

AI and Audiovisual Translation: Evaluating the Roles of GenAI Platforms and Prompts in Automatic Subtitling

1. Introduction

Over the past decades, technology has played an important role in shaping the production, distribution, consumption and reception of audiovisual translation (AVT). As a quick example, on conventional media such as TV, subtitle presentation variables (e.g., speed) need to be carefully controlled so that viewers have sufficient time to read them. However, on streaming platforms, given the technological affordances, viewers can easily pause and re-watch the videos, so they tend to be less concerned about the subtitle presentation variables (Wu & Chen, 2022). This has implications for subtitlers, who may feel less obliged to implement the condensation strategy to accommodate viewers' reading speed. Zooming in on subtitling practice, machine translation has been adopted by companies to produce subtitles that are subsequently post-edited by subtitlers. This practice is supposed to bring productivity gains, but research has found that the quality is unsatisfactory (Hagström & Pedersen, 2022). However, the recent technological advancements have led some scholars to believe that Generative AI (GenAI) can “match or surpass dedicated translation systems like Google Translate and DeepL” (Lee, 2023, p. 2), thus bringing us immense potentials to deliver satisfactory automatic translation. Our study attempts to examine GenAI's performance, as translation quality has been one central issue in evaluating the roles of machine translation in subtitling (Burchardt et al., 2016; Karakanta, 2022). We hope to show the extent to which GenAI platforms and prompt engineering matter in automatic subtitling and draw some implications for subtitlers if they wish to harness the GenAI technology to a greater or lesser degree. In what follows, we will first review some typical issues in interlingual subtitling and GenAI's potentials for automatic subtitling. Then, we will present a case study comparing the subtitles by fansubbers and two GenAI platforms. Based on the findings, we will discuss the roles of GenAI platforms and prompts in automatic subtitling.

2. Subtitling and AI technology

2.1 Common challenges in subtitling

Subtitling is seldom a fully literal translation of the source text (Guillot, 2020). Given the constraints of time and screen space for subtitles, alongside the multimodal complexities of audiovisual content, challenges typically encountered in other text genres would only intensify in subtitling. AVT researchers have extensively studied key subtitling challenges such as puns (Aleksandrova, 2019), cultural references (Pedersen, 2007), humour (Pozo, 2021), and features of oral discourse (Chaume, 2004). It would not be possible to deal with all subtitling issues within a single chapter, so we limit our focuses to three subtitling challenges commonly discussed in the literature: wordplay, multimodal humour, and orality features.

Wordplay, such as puns and polysemous words, poses considerable challenges for subtitlers. Chiaro (2010) highlighted the difficulty of translating humour based on linguistic wordplay. As it is often rooted in the linguistic and culture-specific features of the source

language, an equivalent translation is difficult to find. Schauffler (2015) described the subtitling of wordplay as a “dilemma between the maintenance of formal fidelity to the source text...and the preservation of comic value by selecting the funniest option” (p. 230). In her reception experiment, Schauffler (2015) compared the effect of two subtitle versions of humour: one target-oriented, prioritising the transfer of humour, and the other source-oriented, more faithful to the source dialogue. The results showed that the target-oriented subtitles were better received by the target audience, highlighting the importance of creativity in subtitling wordplay.

In addition to verbal humour from wordplay, multimodal humour presents another hurdle for subtitlers. Multimodal humour generally refers to comic effects arising from the interaction between different semiotic modes, such as verbal and visual elements. For example, Barrera-Rioja (2023) examined how visually restricted wordplay/humour was rendered in the target subtitles of the sitcom *Modern Family*. It was found that literal translation and substitution were the two most common strategies, while recreation was the least common one. Relatedly, Lautenbacher (2012) found that losing the link between verbal and visual humour could alter the gaze patterns of target viewers, suggesting that a target-oriented approach should be considered when subtitling multimodal humour.

When it comes to representing orality features like repetitions or incomplete utterances in subtitles, there are different perspectives among AVT researchers. Written texts, being more formal and requiring longer processing times (Díaz-Cintas & Remael, 2021), often lead subtitlers to omit features of oral discourse for the sake of brevity (Chaume, 2004). However, in fictional genres, such as films and TV series, these orality features are intentionally included to simulate spontaneous conversations and enhance character development. In this context, retaining certain elements of “prefabricated orality” (Baños Piñero & Chaume, 2009) in translations can preserve the authenticity and naturalness of the original dialogues. Thus, while subtitling necessitates a balance between textual brevity and retention of orality features, the decision to include such elements in subtitles should be strategically guided by genre-specific demands.

2.2 Potentials of AI technology in automatic subtitling

The increasing sophistication of Large Language Models, exemplified by GenAI platforms like ChatGPT, has significantly transformed the field of communication and translation. While the application of machine translation to subtitling has been extensively examined in AVT research (e.g., Bywood et al., 2017; Karakanta, 2022), the application of newer GenAI tools in subtitling has received relatively less attention. Tian’s (2023) examination of machine-generated versus human-translated subtitles demonstrates a crucial gap. From a qualitative comparison of the translation efficiency offered by DeepL Translator, Youdao Translation, and ChatGPT (based on GPT-3.5), she observed that these systems often fell short in delivering the accuracy and conciseness expected in high-quality subtitles. This finding supports the ongoing indispensability of human subtitlers in ensuring quality translations. Conversely, Calvo-Ferrer’s (2023) reception experiment presents an interesting counterpoint. The inability of viewers to distinguish between human and ChatGPT-translated subtitles suggests that GenAI can closely mimic human-like translation outputs. However, viewers tend to rate subtitles as lower quality when they believe they are created by non-human translators, highlighting the reception challenges faced by GenAI in AVT tasks.

