IRASS Journal of Multidisciplinary Studies
https://irasspublisher.com/journal-details/IRASSJMS
ISSN *(Online) 3049-0073*

AI-Generated versus Official Subtitles in Documentary Translation: A Literature Review Based on Season 2 of the Potala Palace

Jiayi FAN[1*] & Yang HAN[2]

[*1-2]University of Shanghai for Science and Technology

Corresponding Author: Jiayi FAN

University of Shanghai for Science and Technology

Article History

Received: 22 / 09 / 2026

Accepted: 04 / 11 / 2026

Published: 20 / 11 /2026

Abstract: Comparisons between machine-generated renderings and official subtitles are accumulating, yet it remains unclear where the two diverge in documentary subtitling, a genre in which the image carries part of the meaning. This review examines 25 items bearing on that question: 17 Chinese journal articles, six monographs, one dissertation, and one web source, 12 of them published after 2020, retrieved from CNKI and through reference-list tracing. Four lines of research are distinguished: comparisons of machine and human translation, the constraints of subtitling, documentary subtitle research, and translation quality assessment. Three findings emerge. First, the studies reviewed broadly agree that machine output is no longer inferior to human translation in semantic fidelity, and that its weaknesses concentrate in fluency, terminological consistency, and contextual cohesion. Second, most of these studies argue from illustrative examples rather than item-level counts and explicit criteria, so their conclusions are hard to verify. Third, research on subtitles has described the constraints of the medium in detail, and international work has turned them into computable parameters such as line limits and characters per second. A gap follows from the three. Existing comparisons treat the official translation as the sole standard, without separating genuine mistranslation from legitimate variation. The coordination of subtitles with the image, and the consistency of proper-name renderings across a whole film, has rarely been examined within the scope of this search. The review therefore proposes a three-way alignment that takes the source text and the image as joint reference, and a three-level scheme of judgement — word-and-sentence, contextual, and whole-film — counted level by level. Both are illustrated with three examples from Season 2 of the documentary The Potala Palace, in order to delimit what current large language models can and cannot achieve here.

Keywords: *machine translation; large language model; subtitle translation; translation quality assessment; documentary film*

How to Cite in APA format: FAN, J. & HAN, Y. (2026). AI-Generated versus Official Subtitles in Documentary Translation: A Literature Review Based on Season 2 of the Potala Palace. *IRASS Journal of Multidisciplinary Studies, 3(11), 13-18.* http://doi.org/10.67564/IRASSJMS.v3.i11.0195

Introduction

Season 2 of the documentary *The Potala Palace* was released on Migu Video in February 2025. It comprises seven episodes, narrated by the actress Zhao Liying, and covers the architecture, cultural relics and art of the Potala Palace. The film uses AI animation and virtual reality to reconstruct historical scenes, and was released simultaneously on Migu's overseas YouTube channel and the MiguC client; the first episode attracted more than 160,000 overseas views [1]. The figure shows that the film was not only broadcast domestically but also had a real overseas audience. Mediotranslatology has long stressed that the translation of a work abroad is only the first step; what matters is whether it is accepted [2], and the subtitles are the first gateway to that acceptance.

Watching the film, I noticed a contrast: the images of the Potala Palace on screen may be generated by AI, yet the subtitles are still translated by humans. Within one and the same text, technology occupies two different positions — one deliverable to an algorithm as imagery, the other still requiring human judgement as words. This gives rise to the question examined here: if the same Chinese commentary is handed to a current large language model, can its rendering reach the standard of the official subtitles? If not, where are the differences concentrated?

The aim is not to compute an "accuracy rate" but to examine the boundary of intelligent translation in the genre of documentary subtitles: which positions it can handle, which it cannot, and why.

One term needs defining at the outset. "AI translation" in this review refers to renderings produced directly by general-purpose large language models; "machine translation" in the earlier literature usually denotes neural machine translation systems, which are not the same object. When citing earlier studies, the term "machine translation" is retained in its original sense, referring to the object of those studies.

