

The language learners' reception of fansubs and raw machine-translated subtitles: A pilot study

Juerong Qiu

University of Melbourne

Abstract

This study explores how learners of Chinese react to and engage with fansubs and raw machine-translated subtitles and whether the two modes of subtitling are effective in facilitating vocabulary learning. Sixteen L1 English speakers with middle and high levels of proficiency in Chinese were invited to watch a nine-minute clip extracted from a Chinese period drama. Screen recordings, think-aloud protocols, vocabulary pre- and post-tests, and interviews were used. The results show that the combination of both modes of subtitles had a positive effect on vocabulary learning, regardless of proficiency levels. Learners often underestimated the quality of machine translation and associated erroneous subtitles with machine translation output. This highlights the need for increased machine translation literacy. The study emphasises the importance of learners' motivation and genuine interest for effective language learning. Additionally, it sheds light on the ongoing discussion about the use of subtitles in language learning and can potentially broaden learners' range of learning resources and improve their overall learning outcomes.

Keywords

Machine translation, audiovisual translation, language learning, viewing habits, motivation

1. Introduction

Audiovisual translation (AVT) is an essential tool that enables learners to think about language in an enjoyable and holistic way (Neves, 2004). Research has explored the impact of different subtitle conditions on language learning, including no subtitles, intralingual subtitles, interlingual subtitles, and bilingual subtitles (e.g., Baranowska, 2020; Kruger & Steyn, 2014; Liao *et al.*, 2020). However, much of this research has focused on videos subtitled from English into other languages. Subtitling from other languages into English is less commonly studied.

Moreover, the use of subtitles in language learning has experienced notable transformations in how learners interact with media content, owing to the increasing popularity of fansubs and the continuously improving accuracy of machine-translated subtitles (MTS). Fansubs have emerged as widely accessible and freely available resources on various online platforms, offering learners a diverse range of subtitled content spanning various genres, languages, and cultures. Many language learners and translation students use fansubbing to improve their language skills (Liu & De Seta, 2014; Lakarnchua, 2017). Automated subtitling, on the other hand, has undergone significant advancements. In 2019 Google introduced Live Captions, which provide automatic intralingual subtitles on users' phones. The captioning happens in real-time, completely on-device without using network resources (Tadmor-Ramanovic & Bar, 2019). In the same year, Viki, a video-streaming platform targeting fans of Asian TV shows, collaborated with the Rakuten Institute of Technology to introduce automated subtitling—Vikibot (Zhan, 2022). Given these changes, it has become crucial to investigate the potential learning benefits of fansubs and raw MTS.

A comparison in learning effects between fansubs and raw MTS is made to see whether fansubs have better effects on language learning than raw MTS and whether language learners can tell the difference between the two modes of subtitling. The participants' performance in vocabulary tests as well as their use and perception of raw MTS can provide insights into the didactic potential of raw MTS for language learning. The results of this study also have significant implications for language educators, providing clues about how to incorporate raw MTS into language training and how to help language learners approach machine translation (MT) in a critical and informed way.

2. Fansubs and MTS in language learning

Fansubbing is often regarded as a form of user-led production because fansubs are created by fans and for fans. In fansubbing, the translator's notes are used to explain "untranslatable" cultural references or other culture-specific elements in audiovisual materials (Díaz-Cintas & Muñoz Sánchez, 2006). The presence of the translator's notes not only increases the visibility of fansubbers but also helps the audience who wants to know more about the background or contextual information (Wang, 2017). Previous research on translator's notes in fansubs often explores their effect on comprehension, but the question of whether translator's notes can be helpful in vocabulary learning is not closely examined. For example, Künzli and Ehrensberger-Dow (2011) found that the translator's notes did not have a significant effect on comprehension. Caffrey's (2012) research showed that using a translator's note with a one-line subtitle can facilitate better understanding, but a translator's note with a two-line subtitle does not yield the same beneficial effects.

Benson and Chan (2010) investigated language-related interactions in the comment section of a fansubbed clip of the 2008 Olympics song on YouTube. They found that the fansubs and *pinyin* provided in the subtitle area help language learning. The comment section is also useful for language learners to ask questions about language-related and culture-related issues and receive answers from others. Bolaños García-Escribano (2017) explored the potential benefits

of using fansubs as a didactic resource in English as a foreign language (EFL) education. Views on the application of fansubs in EFL classrooms varied significantly. Recurring translation errors, lack of professionalism, literal translations, and synchronisation errors caused problems in language learning, but fansubs can be used for comparative translation analysis, error-solving, and proofreading.

Similarly, the pedagogical application of MT in subtitling has also not been closely investigated, partly because the subtitling industry is not ready for the comprehensive implementation of automation (Karakanta, 2022). For example, Chan *et al.* (2019) compare the effects of auto-generated intralingual subtitles and corrected intralingual subtitles on cognitive load and learning. The findings indicate that neither auto-generated nor corrected intralingual subtitles had a significant impact on performance in the post-test or self-reported cognitive load. Chan *et al.* (2019) suggest that this could be two factors. First, the low quality of auto-generated intralingual subtitles may cause distractions. Second, the speed of corrected subtitles is significantly faster than automated subtitles, leading to comprehension difficulties. Malakul and Park (2023) compared auto-subtitles, human-translated subtitles, and no subtitles to assess the impact on learning. The study involved 79 Thai secondary school students who took English listening and video content-learning comprehension tests, as well as cognitive load and learner satisfaction scales. The results showed no significant difference between the auto-subtitles and human-translated subtitles groups in terms of comprehension, cognitive load, and satisfaction.

