The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*

Haina Jin
Communication University of China
jinhaina@cuc.edu.cn
https://orcid.org/0000-0003-1848-7489

Zichen Yuan
Communication University of China
zichen.yuan@outlook.com
http://orcid.org/0000-0003-2395-3237

Abstract

The use of artificial intelligence (AI) for audiovisual dubbing has become increasingly popular due to its ability to improve content production and dissemination. In particular, the application of machine translation (MT) to audiovisual content has resulted in more efficient and productive AI content generation. To assess the quality of MT-dubbed videos, this study proposed the use of the new FAS model. This model adapts the FAR model put forward by Pedersen (2017), with the "R (Readability)" parameter replaced with "S" to include synchrony: this amendment responds to research on dubbing quality that identifies the need to explore better methods for synchronization between audio and video. Using *Mulan* (Caro, 2020) – an English-language Disney feature film released in 2020 – as a case study, this article evaluates the quality of automatically dubbed videos generated by YSJ (Ren Ren Yi Shi Jie), an MT platform for audiovisual products in China. By analysing errors in functional equivalence, acceptability, and synchrony, the study assesses whether China's latest MT engine can meet the demand for quality dubbing and improve cross-cultural communication. The findings show that whereas China's present MT platform can generate a moderately acceptable result, there are still semantic errors, idiomaticity problems, and synchrony errors which may lead to incorrect translations and consequently possible misunderstanding by viewers. Overall, this study sheds light on the current state of AI-dubbing technology in China and highlights areas for improvement.

Keywords: automatic dubbing, FAS model, *Mulan*, audiovisual translation, quality assessment

1. Introduction

In the era of artificial intelligence (AI), machine translation (MT) technology has emerged as a powerful tool for content production and dissemination, saving time for translators and improving efficiency. However, in the context of audiovisual translation (AVT), MT has faced challenges owing to the complexity of audiovisual texts, which involve not only linguistic content but also visual and acoustic elements. While MT has been extensively studied in textual contexts, its application in the audiovisual domain is still in its early stages.

Research on AVT and MT has focused predominantly on subtitle translation, with relatively limited attention having been given to dubbing translation. Internationally, numerous studies have investigated the challenges and advantages of applying MT in AVT. Burchardt et al. (2016) explored the difficulties of using MT in AVT, whereas Bywood (2020) provided an overview of the relationship between technology and various aspects of AVT, including subtitling, dubbing, and workflow software. Matamala and Ortiz-Boix (2016) discussed the feasibility of translating audio descriptions and the application of MT in this mode of AVT. Bellés-Calvera and Quintana (2021) examined the quality of English subtitles in the Spanish Netflix series *Cable Girls*, translated by Google Translate and DeepL.

In China, the application of MT in AVT is still relatively limited, with very little research having been conducted on this topic, mainly in the context of subtitle translation. In one study, Xiao and Gao (2020) assessed the quality of MT subtitles and human-translated subtitles using the FAR model, focusing on colloquial dialogues in a TED talk.

One notable model proposed for quality assessment in interlingual subtitling is the functional equivalence, acceptability, and readability (FAR) model developed by Pedersen (2017). This model evaluates translation quality based on three parameters: functional equivalence, acceptability, and readability. To be specific, functional equivalence errors include semantic errors and stylistic errors. Acceptability errors consist of grammar, spelling, and idiomatic errors. Three types of readability error are considered: segmentation and spotting, punctuation and graphics, and the impact of line length on reading speed. The FAR model allows for a detailed analysis of errors and provides constructive feedback to subtitlers, contributing in this way to the improvement of subtitle translations. Errors are classified as "minor", "standard", or "serious", with proposed equivalent scores of 0.25, 0.5, and 1, respectively. Minor errors are of the type that may go unnoticed unless the viewers are very attentive. Standard errors are those that may break the suspension of disbelief and have an impact on the viewing experience of the subtitles for most viewers. Serious errors refer to mistakes that significantly affect the accuracy, clarity, or naturalness of a translation, and may have a negative impact on the overall quality of a translation. The model calculates not only a total score but also the score of each parameter individually – which is useful for identifying

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

the individual shortcomings of MT subtitles. The model is based on error analysis: researchers analyse errors in the subtitles and make corresponding deductions to calculate the final score, making it identify where the subtitles contain problems. Such a model may therefore be used to provide constructive feedback to subtitlers. FAR provides a more objective assessment standard compared to the DQF (Dynamic Quality Framework), MQM (Multidimensional Quality Metrics), LISA (Localization Industry Standards Association), and TEP (Translation Editing Proofreading) models – which all involve a certain degree of subjective assessment because they rely upon human evaluators to make judgments about the quality of a translation. This objectivity is possible because the FAR model is able to provide criteria and penalties that minimize the evaluator's bias toward machine-translated subtitles. However, the model is not entirely immune to bias, as evaluators may still be aware that they are assessing MT subtitles. Even so, by using clear and specific criteria, the model may help to reduce subjectivity and provide a more objective evaluation of both human- and machine-translated subtitles.