These conflicting findings point to significant research gaps in current GenAI applications in subtitling. First, the issue of prompt design is crucial. The simplistic prompts used in these studies may not fully harness the capabilities of GenAI, since prompts with different sets of details (e.g., persona prompts vs. translation brief prompts) may lead to varying translation quality (He, 2024). Hence, it is worth exploring more complex, context-rich prompts for improving subtitling performances. Second, the absence of visual information in automatic subtitling is a notable limitation of the reviewed studies. Since non-verbal cues are integral to the audiovisual narrative (Chen & Wang, 2019), excluding them from GenAI’s translation process constrains the accuracy and multimodal congruence of the target subtitles. Thus, there is a need to compare the performances of AI models capable and incapable of processing dynamic texts so that subtitlers can make informed decisions to maximize the potentials of AI technologies.

3. A comparative case study of fansubbers, ChatGPT and Gemini

3.1 Research design

The primary aim of this study is to assess GenAI’s performances in overcoming common subtitling challenges. To this end, the study adopts a case study design and compares subtitles produced by fansubbers, ChatGPT and Google Gemini for a selected scene from *Modern Family* (see Section 3.2 below). The fansubbed subtitles are from YYeTs, one of the most well-known fansubbing communities in China (Wu, 2017). Their subtitles for *Modern Family* are popular among Chinese audience, as indicated by the large number of downloads from the subtitle sharing website *Zimuku*. The ChatGPT model used in this study is GPT-4, representing one of the most sophisticated language processing and generation models. A third source of subtitles is from Google Gemini 1.5 Pro, recognised for its capability in processing dynamic texts like videos. These three sources can allow us to conduct comprehensive comparative analysis on subtitles generated by humans and GenAI models (in)capable of processing videos.

3.2 Materials

The source material for the case study is from an American sitcom, *Modern Family*. The sitcom is renowned for its witty wordplay and multimodal humour (Barrera-Rioja, 2023), thus presenting common subtitling challenges to subtitlers. The selected scene is from Episode 10, Season 5. It features Cam, Mitch, and their adopted daughter Lily. The gist of the scene is as follows: Lily expresses her excitement about her desired Christmas gift—a “Puppy Pound,” a toy shelter for puppy dolls. Mitch is confused by Lily’s desired gift; and Cam finds out that in the wish list Lily misspelt “Puppy Pound” as “Guppy Pond” (the gift Mitch prepared for Lily). Cam holds up his index finger, emphasising that Lily wants only “one” gift but someone didn’t get it. He slowly bends down the finger and points it to Mitch. Although Mitch interprets Cam’s gesture as accusatory, Cam jests that the finger merely represents the number “one” and humorously claims it just “fell asleep.”

The scene encapsulates several typical interlingual subtitling challenges: (a) wordplay (the mix-up between “Puppy Pound” and “Guppy Pond”); (b) multimodal humour (Cam’s pointing gesture and his explanation); and (c) orality features (e.g., repetitions and incomplete utterances,

especially evident when Mitch is confused). These elements provide a rich context for evaluating GenAI's capabilities in translating complex, humour-infused dialogues.

3.3 Prompt engineering

Prompt engineering, the practice of designing inputs for AI tools to generate desired outputs or responses, is a critical aspect of GenAI interaction and performance (Liu et al., 2023). Effective prompt design has been shown to improve the quality of GenAI translations. This includes integrating contextual information about the source text (Peng et al., 2023) and specifying the translation's purpose and target audience (Yamada, 2023). With reference to previous prompt designs for translation tasks (Calvo-Ferrer, 2023; He, 2024; Tian, 2023), this study developed and tested six prompts. The first prompt (P1B) was the basic one, giving general translation instructions and quality standards typical in subtitling, such as naturalness, conciseness, and coherence (Künzli, 2021). The second prompt (P2C) was more detailed and provided contextual information about the source material. Built on the basic prompt, the third one (P3F) instructed GenAI to function as a fansubber, while the fourth one (P4P) set the role of a professional subtitler, both detailing the purpose and target audience of the translation. The fifth prompt (P5CF) combined P2C and P3F, while the sixth prompt (P6CP) integrated P2C and P4P. These combinations aim to explore if additional contextual and role-specific information within a single prompt can lead to higher-quality subtitles. As we tested both ChatGPT and Gemini, the prompts were identical except that the Gemini prompts had one additional requirement to consider the visual information. The full prompts are presented in the Appendix.

3.4 Analytical procedures

This comparative case study was carried out in three main stages: subtitle preparation and generation, *etic* comparative analysis, and *emic* comparative analysis. The difference between the latter two is that *etic* analysis was conducted from the researchers' (outsider-like) perspective, while *emic* analysis was performed by the researchers assuming the role of fansubbers (an insider-like perspective).¹ The two perspectives combined can afford a fuller understanding of GenAI's potentials for automatic subtitling.

To ensure comparability, English subtitles with time codes in the SRT format were prepared from YYeTs' fansubbed version. To generate automatic subtitles, each designed prompt was tested separately in a new chat interface, minimising carryover effects from previous interactions. For ChatGPT, the subtitling procedure involved initialising a new chat session for each prompt and inputting the English subtitles paired with their time codes as one continuous, self-contained text block. This was done to allow ChatGPT to consider the entire scene when translating the subtitles (Calvo-Ferrer, 2023). Gemini followed a similar procedure, but an additional element tailored for Gemini was added to the ChatGPT prompts (see the Appendix). Accordingly, the source video file was uploaded to help Gemini understand the visual context.

After obtaining the automatic subtitles, we conducted *etic* comparative analysis by focusing on three subtitling challenges (wordplay, multimodal humour, and orality features).