Materials and Methods

Corpus scope

The film has now run for three seasons. Season 1 went online in August 2024, and Season 3, *Spacetime Decoding*（时空解码）, was broadcast in October 2025. The three seasons differ in

This is an open access article under the CC BY-NC license

narrator, production batch and platform, and it cannot be confirmed whether their official subtitles came from the same team of translators. If the three seasons were examined together, the official renderings used as reference would themselves be inconsistent, and the baseline of comparison would collapse. The seven episodes of Season 2, by contrast, were produced in one batch and distributed on one platform, and are known to carry official English subtitles; this review therefore takes Season 2 only.

Literature search and screening

The literature was retrieved from CNKI (China National Knowledge Infrastructure) with the search terms "machine translation" (机器翻译), "large language model" (大语言模型), "human–machine comparison" (人机对比), "subtitle translation" (字幕翻译), "documentary" (纪录片) and "translation quality assessment" (翻译质量评估). Screening applied three criteria: whether the study works on concrete data or a specific object of analysis; whether it states how judgements are made; and whether it can locate a problem at a specific level. Purely technical introductions and experience summaries were generally excluded.

Distribution of the retrieved literature

The size distribution of the two fields is itself telling. According to two bibliometric studies [3,4], the volume of literature in the two fields differs by a factor of about six, as shown in Table 1.

Table 1. Distribution of publications in the two fields

Search object	Data source and period	Number of items	Trend
Subtitle translation	CNKI core journals, 1999 – Sept. 2023	380 papers	Peak of 76 papers in 2013, then flat
Machine translation	CNKI, all journals, 2013 – 2022	2,222 papers	Rising year by year
Cross-section: AI renderings vs. official subtitles	CNKI, no time limit	2 items	One short paper and one dissertation

Note. The first two rows use different counting bases (core journals vs. all journals) and should not be compared directly. The present review draws on 25 items in total: 17 Chinese journal articles (7 from *Chinese Translators Journal*, 2 from *Shanghai Journal of Translators*), 6 monographs (2 Chinese, 4 foreign), 1 dissertation and 1 web source; by year, 12 items postdate 2020, 6 of them from 2025 alone.

The pattern in Table 1 is clear. Core-journal research on subtitle translation totals fewer than 400 papers and stopped growing after its 2013 peak; journal papers on machine translation run into the thousands and keep increasing; and the intersection of the two strands is where coverage is thinnest — only Luo Yan's short paper and Li Maocairang's dissertation were found. Discussion has indeed grown over the past three years, but it concentrates on human–machine comparison and post-editing; dedicated treatment of subtitling remains scarce.

Results and Discussion

Machine translation versus human translation: findings and methods

This is the most productive area, and traced through the evolution of its research questions, it falls roughly into three strands.

The earliest strand is error statistics. Luo and Li aligned machine renderings of an automotive technical corpus with human translations sentence by sentence, counted total errors, and filed them under vocabulary, syntax and symbols [5]. The strength of this approach is that its conclusions are replicable: anyone recounting under the same criteria should reach similar figures. Its premise, however, is that a "correct translation" must be fixed in advance; otherwise there is nothing to count against.

The second strand turned to overall quality comparison. Li applied an error-scoring method along five dimensions — fidelity, fluency, terminology, style and cultural acceptability — to compare English-major student translators with neural machine translation. Total errors and fidelity errors were markedly fewer in the machine output; fluency errors were more numerous; terminology, style and cultural acceptability showed no significant difference [6]. The result deserves attention: at the level of semantic fidelity, the machine is no longer inferior to students trained in translation, and its weakness lies elsewhere. Hu and Li summarised five features of machine translation: automation, mechanicalness, the sentence as the translation unit, secondary imitation, and limited contextual constraint [7]. Two of these — the sentence-bounded unit and limited context — explain precisely the difficulties such systems meet in texts as context-dependent as subtitles. Li et al. subsequently changed the objects of comparison to machine versus machine, testing two large language models against two neural systems on legal texts, and concluded that the LLMs had caught up with, or overtaken, the neural systems, while the errors of all four systems concentrated in accuracy, fluency and terminology; specialised texts still require human checking [8].

The third strand focuses on the specific defects of large language models. Drawing on post-editing experience, Ren identified four types of problems: mistranslation of technical terms, loss of cultural meaning, generation of content absent from the source, and inability to understand multimodal context [9]. The first three occur within the sentence; the fourth already exceeds it. Taking metaphor as his entry point, Ye argued that some content machines cannot handle, because understanding a metaphor requires life experience beyond the text, which literal correspondence cannot supply [10].