To redress the lack of a quality assessment model for pre-prepared interlingual dubbing, this study used the FAS model. Inspired by the FAR model, and factoring in the importance of synchrony in dubbing, the FAS model introduces an additional criterion with which to assess the synchronization of dubbed content. The R (readability) criterion of the FAR model was eliminated because it is considered less relevant in the context of dubbing. Readability is an important consideration when evaluating subtitles for traditional subtitling, where the text appears on screen and must be read quickly and easily by viewers. However, in dubbing, the subtitles are typically spoken by dubbing actors and may not appear on screen at all; for this reason, readability may be of less concern. Therefore, synchrony becomes a key factor to consider in dubbing, as it is essential to creating a seamless viewing experience for the audience. Synchrony errors are based on the categorization proposed by Chaume (2004): phonetic or lip synchrony, kinesic or kinetic synchrony, and isochrony:

- Phonetic or lip synchrony refers to the accurate timing of the subtitles to align them with the movements of an actor's lips and mouth. In dubbing, it is important to ensure that the utterances are synchronized with the lip movements of the actors in order to create more natural and believable dubbing.

- Kinesic or kinetic synchrony refers to the synchronization of the utterances with the body movements and gestures of the actors on screen. This type of synchrony is especially important in conveying the emotional and physical nuances of the original dialogue.

- Isochrony refers to the synchronization of utterances with the rhythm and pace of the original dialogue. In dubbing, it is important to maintain the natural speech rhythm and pace in order to preserve its emotional impact and meaning.

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

By assessing synchrony, the FAS model aims to provide a comprehensive evaluation of automatically generated dubbed videos. In the Chinese context, the focus of this research is on assessing the quality of videos that have been automatically dubbed by the MT platform YSJ (人人译视界 https://www.1sj.tv). YSJ is a leading Chinese company that specializes in AVT services and technology. With its advanced AI-driven platform, YSJ offers one-stop AVT solutions, including subtitling, dubbing, and video content. By examining the raw output of the MT engine employed by YSJ, this study intended to evaluate the quality of the automatically generated dubbed videos. Recently, YSJ has developed a new function – automatic dubbing. Based on powerful and intelligent voice technology, automatic dubbing generates the sound of the voice based on subtitle files which are designed by the company to sync accurately with the utterances and in this way to ensure efficient synthesis and output. Following automatic subtitling through the system NeteaseSight, users can select and modify various parameters, such as the speakers' accent, style, speed, volume, and tone of voice, through the YSJ interface. This allows for natural and realistic multi-character dubbing that fits the videos. At the same time, by intelligently eliminating overlaps, this process ensures that sounds and pictures are well synchronized. Furthermore, by merging lip-syncing and reverberation, YSJ provides a one-stop service for dubbing optimization. On the YSJ platform, lip-synching refers to the process of re-synthesizing a video to match the translated text rather than adapting the translation to the lip movements in the original video. This means that the platform uses advanced technology to match the timing of the translated audio to the movement of the characters' lips on screen, resulting in a more natural and realistic viewing experience for the audience. To resolve the issue of time overlaps after MT, YSJ proposes a series of sequential steps that automate the removal of overlaps: panning subtitles to eliminate any collision with the video content without shortening them, adjusting the speech speed, and slowing down the video clip to extend the subtitle timeline. Backed by MT and automatic dubbing on YSJ, this study investigated directly the raw output of the MT engine.

2. Quality assessment of automatic dubbing

The study analysed the errors in the automatic dubbing of an American film into Chinese through the FAS model according to the criteria of functional equivalence, acceptability, and readability, as well as synchronies, without post-editing.

The audiovisual sample for the study was *Mulan* (Caro, 2020), an English-language Disney feature film, which was officially distributed in both the English and the Chinese language and has gained popularity among viewers unfamiliar with *Mulan*. This study evaluates the quality of the machine-translated subtitles and the dubbed audio together with the quality of the official theatrical dubbed Chinese version, which also includes the official subtitles. The official theatrical version was released in theatres, where the subtitles and dubbed audio contained the same text – here, the Chinese subtitles were adapted directly from the Chinese dubbing

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

text. To enable a comparison between the quality of the machine and that of the human translations, the subtitles were used as the reference for human translation because they reflect the same text as the dubbed audio. As *Mulan* portrays a specific cultural background, this study was able to both evaluate the effectiveness of MT in automatic dubbing and examine the ways in which MT deals with historical and cultural nuances in language. The goal was to assess the strengths and limitations of current MT technology, even in the face of such challenges.