¹ For the discussion about *etic* and *emic* perspectives, see Rose et al. (2020).

We aligned the English source texts with their corresponding target texts produced by fansubbers, ChatGPT and Gemini. Then, we compared the strategies and communicative effects in each target text and tally the number of acceptable translation by prompt and by AI model. Both authors/researchers discussed the results until complete agreement was reached.

Upon the evaluation of GenAI performances, we noticed that wordplay was a difficult issue for both human and GenAI translators. Thus, we assumed the role of fansubbers and conducted emic analysis on the possibility of GenAI-assisted creative solutions. We used one of the best performing prompts in the better model (i.e., Gemini P3F) as the baseline. Then, we asked specific follow-up questions to elicit possible ideas to deal with the wordplay. Again, we discussed the subtitling suggestions provided by Gemini and reached complete agreement on our evaluation. As a final step of the analysis, we translated the wordplay by ourselves to illustrate how human intelligence may differ from artificial intelligence.

4. Results

4.1 Wordplay

In the selected scene, the term “Puppy Pound” presents a subtitling challenge. “Pound” is a polysemous word, commonly referring to the standard unit of money in the UK or a unit for measuring weight. However, in this context, it means a toy shelter for puppy dolls. Moreover, “Puppy Pound” as a toy concept might not be familiar to the Chinese audience, potentially requiring additional explanation or a culturally equivalent reference. The translation of this term was handled differently by fansubbers, ChatGPT, and Gemini, as presented in Tables 8.1 and 8.2.²

Table 8.1. Source text and fansubber’s translation in Example 1

Subtitle Line	Source text	Fansubber’s translation
1	[Mitch] I-it’s not on your list.	礼物单上没有这个
2	[Lily] Yes, it is. “Puppy Pound.”	有啊 在这儿 “胖胖狗之家”
3	[Cam] Oh, no! She put a “P” backwards.	不是吧 她把“P”写反了
4	[Cam] And, honey, there’s a “u” in “Pound.”	而且亲爱的 “Pound”里有字母“u”
5	[Cam] This says “Guppy Pond”!	这写的是“古比鱼池”
6	[Mitch] “Guppy Pond”!	古比鱼池

Table 8.2. GenAI-translated versions of Line 2 in Example 1

Prompts	GPT-4	Gemini 1.5 Pro
P1B	有的，“小狗磅”。	在呀。“Puppy Pound” (小狗乐园)。
P2C	有的，就是“小狗磅”。	有的。“小狗乐园”。
P3F	有的，就是“小狗磅”。	有的，就在这。“小狗窝”。
P4P	是的，有的。“小狗磅”。	有的，“Puppy Pound”。

² Due to space limits, back translation of the entire Chinese texts is not provided in Tables 8.1-8.8, but the textual segments containing the focal points of comparison are back-translated in the main text.

P5CF	有的，“小狗磅”。	是的，它在。“小狗乐园”。
P6CP	有的，“小狗磅”。	有的，“幼犬集中营”。

As shown in Line 2 of Table 8.1, the fansubbers opted for a creative adaptation by translating “Puppy Pound” to “胖胖狗之家” (Home for Chubby Puppies). The rendition not only adjusted the cultural reference to make it more accessible to the Chinese audience but also captured the toy-like, endearing aspect of the term as an animal shelter.

All the translations generated by ChatGPT across six prompts interpreted “Pound” incorrectly as a measure of weight (see Table 8.2), resulting in “小狗磅” (Puppy the Weight). This literalism approach failed to grasp the contextual implications of the term, leaving the target viewers confused. In contrast, Gemini’s translations varied depending on the prompts. P1B adopted a borrowing approach by providing the original English term alongside a culturally adapted translation, “小狗乐园” (Puppy Happy Land). P2C and P5CF provided the same culturally appropriate translation as P1B, but they omitted the English original term. P3F, rendering the term as “小狗窝” (Puppy House), also effectively conveyed the concept of a homely place for puppies. P4P did not translate the term, while P6CP produced a potentially offensive translation “幼犬集中营” (Young Canine Concentration Camp).

Another difficult issue is the wordplay of “Puppy Pound” and “Guppy Pond”. The mixed-up spellings leading to a humorous misunderstanding in English require a creative rendition in Chinese. Since the translation of reversing “p” does not differ among fansubbers’ and GenAI’s versions, it is not discussed here. Regarding the missing “u”, as shown in Tables 8.1 and 8.3, the translations produced by fansubbers and most GenAI prompts incorporated the English words directly into the target subtitles. Two exceptions were observed from P3F and P6CP with Gemini. These two subtitle versions tried to explain the missing of “u” in the Chinese words “窝” (House) and “集中营” (Concentration Camp) but fell short of clearly conveying the original wordplay. It should also be noted that fansubbers’ and GenAI’s renditions (even the semantically faithful ones) did not cohere with the co-text in the Chinese translations.