Similar discussions exist abroad. The collection edited by Kenny addresses users, covering the evaluation of machine output, the preprocessing of source texts, post-editing, and the customisation of systems on self-built corpora; it repeatedly makes one point: the same system, with different prior processing, yields different results [11]. This prompts a caution for the present study: asking "how well does AI translate" requires first stating how it was used and under what conditions.

As for method, three paths can be distinguished. The first compares several systems to decide whose output is better; the second places machine and human translations side by side and scores them along several dimensions; the third sets a single LLM

rendering against the official translation item by item. The three answer different questions. The first two can rank systems, but the scoring dimensions must be fixed before the data are seen; if the dimensions are ill chosen, phenomena in the data cannot be accommodated. The third is the least laborious and the most common in subtitle research, but it argues mostly from examples, without statistics.

The conclusions are fairly consistent: machines are no longer inferior to humans in semantic fidelity; what they lack concentrates in fluency, terminological consistency and contextual cohesion. Methodologically, however, there is a common weakness: most studies illustrate with a handful of examples, rarely reporting the total number of items or the proportion of each problem type, so readers must take the author's judgement on trust. Luo Yan's study illustrates the point. She placed the official English subtitles of the documentary *Shaolin Gravity* beside a GPT-4 rendering and compared the two in meaning construction and translation strategy at the cultural, contextual and content levels [12]. The topic and the question are both worthwhile, but the paper runs to only three pages and relies on typical examples, without total item counts or decision criteria; it can show that "differences exist" but not their scale or distribution.

The constraints of subtitling and documentary subtitle research

Subtitle translation is bound by a set of basic constraints. Qian listed five: listenability, comprehensiveness, transience, popular register, and no annotation [13] — the viewer must watch images and hear sound at once, the subtitle vanishes in seconds, and there is no way to gloss what is unclear. Li proposed a different scheme, discussing space–time constraints, informative function and cultural factors [14]; Ma dealt with principles — colloquialism, characterisation, affect, lip synchronisation and popular register [15]; Zhang stressed only immediacy and popularity [16]. The formulations differ but point the same way: subtitles must be compressed, grasped at a glance, and cannot carry notes.

International research has gone further and converted these constraints into computable parameters. The textbook by Díaz Cintas and Remael specifies that a subtitle has at most two lines, that adult viewers read at about fifteen characters per second, and that a full two-line subtitle should be readable within six seconds; compression proceeds either by deleting irrelevant information from the whole subtitle or by rewriting the sentence [17]. With such parameters, "this subtitle is too long" is no longer an impression but a checkable judgement.

Documentary subtitles have features of their own. Cheng Wei's study of the translation of *Beautiful Nostalgia* showed that documentary subtitling is not a purely verbal conversion: the translator must complete the narrative with images, sound and music as "intermediate texts", while keeping pace with the rhythm of expression [18]. A documentary carries knowledge through narration; its information density is higher than that of drama, and it lacks the pauses of dialogue as a buffer, so every sentence must be more compact.

Culture-specific items are the hardest class in documentary subtitling. Wang's notion of "cultural default" refers to the cultural background shared by source author and readers and therefore left unwritten; without compensation, a comprehension gap opens for the target reader [19]. This explains why cultural terms are hard: the difficulty often lies not in the word itself but in the layer of tacit knowledge behind it. Pedersen calls such items "extralinguistic cultural references" and classifies their treatment into retention, substitution, generalisation, omission, adoption of official renderings and others; counting the uses of each strategy shows whether the translator leans conservative or liberal in cultural items [20]. The merit of this classification is that it can be counted; it will be used in the present corpus work.