Table 1 presents a summary of the types and number of errors according to several parameters. The study aimed to assess whether, in line with the three criteria of the FAS model, the quality of the raw dubbing output generated by the latest MT engine met the demands of viewers. As seen in Table 1, based on error analysis typology, the study analysed the errors in MT subtitles and dubbing, then calculated the score in each category. As the data demonstrate, the most prominent errors are clearly semantic, idiomaticity, and synchrony or isochrony, with each scoring 36, 11.75, and 9, respectively. Drawing from the examples of specific errors illustrated in the following sections, several suggestions are made as to how the performance of the MT platform may be improved and light is shed on the post-editing of machine-translated audiovisual text.

Table 1

Types and number of errors

	Error type	Minor	Standard	Serious	Total	Overall score
Functional equivalence	Semantic	30	21	18	69	36
	Stylistic	6	3	2	11	5
Acceptability	Idiomaticity	1	15	4	20	11.75
Synchrony	Phonetic or lip-synchrony	0	0	7	7	7
	Kinesic/kinetic synchrony	11	2	2	15	5.75
	Isochrony or synchrony between utterances and pauses	2	1	8	11	9

2.1 Functional equivalence

According to Pedersen (2017, p. 210), functional equivalence errors include two categories of error: semantic and stylistic:

- Serious semantic errors may affect contextual coherence and cause misunderstanding.

- Standard semantic errors relate to the absence of information; however, they do not affect the viewers' understanding of the remainder of a subtitle.

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

- Minor semantic errors relate to general problems with wording but do not affect the viewers' understanding of the rest of the subtitles.

2.1.1 Semantic errors

Table 2

Example 1

Time	Source text (ST)	Target text (TT)	Official theatrical version
00:00:41,500 --> 00:00:45,150	There have been many tales of the great warrior, Mulan.	关于伟大的战士木兰的故事有很多 (There have been many tales of the great warrior, Mulan.)	有关花木兰的故事有很多 (There have been many tales of Hua Mulan.)
00:00:47,750 --> 00:00:50,250	But, ancestors, this one is mine.	但是 祖先们 这一个是我的 (But, ancestors, this one is mine.)	今天 我想说 (Today, I want to say)
00:00:52,700 --> 00:00:53,900	Here she is.	她来了 (She's coming.)	小时候 (When she is young,)
00:00:54,600 --> 00:00:57,200	A young shoot, all green …	一颗年轻的嫩芽 满身的绿色 … (A young shoot full of green …)	她像麦苗一样 青涩稚嫩 (Mulan was as immature as young shoot)
00:00:57,650 --> 00:00:59,100	unaware of the blade.	不知道这把刀 (unaware of the blade.)	不知道镰刀的锋利 (unaware of the sharp blade.)

As a prelude, the movie introduces Mulan, the film's protagonist, and briefly presents Mulan's character to pave the way for the storytelling. In the second unit of subtitles, the machine translation of "但是 祖先们 这一个是我的(But, ancestors, this one is mine.)" has some wording problems. The use of "但是(but)" in the translation may not accurately convey the intended meaning of contrast or opposition in the source text (ST). In addition, the phrase "祖先们 (ancestors)" may not fully capture the meaning of "ancestors" as it is commonly understood. Yet this does not affect the viewers' understanding. This is therefore a minor error.

The last three units of the subtitles describe Mulan; because the English subtitles often omit the subject, the subtitles may describe only the metaphor, and the MT may overlook the intended vocal expressions. The last sentence mentions that Mulan is young and naive, indicating that she lacks awareness of worldly affairs. The translation is erroneous because it does not consider the original context in which to convey the meaning of the original text accurately or to describe the text effectively for rhetorical purposes. Because of the likelihood of viewers' misunderstanding and some contextual incoherence, the last three units of the subtitle should be classified as serious semantic errors.