Table 8.3. GenAI-translated versions of Line 4 in Example 1

Prompts	GPT-4	Gemini 1.5 Pro
P1B	亲爱的，“Pound”里有个“u”。	还有，宝贝，“Pound”里面有个“u”。
P2C	亲爱的，“Pound”里有个“u”。	还有，亲爱的，“Pound”里有个“u”。
P3F	而且，亲爱的，“Pound”中有一个“u”。	还有，亲爱的，“窝”里有个“u”。
P4P	而且，亲爱的，“Pound”中有一个“u”。	还有，宝贝，“Pound”里面有个“u”。
P5CF	而且，亲爱的，“Pound”中有一个“u”。	还有，亲爱的，“Pound”里有个“u”。
P6CP	亲爱的，“Pound”中有一个“u”。	而且，亲爱的，“集中营”里有个“u”。

The translations of “Guppy Pond” showcase different strategies adopted by fansubbers, ChatGPT, and Gemini, as shown in Tables 8.1 and 8.4. The fansubbers made a phonetic translation: “古比鱼池” (Gubi Fish Pond). Within ChatGPT, the translations from P1B and P5CF, “小鱼池” (Little Fish Pond), focused on the size, suggesting a miniature version of a pond, which was contextually appropriate. From P2C and P6CP, the target terms “鳅鱼池”

(Loach Pond) and “金鱼池” (Goldfish Pond) misrepresented the specific type of fish. P3F used the term “鱼缸” (Fish Tank) that indicated a different container for fish, potentially misleading as it suggested an indoor aquarium rather than a pond. The rendition as “鱼苗池” (Fish Hatchery) in P4P could imply a place where fish are hatched, which could sound too serious for a gift. Gemini consistently translated the term into “孔雀鱼池塘” (Peacock Fish Pond), the semantically Chinese equivalent; but it overlooked the wordplay of the original term. It appeared that the translations by both fansubbers and GenAI generally failed to recreate the wordplay effectively, mostly focusing on the literal meaning.

Table 8.4. GenAI-translated versions of Line 5 and Line 6 in Example 1

Prompts	GPT-4	Gemini 1.5 Pro
P1B	- 这写的是“小鱼池”! - 小鱼池!	- 这写的是“Guppy Pond” (孔雀鱼池塘)! - Guppy Pond!
P2C	- 这写的是“Guppy Pond”! - 鳅鱼池!	- 这写的是“孔雀鱼池塘”! - 孔雀鱼池塘!
P3F	- 这写的是“Guppy Pond”! - 鱼缸!	- 这写的是“孔雀鱼池”! - 孔雀鱼池!
P4P	- 这写的是“Guppy Pond” - 鱼苗池!	- 这写的是“Guppy Pond”! - “Guppy Pond”!
P5CF	- 这写的是“小鱼池”! - 小鱼池!	- 上面写的是“孔雀鱼池塘”! - 孔雀鱼池塘!
P6CP	- 这写的是“金鱼池”! - 金鱼池!	- 这里写的是“孔雀鱼池塘”! - 孔雀鱼池塘!

4.2 Multimodal humour

In the selected scene, Cam’s remarks and visual gestures are at play, working towards the multimodal humour. This is a typical challenge for subtitling practice that requires skilful transfer of both visual and verbal information. This is where human and GenAI models differ in dealing with the multimodally rich text (see Tables 8.5 and 8.6).

Table 8.5. Source text and fansubber’s translation in Example 2

Subtitle Line	Source text	Fansubber’s translation
1	[Mitch] You’re pointing fingers.	你在“指”责我
2	[Cam] No. It’s still a “1.”	不 这还是个“1”
3	[Cam] It just fell asleep on the job.	只不过它偷懒睡着了

Table 8.6. GenAI-translated versions of subtitles in Example 2

Prompts	GPT-4	Gemini 1.5 Pro
P1B	- 你在指责别人。 - 不。这仍然是“1”。 - 它只是在工作时睡着了。	- 你就是在指责我。 - 没有，这还是“1”。 - 它只是在工作时睡着了。

P2C	- 你在指责人。 - 不，这还是“1”。 - 它只是在工作中睡着了。	- 你就是在指责我。 - 没有，这还是“1”。 - 它只是打了个瞌睡。
P3F	- 你在指责我。 - 不。这仍然是“1”。 - 它只是在工作时睡着了。	- 你就是在指责我。 - 不，还是“1”。 - 它只是睡着了而已。
P4P	- 你在指责别人。 - 不。这仍然是“1”。 - 它只是在工作中睡着了。	- 你在指责我。 - 没有啊，它还是“1”啊。 - 只是打瞌睡，没做好本职工作而已。
P5CF	- 你在指责别人。 - 不。这仍然是“1”。 - 它只是在工作时睡着了。	- 你在指责我。 - 不，它仍然是一个。 - 它只是在工作中睡着了。
P6CP	- 你在指责别人。 - 不。这仍然是“1”。 - 它只是在工作时睡着了。	- 你在指名道姓。 - 没有，还是“一”， - 它只是在工作时睡着了。

The fansubbers’ translation showed a keen awareness of the visual context and the importance of multimodal cues. As shown in Line 1 Table 8.5, by using double quotes around the Chinese character “指” in ‘你在“指”责我’ (You are “pointing” and blaming me), the fansubbers cleverly highlighted the dual meanings of the Chinese word — both “to point with a finger” and “to blame.” This not only preserved the humour in the original dialogue but arguably enhanced it by making the visual gesture explicitly relevant to the verbal accusation.

In contrast, most translations provided by ChatGPT, such as “你在指责别人” (You are blaming other people), missed the specific target of the blame (Mitch) and diluted the humorous effect of Cam’s accusation. The prompt P3F within ChatGPT and most prompts within Gemini (from P1B to P5CF) resulted in better interpretation of the context by specifying Mitch as the subject of the blame, such as “你在指责我” (You are blaming me) and “你就是在指责我” (You are exactly blaming me). However, Gemini’s P6CP translation, “你在指名道姓” (You are calling names), did not align well with the context.