Three examples from Season 2 show what these difficulties look like in particular sentences. Episode 6, "The Red Palace," contains this line of commentary: ″红宫的建造是众多工匠通力合作的成果，其最高处矗立着七座鎏金铜瓦顶，无数的金色自顶端流淌下来，跳跃着，闪动着，灯灯相续，熠熠生辉。″ ("The Red Palace was built through the concerted work of countless craftsmen; at its highest point stand seven gilded copper-tile roofs, and countless streams of gold flow, leap and glimmer from the summit, lamp transmitting lamp, radiant."). The sentence contains four decision points: "seven"（七座）is a number; "gilded copper tiles"（鎏金铜瓦）is a compound term of craft and material — gilding（鎏金）is fire gilding, rendered as *gilded*, and the tiles are copper tiles, the material "copper" is precisely what compression most easily loses; ″灯灯相续″ alludes to the Buddhist transmission of the lamp; ″熠熠生辉″ is a four-character idiom, pictorial but informationally light, usually deleted as a whole under compression. At the reading-speed parameter of fifteen characters per second, the display time of this subtitle is generally under six seconds (not measured); the translator must first decide what to keep, then how to compress. If line length and display time are not stated to the model, AI is the more likely to translate all four items in full; whether the constraints are given, and how, is itself a variable to be compared in what follows.

The opening of Episode 1, "New Looks," is of another type: ″赭红宫墙，飞檐挂雪，这里是最接近苍穹的地方。″ ("Ochre palace walls, upturned eaves hung with snow — here is the place closest to the firmament."). Four four-character phrases run in succession, a common pattern of Chinese narration. English has no matching structure: ″赭红″ must settle on the concrete pigment name *ochre*; ″飞檐″ on the architectural term *upturned eaves*; ″苍穹″ should not be conflated with *sky*. If any of the three goes wrong, what the viewer sees on screen and what the subtitle says fall out of joint. Such deviation involves neither grammar nor vocabulary size; the point is whether the translator commands the established usage of the field.

The third case is more hidden. The closing episode calls the Potala Palace ″一座落入凡间的天空之城″ — ″a city in the sky fallen to earth." In Chinese, ″天空之城″ is an image, but the English phrase *Castle in the Sky* has long been taken by an animated film; a literal rendering would point a palace documentary somewhere else entirely. This is neither a grammatical nor a terminological problem; it is the translator's knowledge that an English phrase is already occupied by something else in the target culture. Whether a model can notice this is a question the present review cannot yet answer, but it is a class of phenomenon worth searching for specifically in the corpus.

These three examples are not offered to prove that AI mistranslates — this review has not reached that step — but to

show what a "translation decision point" is: a position in a line of commentary where the translator must stop and choose. What the later analysis marks is precisely these positions.

Documentary subtitles also coordinate with images. When an object is mentioned, the camera is showing its shape; when a process is described, the shot is displaying its sequence. Since the image has already presented the object to the viewer, the subtitle need not repeat it — one more reason why subtitling compresses. The AI model's input is text only; it cannot see the image and cannot judge what the shot has already conveyed. One may therefore conjecture that its renderings tend to say everything rather than just enough. Such renderings may not be wrong in meaning, may even be more complete, but the viewer must re-read the same information in limited time, and the viewing experience suffers. This conjecture must be tested on the actual corpus, which is why the present study brings images into its judgements.

One further point is consistency across the whole film. The same monastery and the same name should be rendered the same way throughout. Human translators usually prepare a glossary first; AI processes each subtitle in isolation, without memory of what came before, so one name may take several spellings. In their study of the English translation of Yu Hua's *The Seventh Day*, Luo and Huang first counted the strategies and frequencies for five classes of culture-specific items and only then characterised the translator's overall tendency [21]. The method transfers directly: counting how many renderings one proper name receives across the film is enough to judge its consistency. Corpus-based translation studies follows the same path — describe the distribution of linguistic features first, then explain [22]. Such data are objective: two people counting separately should reach the same result, which makes this the most reliable class of judgement.

Setting the strands side by side: subtitle research mostly says "how one ought to translate," machine translation research mostly says "how the machine translates," and the two are rarely joined. The constraints of length, display time and no annotation are largely absent from machine translation evaluation; the error types catalogued by machine translation research do not include the special requirements of subtitles.

Standards for judging translation quality

The received standards of translation are "faithfulness" and "fluency". Si divides them into two stages: the first examines whether the translation is faithful to the original in function and meaning; the second, whether the translation itself is fluent [23]. For literary translation the scheme is uncontroversial; for subtitles it is hard to operate, because subtitling must compress and rewrite, and deletion of part of the content is legitimate behaviour, cannot automatically be judged as unfaithful. House's m

House's method is finer. Original and translation are set together and examined for shifts in field, tenor and mode, and in genre; shifts accumulating beyond a certain degree constitute translation quality problems [24]. The method was designed for literary and applied translations and needs adjustment for subtitles: what it compares are still two texts, and the images are not among them.

