Author/s:
Qiu, J;Pym, A

Title:
Fatal flaws? Investigating the effects of machine translation errors on audience reception in the audiovisual context

Date:
2024-03-13

Citation:
Qiu, J. & Pym, A. (2024). Fatal flaws? Investigating the effects of machine translation errors on audience reception in the audiovisual context. Perspectives: Studies in Translation Theory and Practice, 33 (5), https://doi.org/10.1080/0907676x.2024.2328757.

Persistent Link:
https://hdl.handle.net/11343/340612

Routledge
Taylor & Francis Group

Perspectives
Studies in Translation Theory and Practice

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rmps20

Fatal flaws? Investigating the effects of machine translation errors on audience reception in the audiovisual context

Juerong Qiu & Anthony Pym

To cite this article: Juerong Qiu & Anthony Pym (13 Mar 2024): Fatal flaws? Investigating the effects of machine translation errors on audience reception in the audiovisual context, Perspectives, DOI: 10.1080/0907676X.2024.2328757

To link to this article: https://doi.org/10.1080/0907676X.2024.2328757

Published online: 13 Mar 2024.

PERSPECTIVES
https://doi.org/10.1080/0907676X.2024.2328757

Routledge
Taylor & Francis Group

Fatal flaws? Investigating the effects of machine translation errors on audience reception in the audiovisual context

Juerong Qiu [a] and Anthony Pym [b]

[a]Asia Institute, University of Melbourne, Melbourne, Australia; [b]School of Languages and Linguistics, University of Melbourne, Melbourne, Australia

ABSTRACT

This study reports on an experiment where machine translation errors in subtitling are evaluated from the perspective of nine viewers who did not know the source language and seven viewers who were studying the source language. Screen recordings, think-aloud protocols, comprehension tests, and interviews were employed to explore participants' responses and reactions to erroneous subtitles and to investigate how specific errors impacted comprehension and immersion in the viewing experience. The analysis identifies which errors were most noticed and to what extent those errors affected viewers' trust in the subtitles. Errors causing significant misunderstanding and distrust are initially considered 'fatal', as they may halt viewer immersion and prompt disengagement from the audiovisual product. However, the findings highlight a remarkable tolerance of the uncertainty that results from errors, as viewers filter out misinformation or draw on other sources of information to construe and rectify their interpretations. This tolerance is explained in terms of a general trade-off with the enjoyment of the viewing experience, which varies in accordance with the viewer's knowledge of the source language.

ARTICLE HISTORY
Received 1 August 2023
Accepted 4 March 2024

KEYWORDS
Machine translation errors; translation reception; misunderstanding; tolerance of uncertainty; trust; trade-offs

1. Introduction

Recent years have seen significant progress in the quality and use of online machine translation (MT), to the point where professional human translators now produce less than 1 percent of the words translated in the world (Pym & Torres-Simón, 2021). However, machine translation algorithms are not infallible. In the evaluation of machine translation errors, automatic and manual approaches have traditionally focused on assessing semantic accuracy and grammatical correctness. Linguistic-based taxonomies (e.g., Costa et al., 2015; Vilar et al., 2006) were useful when the accuracy of machine translation was relatively low. However, with the advances in neural machine translation since 2016, errors are caused by probabilistic filtering rather than linguistic-based algorithms. It is crucial to consider that meaning is created through

CONTACT Juerong Qiu Juerong.qiu@unimelb.edu.au, juerongq@student.unimelb.edu.au

the receiver's active combining of texts and contexts. Audiovisual communication encompasses visual, verbal, acoustic, and textual channels, all of which can contribute to the viewer's meaning-making. Gambier (2018) highlights that although subtitles are an important part of meaning-making, there are instances where non-verbal signs take precedence. This implies that a small language error can become critical when subtitles conflict with on-screen images. Conversely, an apparently major linguistic omission may not confuse viewers if other channels help bridge the information gap. In our research, we are interested in the consequences of translation errors that can be called 'fatal' in the sense that audiences misunderstand and/or distrust the subtitles to the extent that they ultimately abandon the audiovisual product.

Given these considerations, we explored how sixteen participants negotiated MT errors without being told beforehand that MT had been used or that there were specific problems in subtitles. In real-life online viewing experiences, viewers are often unaware that machine translation might have been used. They are moreover free to engage in screen-based activities such as pausing, rewinding, fast-forwarding, and skipping. Replicating those conditions, we investigated how viewers reacted to machine translation errors and the effects the errors had on their subtitle reading, comprehension, and engagement. We thus identify which errors were relatively detrimental or disruptive of the reception process. Most importantly, our experiment did not measure machine-translated subtitles against human-translated subtitles, since our focus is on the actual viewing experience.

We hope that awareness of how audiences negotiate MT errors will help post-editors efficiently revise machine-translated subtitles, directing their attention to the errors that are likely to have negative effects on cultural understanding and appreciation.

2. The reception and evaluation of audiovisual translation

Recent studies have gradually explored the challenges of implementing MT in audiovisual translation from the viewer's perspective. For instance, Hu et al. (2020) examined Chinese viewers' responses to raw machine-translated subtitles for MOOCs, using the '3Rs framework' – *reactions* measured through eye-tracking, *responses* gauged by eye-tracking and comprehension tests, and *repercussions* evaluated through interviews. This framework was first proposed by Chesterman (2004) and was introduced into audiovisual translation by Gambier (2018). Hu et al. (2020) conclude that using MT to render English subtitles into Chinese is acceptable for most Chinese viewers and may also contribute to the progress of MOOCs in the long run. More recently, Tuominen et al. (2023) explored the potential of automated subtitling to support linguistic accessibility for public broadcasts. They incorporated focus groups and questionnaires to examine English viewers' comprehension, acceptability, and self-reported cognitive load when watching clips of Finnish-language news. They found that, although the participants were able to understand the gist of a programme with automated subtitles, out-of-synchrony subtitles and MT errors disrupted their processing of the clips and led to significant cognitive load. However, a participant's interest and motivation can affect their use of automated subtitles: Viewers have a higher tolerance for the lower quality of subtitles if they genuinely want to or need to access the audiovisual material.