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

Table 3

Example 2

Time	Source text (ST)	Target text (TT)	Official theatrical version
00:05:05,980 --> 00:05:07,010	Do you know why the phoenix	你知道为什么凤凰 (Do you know why the phoenix)	你知道为什么 (Do you know why)
00:05:07,010 --> 00:05:09,410	sits at the entrance of our shrine?	坐在我们神社的门口吗 (sits at the entrance of our shrine?)	我们的祠堂门口有只凤凰吗 (a phoenix sits at the entrance of our ancestral hall?)
00:05:11,650 --> 00:05:14,780	She is the emissary for our ancestors.	她是我们祖先的使者 (She is the emissary for our ancestors.)	凤凰是我们花家的吉祥瑞兽 (Phoenix is the mascot of our Hua family and ancestors.)

In example 2, the MT is too literal, which is not in line with Chinese cultural uses of language. This may confuse viewers because the word "shrine" in the original context signifies a place where people come to worship, namely, a deity or a religious event. The official theatrical version translated shrine as "祠堂 (ci tang)" – meaning "ancestral hall" – which is a place where Chinese people worship their ancestors; this is more in line with Chinese culture. This translation choice is better localized and aligns with Chinese cultural norms. In addition, in the last unit of the subtitle, the term "emissary" is literally translated, which could weaken its association with the phoenix depicted in the film and leave viewers questioning the connection between the two. In the film, the phoenix is the mascot of the Hua family and their ancestors, which will bring them fortune, safety, and strength.

Table 4

Example 3

Time	Source text (ST)	Target text (TT)	Official theatrical version
00:11:05,050 --> 00:11:06,410	We'll protect our beloved people ...	我们将保护我们亲爱的人民... (We'll protect our beloved people ...)	保边境安宁 百姓平安 (We'll keep the border safe for all beloved people ...)
00:11:16,980 --> 00:11:18,820	and crush these murderers.	并粉碎这些杀人犯 (and crush these murderers.)	击溃北方的来犯者 (and crush these murderers from the north.)
00:11:20,810 --> 00:11:23,100	Deploy the Imperial Army.	部署帝国军队 (Deploy the Imperial Army.)	大军即日向北进发 (The Imperial Army will march northward today.)
00:11:24,540 --> 00:11:27,450	The dynasty will not be threatened.	王朝将不会受到威胁 (The dynasty will not be threatened.)	蛮夷休想侵占我一寸国土 (The dynasty will not be threatened by Rourans.)

In the film, the emperor learns that the Rourans are invading from the north and is readying troops to garrison the border and defend the country. The original text is more concise, mostly comprising short sentences, reflecting the emperor's anger and the confidence of his superiors; this is more in line with the cultural context of ancient China. However, the literal translation is not coherent enough and so for Chinese viewers it is somewhat unclear – Where is the army to be dispatched, and for what reason? Who is threatening the kingdom? Therefore, the official theatrical version uses an additional translation to complete the information in the film, which is also an embodiment of the Northern Wei dynasty as a great power. The absence of information in the MT subtitles consequently results in standard semantic errors.

2.1.2 Stylistic errors

Table 5

Example 4

Time	Source text (ST)	Target text (TT)	Official theatrical version
00:09:50,900 --> 00:09:52,430	If we allow this to continue,	如果我们允许这种情况继续下去 (If we allow this to continue,)	倘若放任其不管 (If we allow this to continue,)
00:09:52,630 --> 00:09:54,580	it could be the end of the kingdom.	这可能是王国的末日 (it could be the end of the kingdom.)	臣恐国将危矣 (I'm afraid that it could be the end of the kingdom.)
00:09:55,140 --> 00:09:56,360	And my citizens?	我的公民呢 (And my citizens?)	百姓如何 (How about the common people?)

Throughout the video, there are many stylistic errors regarding terms of address, as example 4 demonstrates. In this set of subtitles, the word "citizen" appears in the film for the first time. In this scene, the witch, disguised as a prime minister, reports to the emperor that the Rourans are attacking, and the emperor asks, "And my citizens?" The Chinese subtitles do not directly translate the word "citizens": whereas in the original movie subtitles the word "citizens" was used, this was translated into the term "百姓(common people)" for the official Chinese version. This translation choice was made to ensure that the dialogue aligned with the cultural context of the time, which did not have a comparable concept for the term "citizens".

In addition, in the scene where the emperor sends people to recruit soldiers and the recruiting officer calls out the word "citizens", the official theatrical version opted to translate this term as "common people" in order to bridge the cultural gap and to provide dialogue that was easier for Chinese viewers to understand. Similarly, in the film, stylistic errors related to terms of address are also present. For example, in the English ST, the little sister Xiu calls Mulan by her name; however, in the target Chinese text, this term should become the word "sister" rather than the name "Mulan", which is in line with the traditional conventions of address in

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

ancient China. Ultimately, without possessing the ability to analyse the cultural context, the MT engine fails to generate an appropriate term of address.