The present study investigates a relatively underexplored aspect: subtitling from Chinese into English. Furthermore, this paper presents an innovative examination of how language learners' perceptions of MT and their consumption habits of subtitled audiovisual content influence their use of MTS.

3. Methods

The questions driving this research are: (1) How do fansubs and MTS affect vocabulary learning?; and (2) How do language learners perceive MT in subtitling?

I hypothesise that fansubs facilitate better vocabulary learning outcomes than MTS. I assume that errors and other problematic aspects have the potential to divert the viewer's attention and create difficulties in construing the contextual meanings of certain vocabulary items. Since pure MTS contain more errors than fansubs (see examples in Table 1 below), MTS are more likely to disrupt the vocabulary learning process.

3.1. Design

The study employed a comprehensive research methodology, which included the administration of questionnaires, vocabulary pre-tests and post-tests, think-aloud protocols, screen recordings, and post-hoc interviews (see sections below for details). The research design and methodology have undergone a rigorous review and received approval from the Human Research Ethics Committee at the University of Melbourne.

3.1.1. Materials

The clips shown to participants were extracted from the Chinese period drama *Nirvana in Fire* (Lang Ya Bang, 琅琊榜) Season One (Kong & Li, 2015). The clip used in the pre-experiment questionnaire was extracted directly from the original show, with Chinese subtitles only and the original soundtrack. The clip lasts 100 seconds. The plot is carried out by conversation between characters, so audiences need to have at least middle levels of proficiency in Mandarin to understand it. Further, the images and scenes in the video offered little help when the viewers had to choose the correct answer in the pre-experiment survey.

The clip used in the main experiment lasts 9 minutes and 19 seconds. This clip introduces the background of the entire drama. The subtitle content of the clip is divided into two, based on duration and plot. A logical break in the storyline was selected as close as possible to the halfway point (04:30) of the overall duration of the clip. One half of the clip is the online fansubbed version and the other half is the machine-translated version. Two separate groups were given two different versions of the combined modes of subtitles (see Table 3 below for details). The fansubbed version I used in the research was subtitled and timed by the Viki fansubber team, and the MTS were generated by DeepL.

3.1.2. Participants

Snowball and convenience sampling methods were used to select participants. Sixteen students at the University of Melbourne were selected after pre-experiment tests. They were intermediate or advanced learners of Chinese who were native English speakers aged between 18 and 34. Apart from A01 and A08, the remaining fourteen participants were currently enrolled in Chinese courses at the University of Melbourne (Chinese 3, Chinese 5, Chinese 7, and Master of Translation in the Chinese language).

To test the possible correlation between scores in the pre-experimental listening comprehension test and the language levels of Chinese indicated by the courses that respondents currently took, I excluded test results given by A01 and A08 who were not currently enrolled in any Chinese courses at the University of Melbourne because these two participants' language proficiency was not measured by the same standards set by the University of Melbourne. It turns out that higher scores in the pre-experiment language test correlate with higher levels of education in Chinese ($r(13) = .60, p = .022$).

3.1.3. Pre-experiment survey

The pre-experiment questionnaire was used to select potential participants for the experiment. The pre-experiment questionnaires first asked all respondents about their age, gender, educational background, and language background. This was followed by a listening comprehension test.

The language learners watched a 100-second clip first and then answered five multiple-choice questions based on what they had heard in the soundtrack and what they had read in the Chinese subtitles. Once a participant clicked Next to see the questions, the action was irreversible. Any form of screen-based activity while watching was prohibited. If they did not follow the instructions, their test results were considered invalid. The questions had three options: one correct answer, one incorrect option, and an I-don't-know option. Each question was worth one point, making five points in total. Based on the number of correct answers, the participants were divided into three groups: low (0-2 correct answers), intermediate (3-4 correct answers), and high (5 correct answers). Those who scored between 3 and 5 were selected to participate in the experiment: they knew the basic vocabulary of Chinese and could pick up some information from the original Chinese soundtrack, but they still needed English subtitles to fully understand the storyline and plot.

Eighty-nine learners of Chinese provided complete answers to the pre-experiment questionnaires. Sixty-five were native English speakers aged between 18 and 34. Twelve responses were excluded because they had screen-based activities during the test. That is, their scores on the language test may have failed to represent their actual listening comprehension skills. The remaining fifty-three responses were considered valid. Thirty-seven respondents scored between 3 and 5 in the listening comprehension test. Of the 37 respondents, 25 left their email addresses and expressed their willingness to participate in the experiment. Sixteen of them agreed and took part in it.

3.1.4. Screen recordings and think-aloud protocols

Screen recordings are used to capture the participants' screen-based activities. Think-aloud protocols can provide clues about their thought process and sense-making process at certain points in time during viewing. Some of their verbalisations can also showcase their reactions and responses to problematic subtitles. The entire viewing process from the participant's situated perspective was recorded via Zoom. Meanwhile, when participants enabled the audio and video functions on Zoom, their verbalisations and facial expressions could be captured at the same time.

The participants were asked to start screen-sharing when they were ready to watch the clip provided to them. The screen-based activities were limited to pauses, skips, going back, and fast-forwarding. Other activities, such as rewatching the entire clip or posting comments, were prohibited. The participants were informed that the screen-based activities were not mandatory. They could take those actions whenever they wanted.