Most studies in audiovisual translation reception adhere to the natural pace of the content, similar to cinema viewing. This means we know little about non-linear reception

activities – pauses, rewinds, and skips – even though in online video streaming, audiences increasingly control the pace of the product in these ways. There is thus a need to investigate viewing behaviour based on the viewer's pace rather than the product's pace. To address this, alternative data-collecting methods such as think-aloud protocols and screen recording offer insights into receptive activities.

Reception studies of machine-generated subtitles also tend to investigate the audiovisual text as a whole, without a detailed analysis of translation errors. Some error typologies can nevertheless offer insights. In the 1980s, the word-error rate was used in speech recognition assessment focusing on deletion, insertion, and substitution, but the accuracy of the error rate was questionable due to the subjective determination of information loss, the ambiguity in identifying the cause of deletions, and a lack of attention to language-specific factors (Romero-Fresco & Pérez, 2015). The NER model (N for the number of words, E for editing errors, and R for recognition errors) was then designed by Romero-Fresco and Pérez (2015) to assess intralingual live subtitling. This model categorises subtitle errors into three types: serious errors, which introduce a new meaning that may go unnoticed in the specific context; standard errors, which involve information loss, disruptive flow, and confusion; minor errors, which allow viewers to comprehend the original meaning and occasionally reconstruct the original words. Inspired by the NER model, Pedersen's (2017) FAR model provides a more detailed examination of issues in interlingual subtitles, evaluating them for functional equivalence (F), accuracy (A), and readability (R). Functional errors include semantic errors and stylistic errors; accuracy concerns the extent to which the target text adheres to the norms of the target language; readability concerns the technical features that affect reading effort. Both the NER and FAR models assess errors on the basis of the *projected* severity of their impact on hypothetical viewers. For example, Hagström and Pedersen (2022) applied the FAR model to assess two subtitle corpora from 2010 and 2020. The assignment of penalty levels (serious, standard, and minor) was determined by one of the authors of the study, without analysing any actual reception process.

In empirical experiments with actual reception processes, viewers' perceptions of translation errors are affected by different contexts, as shown in Hu et al. (2020) and Tuominen et al. (2023). Our audience-based approach to subtitle errors has therefore looked for interactions between multiple on-screen information sources and their impacts on the viewers' interpretations of translation errors. Inspired by the above research on reception and error typology, we chose to employ specific categories: (1) *subtitle reading* looks at whether problematic subtitles trigger screen-based activities, (2) *comprehension* concerns how linguistic errors affect understanding of the plot, and (3) *engagement* refers to the participants' perceptions and attitudes towards subtitle errors. Additionally, we consider the extent to which each viewer's relative proficiency in the source language can influence their reaction to problematic subtitles, as tested by Orrego-Carmona (2015) and Burczynska (2017). This also allows us to investigate the impact of viewing in a language-learning context.

3. Research design

Our research question is: To what extent do translation errors affect the audiences' reception processes? This question is examined from two aspects: whether an error is noticed by the participants, and whether it causes confusion, misunderstanding, and/or distrust in the subtitles.

3.1. Research material

Our research material was extracted from the Chinese period drama *Nirvana in Fire* (Lang Ya Bang, 琅琊榜) Season One (Kong & Li, 2015). The sophisticated register of the language and the references to numerous literary works and historical events make this drama more difficult to translate than dramas set in modern times. We were interested in seeing how machine translation fares when the source text is difficult to process. Also, can the participants identify problematic subtitles when they are not informed beforehand about the use of MT?

Two short scenes were extracted from the first episode. The resulting clip introduces the background of the entire drama: Changsu Mei's two friends (Jingrui and Yan) go to Langzhou to pick up Mei, and then the three of them go to the capital city Jinling.

We used Aegisub to insert Chinese subtitles, ensuring precise synchronisation with the original timestamps without using auto-transcription or auto-spotting. The output was exported in.txt format. The original subtitles were then fed into DeepL, and the machine-generated English subtitles were directly incorporated into Aegisub. Video-editing software was used to embed the English subtitles into the clip, ensuring synchronisation and proper placement. This means that the English subtitles provided to participants were raw machine translations without any pre-editing or post-editing.

3.2. Errors in the machine-translated subtitles

We identified eighteen translation errors in total: six serious errors, eight standard errors, and four minor errors (the same error in different lines is counted as one) using the NER and FAR models (see above). This pre-categorisation was made by one author of this paper and reviewed by the other to ensure consistency and alignment with the models. These encompass problems in functional equivalence, accuracy, and readability. Of the 72 lines in total, seventeen (23.6%) have a character per second (cps) rate equal to or greater than 20, with the maximum cps recorded at 38. According to Pedersen (2017), a cps of 20 or above can be considered a standard error.

This pre-categorisation was neither for textual analysis nor to prompt participants to search for the errors. We instead needed the initial error identification to design the comprehension test, which would then enable us to compare our analysis with the errors actually identified by the participants.

3.3. Participants

All participants were L1 English speakers aged between 18 and 34 and studying at the University of Melbourne.[1] As mentioned, two groups of participants were recruited. Group A consisted of nine participants who did not know Mandarin or any other Chinese varieties. Group B comprised seven learners of Chinese: they were enrolled in Chinese 5 (B01, B02, B03, and B04), Chinese 7 (B05), or the Master of Translation (Chinese-English) programme (B06 and B07). Chinese 5 and 7 are undergraduate language subjects that require intermediate language skills, as indicated in the university handbook. The Master of Translation programme requires native or near-native speaker

competence in Chinese. This means the participants in Group B were intermediate to advanced learners of Mandarin.