As for the translation of Cam’s explanation of his pointing gesture, the fansubbers rendered it as “只不过它偷懒睡着了” (It was just loafing and fell asleep), which creatively conveyed the humour by describing the finger as if it had its own personality and intentions. ChatGPT translated it consistently as “它只是在工作中睡着了” (It just fell asleep at work). Although the translation conveyed the main information of the original, it sounded slightly unnatural because the Chinese phrase “在工作中” (at work) did not fit into the context of a finger. Similar subtitle versions were also provided by some of the Gemini prompts (P1B, P4P, P5CF, and P6CP). Two contextually appropriate translations from Gemini were generated by P2C as “它只是打了个瞌睡” (It just dozed off) and P3F as “它只是睡着了而已” (It just fell asleep).

4.3 Orality features

When it comes to the subtitling of orality features, three elements were observed: repetitions, interjections and incomplete utterances. To translate these spoken features, the fansubbers and

GenAI employed different strategies. As shown in Table 8.7, fansubbers translated the repeated “I’m” into “我 抱歉”, spacing the Chinese characters “我” (I) and “抱歉” (sorry) to replicate Mitch’s hesitation. The six ChatGPT prompts produced similar subtitles as “我——我很抱歉” (I--I am sorry), preserving the repetition of “I”. The six translations from Gemini generally rendered the phrase as “我...抱歉” (I...am sorry), using an ellipsis to signify the oral nuances in the source text. However, when handling the repeated “a-a-a”, neither the fansubbers nor GenAI reproduced the repetition in the target subtitles. It appears that the decision to represent repetitions or hesitations in the subtitles was made selectively, depending on the context.

Table 8.7. Source text and target subtitles for repetitions

Texts	Repetition 1	Repetition 2
ST	[Mitch] I’m – I’m sorry. Your what?	[Mitch] Honey, you never mentioned a-a-a puppy pound.
Fansubber	我 抱歉 你说你的什么	亲爱的 你从没提过什么胖胖狗之家啊
ChatGPT	我——我很抱歉。你的什么？	亲爱的，你之前没提过小狗磅。
Gemini	我...抱歉，你的什么？	宝贝，你从没提过什么 Puppy Pound 啊。

Table 8.8 presents an example of subtitling interjections and incomplete utterances. Mitch began with the interjection “um” and ended with “C” before he finished the word “Christmas”, indicating his frustration. The fansubbers chose to omit the interjection but captured the incomplete nature of the sentence with “亲爱的 今天是圣” (Honey, today is *Sheng* [the first Chinese character of “Christmas”: *Sheng Dan Jie*]). The subtitles from the six ChatGPT prompts handled the interjection by using the Chinese equivalent “呃” for “um,” but they kept the letter “C” in the translation, which may confuse the target viewers. Gemini’s translations, consistent across all prompts, also used “呃” for “um” but mistranslated “it’s” as “这叫” (it’s called...).

Table 8.8. Source text and target subtitles for interjections and incomplete utterances

Texts	Example
ST	[Mitch] Um, honey, it’s C--
Fansubber	亲爱的 今天是圣
ChatGPT	呃，亲爱的，那是 C——
Gemini	呃，亲爱的，这叫...

4.4 Possibility of GenAI-assisted creativity

Upon the comparative analysis of fansubbers and AI models, we found that wordplay was a subtitling challenge worthy of further exploration. As explained in Section 3.4, we conducted an emic analysis on the possibility of GenAI-assisted creativity. Realising that the wordplay (“Puppy Pound” and “Guppy Pond”) was about the misspelt words, we wanted to use the quintessential features of the Chinese writing system (e.g., strokes and radicals) to recreate the wordplay. Thus, on the better platform (Gemini) and based on the output prompted by P3F (one of the best performances for wordplay and multimodal humour), we asked three follow-

up questions with increasingly specific information to simulate the brainstorming process. The results are summarised in Table 8.9. The ideas suggested by Gemini were not particularly useful. For example, “泡泡狗乐园” and “古灵鱼塘” do not look similar to each other in the Chinese word forms, thus failing to reproduce the wordplay. The third version in Table 8.9 is tantamount to AI hallucination because for common viewers, the characters 囿 and 混 are as perplexing as garbled characters.

Table 8.9. Follow-up prompts to solicit ideas

Follow-up prompt	Example output
Can you use the meaning-making resources of the Chinese language to recreate the wordplay in “Puppy Pound” and “Guppy Pond”?	泡泡狗乐园 古灵鱼塘
Can you take advantage of the Chinese writing system (e.g., strokes, radicals) to recreate the wordplay in “Puppy Pound” and “Guppy Pond”?	小犬园地 鱼苗水池
Can you take inspiration from the following translation of the poem and recreate the wordplay in “Puppy Pound” and “Guppy Pond”? Original English version: l(a / le / af / fa / ll / s) / one / l / iness Chinese translated version: 子(一 / 口 / 十 / 卅 / 洛一 / 卜) / 瓜 / 豸 / 虫*	犬囿口囿 鱼苗水混

*Note. These are the poem by E. E. Cummings and the translated version by Li (1998). When entered into Gemini, they were presented with line breaks, indicated by slashes here to save space (see Li, 1998, pp. 205-206 for the original presentation forms).

As a final step of the analysis, we relied on human intelligence and came up with our humble attempt (see Table 8.10). We are not saying that our version is perfect but it serves to illustrate how human intelligence and artificial intelligence may differ in experimenting domestication strategies. As explained previously, we wanted to make use of the Chinese writing system to recreate the wordplay. “狗崽崽” and “钩鱼鱼” look similar in their forms and even more so, if the radicals are mixed up and the handwriting is cursive. Admittedly, our rendition does not convey the original meanings of “pound” and “pond”, but we maintain that it reasonably balances “the semantic and humorous load” (Pozo, 2021, p. 133) because the core meanings of the animals (dog and fish) and the misspelt wordplay are both preserved. In addition, the use of duplicate characters sounds catchy and endearing, stylistically consistent with the father-daughter dialogue about gifts.