The time for thinking aloud was unlimited. To mitigate the participant's possible anxiety, I provided a demonstration accompanied by a practice session, showing them what kinds of screen-based activities were allowed and how to verbalise their thoughts while watching. The practice sessions helped participants to familiarise themselves with the research process. Therefore, the practice session with demonstrations was provided to (1) make sure the screen-sharing function on Zoom worked well on the participant's device, and (2) remind participants that they needed to verbalise their thoughts while viewing the clip.

The participants communicated their thoughts to the researcher, making the thinking-aloud process an ongoing conversation between the researcher and the participant. Prompting ("yes", "okay", or "oh") and nonverbal responses (smiling or nodding) were used simply to encourage participants to talk more, instead of pushing them to verbalise their thoughts. Sometimes, participants asked me a question regarding the plot or the meaning of a specific phrase during viewing. If it was a simple yes-or-no question and was not related to the vocabulary test, I would answer them directly; if the question requires a detailed explanation, I would say "We can talk about this later" or "What do you think".

3.1.5. Vocabulary tests

Vocabulary pre-tests and post-tests are widely used in accessing language learning in AVT (Baranowska, 2020; Chan *et al.*, 2019; Peters *et al.*, 2016). However, the vocabulary test in the present study serves as a prompt to encourage active attention from the language learners towards both the subtitles and the original soundtrack, rather than solely testing whether the learners can pick up new vocabulary items.

The questions in the pre-test and post-test were the same: eight Chinese vocabulary items in the clip. The first four vocabulary items appear in the first half of the clip, and the remaining four are in the second half (see Table 1).

No.	Time	Vocabulary items	Literal translation	FAS ¹	MTS ²
1	00:03	文	unit of currency	wen-unit of currency	word
2	00:48	大	great	da- honorific, lit. the Great	da
3	01:42	诊脉	check pulse	check pulse	feel pulse
4	03:21	哥哥	term of endearment for older males	gege-endearment for older males	brother

¹ FAS: fansubs.

² MTS: raw machine-translated subtitles.

5	04:32	姐姐	term of endearment for older females	jiejie-endearment for older females	sister
6	05:51	长进	progress	progress	improve
7	06:28	休养	recuperate	recuperate	rest
8	07:46	江湖	rivers and lakes	the pugilist world	around the world

Table 1. Eight Chinese phrases presented in the vocabulary test (source: author)



Figure 1. A translator's note (source: Viki)

“文” (1), “诊脉” (3), “休养” (7), and “江湖” (8) are mistranslated in the MT, although the translation of the last phrase in fansubs could also confuse viewers. The MT of “大” (2) uses the character's *pinyin* without further explanation. “哥哥” (4) and “姐姐” (5) literally mean “older brother” and “older sister” respectively, but they do not necessarily indicate a familial relationship. The two terms in the clip are suffixes after a person's given name to show admiration and proximity. The machine translations of the two terms are inadequate in the context. Thus, except for “长进” (6), the translations of the remaining seven phrases in the MTS are inadequate or even incorrect.

Vocabulary items 1, 2, 4, and 5 are selected because they are explained in the translator's notes (see an example in Figure 1). The purpose is to assess whether these translator's notes aid language learners in grasping the contextual meanings of these four vocabulary items. Since vocabulary tests have the potential to naturally increase the language learners' attention towards the subtitles, problems in subtitles may be more likely to be identified. Thus, adding mistranslations in the vocabulary tests can not only examine how mistranslations impact the interpretation of specific phrases but also explore how language learners react to problematic subtitles.

Further, repetition can help language learners pick up new vocabulary items. Vocabulary items 3, 6, 7, and 8 are chosen from the video specifically because they appear at least twice within one back-and-forth conversation. The last vocabulary item “江湖” (8) is special because it appears in both the first half and the second half of the video, so all language learners encountered fansubbed and machine-translated versions. The fansubbed version has been widely discussed in the viewers' comments on Viki. This cultural-specific item is thus selected to explore whether language learners can make sense of “the Pugilist world” with the help of MTS.

The pre-test was given to participants before they watched the clip. In this test, the participants were presented with the vocabulary items one at a time, on the screen. They were required to

verbally explain the meaning of each item. The language learners were aware that they would be asked the meaning of the eight words again after they had finished watching.

The questions were adopted from the Revised Vocabulary Knowledge Scale (Zhao & Macaro, 2016). A participant's knowledge of a vocabulary item is categorised according to three levels: (1) they have not seen the word before, (2) they have seen the word and know the word is related to a certain field, and (3) they know the meaning of the word. One point is given to level one, two points to level two, and three to level three. Of the eight vocabulary items, four have specific meanings (“文”, “大”, “哥哥”, and “姐姐”) in the context. One extra point would be given to participants who identified the contextual meanings. Thus, the total score for both pre-test and post-test was 28.

3.1.6. Interviews

All the interviews were conducted individually and recorded on Zoom. The interview sessions not only supported and complemented the screen recordings and verbal reports, but also delved into language learners' attitudes toward MTS and their effectiveness in language learning.

The interview was semi-structured. I asked participants open-ended questions and raised some follow-up questions based on their responses, so this phase lasted from 10 to 20 minutes. I arranged the interview based on two themes:

- (1) Can they identify which part is subtitled by MT and which part by humans?
- (2) Would they consider using MTS to learn Chinese?

3.2. Procedure

Participants in the experiment were divided into two groups based on their scores in pre-experiment tests. The number of participants in the two groups is uneven (nine in Group A and seven in Group B) because I need to make sure the mean of the scores in both groups is the same (see Table 2). This is to avoid the possible effects of differences in listening comprehension skills on the group comparison. Participants in the two groups would encounter the clip with the same content and two modes of subtitles in reversed sequence (see Table 3 for details).