3.4. Methods

The data-collecting methods included screen recordings, think-aloud protocols, comprehension tests, and interviews. Each instrument focuses on how problematic subtitles affect different aspects of reception: screen recordings and think-aloud protocols can provide clues about how viewers interact with problematic subtitles during viewing; comprehension tests investigate whether specific errors cause misunderstanding; interviews explore each participant's perception of and attitude toward problematic subtitles. This enabled us to study the richness and complexity of receptive behaviour by analysing quantitative and qualitative data.

In the experiment, the participants were allowed to pause, go back, or drag the progress bar to skip during viewing. These screen-based actions provided participants with opportunities to revisit and reassess problematic subtitles, potentially revealing their responses to translation errors. This should increase the likelihood of the participants identifying and processing machine translation errors.

During viewing, participants were not informed that the subtitles were generated using machine translation, nor were they directed to search for translation errors. This approach was inspired by Pedersen's (2011) proposal that audiences, in their viewing processes, activate the 'willing suspension of disbelief' identified by Coleridge (1817/2014, p. 208) and they extend that suspension to the subtitles, producing 'a contract of illusion' where viewers take the translated subtitles to be what is said by the characters. This is similar to the 'presumption of complete interpretative resemblance' conceptualised by Gutt (1991/2014, p. 186) as a feature of how translations are received. If this illusion is operative, it would be unrealistic to expect participants to meticulously analyse each line and identify every error. The focus of our experiment was thus to collect data on the participants' sense-making processes while viewing, rather than impose any specific error-detection task.

3.4.1. Screen recordings and think-aloud protocols

Screen recordings were used to capture everything happening on screen. The participants' concurrent verbalisations gave clues as to how they responded to the subtitles and how they construed any misinformation caused by machine translation errors. This combination helped to locate noticeable errors (viewers make statements like 'This is a mistranslation'), points of confusion (e.g., 'What does this mean?'), and lines that were too fast to read.

3.4.2. Comprehension tests

A comprehension test assesses the impact of specific translation errors on comprehension. The test comprises four questions, each addressing what we had identified as a specific error in the subtitles. These four serious errors (numbers 1, 2, 3, and 6 in Table 1) were selected based on our assumption that they were more likely to introduce new meanings. The remaining two serious errors (numbers 4 and 5) were excluded from the test for two reasons. First, both errors involve the misuse of pronouns and appear in

Table 1. The six serious errors as identified by the FAR model.

No	Time	Error	Literal translation	Machine translation
1	00:03	十文	Ten wen [unit of currency].	Ten words.
2	00:18	什么时候才能回到金陵啊	When can go back to Jinling?	When are we going to return to Golden?
3	01:29	传闻此人手无缚鸡之力	It is said that this man has no strength to bind a chicken.	Rumour has it that this man has no hands.
4	02:12	你能不能别每次诊完脉	After you check my pulse, can you stop	After you check your pulse, can you stop
5	02:21	我诊完脉可是什么都没说	After I checked your pulse, I didn't say anything	I finished my pulse and didn't say anything
6	03:21	你苏哥哥就要丢下你去金陵了	Your Su Gege is going to leave you for Jinling.	Your brother Su is about to leave you and go to Jinling

the same conversation between the male lead and his doctor. Given the clear context of the scene, the subtitles were unlikely to add new meanings to the plot. Second, even if misunderstanding or confusion occurred, it would be difficult to discern which one had contributed to it.

In the comprehension test (see Appendix 1), the first three questions were multiple-choice, with one correct option, two incorrect options, and an I-don't-know option. The last question was a true-or-false question with an I-don't-know option.

The first question asked for the meaning of the term'文' (*wen*). This was machine-translated as 'words' because *wen* can be interpreted as '文字' (word) in some contexts. In the clip, *wen* as 'words' appears twice, presenting a potential source of confusion and allowing for novel interpretations of the scene.

The second question addressed the inconsistent translations of '金陵' (Jinling), denoting the capital city of Liang, which occurs three times in the clip. The different translations are 'Golden' (00:18), 'Golden Tomb' (03:30), and 'Jinling' (03:21). The participants were unlikely to establish a connection between these three terms and realise that the first two translations were incorrect.

The third question asked about the misinterpretation of the idiom '手无缚鸡之力' as '… man has no hands' rather than as a reference to physical weakness. Given the absence of supplementary visual cues about the person mentioned in that scene, the participants might be misled by the line and not detect the error in the dialogue.

Similarly, the final question concerned the mistranslation of '哥哥' (*gege*). This term encompasses two meanings: biological elder brother and a term of endearment with older males within the same generation. However, the machine translation rendered it as 'brother', potentially leading to misinterpretations regarding the relationship between the two characters.

3.4.4. Interviews

Our one-on-one interviews invited the participants to narrate how they interpreted the specific lines and how they were possibly misled by problematic subtitles. The participants first explained how they had made decisions while answering the comprehension test. Why had they chosen the incorrect option? How did they filter out the misinformation? The interviews then moved on to questions about the participants' perceptions of possible translation errors. They were asked whether they noticed any errors in the subtitles. If so, how many could they identify? Did presumed translation errors affect their

trust in the subtitles? None of the participants was informed that the subtitles had been generated by raw machine translation. They were told this at the end of the interview.

3.5. Procedure

The entire experiment was conducted in English on Zoom, which allows for screen-sharing, real-time conversations, and recording. The participants did the experiment non-concurrently. Before watching the clip, they were given a five-minute demonstration of think-aloud verbalisation, with a practice session and guidance on what should be verbalised. No limit was imposed on how long the participant could watch the clip for. The participants could initiate screen-based activities whenever they felt it necessary. In this sense, they were watching the video at their own pace and verbalising in their chosen manner.