Machine translation evaluation has its own system. Early studies scored texts with automatic metrics — fast, but measuring only surface overlap with a reference translation, so any legitimate rewriting that differs from the reference is penalised as an error. Human assessment is more accurate but far more costly; most recent human–machine comparisons have turned to human scoring plus error classification for exactly this reason.

The present study must face one concrete difficulty: subtitles have no single standard answer. The official translation is itself compressed and may contain genuine errors. If "agreement with the official translation" is taken as the criterion of correctness, official compression is charged to AI as over-translation, official errors are charged to AI as mistranslation, and the cases where AI handles the passage better than the official version cannot be recorded at all. Judgement therefore cannot compare two translations alone; it must return to the source text and the images and make a three-way alignment.

Synthesis: convergence, methods and the research gap

The preceding sections can be condensed into three points. First, machine translation research has moved from error statistics to overall quality comparison and lately to the specific defects of large language models; the conclusions converge — at the level of semantic fidelity the machine now approaches or surpasses student translators, and what it lacks is fluency, terminological consistency and contextual cohesion. What Ren calls "multimodal context" and Ye's metaphors both exceed the sentence, showing that the weakness lies beyond word and sentence. Second, subtitle research has said enough about compression, immediacy and no annotation, and international work has converted these constraints into lines and characters per second, so that "is this subtitle too long" becomes a checkable judgement rather than an impression. Third, Si's two-stage scheme and House's register and genre parameters both take the source text and the translation as their objects of comparison — a premise that hardly holds for subtitles.

A gap thus appears: most studies take the official translation as the sole standard and do not distinguish mistranslation from legitimate variation; the medium constraints of subtitling rarely enter machine translation evaluation; and the coordination of subtitles with images and the consistency of renderings across a full film have rarely received dedicated treatment within the scope of this search. Li's study of the subtitles of *The Potala Palace* follows the path of eco-translatology and cross-cultural adaptation — strategy description, without AI [25]; Luo did set official subtitles against AI renderings, but at three pages and by examples alone, her study cannot show the scale or distribution of the differences.

Conclusion

The present study is designed against these three points. It takes both the source text and the images as references and makes a three-way alignment, so that cases where the official rendering is wrong while AI handles the passage well are listed separately and not counted as losses. Judgements are divided into three levels — word and sentence, context, and whole film — and counted level by level; the layering echoes House's three register variables and Ren's "multimodal context": the word-and-sentence level corresponds to language itself, the contextual level to register shifts, and the whole-film level to image coordination and cross-film consistency of proper names. Finally, "what AI failed to do" is further distinguished into "could have been done" and "cannot be done".

The distinction matters for the usability of the conclusions. If the problems concentrate at the word-and-sentence level, they can be remedied with terminology lists, prompts and post-editing — a matter of use; if they concentrate at the contextual and whole-film levels, no change of tool will help — a matter of capability. The division of labour in subtitling will henceforth be, perhaps, not a choice between human and machine but a question of which segment is delivered to whom. What this review has tried to mark is the approximate position of that line.

Acknowledgement Statement

This research was supported by the Research Project Course Construction Fund of the University of Shanghai for Science and Technology.