Group	Participants	Test score	Mean
A	01	3	4.0
	02	3	
	03	4	
	04	4	
	05	4	
	06	4	
	07	4	
	08	5	
	09	5	
B	01	3	4.0
	02	3	
	03	4	
	04	4	
	05	4	
	06	5	
	07	5	

Table 2. Participants' scores in the listening comprehension test (source: author)

Participants	Group A	Group B
Number of participants	9	7
Subtitle sequence	FAS+MTS	MTS+FAS

Table 3. Number of participants and the sequences of subtitle modes used (source: author)

When the participant was ready to proceed, I started recording the process. Each participant responded to the vocabulary pre-test before watching the clip. They could pause, skip, go back, and fast forward whenever they felt necessary during viewing. In terms of verbalisation, they did not need to click “Pause” before articulating their thoughts, but they were allowed to do so if they wanted to. In this sense, participants were watching the clip at their own pace and verbalising in their chosen manner. After they finished watching the clip, it took approximately five minutes for language learners to complete the vocabulary post-test. They then began the interview sessions, which lasted 10 to 15 minutes and were audio-recorded.

4. Results

4.1. Test results

A t-test was used to explore whether the research material helped the participants to learn new Chinese vocabulary items. There was a significant difference in vocabulary scores between before watching the research material ($M = 19.50$, $SD = 3.12$) and after watching it ($M = 24.06$, $SD = 2.69$); $t(15) < -.001$, $p < .001$). It means that the clip shown to participants in the main experiment had a positive effect on vocabulary learning. As shown in Table 4, the highest progress rate was 29.17% (A03 and A06) and the lowest progress rate was 7.14% (B06).

No.	Pre-test	Post-test	Improvement (%)
A01	18	24	25.00
A02	19	24	20.83
A03	17	24	29.17
A04	19	25	24.00
A05	15	21	28.57
A06	17	24	29.17
A07	24	27	11.11
A08	22	27	18.52
A09	23	25	8.00
B01	17	19	10.53
B02	18	22	18.18
B03	18	25	28.00
B04	17	19	10.53
B05	19	24	20.83
B06	26	28	7.14
B07	23	27	14.81

Table 4. Language learners' scores and improvement in the vocabulary test (source: author)

The pre-experiment test for the language learners was used to measure the participant's listening comprehension skills, and performance in the pre-experiment test also correlated with their levels of proficiencies in Chinese indicated by the Chinese course they enrolled in (see 3.1.4 for details). There are significant positive correlations between language learners' test results in the pre-experiment language test and their performance in the vocabulary pre-test ($r(14) = .73$, $p = .001$) and post-test ($r(14) = .67$, $p = .004$). In other words, participants with higher levels of proficiency in Chinese were more likely to receive higher scores in the vocabulary pre-test, and a positive linear correlation also occurred in the post-test results.

However, there was no significant correlation between the progress rate and performance in the listening comprehension test ($r(14) = -.37, p = .152$). This does not necessarily indicate that the vocabulary-learning effects of watching the video are not associated with language learners' levels of proficiency in Chinese. For instance, participant B206 achieved the highest scores in both the pre-test and post-test, resulting in a lower progress rate due to limited room for improvement. Conversely, participants with lower performance in the vocabulary pre-test had greater potential for improvement.

No.	First half	Second half
A01	15.38	36.36
A02	7.69	36.36
A03	21.43	40.00
A04	15.38	33.33
A05	30.77	25.00
A06	15.38	45.45
A07	13.33	8.33
A08	13.33	25.00
A09	0.00	16.67
B01	8.33	14.29
B02	15.38	22.22
B03	15.38	41.67
B04	8.33	14.29
B05	7.69	36.36
B06	0.00	15.38
B07	6.67	25.00

Table 5. Language learners' progress rates (%) in the first and second half of the clip (source: author)

For inter-group comparison, two t-tests were conducted for the first half of the video and the second half respectively (see Table 5 for details). In the first half, there was no significant difference in the mean progress rate between Group A ($M = 14.75\%$, $SD = .08$) and Group B ($M = 8.83\%$, $SD = .05$; $t(14) = -1.71$; $p = .109$). In the second half, there was still no significant difference in the mean progress rate between Group A ($M = 29.61\%$, $SD = .12$) and Group B ($M = 24.17\%$, $SD = .08$; $t(14) = -.95$; $p = .361$). This suggests that there is no statistically significant difference in the progress rates between the two groups neither in the first half of the vocabulary test nor the second half. There is no evidence suggesting that one mode of subtitling is more effective in facilitating vocabulary learning than the other. However, no matter whether for Group A or Group B, the mean progress rate in the second half is higher than the first half ($t(8) = 1.86$, $p = .008$; $t(6) = 1.94$, $p = .006$). It is possible that the language learners were inclined to remember more vocabulary items in the second half, as restrictions in working memory may reduce a person's processing capacity (Sweller *et al.*, 2011).

4.2. Screen-based activities and verbal reports

Six participants had screen-based activities and initiated 34 screen-based activities in total. In the main experiment, B03 did not verbalise her thoughts and three participants (A05, A07, and A09) had minimal verbalisations (uttered fewer than five statements) during the viewing process.