After they finished watching the clip, each participant received a link to access the comprehension test. Each character in the clip was identified with a headshot and their name to ensure the participants could match the characters' names with their faces. It took the participants approximately five minutes to finish the test. They could see their test scores after they submitted all their answers.

Next, the interview sessions lasted 10–20 min and were audio-recorded. In the interviews, the participants explained their decision-making process during the comprehension test. They then indicated what they thought were translation errors and what affected their trust in the general quality of the subtitles.

To summarise, we initially identified eighteen errors in the clip. All or any of these might be expected to prompt participants' screen-based activities, verbalisations, or be addressed in interviews. Importantly, participants were not given an error typology, thus allowing them to define and construe 'error' as they saw fit.

4. Results

4.1. Translation errors identified or commented on by the participants

We extracted the utterances that address problematic subtitles through concurrent verbalisations and interviews. Some errors were effectively identified as mistranslations, while others were commented on by the participants who were unaware of errors but expressed uncertainty.

Twelve of our eighteen identified errors prompted participants to pause, go back, and/or make verbalisations while viewing and/or were later recognised as translation errors during the interviews. Of those twelve, all six of the errors we had classified as 'serious' were identified or commented on by different participants, along with four standard errors and two minor errors. We conducted a multiple regression test to model how the speed of subtitles and the type of error (minor, standard, serious) had influenced the participants' detection of the errors. Each error type was assigned a numerical value (1 for minor, 2 for standard, and 3 for serious). The results indicate that both variables affected error noticeability ($F = 7.72, R^2 = 0.51$). Table 2 demonstrates that greater error severity and slower subtitle speeds made errors easier to identify or comment on. This in turn indicates the validity of the FAR and NER models on which we had based our original identification.

Table 2. The severity of the error and speed of the subtitle (characters per second).

	Coefficients	Standard Error	*P*-value
Error type	1.94	0.54	.003
Cps	−0.26	0.10	.02

The three translation errors that elicited the highest occurrence of comments were as follows:

- Eight participants commented on the mistranslation of *wen* as 'words'. The mistranslation appears twice when Yan asks the girl standing beside the market stall about the prices of two different items. A04, A06, and A07, all with no studies of Chinese, first considered 'words' a mistranslation but then convinced themselves that the translation was intentional. They inferred that 'words' carried a special meaning in this context: as payment for each item, the girl would allow Yan to have a short conversation with her. Although this may seem like a creative overinterpretation, the participants construed all this information from the scenes and their reading does make sense in the context. Unlike non-Chinese-speaking participants, the language learners' (B05, B06, and B07) proficiency in Chinese helped them quickly identify the translation error; B02 even provided her translation version as 'yuan', the official currency of the People's Republic of China.
- Five participants noticed the problematic place name 'Golden'. A01, A04, andA06 were confused by this name without realising it was a mistranslation, while language learner B07 said it did not make sense and B06 was correct in identifying Jinling as the old name for the Chinese city Nanjing. Comparing the statements elicited by Groups A and B, it is evident that the language learners were more likely to demonstrate confidence when they encountered problems in the subtitles, instead of questioning their ability to reduce uncertainty.
- In the 'man with no hands' case, Yan uses '手无缚鸡之力' to describe the male lead Mei Changsu, who is physically weak. A04 and A06 were misled by the translation error. A06 went back to search for a character who had lost his hands – but he did not find one. This reaction suggests that the translation error resulted in misunderstandings. Further, the participants remained unaware of exactly where the subtitle went wrong and how they had been misled. Such errors have the potential to impact their comprehension of the subsequent storyline. A08 commented that the line was 'misleading' but failed to figure out what it meant. However, language learner B07 listened to the original soundtrack again and tried to figure out what was wrong with the subtitles.

In terms of subtitle speed, only two participants paused or went back to re-read because they did not have enough time to process the information. A02 went back three times due to fast subtitles. This may be because she had never watched any Chinese audiovisual products previously, as she explained in the interview. Interestingly, the characters per second (cps) measurement for the four lines that prompted pauses and revisits ranged from 14 to 24 cps, while the fastest line (38 cps) did not elicit pausing or revisiting.

4.2. Performance in the comprehension test

A t-test shows that there was no significant difference in the comprehension test scores between the means of Group A (M = 1.67, SD = 0.82) and Group B (M = 1.57, SD = 0.73; t (14) = 0.22, p = 0.42). However, the quantitative data obtained from the test results may have certain limitations due to the small sample size and the limited number of questions. It is crucial to supplement the quantitative data with qualitative insights from the interviews.

> Q1: When Yan asks the girl how much a guqin is, she says "twenty words". What does "words" mean?

Only one learner of Chinese, B01, failed to provide the correct answer here. She chose 'I don't know' because she did not know what *wen* meant in English. All the participants in Group A (non-Chinese-speaking participants) chose the correct answer ('unit of currency') because, as they explained, this choice aligned better with the context than did the two incorrect options ('unit of language' and 'time of the day'). As we have seen above, three viewers in Group A nevertheless construed 'words' as a payment in the exact number of words spoken to the girl. Although this interpretation made during the viewing process was erroneous, these viewers answered the question correctly in the subsequent comprehension test. This demonstrates viewers' conscious acts of rectification and the dynamic nature of sense-making, which will be further discussed below.

> Q2: The name of the capital of Liang is … ?

In the machine-translated subtitles, the capital city Jinling was translated as 'Golden', 'Golden Tomb', and 'Jinling'. Seven participants in the two groups selected either 'Golden' or 'Golden Tomb' since they saw the two terms in the subtitles. One language learner (B05) said that even though she heard 'Jinling' in the soundtrack, she decided to trust the subtitles instead of what she heard. Five participants picked 'I don't know'. A08 reported that he intentionally did not read the place names because he did not think they were important; A07 and B01 did not remember seeing anything related to the capital of Liang; A04 and A09 were confused about the three terms ('Golden', 'Golden Tomb', and 'Jinling') because they did not see how they were related and they did not know what 'Jinling' refers to.