Table 8.10. Authors’ attempted domestication of the wordplay

Source text	Authors’ attempt	Back translation
[Lily] Yes, it is. “Puppy Pound.”	有啊 “狗崽崽”	There is. “dog pup pup”
[Cam] Oh, no! She put a “P” backwards.	她写错豸字旁了	She got the dog-radical wrong.
[Cam] And, honey, there’s a “u” in “Pound.”	而且你字也写潦草了	And your handwriting is cursive.
[Cam] This says “Guppy Pond”!	我看到的是“钩鱼鱼”	What I read is “hook fish fish”

5. Discussion

5.1 Comparative strength of fansubbers

Comparing fansubber- and GenAI-translated subtitles, our case study found that the fansubbers demonstrated creativity and keen awareness of cultural and multimodal nuances that the GenAI models struggled to replicate. As a notable example, the fansubbers added double quotes to “指” in ‘你在“指”责我’ (You are “pointing” and blaming me; see also Table 8.5). Clearly, the fansubbers exercised their creative agency and used the double quotes as “orthotypographic resources” (Díaz-Cintas, 2018, p. 142) to draw viewers’ attention to the interplay between the verbal text and the visual gesture. In another example, we instructed GenAI to help us (assuming an emic perspective of fansubbers) come up with creative solutions to subtitling wordplay, but the Gemini output was not useful. We had to fall back on human intelligence and used radicals and word forms in the Chinese writing system to domesticate the wordplay. The orthotypographic resources deployed by YYeTs’ fansubbers and the logographic resources exploited by us are agentive attempts to go beyond the confine of the source material. These echo Pozo’s (2021) observation that “creativity plays a prominent role” in AVT (p. 133). Subtitling complex linguistic and/or multimodal elements may still require intense human effort. Fansubbers, often prioritising the viewing experience over mere interlingual transfer (Pérez-González, 2014), possess a more profound contextual understanding and are more capable of crafting translations that resonate better with the target audience culturally, multimodally, and emotionally.

5.2 Performances of GenAI platforms and prompts

The performances of ChatGPT (GPT-4-based) and Google Gemini Pro 1.5 across the six designed prompts are summarised in Table 8.11. It should be noted that the orality features are not included in this table because ChatGPT and Gemini had the same performances: they preserved the hesitation and interjection, omitted the repetition, and misrepresented the incomplete utterance. This pattern is in line with previous observation that more orality features were preserved by machine translation than subtitlers (Hagström & Pedersen, 2022). Thus, more specific instructions need to be included in the prompts if we want GenAI to follow in-house guidelines and condense orality features.

Table 8.11. Subtitling performances of ChatGPT and Gemini under different prompts

Prompts	Polysemy (“Pound”)		Wordplay (“Guppy Pond”)		Visual context (“Pointing fingers”)		Multimodal humour (“1” fell asleep)	
	ChatGPT	Gemini	ChatGPT	Gemini	ChatGPT	Gemini	ChatGPT	Gemini
P1B		*	*	*		*		
P2C		*		*		*		*
P3F		*		*	*	*		*
P4P						*		
P5CF		*	*	*		*		
P6CP				*				

Note. The asterisk (*) indicates a contextually acceptable translation.

Table 8.11 shows that Gemini generally outperforms ChatGPT in addressing typical subtitling challenges, such as wordplay and multimodal humour. This is possibly attributed to Gemini’s

multimodal training, which underscores the speculated advantage of incorporating multimodal information alongside textual content to enhance translation quality (Liu et al., 2021; Sulubacak et al., 2020). It also reaffirms the importance of non-verbal cues in subtitling audiovisual content (Díaz-Cintas & Remael, 2021; Lautenbacher, 2012). If verbal information alone is considered, the translation is removed from the multimodal whole and overlooks the verbal-visual interplay, thus giving rise to misleading, inappropriate subtitles generated by ChatGPT.

Regarding the effectiveness of the six prompts, P2C and P3F yielded higher-quality subtitles. This suggests that providing GenAI with detailed contextual information about the source material or specifying the fansubbers' role, purpose, and target audience could improve subtitling quality (see Peng et al., 2023; Yamada, 2023 for similar observations). This improvement might be attributed to the increased relevance and domain specificity for the language processing (Liu et al., 2023). However, it does not necessarily mean that incorporating all the elements in one prompt can lead to the optimal output. For example, prompt P6CF, the integration of P2C and P3F, did not generate the best outcomes. This suggests that there might be a threshold to the benefits of combined prompting, possibly due to over-specification or conflicting directives within the prompts. Combined prompting may also cause information overload or diminished flexibility in creative translation choices to the AI. It is also interesting to note that the prompt with a fansubber's persona (P3F) outperformed the one characterised as a professional subtitler (P4P). This might be because fansubbers' translations are more readily available for training the GenAI models. The presumably larger size of language data might improve the performance of automatic subtitling. Additionally, the role of fansubbers, who are recognised for their creativity (Díaz-Cintas, 2018), may have prompted the GenAI models to use more innovative subtitling strategies that are outside the conventional repertoire of professional subtitlers. Of course, these two interpretations need to be substantiated by further studies with more systematic empirical evidence.

5.3 Practical implications for automatic subtitling

Our study has three practical implications for automatic subtitling facilitated by GenAI. First, subtitlers shall choose an GenAI model capable of processing dynamic texts for better quality of automatic subtitles. Bywood et al. (2017) noted "the apparent lack of interest from the subtitling industry in embracing [machine translation] technology" (p. 494). Our findings about ChatGPT can further explain subtitlers' nonchalant attitude: the GenAI model limited to static texts, like typical machine translation technology, fails to consider visual information alongside verbal information. The subtitling quality would be unsatisfactory if the multimodal nature of the audiovisual text is not given due attention.