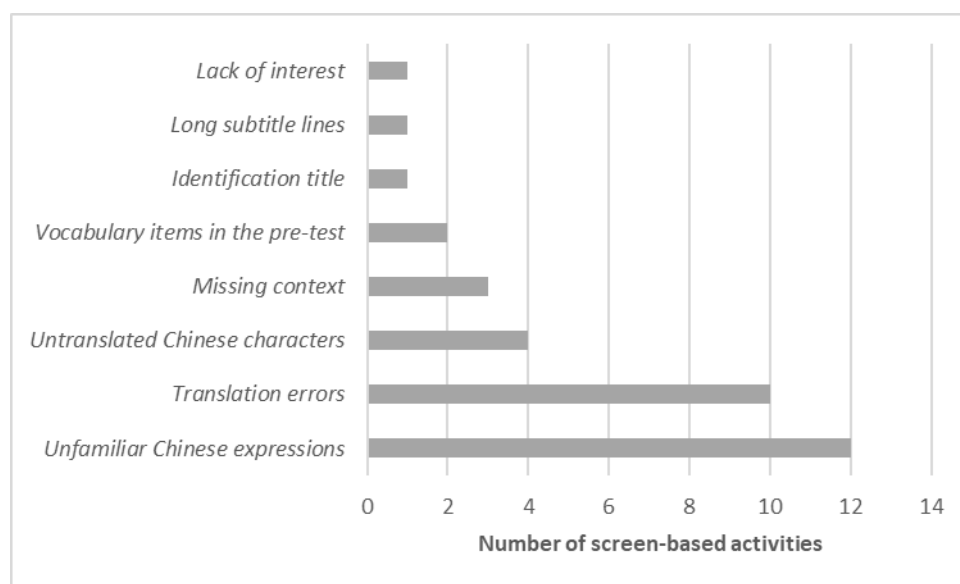


Figure 2. Reasons for screen-based activities by participants (source: author)

Figure 2 presents the eight self-reported reasons that trigger screen-based activities. The two screen-based activities regarding the two vocabulary items (“休养” meaning recuperate and “长进” meaning improvement) and in the pre-test were initiated by the same participant (B07). The other vocabulary item “文” (unit of currency) is mistranslated as “word” in the MTS. Although it did not trigger screen-based activities, three participants commented on it during viewing:

B04: Is she really saying it costs 20 words? Because I heard 文. Does she mean *yuan* [the official currency of the People’s Republic of China]?

B06: Looks like there was a bit of a mistake they made with the subtitles 10 words. But let’s pretend it wasn’t.

B07: Ten words. That doesn’t make any sense.

The scene is where Yan asks the girl standing beside the market stall how much a musical instrument is and the girl replies, “ten wen”. Clearly, *wen* must be a unit of currency. In the machine-translated version, *wen* (文) was translated as “word” because “文” can be interpreted as “文字” (word) in some contexts. Although the three participants did find the term “word” out of place, they did not have problems making sense of the term.

Unfamiliar expressions in the original soundtrack triggered most screen-based activities. The twelve pauses and rewinds were initiated by four participants. Some of their statements are as follows:

A08 [pausing 2 seconds at 00:48]: What... 大渝 [Country name]. Okay, I have no idea...

B02 [going back from 08:32 to 08:27]: When they say 请, the subtitle is this way. I don’t know 请 can mean this.

B07 [going back from 00:20 to 00:17]: I didn’t catch that part. “午膳时分” [during lunchtime] Okay, it seems that he is using an old way to talk about time here.

Their verbalisations indicate that pauses and rewinds gave them a second chance to listen to and respond to unfamiliar expressions in semi-classical Chinese. After they had picked up

what was uttered in the dialogue, with or without the assistance of the researcher, they tried to locate the translation of the expression in the subtitles.

4.3. Interview results

In the experiment, the participants did not know that there were two modes of subtitling in the clip and, therefore, which of the two modes were combined. In the interviews, they were informed that one-half of the clip was subtitled by fansubbers and the other half was by MT. Then they were asked to identify which part had been subtitled by MT and which part by humans.

Of the sixteen participants, twelve (75%) were able to identify which half had been subtitled by MT. Two aspects of the subtitles helped them locate the MT half. First, they tried to recall when the translator's notes appeared in the clip. Since they were aware that MT systems can only generate what is in the original text, the part containing the translator's notes must have been subtitled by human translators, and thus, the other half would be by MT. Second, participants who picked up translation errors in the subtitles automatically classified the part with errors as MTS.

However, of the twelve participants, A05 stated that she did not have much knowledge of MT. Her correct selection of the MTS was a fluke. A09 chose the incorrect half. He believed MT tended to make mistakes in complex phrases, while fansubbers were more likely to make errors when dealing with grammatical and syntactic issues. Since he did not find grammatical errors in the first half of the clip, he thought the first half had been subtitled by MT. In fact, the first half he watched was subtitled by fansubbers and the MTS in this clip contained translation issues related to lexicon, grammar, and syntax. Three participants said "I don't know" when they were asked to locate the half with MTS. A02, B02, and B04 reported that they could not tell the difference between the subtitles in the two halves because the subtitles read well. The participants' answers to this question not only reflected their opinions on the subtitle quality but also manifested their perceptions of MT output in general. It seems that the quality of MT was underestimated by many. Conservative attitudes toward MT predominate in the two groups.

The last question for language learners was: Would you consider using MTS to learn Chinese? As the language learners had some knowledge of the Chinese language, MTS may have provided a reference, especially when the participants encountered unfamiliar phrases. However, most language learners voiced conservative views on MT. Only two learners of Chinese, A03 and B05, reported that they used MTS in daily viewing. A03 said she used Google Translate frequently. However, a problem she had when using MT was that her level of proficiency in Chinese enabled her to sense there was something wrong with the machine-translated text, but she did not know exactly *what* was wrong. In those cases, she needed to ask for help from someone with a higher level of proficiency in Chinese. Another language learner, B05, was a seasoned watcher of Chinese TV shows. She watched Chinese dramas on Netflix with Language Reactor. It is a language-learning Chrome extension that provides viewers with an advanced subtitle panel, supporting different subtitle conditions to meet language learners' needs. This enables her to include *pinyin*, intralingual subtitles, and side-by-side MT during viewing. She believed it helped improve her listening skills.