However, four participants chose the correct option. A02 said, 'Although they mentioned those two places [Golden and Golden Tomb], based on which context they use it in, I thought it was this one [Jinling]'. B03 made the correct choice because she reported that she did not remember seeing 'Golden' or 'Golden Tomb' in the clip.

> Q3: Why does Yan look excited when they are going to meet Mei Changsu?

The first option is 'Yan heard that Mei Changsu has no hand, so he wants to see what he looks like'. Four participants in Group A and two participants in Group B chose this option because they reported that it was what they had seen in the subtitle, even though they found it confusing. This choice also indicates that the participants still somehow trusted the subtitles, despite this option making no sense. We will return to this act of trust below.

Six participants chose 'I don't know' because they claimed that they had skipped that line. Interestingly, two participants selected another incorrect option: 'Yan has not seen Mei Changsu for two years and he misses Mei Changsu a lot' because it seemed to make more sense in the context.

Only A01 and A08 dodged the traps in the question. A01 reported not noticing the incorrect subtitle and thus her answer was a fluke; A08 said that he saw the incorrect subtitle but he believed there were hidden meanings in the conversation, so he inferred from the previous scenes and reached a correct interpretation. However, no language learners selected the correct option, mainly because none of them knew the meaning of the idiom.

Q4: Mei Changsu is Fei Liu's elder brother. True or false?

Since the machine translation mistranslated *gege* as 'brother', we assumed that the participants were likely to select 'True', thinking that Mei Changsu was indeed Fei Liu's brother. Most participants' explanations of their decision-making process concorded with this prediction. However, three participants still chose the correct option. Their answers unveil how viewers can construe the implicatures in conversations. A05 realised that they were not brothers because they did not have the same surname. B07 worked out their relationship based on their interaction: 'He [Fei Liu] was on the roof before he left, so he may be the bodyguard or something'. And A01 was right for the wrong reason: she assumed that Fei Liu was Mei Changsu's elder brother.

4.3. Translation errors and trust

The 'translation errors' discussed below are the errors explicitly pointed out by the participants, either in their verbalisations, as in 'this is a mistranslation', or in their responses to the interview question 'How many translation errors can you recall in the video?'.

In Group A, six participants (66.67%) explicitly pointed out at least one error; in Group B, four (57.14%) identified at least one error. Of the ten participants who pointed out translation errors, B06 detected six, B07 discovered five, A09 identified three, and A01, A04, and A05 observed two each. The remaining four participants each spotted one.

In the interviews, the ten participants who identified at least one error were asked whether the errors they spotted affected their trust in the subtitles. Three reported that the presence of translation errors undermined their trust in the subtitles. These three were the ones who identified the highest number of errors. A09 found that the translation errors disrupted her immersion in the content. Initially, B06 distrusted the subtitles due to the inherent challenges of translating Chinese historical dramas. The translation errors he identified then served to reinforce his initial scepticism. B07's trust in the subtitles swiftly eroded because, in his words, 'they're no good from the start'. Despite their distrust, theirverbalisations indicated that they tried to match the subtitles with the original soundtrack and located the meanings of unfamiliar phrases, such as '午膳' (lunch) and '诊脉' (checking pulse). In other words, both B06 and B07 used the subtitles as vocabulary aids when expressions were correctly translated.

However, six participants stated that their trust in the subtitles was *not* affected by the translation errors because (1) the number of errors they identified was within an acceptable range, and (2) their previous exposure to subtitled shows made them relatively accustomed to inadequate translations. A01 said: 'It depends. If I'm already invested in it, then it's not going

to matter'. This indicates a willingness to accept the risks of any misunderstandings that may arise from such errors. The participant's level of interest in the audiovisual product ultimately outweighed the possible distrust that might ensue from erroneous subtitles.

5. Discussion

5.1. Misunderstanding as a chain reaction?

Pedersen (2017) points out that serious errors not only affect viewers' comprehension of a subtitle line but also cause problems with later lines. In this view, the seriousness of an error directly concerns the 'chain effect' the misunderstanding can provoke.

This echoes Levý's application of game theory to literary translation, formulated from the translator's perspective. Levý (1967, p. 1172) sees translating as 'a [chess] game in which every succeeding move is influenced by the knowledge of previous decisions and by the situation which resulted from them'. So can subtitle errors be seen as bad moves that might lead to a lost game? In some cases, yes, and most clearly when the viewer loses trust in the subtitles and gives up on the show. In other cases, though, the chains of potential misunderstandings are of limited extension (it is not a question of '*every* succeeding move') and are subject to processes of checking and self-correction, since the narrative structure and other channels of information also play a significant role in framing interpretations. To continue the analogy, there is more than one chess-board in play. This would be the case of 'word' as a unit of currency, 'Golden' as a place name, or *gege* as 'brother', none of which bears on anything down the line.

However, the effect of the visual channel on interpretation is not always positive. The official title of the female lead, *Junzhu*, ('郡主') was machine-translated as 'sheriff'. Some viewers may associate this with the Sheriff of Nottingham from various versions of Robin Hood, interpreting the term as a royal or noble official with authority. Viewers like A07could thus profitably construe the female lead's authority from the term. However, for anyone who associates sheriffs with men in Westerns, the translation might seem a false move. A08 and A09 erroneously thought that the 'sheriff' was the *male* lead Mei Changsu, since they assumed that a sheriff should be a man. We did not realise they had made this construal until they called Mei Changsu 'the sheriff' in the interviews. In this case, the subtitle blended in so well with the visual clues that it became difficult for viewers to notice any misinformation.