References

1. China Industry and Information Technology News Network. (2025, April 18). How does an "ancient" documentary attract large numbers of young viewers? Season 2 of *The Potala Palace* innovates with technological expression [一部"古老"纪录片何以吸引大批年轻观众？《布达拉宫》第二季创新科技表达].
2. Xie, T. (2013). *Mediotranslatology* (rev. ed.) [译介学（增订本）]. Yilin Press.
3. Lyu, L., & Cai, C. (2024). A CiteSpace-based visualised analysis of subtitle translation in film and television [基于CiteSpace的影视剧字幕翻译文献可视化分析]. *Modern Linguistics*, *12*(7), 730–743. DOI: https://doi.org/10.12677/ml.2024.127618
4. Guo, X., Li, Z., Liu, Q., et al. (2025). A bibliometric review of machine translation research [基于文献计量的机器翻译研究综述]. *Intelligent Computer and Applications*, *15*(9), 107–111. DOI: https://doi.org/10.20169/j.issn.2095-2163.250917
5. Luo, J., & Li, M. (2012). Error analysis of machine-translated texts [机器翻译译文错误分析]. *Chinese Translators Journal*, *33*(5), 84–89.
6. Li, F. (2022). A comparative study of human and machine English–Chinese translation quality in the era of artificial intelligence [人工智能时代人机英汉翻译质量对比研究]. *Foreign Language World*, 2022(4), 72–79.
7. Hu, K., & Li, Y. (2016). Features of machine translation and its relationship with human translation [机器翻译特征及其与人工翻译关系的研究]. *Chinese Translators Journal*, *37*(5), 10–14.
8. Li, F., Zhang, Y., & Ding, L. (2025). Comparing the translation quality of large language models and neural machine translation systems for specialised texts: The case of legal Chinese–English translation [大语言模型与神经网络机器翻译系统专业文本翻译质量对比——以法律汉英翻译为例]. *Shanghai Journal of Translators*, 2025(6), 62–67.
9. Ren, K. (2025). The art of using tools by human translators in the era of large language models [大语言模型背景下人类译者的用"器"之"术"]. *Chinese Translators Journal*, *46*(6), 113–121.
10. Ye, Z. (2025). Metaphor analysis and AI translation: Starting from a metaphor in the Han Suyin International Translation Competition [隐喻分析与人工智能翻译——从韩素音国际翻译大赛的一个隐喻说起]. *Chinese Translators Journal*, *46*(6), 122–131.
11. Kenny, D. (Ed.). (2022). *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Language Science Press.
12. Luo, Y. (2026). A contrastive study of official and AI publicity subtitle translation from a multimodal discourse analysis perspective: The English rendering of the documentary *Shaolin Gravity* [多模态话语分析视角下官方与AI外宣字幕翻译对比研究——以纪录片《少林引力》英译为例]. *Overseas English*, 2026(11), 33–35.
13. Qian, S. (2000). Film and video translation: An increasingly important field in translation studies [影视翻译——翻译园地中愈来愈重要的领域]. *Chinese Translators Journal*, 2000(1), 61–65.
14. Li, Y. (2001). Subtitling strategies [字幕翻译的策略]. *Chinese Translators Journal*, *22*(4), 38–40.
15. Ma, Z. (1997). On the basic principles of film and video translation [论影视翻译的基本原则]. *Modern Communication*, 1997(5), 81–84.
16. Zhang, C. (1998). A preliminary study of film and video translation [影视翻译初探]. *Chinese Translators Journal*, 1998(2), 49–52.
17. Díaz Cintas, J., & Remael, A. (2021). *Subtitling: Concepts and practices* (2nd ed.). Routledge.
18. Cheng, W. (2014). Chinese–English subtitle translation of documentaries: The translation practice of the documentary *Beautiful Nostalgia* [纪录片字幕汉英翻译研究——以纪录片《美丽乡愁》的翻译实践为例]. *Shanghai Journal of Translators*, 2014(2), 24–27.
19. Wang, D. (1997). Cultural default and the reconstruction of coherence in translation [文化缺省与翻译中的连贯重构]. *Foreign Languages*, 1997(6), 55–60.
20. Pedersen, J. (2011). *Subtitling norms for television: An exploration focussing on extralinguistic cultural references*. John Benjamins.
21. Luo, X., & Huang, D. (2025). Bai Yaren's translation strategies for culture-specific items: The English translation of Yu Hua's *The Seventh Day* [白亚仁对文化专有项的翻译策略研究——以余华《第七天》英译为例]. *Technology Enhanced Foreign Language Education*, 2025(1), 28–34.
22. Wang, K. (2012). *Explorations in corpus-based translation studies* [语料库翻译学探索]. Shanghai Jiao Tong University Press.

23. Si, X. (2016). A further study of translation quality assessment models [翻译质量评估模式再研究]. *Foreign Language Journal*, 2016(3), 84–94.

24. House, J. (2015). *Translation quality assessment: Past and present*. Routledge.

25. Li, M. (2026). Cross-cultural adaptation strategies in the subtitle translation of humanistic documentaries: A case analysis of *The Potala Palace* [人文纪实类纪录片字幕翻译跨文化适配策略研究——基于《布达拉宫》的案例分析] [Doctoral dissertation, Tibet University].