B04 and B06 said they would use MTS only when no human translation was available. A05, B01, and B02 stated they would read several machine-translated lines first to see if the subtitles were accurate and then decide whether to use them. B02 further added, "If you go on YouTube, for English [translation], they're pretty inaccurate. So, I wouldn't trust it very much." Her previous experience with watching YouTube videos with closed captions made her distrust MT in general.

Nine language learners (56%) stated that they refused to use MTS, mainly because MT is not always accurate. Of those nine, A02, A07, and A08 claimed that they watched Chinese TV shows for entertainment only. They did not view subtitles as learning materials, let alone MTS. Moreover, A04 and B07 were graduate students majoring in translation. Although they took courses related to MT and knew about different MT systems, they still held pessimistic views of MTS. They reported that using MT in translation practice is different from using it to learn a language. They still preferred reading non-translations in language learning.

Clearly, the language learners' perception of MT and of subtitles as a learning tool and their levels of proficiency in Chinese affected how they read and use subtitles. Most language learners had negative attitudes toward MT and thus would not use it to learn a language.

5. Discussion

5.1. Accuracy in subtitles: a key to vocabulary acquisition?

My initial hypothesis was that fansubs lead to better outcomes in vocabulary learning than MTS. However, the quantitative results show that there is no significant correlation between the modes of subtitles and improvement in the vocabulary test. In other words, the effect of MTS and fansubs are comparable and both have a positive effect on vocabulary learning. I will combine participants' test results with their verbal reports and screen-based activities to sketch out why the hypothesis is refuted.

First, the language learners often read subtitles selectively. The language learners' verbalisations indicate that they tended to skip subtitles when they thought they understood what had been said in the original soundtrack. For example, most participants did not pay attention to the translations of “哥哥” and “姐姐” because they already knew that the former means the elder brother and the latter is the elder sister. However, they did not realise that the two terms do not always refer to blood-related siblings, so they failed to receive extra points in the post-test even though they encountered the part in fansubs with the translator's note. This echoes Orrego-Carmona's (2015) and Malakul and Park's (2023) finding that viewers with higher proficiency in the language have more flexible reading processes, without fully relying on the subtitles, whereas those with lower proficiency tend to follow the subtitles more closely.

Second, other channels of information, such as the acoustic (soundtrack) and visual (images) channels, can make up for misinformation in the problematic subtitles. For example, when B06 saw “诊脉” (checking pulse) in the pre-test, he could gather that the term is related to a practice in Chinese medicine. When he encountered the scene where a doctor is checking the male lead's pulse, he shouted, “This is 诊脉!” before the subtitle line with the term appeared.

Third, although the fansubs contain fewer linguistic errors than MTS, fansubbers are still unable to provide a detailed explanation of culture-specific items within a few lines. Even when the participant's language proficiency in Chinese was good enough to gain a gist understanding of the clip without English subtitles, there were culturally embedded terms and idiomatic expressions in Chinese that they would rarely encounter in a textbook or daily conversations. For instance, the term “江湖” (*jianghu*) is translated as “the Pugilist world” in the fansubbed version. B06 commented on the translation in his verbalisation: “The pugilist world? That's an interesting translation for ‘江湖’”. Later, he figured out the contextual meaning of *jianghu* in the post-test, based on an educated guess. B07 also tried to construe the meaning of “the Pugilist world”, saying “‘江湖’ the pugilist world. Like a commoner? I must say the subtitles are not very helpful.” B07's interpretation of “江湖” was limited. In *Nirvana in Fire*, *jianghu* refers to the world outside of the imperial court: the people who live in *jianghu* are not bound by the imperial laws and rules. Although “the Pugilist world” fails to convey the meaning and

implication of the notion of *jianghu*, this translation is commonly accepted and used in fansubs and fanfictions. For newcomers like B06 and B07, it was difficult to construe even though they are advanced learners.

5.2. What affects the effectiveness of vocabulary learning?

Based on the results of the vocabulary tests, all learners of Chinese picked up at least one new vocabulary item. The clip extracted from *Nirvana in Fire* had a positive effect on vocabulary learning ($p < .001$). This is in line with previous research in the field: exposure to subtitled audiovisual materials correlates positively with foreign language acquisition, no matter what kind of subtitle is used (intralingual, interlingual or bilingual) (Caimi, 2006; Frumuselu *et al.*, 2015; Liao *et al.*, 2020; Baranowska, 2020). However, since the vocabulary test only included eight Chinese vocabulary items selected by the researcher, the phrases that the language learner acquired might not appear in the vocabulary test.

The participants' verbal reports and interview responses also provided clues about other elements that affect language learning. Firstly, different language learners' viewing habits can lead to different effects of subtitles on the degree of vocabulary learning. Of the sixteen language learners in the experiment, five (A03, A04, B02, B06, and B07) learned vocabulary items in the clip with *conscious* effort. They asked the researcher questions about the meanings of the vocabulary items. When they paused or went back, they re-listened to the original soundtrack, matching each phrase in the Chinese segment with the English subtitles. Then they were able to locate the translation of the unknown phrase. Alternatively, when they looked at the English subtitles, they would mentally translate them back into Chinese. If the Chinese soundtrack was not what they expected, they went back to listen again. This process helped them pick up the new expression in Chinese. For example, A04 reported, "When I saw the subtitle saying 'improvement', I thought the word for improvement would be '进步'. I was listening for that, and I didn't hear it, so I... [went back to listen again] Turns out it was '长进'." However, intentional learning did not necessarily indicate better learning outcomes. Not all five of these participants are among the top five in terms of test performance or the ones who have made the greatest progress. This could be related to the fact that they had different levels of proficiency in Chinese.