Levý (1967) saw translating as a game with finite information and a continuous chain of information, where each new piece had to be interpreted in sequence. Viewers, however, can rectify misinterpretations as the plot unfolds. Further, Levý's linear model does not account for the screen-based activities by which online viewers engage with the audiovisual content in a non-linear way. Although translation errors like 'words', 'Golden', 'brother', and 'sheriff' did cause misunderstandings, they did not block communicative cooperation (Pym, 2015). The viewers were mostly still willing to keep watching and continue their engagement.

5.2. Misunderstanding as a challenge to credibility

Narrative chains are not the only way a machine translation error can be of consequence. None of the three errors that received the most comments in the think-aloud protocols

('word', 'man with no hands', and 'Golden') makes sense in its immediate context; they were all able to provoke momentary perplexity; and that perplexity was in all cases explained away or otherwise neutralised. Such errors are clearly not fatal in themselves – they can be tolerated, in accordance with the variable appetences of the viewer.[2] That said, an accumulation of identified errors can eventually outweigh tolerance and challenge the credibility of the translation. That is, there is a point where the subtitles will be considered so untrustworthy that the viewer will refuse Pedersen's 'contract of illusion'. Such a loss of credibility can be tantamount to a translation ceasing to function as a translation (Pym, 2015).

An interesting question here is that loss of credibility is not a general phenomenon. For most of the participants, there was no point at which the subtitles risked being dismissed as a whole. In the interviews, the ten participants who identified errors were asked if those errors affected their trust in the subtitles. Only three answered 'yes'. It was more frequently reported that the errors were not severe enough to undermine trust, even though, as we have seen, other statements indicate that the identified errors had the potential to erode trust and lead to a negative evaluation of the audiovisual product *en bloc*. The viewers' capacity to make sense of the errors or otherwise tolerate uncertainty is thus more remarkable.

5.3. Why are some translation errors not identified?

The reason some translation errors go unnoticed can be understood from a different perspective. As indicated above, there is a correlation between the frequency of error identification and both the severity of the error and the subtitle speed. Thus, the visibility of a translation error ensues not solely from how well it aligns with the textual context but also from the cognitive effort required to process the line – since faster subtitles demand increased cognitive effort.

That said, our participants were allowed to pause or rewind, so they could have revisited the fast and/or confusing subtitles if they wanted to. Why did most participants not do so? The answer may concern the amount of effort they were willing to invest. On the one hand, the viewer might distrust the subtitle to the extent that the processing effort is not considered worthwhile. Alternatively, they might trust that the subtitle is correct and assume that the missing information can be picked up through other communicative clues – a regime of tolerance is thereby extended as blind faith. Either way, if the viewer keeps watching, any added cognitive load caused by possible errors must be in a trade-off situation with some alternative positive value. If the viewer invests more effort in a problematic subtitle, it must be in the hope of gaining some value: perhaps comprehension or pleasure (Pym & Hu, forthcoming). But could the added effort in itself have negative consequences for the viewing experience?

Particularly in narrative fiction, added cognitive effort can disrupt the viewer's immersive experience. When a character says something that does not make sense in the context, the viewer may initially fail to realise it is an error and instead perceive it as the character behaving somehow out of character. Consequently, the viewer risks being taken out of the fictional world. The seriousness of a noticeable error thus not only lies in the possible loss of the translation's credibility but also in the consequent risk of disrupting the willing suspension of disbelief (Pym, 2015).

There is nevertheless a further consideration here, particularly given that the setting of this show is a long way from the viewers' home cultures. Part of the appeal of the ancient Chinese setting lies in the value of the exotic, of the foreign milieu that remains partly mysterious, unknown, and thus awaiting discovery. Aesthetic pleasure arises from exploring new points and embracing imperfections in our understanding. This creates tension as part of the aesthetic experience (cf. Ingarden, 1937/1973). Indeed, the exotic appeal of a mysterious but unknown culture has elsewhere been found to add a sense of adventure that can at the same time justify limited cognitive expenditure: 'The effort needed to process an unfamiliar concept should be less than the enjoyment generated by exoticism' (Hu, 2022, p. 211).

This more complex sense of aesthetic enjoyment, which goes beyond a simple 'contract of illusion', might help explain the remarkable tolerance that most viewers displayed concerning the erroneous subtitles. Rather than simply diminishing their engagement in the plot, touches of the unknown seem to have marginally enhanced their viewing experience in some cases.

5.4. Is trust the same for everyone?

As we have seen, different viewers indicate different degrees of trust and distrust, with different kinds of reasons, making trust a dynamic variable relationship between viewers and subtitlers. We are not talking about a set of subtitles being intrinsically more or less trustworthy, but we are instead asking whom viewers (dis)trust and to what extent. The nature of the trust relationship has been studied in many disciplines, from behaviourist psychology to system sociology. It has also been recognised as an essential relationship for the success of translation as a communication act. Most notably, Chesterman (2016, p. 179)claims that without trust, 'the[translation] profession would collapse, and so would its practice'; or in our terms, errors are deemed 'fatal' when they incur distrust to the extent that communication is curtailed. In our study, though, we did not observe a collapse of trust when errors were identified.

In conceptualising the relationship between viewers and subtitlers, we draw particularly on Luhmann's (1968) trust as a 'reduction of complexity'. It is a risk-management strategy, not to be confused with the kind of trust that is based on familiarity, as when we trust a friend or a family member (Luhmann, 1988). That complexity could be reduced by language learning, but for most receivers, it is more efficiently reduced by trust in the translation – this kind of trust makes sense in an economy of effort rather than as an affective disposition. In such cases, viewers may trust subtitlersas long as the text works well enough or there is no contrary information that indicates the trust has been misplaced (Pym, 2015, p. 70). Once that trust is established, errors might not matter much if viewers are already invested in the drama. As noted, there is a trade-off between how much the viewer enjoys the show and what levels of translation-led misunderstandings they are willing to risk. As observed in Tuominen et al. (2023), viewers can accept imperfect machine-translated subtitles if the translated text gives them access to something they genuinely need or want. This provides a rationalist model of the remarkable tolerance that we have seen.