Second, prompt engineering can affect translation quality to a great extent. Our study has found that prompts furnished with contextual information or a fansubber's persona result in better subtitles. However, more information provided does not mean better quality produced. Between the contextual prompt and the fansubber-persona prompt, subtitlers can start with the latter in generating automatic subtitles. This saves time from (a) experimenting with different prompts and (b) viewing the whole video to provide the contextual information as part of the prompt. It should be noted that this practical advice is tentatively offered based on our case

study. A larger sample size of empirical data from varying audiovisual genres shall be collected to substantiate its validity.

Third, GenAI can produce fit-for-purpose subtitles but cannot replace human translators. In line with Bywood et al. (2017) and Sandrelli (2024), we recognise that GenAI can speed up the subtitling process and increase productivity. Its application is particularly relevant to fansub groups who may race to publish their subtitles within 24 hours of the broadcast of the audiovisual content (Wu, 2017). When speed is prioritised, Gemini can be a fit-for-purpose solution because the subtitles it produces are generally acceptable and require light post-editing. However, if subtitles are meant to be published in a DVD or to maximise viewing experience, the issues of wordplay and multimodal humour should be carefully checked and creatively solved by subtitlers.

6. Conclusion

Our case study has compared the performances of fansubbers, ChatGPT and Gemini in subtitling a selected scene from *Modern Family*. We have found that (a) Gemini is better than ChatGPT in handling typical subtitling challenges and (b) prompts furnished with contextual information or a fansubber's persona appear to result in better subtitles. However, both GenAI models fall short of the fansubbers' creative agency in handling typical subtitling difficulties. While we do not aim to generalize these case-based, qualitative findings to all automatic subtitling performances, we believe that they are reasonably helpful in formulating hypotheses to be tested in future empirical studies. For example, since our study only used GenAI for subtitle translation, it will be interesting to know whether a higher degree of automation will further increase the productivity (see Sandrelli, 2024). Typically, a subtitling workflow involves transcription, spotting, and translation. Future studies can compare whether Gemini is also better than ChatGPT in facilitating transcription (with automatic speech recognition) and spotting (with automatic segmentation, see Karakanta, 2022). The comparative results will help us decide the amount of effort required by subtitlers in the non-, semi- and fully-automatic workflows. These empirical findings can contribute to a more nuanced understanding of when and how to use artificial intelligence in tandem with, not in place of, human intelligence in subtitling practice.

References

- Aleksandrova, E. (2019). Audiovisual translation of puns in animated films. *The European Journal of Humour Research*, 7(4), 86-105.
- Baños Piñero, R., & Chaume, F. (2009). Prefabricated orality: A challenge in audiovisual translation. *inTRAlinea*, 11. <https://www.intraline.org/specials/article/1714>
- Barrera-Rioja, N. (2023). Visual restrictions in the translation of audiovisual wordplays: The case of *Modern Family*. *New Voices in Translation Studies*, 28(1), 79-104. <https://doi.org/10.14456/nvts.2023.5>
- Burchardt, A., Lommel, A., Bywood, L., Harris, K., & Popović, M. (2016). Machine translation quality in an audiovisual context. *Target*, 28(2), 206-221. <https://doi.org/10.1075/target.28.2.03bur>
- Bywood, L., Georgakopoulou, P., & Etchegoyhen, T. (2017). Embracing the threat: Machine translation as a solution for subtitling. *Perspectives*, 25(3), 492-508. <https://doi.org/10.1080/0907676X.2017.1291695>