Apart from these observable engagements, A05 and A09 claimed that they intentionally focused on the audio channel, trying not to look at the translated subtitles. The language learners' positive perceptions of exposure to Chinese TV shows motivated them to use Chinese audiovisual products to learn the language. Their progress rates in the vocabulary tests were nevertheless relatively low, which may be attributed to their viewing pattern.

However, not every language learner regarded Chinese audiovisual materials as learning tools. A02, A07, and A08 said that they watched Chinese TV shows mainly for entertainment. Although these three participants generally watched Chinese shows for fun, they still actively looked for unfamiliar Chinese phrases and tried to learn them in this experiment. Their progress rates in the vocabulary test were 21%, 11%, and 19% respectively, not the lowest among all language learners. However, according to Hulstijn (2001), the announcement of a test (as when I told the participants that they were going to be tested) is a sign of intended vocabulary learning. In the testing environment, regardless of the language learner's level of interest in subtitles and language learning, their primary objective might be to achieve a good performance in the vocabulary post-test. Therefore, language learners may make a conscious effort to perform well in the tests, but this can disrupt the immersive experience and may not fully reflect language learning in the daily viewing experience.

5.3. Are language learners ready to embrace MTS?

Although most of the participants agreed that the MTS were acceptable, they still focused on the negative aspects. The accuracy of MTS was thus possibly underestimated. Moreover, MT tools are free and easily accessible, making it simple to use MT in a non-reflective way. If language learners are not fully aware of the evolving capacity of MT, they may not use MT in an informed and critical way.

Interestingly, the participants' sceptical views on MT stem from completely different levels of knowledge of it. A04, studying translation at the postgraduate level, did not trust MT because he knew how a MT system processes texts. Although he praised the MT of the phrase “可不是一般人能比的” as “it's not for the faint-hearted”, he still insisted that he trusted human translators more than MT. He did not think the algorithms for neural MT could always arrange those segments of human translation in a sensible order. Another language learner, A05, did not trust MT due to her limited knowledge of it. Since the participants were not aware of the presence of MTS during viewing, they used the same standard in mind to evaluate all subtitles. The two cases show that the participants expected the translated subtitles to be error-free and be able to fully convey the meanings in the original text. Therefore, for them, MT can be a risky undertaking, especially if unfamiliar vocabulary items or phrases are mistranslated. Since they presumed that MT is less accurate than human translation, they did not want to take the risk.

Further, the use of MT can also be associated with the user's level of proficiency in the language being learned as indicated in 4.3. Beginners may be easily misled by machine-translated texts since they do not have enough knowledge to identify whether the translation is acceptable. For intermediate learners, MT in the subtitles may serve as a reference to help them pick up unknown phrases, but they can be misled by mistranslations without realising it. In terms of advanced learners, non-translations seem more effective.

In order to equip language learners with the necessary skills to utilise MTS effectively and critically, it is imperative to cultivate their MT literacy. Williams (2006) posits that language teachers should introduce MT tools to their students as a means of promoting both language awareness and electronic literacy. Bowker and Ciro (2019) propose that language learners need to understand how MT systems process text, and when and how to use MT appropriately. Additionally, learners should know what types of texts can be easily translated by machines and possess the necessary skills for pre-editing and post-editing. By developing proficiency in these areas, language learners can be more flexible towards the use of automated subtitling in didactic scenarios.

6. Conclusion

This study has explored the didactic potential of raw MTS by comparing them with fansubs in terms of their effect on vocabulary learning and investigating language learners' perception of MT in subtitling. The results show that the combination of both types of subtitles had a positive effect on vocabulary learning, regardless of language learners' proficiency levels in the language being learned. Moreover, no significant difference was found between the effects of fansubs and MTS on vocabulary learning. This suggests that language learners can benefit from watching audiovisual content with either type of subtitles, and that the number of linguistic errors in subtitles is not a critical factor for vocabulary acquisition. Language learners who were motivated to learn the language through watching subtitled shows and who were interested in the audiovisual genre were more likely to initiate screen-based activities while watching the clip.

The study also reveals that language learners often underestimate the quality of MT and

associate erroneous subtitles with MT outputs. Only a few language learners reported using MT while watching Chinese audiovisual products. The participants' sceptical views on MT seemed to stem from their different levels of knowledge about it, highlighting the need for increased MT literacy among language learners.

Due to the small sample size and the limited number of vocabulary items in the language tests, my findings are preliminary and further studies with larger samples are warranted to validate the initial hypothesis on a larger scale. Another limitation of this study is that it did not include a later vocabulary post-test to determine whether the learned vocabulary items entered the learners' long-term memory. Therefore, the study cannot conclusively assert that watching audiovisual products is an effective method for stable vocabulary acquisition. Future studies could address this limitation and investigate the potential of using both types of subtitles in long-term vocabulary retention. The two vocabulary tests were initially designed to encourage language learners to pay more conscious attention to the original soundtrack and to the English subtitles, and even to provoke screen-based activities such as going back to check the meaning of a phrase or pausing to read certain phrases. Some participants might opt to initiate more screen-based activities than they actually needed to have better performance in the vocabulary test. Their screen-based activities could affect their scores on vocabulary post-tests.


