Research on Chinese Idioms*. Sichuan: Sichuan People's Publishing Co., Ltd.
10. Sun Lin. (2023). On the assessment of translation quality. *Shanghai Journal of Translators*.
11. Wen Xu, Si Weiguo (2020). Category Shift in Translation Related to Embodied Cognition and Iconicity. *Shanghai Journal of Translators*.
12. Wang Yin. (2021). *Embodied-Cognitive Translatology*. Beijing: Peking University Press.
13. Wang Wei. (2020). Analysis on the overseas communication of Chinese films and TV dramas. *China Radio & TV Academic Journal*, 20, 85-88.