- Calvo-Ferrer, J. R. (2023). Can you tell the difference? A study of human vs machine-translated subtitles. *Perspectives*, 1-18. <https://doi.org/10.1080/0907676X.2023.2268149>
- Chaume, F. (2004). Discourse markers in audiovisual translating. *Meta*, 49(4), 843-855. <https://doi.org/10.7202/009785ar>
- Chen, Y., & Wang, W. (2019). Semiotic analysis of viewers' reception of Chinese subtitles: A relevance theory perspective. *The Journal of Specialised Translation*, (32), 194-216.
- Chiaro, D. (2010). Translation and humour, humour and translation. In D. Chiaro (Ed.), *Translation, humour and literature* (pp. 1-29). Continuum.
- Díaz-Cintas, J. (2018). 'Subtitling's a carnival': New practices in cyberspace. *The Journal of Specialised Translation*, 30, 127-149.
- Díaz-Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices*. Routledge.
- Guillot, M. N. (2020). The pragmatics of audiovisual translation: Voices from within in film subtitling. *Journal of Pragmatics*, 170, 317-330. <https://doi.org/10.1016/j.pragma.2020.09.015>
- Hagström, H., & Pedersen, J. (2022). Subtitles in the 2020s: The influence of machine translation. *Journal of Audiovisual Translation*, 5(1), 207-225. <https://doi.org/10.47476/jat.v5i1.2022.195>
- He, S. (2024). Prompting ChatGPT for translation: A comparative analysis of translation brief and persona prompts. *arXiv*. <https://doi.org/10.48550/arXiv.2403.00127>
- Karakanta, A. (2022). Experimental research in automatic subtitling: At the crossroads between machine translation and audiovisual translation. *Translation Spaces*, 11(1), 89-112. <https://doi.org/10.1075/ts.21021.kar>
- Künzli, A. (2021). From inconspicuousness to flow-the CIA model of subtitle quality. *Perspectives*, 29(3), 326-338. <https://doi.org/10.1080/0907676X.2020.1733628>
- Lautenbacher, O. P. (2012). From still pictures to moving pictures: Eye tracking text and image. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 135-155). Aracne.
- Lee, T. K. (2023). Artificial intelligence and posthumanist translation: ChatGPT versus the translator. *Applied Linguistics Review*. <https://doi.org/10.1515/applirev-2023-0122>
- Li, O. (Ed.). (1998). *World poetic schools*. Foreign Language Teaching and Research Press.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1-35. <https://doi.org/10.1145/3560815>
- Liu, X., Zhao, J., Sun, S., Liu, H., & Yang, H. (2021). Variational multimodal machine translation with underlying semantic alignment. *Information Fusion*, 69, 73-80. <https://doi.org/10.1016/j.inffus.2020.11.011>
- Pérez-González, L. (2014). *Audiovisual translation theories, methods and issues*. Routledge.
- Pedersen, J. (2007). Cultural interchangeability: The effects of substituting cultural references in subtitling. *Perspectives: Studies in Translatology*, 15(1), 30-48. <https://doi.org/10.2167/pst003.0>
- Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). Towards making the most of ChatGPT for machine translation. *arXiv*. <https://doi.org/10.48550/arXiv.2303.13780>
- Pozo, M. D. M. O. (2021). A creative approach for subtitling humour: A case study of the political comedy *Veep*. In C. Wang & B. Zheng (Eds.), *Empirical studies of translation and interpreting* (pp. 114-136). Routledge.
- Rose, H., McKinley, J., & Baffoe-Djan, J. B. (2020). *Data collection research methods in applied linguistics*. Bloomsbury.
- Sandrelli, A. (2024). Integrating ASR and MT tools into cloud subtitling workflows: The ¡Sub! and ¡Sub!2 projects. In Y. Peng, H. Huang, & D. Li (Eds.), *New advances in translation technology: Applications and pedagogy* (pp. 79-97). Springer.
- Schauffler, S. (2015). Wordplay in subtitled films—An audience study. In A. Zirker & E. Winter-Froemel (Eds.), *Wordplay and metalinguistic/metadiscursive reflection: Authors, contexts, techniques, and meta-reflection* (pp. 229-244). De Gruyter.

- Sulubacak, U., Caglayan, O., Grönroos, S. A., Rouhe, A., Elliott, D., Specia, L., & Tiedemann, J. (2020). Multimodal machine translation through visuals and speech. *Machine Translation*, 34, 97-147. <https://doi.org/10.1007/s10590-020-09250-0>
- Tian, L. (2023). Applying machine translation to Chinese–English subtitling: Constraints and challenges. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*, 22, 95-113. <https://doi.org/10.52034/lans-tts.v22i.752>
- Wu, Z. (2017). The making and unmaking of non-professional subtitling communities in China: A mixed-method study. In D. Orrego-Carmona & Y. Lee (Eds.), *Non-professional subtitling* (pp. 115-144). Cambridge Scholars Publishing.
- Wu, Z., & Chen, Z. (2022). Towards a corpus-driven approach to audiovisual translation (AVT) reception: A case study of YouTube viewer comments. *The Journal of Specialised Translation*, 38, 128-154.
- Yamada, M. (2023). Optimizing machine translation through prompt engineering: An investigation into ChatGPT's customizability. In *Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track* (pp. 195-204). Macau SAR, China.

Appendix

Prompts developed and tested in this study:

Prompt 1 Basic (P1B)

You will be given a set of subtitles with time codes. Translate the subtitles from English into simplified Chinese without modifying the time codes. Make sure that the translation is natural, comprehensible, coherent, and concise.

Prompt 2 Contextual (P2C)

You will be given a set of subtitles with time codes. They are excerpted from the American sitcom, *Modern Family*. Some typical features of this sitcom include puns, wordplay, cultural references, irony, and verbal and visual humour. Translate the subtitles from English into simplified Chinese without modifying the time codes. Make sure that the translation is natural, comprehensible, coherent, and concise.

Prompt 3 Fansubber (P3F)

You are a fan subtitler. You will be given a set of subtitles with time codes. Translate the subtitles from English into simplified Chinese without modifying the time codes. Make sure that the translation is natural, comprehensible, coherent, and concise. The translation is intended for the general Chinese audience to understand and enjoy the audiovisual content.

Prompt 4 Professional subtitler (P4P)

You are a professional subtitler. You will be given a set of subtitles with time codes. Translate the subtitles from English into simplified Chinese without modifying the time codes. Make sure that the translation is natural, comprehensible, coherent, and concise. The translation is intended for the general Chinese audience to understand and enjoy the audiovisual content.

Prompt 5 Contextual + Fansubber (P5CF)

You are a fan subtitler. You will be given a set of subtitles with time codes. They are excerpted from the American sitcom, *Modern Family*. Some typical features of this sitcom include puns,

wordplay, cultural references, irony, and verbal and visual humour. Translate the subtitles from English into simplified Chinese without modifying the time codes. Make sure that the translation is natural, comprehensible, coherent, and concise. The translation is intended for the general Chinese audience to understand and enjoy the audiovisual content.

Prompt 6 Contextual + Professional subtitler (P6CP)

You are a professional subtitler. You will be given a set of subtitle lines with time codes. They are excerpted from the American sitcom, Modern Family. Some typical features of this sitcom include puns, wordplay, cultural references, irony, and verbal and visual humour. Translate the subtitles from English into simplified Chinese without modifying the time codes. Make sure that the translation is natural, comprehensible, coherent, and concise. The translation is intended for the general Chinese audience to understand and enjoy the audiovisual content.

Gemini-specific prompt

You also need to make sense of the visual context in the provided video and ensure that the translated subtitles are congruent with the on-screen information in the video, such as actions and facial expressions.