That said, self-confidence can also play a significant role in guiding trust formation and the decision-making process (Hoff & Bashir, 2015). In our study, certain Chinese-speaking participants' responses to errors highlight the interplay between trust in subtitles and

confidence in one's proficiency in the source language. Generally speaking, the learners of Chinese could fall back on their proficiency in Chinese to mitigate the risks of misunderstanding. As shown in their verbalisations and responses to interview questions, they were more given to re-listening to the original soundtrack to check what was wrong with the subtitles. However, some language learners trusted the translated subtitles more than they trusted themselves, displaying a lack of self-confidence. B05's decision-making process when answering Question 2 (see 4.2) indicated that her trust in the subtitles largely came from a lack of confidence in her own proficiency in Chinese. And none of the translation errors was critical enough to remind her that her trust had been misplaced. Thus, when participants' trust was low and self-confidence was high, the accuracy of subtitles was questioned more frequently; the opposite effect occurred when trust was high and self-confidence was low. These particular relationships are nevertheless not extreme enough to bring about the complete shift from trust to distrust.

Two other cases illustrate a more complex dual relationship of trust and distrust, where both coexist independently or influence each other simultaneously, as outlined by Zand (2016, p. 70). The postgraduate translation students B06 and B07 reported a lack of trust in the translations from the outset. Their distrust was rooted in experiential aspects tied to the audiovisual content and subtitle conventions. With extensive knowledge of the constraints of subtitles and awareness of the challenges faced when translating period dramas, they exhibited what Yu et al. (2018) term 'initial learned trust' – knowledge acquired before interaction. This initial state of learned (dis)trust evolves dynamically as users interact with the text, developing new experiential knowledge of its variable reliability and usefulness. Consequently, we observed the coexistence of trust and distrust in these cases, where subtitles served as a vocabulary learning tool and participants treated subtitle reading as an error-detection game.

6. Conclusions

We find that in audiovisual reception, viewers use multiple channels of information to construe meanings, such that a translation error can be overridden by other communicative clues. A misinterpretation can still work successfully in context (e.g., *wen* as word) and a correct interpretation can be for the wrong reasons (e.g., *gege* as older brother), with no particular consequence. Errors are more likely to go unnoticed when they cause little or no confusion or misunderstanding and/or are in fast subtitles. The impact of an error can therefore depends on the semantic network it is in (as in Levý, 1967), the degree to which it can be tolerated in the local context, and the extent to which accumulated errors challenge trust. Needless to say, all these findings merit investigation on a wider range of cultural projects and with larger samples of viewers.

If a general explanation can be advanced based on our relatively small data set, it would be that our viewers' remarkable tolerance of subtitle errors can be understood in terms of a trade-off with their enjoyment of the show. That is, when the viewing experience is pleasurable because of immersion in the narrative world and/or the appeal of the exotic (for which a certain degree of doubt and intrigue is necessary), that pleasure can outweigh the added effort required to negotiate translation errors. The general trade-off would therefore be: *The subtitles might be dubious but the show is enjoyable*. When this holds, there need be no fatal errors.

On the other hand, language learners cannot be expected to seek enjoyment as the sole value to place on the scales. If they also hope to improve their language skills, that becomes another value that enters a rough three-way calculation. Their most positive trade-off could be along the lines of: *The subtitles are dubious, but I can still learn something and the show is still enjoyable, even if as an error-detection game*. Again, the errors need not be fatal.

Notes

1. This study was approved by the University of Melbourne Humanities and Applied SciencesHuman Ethics Sub-Committee (ID: 14177). Participants were assured that the privacy and confidentiality of the collected information would be maintained at all times.
2. 'Tolerance of ambiguity' is a psychological variable that has been studied since Frenkel-Brunswik (1949), sometimes analysed as 'uncertainty avoidance'. In translation studies, it has been applied to the way different translators negotiate uncertainty in the source text (cf. Fraser, 2000) but not, as far as we know, to the reception of translations. Since our study includes no data on the personalities of the viewers, we do not analyse tolerance of ambiguity as a variable in itself.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Juerong Qiu holds a PhD in translation studies from the School of Languages and Linguistics at the University of Melbourne. Her doctoral project delves into how target-language viewers and language learners read, understand, and appreciate subtitled shows. She teaches subjects related to translation technologies and specialised translation at the University of Melbourne. Email: juerong.qiu@unimelb.edu.au, qiu_juerong@163.com.

Anthony Pym is Professor of Translation Studies at the University of Melbourne, distinguished Professor of Translation and Intercultural Communication at Universitat Rovira i Virgili in Spain, and Extra-ordinary Professor at Stellenbosch University in South Africa. He was President of the European Society for Translation Studies from 2010 to 2016. He holds a doctorate in sociology from the École des Hautes Études en Sciences Sociales, Paris. He has authored or edited some 30 books and more than 230 articles and book chapters in the field of translation and intercultural studies. Email: anthony.pym@unimelb.edu.au.

ORCID

Juerong Qiu http://orcid.org/0000-0002-1011-7988
Anthony Pym http://orcid.org/0000-0002-9440-0886

References

Burczynska, P. (2017). *Investigating the multimodal construal and reception of irony in film translation: An experimental approach* [Unpublished doctoral dissertation]. University of Manchester.