As Baranowska (2020) proposes, audiovisual materials should be available in different subtitle conditions to meet diverse learning objectives. The materials may have added value in language learning and entertainment if language learners are able to access different kinds of subtitles and combine two or three conditions as they want.

7. References

- Baranowska, K. (2020). Learning most with least effort: Subtitles and cognitive load. *ELT Journal*, 74(2), 105–115. <https://doi.org/10.1093/elt/ccz060>
- Benson, P., & Chan, N. (2010). TESOL after YouTube: Fansubbing and informal language learning. *Taiwan Journal of TESOL*, 7(2), 1–23.
- Bolaños García-Escribano, A. (2017). The effects of fansubs on EFL education for translation and interpreting students: An empirical approach. *The Journal of Specialised Translation*, 28, 122–163. https://www.jostrans.org/issue28/art_bolanos.pdf
- Bowker, L., & Ciro, J. B. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing.
- Caffrey, C. (2012). Using an eye-tracking tool to measure the effects of experimental subtitling procedures. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 223–258). Aracne.
- Caimi, A. (2006). Audiovisual translation and language learning: The promotion of intralingual subtitles. *Journal of Specialised Translation*, 6, 85–98. https://www.jostrans.org/issue06/art_caimi.php
- Chan, W. S., Kruger, J. L., & Doherty, S. (2019). Comparing the impact of automatically generated and corrected subtitles on cognitive load and learning in a first- and second-language educational context. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 18, 237–272. <https://doi.org/10.52034/lanstts.v18i0.506>
- Díaz-Cintas, J., & Muñoz Sánchez, P. (2006). Fansubs: Audiovisual translation in an amateur environment. *The Journal of Specialised Translation*, 6, 37–52. https://www.jostrans.org/issue06/art_diaz_munoz.pdf
- Frumuselu, A. D., De Maeyer, S., Donche, V., & Gutiérrez Colon Plana, M. M. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, 32, 107–117. <https://doi.org/10.1016/j.linged.2015.10.001>
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (1st ed., pp. 258–286). Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.011>
- 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>
- Kong, S., & Li, X. (Directors). (2015). *Nirvana in Fire (Season One)* (No. 1). <https://www.viki.com/tv/22943c-nirvana-in-fire>

- Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49(1), 105–120. <https://doi.org/10.1002/rrq.59>
- Künzli, A., & Ehrensberger-Dow, M. (2011). Innovative subtitling: A reception study. In C. Alvstad, A. Hild & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 187–200). John Benjamins. <https://doi.org/10.1075/btl.94.14kun>
- Lakarnchua, O. (2017). Examining the potential of fansubbing as a language learning activity. *Innovation in Language Learning and Teaching*, 11(1), 32–44. <https://doi.org/10.1080/17501229.2015.1016030>
- Liao, S., Kruger, J. L., & Doherty, S. (2020). The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation*, 33, 70–98. https://www.jostrans.org/issue33/art_liao.pdf
- Liu, X., & De Seta, G. (2014). Chinese fansub groups as communities of practice: An ethnography of online language learning. In D. K. Herold & P. Marolt (Eds.), *China online* (pp. 125–140). Routledge.
- Malakul, S., & Park, I. (2023). The effects of using an auto-subtitle system in educational videos to facilitate learning for secondary school students: Learning comprehension, cognitive load, and satisfaction. *Smart Learning Environments*, 10(1), 4–21. <https://doi.org/10.1186/s40561-023-00224-2>
- Neves, J. (2004). Language awareness through training in subtitling. In P. Orero (Ed.), *Topics in audiovisual translation* (pp. 127–140). John Benjamins. <https://doi.org/10.1075/btl.56.14nev>
- Orrego-Carmona, D. (2015). *The reception of (non)professional subtitling*. [Doctoral dissertation, Universitat Rovira i Virgili]. TDX Xarxa. <https://www.tdx.cat/handle/10803/306439#page=1>
- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148. <https://doi.org/10.1016/j.system.2016.10.002>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
- Tadmor-Ramanovic, M., & Bar, N. (2019, October 29). *On-device captioning with live caption*. <https://ai.googleblog.com/2019/10/on-device-captioning-with-live-caption.html>
- Wang, D. (2017). Fansubbing in China: With reference to the fansubbing group YYeTs. *Journal of Specialised Translation*, 28, 165–190. https://www.jostrans.org/issue28/art_wang.pdf
- Williams, L. (2006). Web-based machine translation as a tool for promoting electronic literacy and language awareness. *Foreign Language Annals*, 39(4), 565–578. <https://doi.org/10.1111/j.1944-9720.2006.tb02276.x>
- Zhan, J. (2022, May 31). *Racing against rivals and robots on the translation assembly line*. vulture. <https://www.vulture.com/article/rakuten-viki-volunteer-stories.html>
- Zhao T., & Macaro E. (2016). What works better for the learning of concrete and abstract words: Teachers' L1 use or L2-only explanations? *International Journal of Applied Linguistics*, 26(1), 75–98. <https://doi.org/10.1111/ijal.12080>
-



 Juerong Qiu

University of Melbourne
Sidney Myer Asia Centre, Building 158
3052 Melbourne
Australia

juerong.qiu@unimelb.edu.au

Biography: 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 specialized translation at the University of Melbourne.



This work is licensed under a Creative Commons Attribution 4.0 International License.