Chesterman, A. (2004, May 17). *Functional quality* [Video]. YouTube. https://www.youtube.com/watch?v=IJW1Y6rAB1I

Chesterman, A. (2016). *Memes of translation. The spread of ideas in translation theory* (Revised version). John Benjamins.
Coleridge, S. T. (2014). *Biographia literaria*. Edinburgh University Press (Original work published 1817).
Costa, Â, Ling, W., Luís, T., Correia, R., & Coheur, L. (2015). A linguistically motivated taxonomy for machine translation error analysis. *Machine Translation*, *29*(2), 127–161. https://doi.org/10.1007/s10590-015-9169-0. http://www.jstor.org/stable/44113801
Fraser, J. (2000). What do real translators do? Developing the use of TAPs from professional translators. In S. Tirkkonen-Condit & R. Jääskeläinen (Eds.), *Tapping and mapping the processes of translation and interpreting* (pp. 111–122). John Benjamins. https://doi.org/10.1075/btl.37.11fra
Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. *Journal of Personality*, *18*(1), 108–143. https://doi.org/10.1111/j.1467-6494.1949.tb01236.x
Gambier, Y. (2018). Translation studies, audiovisual translation and reception. In E. Di Giovanni & Y. Gambier (Eds.), *Translation studies, audiovisual translation and reception* (pp. 43–66). John Benjamins. https://doi.org/10.1075/btl.141.04gam
Gutt, E. (2014). *Translation and relevance. Cognition and context* (2nd ed.). Routledge (Original work published 1991).
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
Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570
Hu, B. (2022). Feeling foreign: A trust-based compromise model of translation reception. *Translation Studies*, *15*(2), 202–220. https://doi.org/10.1080/14781700.2022.2032306
Hu, K., O'Brien, S., & Kenny, D. (2020). A reception study of machine translated subtitles for MOOCs. *Perspectives:Studies in Translation Theory and Practice*, *28*(4), 521–538. https://doi.org/10.1080/0907676X.2019.1595069
Ingarden, R. (1973). *The cognition of the literary work of art*. R. A. Crowley, & K. R. Olson (Trans.). Northwestern University Press (Original work published 1937).
Kong, S., & Li, X. (Directors). (2015). *Nirvana in fire (Season One)*. https://www.viki.com/tv/22943c-nirvana-in-fire
Levý, J. (1967). Translation as a decision process. In *To Honor Roman Jakobson: Essays on the Occasion of His 70. Birthday, 11. October 1966: Vol. 1* (pp. 1171–1182). https://doi.org/10.1515/9783111349121-031
Luhmann, N. (1968). *Vertrauen. Ein Mechanismus der Reduktion sozialer Komplexität*. Ferdinand Enke. Blackwell.
Luhmann, N. (1988). Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta (Ed.), *Trust making and breaking. Cooperative relations* (pp. 94–108).
Orrego-Carmona, D. (2015). *The reception of (non)professional subtitling* [Unpublished doctoral dissertation]. Universitat Rovira i Virgili.
Pedersen, J. (2011). *Subtitling norms for television*. John Benjamins.
Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *Journal of Specialised Translation*, *28*, 210–229. https://www.jostrans.org/issue28/art_pedersen.pdf
Pym, A. (2015). Translating as risk management. *Journal of Pragmatics*, *85*(August), 67–80. https://doi.org/10.1016/j.pragma.2015.06.010
Pym, A., & Hu, K. (forthcoming). Trade-offs in translation effects. Illustrations and methodological concerns. *Target*.
Pym, A., & Torres-Simón, E. (2021). Is automation changing the translation profession? *International Journal of the Sociology of Language*, *2021*(270), 39–57. https://doi.org/10.1515/ijsl-2020-0015
Romero-Fresco, P., & Pérez, J. M. (2015). Accuracy rate in live subtitling: The NER model. In R. BañosPiñero & J. Díaz Cintas (Eds.), *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28–50). Palgrave Macmillan.
Tuominen, T., Koponen, M., Vitikainen, K., Sulubacak, U., & Tiedemann, J. (2023). Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution. *Journal of Specialised Translation*, *39*, 77–98. https://jostrans.org/issue39/art_tuominen.pdf

Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). Error analysis of statistical machine translation output. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, 697–702.

Yu, K., Berkovsky, S., Conway, D., Taib, R., Zhou, J., & Chen, F. (2018). Do I trust a machine? Differences in user trust based on system performance. In J. Zhou & F. Chen (Eds.), *Human and machine learning* (pp. 245–264). Springer International Publishing. https://doi.org/10.1007/978-3-319-90403-0_12

Zand, D. E. (2016). Reflections on trust and trust research: Then and now. *Journal of Trust Research*, *6*(1), 63–73. https://doi.org/10.1080/21515581.2015.1134332

Appendix 1. Questions in the comprehension test

Thank you so much for participating in this research!

In this comprehension test, there will be 4 questions for the video you just watched. Please answer the questions as accurately as possible based on your understanding.

At the end of the survey, you will know your final score.

1. When Yan asks the girl how much a guqin is, she says 'twenty words'. What does 'words' mean?

[Yan's name and headshot here]

(A) Time of the day
(B) Unit of currency
(C) Unit of language

(D) I don't know.

2. The name of the capital of Da Liang is:
(A) Golden
(B) Golden Tomb
(C) Jinling

(D) I don't know.

3. Why does Yan look excited when they are going to meet Mei Changsu.

[Yan's name and headshot here] [Mei Changsu's name and headshot here]

(A) Yan heard that Mei Changsu has no hand, so he wants to see what he looks like.
(B) Yan heard that Mei Changsu has no ability to fight, so he wants to know how he can lead an alliance.
(C) Yan has not seen Mei Changsu for two years and he misses Mei Changsu a lot.

(D) I don't know.

4. Mei Changsu is Fei Liu's elder brother. True or false?

[Mei Changsu's name and headshot here] [Fei Liu's name and headshot here]

(A) True
(B) False
(C) I don't know.