Similarly, in the film, there is a scene where Mulan's sister, Xiu, calls her by name in English. In the Chinese translation, however, Xiu addresses Mulan as "*jie jie* (姐姐)", which means "older sister" or "sister" in English. The difficulty for MT in this case lies in the cultural significance of familial titles in the Chinese language and culture. In the Chinese language, it is common to use familial titles such as "*jie jie* (姐姐)" or "*di di* (弟弟 younger brother) " instead of names when addressing siblings or other family members. This convention contrasts with English language conventions, where it is more common to address someone by their name. When translating this scene from English to Chinese, an MT system may not appreciate the cultural significance of familial titles in Chinese and may simply translate the English name "Mulan" into the Chinese characters "木兰".

In addition, there is a scene where Mulan returns home and is reunited with her family. In the original English version, Mulan addresses her mother as "mother" and her father as "father". However, in the official Chinese version, Mulan addresses her mother as "*niang* (娘)" and her father as "*die* (爹)", both of which are common familial titles in Chinese culture. This would result in a translation that is technically accurate yet culturally inappropriate.

The automatic dubbed output is prone to producing a number of errors in functional equivalence thanks to the following reasoning: without added notes for translation in English subtitles and an understanding of cultural context, the MT TT never considers the ST without any accompanying aural and visual elements, therefore missing the information in the ST. In addition, as demonstrated in examples 2 and 4, not understanding scenes and with translations possibly having multiple meanings, the MT engine is unable to translate the original word into the expected meaning, let alone translate culturally loaded words which play an important role in achieving functional equivalence. Moreover, as for stylistic errors, it is difficult for the MT engine to generate correct names and terms of address. In example 2, the term "citizens" should not be literally translated into its contemporary meaning while ignoring the cultural background of *Mulan*.

To avoid these errors, it is vital to set up a corpus containing culturally loaded words, terms of address, and names of characters, places, titles, etc., before engaging in automatic dubbing.

2.2 Acceptability

The automatically dubbed subtitles performed relatively better on acceptability than on other parameters, with no grammatical or spelling errors being included. This indicates that the parallel corpus used by the platform is strictly screened and of a high quality. At the same time, compared to English and some other languages, Chinese has looser grammatical rules; accordingly, there are fewer grammatical errors. In addition, Chinese characters have a square

character structure and do not possess the same potential to commit spelling errors as in the English language; this contributes to a higher level of acceptability of the quality. For this reason, the analysis in this part focuses on idiomatic errors.

Table 6

Example 5

Time	Source text (ST)	Target text (TT)	Official theatrical version
00:14:09,270 --> 00:14:10,850	We have excellent news.	我们有极好的消息 (We have excellent news.)	告诉你一件大喜事 (We want to tell you a big happy event.)
00:14:11,470 --> 00:14:14,100	The matchmaker has <u>found you an auspicious match</u>.	媒人已经为你找到了一个吉利的对象 (The matchmaker has found you an auspicious man.)	媒人跟你说了一门亲事我很满意 (I'm satisfied about the marriage arranged by the matchmaker.)
00:14:16,270 --> 00:14:19,140	Yes, Mulan, it is decided.	是的 木兰 已经决定了 (Yes, Mulan, it is decided.)	是啊 木兰 已经定下来了 (Yes, Mulan, it is decided.)

In example 5, the sentence "We have excellent news" should be understood as "We have excellent news to tell you". In addition, the following subtitles indicate that the concept of "excellent news" actually refers to an engagement, which is "a big happy event". In the Chinese language, this sentence should be rephrased in a way that brings to the fore the intended message.

As shown in example 5, the underlined subtitle is translated into "吉利的对象 (auspicious match)", which is a literal translation. The officially published version is commonly used in ancient Chinese culture, with the collocation of "说了一门亲事 (arrange a marriage)". Most of the translations of idioms in this film are related to culturally loaded words. The inappropriate translation of such words in subtitles may hinder cultural exchange, resulting in TV dramas and films that are popular in China being less popular when exported overseas. As a result, the translation of culturally loaded words should not only convert words between the English and Chinese languages but also convey cultural connotations and reflect distinctive national characters. Owing to the differences between the two cultures, the MT engine must choose an appropriate mode of expression to convey the essence of Chinese culture faithfully. However, the limitations of the MT engine become apparent when it fails to recognize the cultural context of the original text or provide accurate translations of words found in films depicting ancient Chinese life.

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

Table 7

Example 6

Time	Source text (ST)	Target text (TT)	Official theatrical version
00:01:01,380 --> 00:01:03,010	If you had such a daughter …	如果你有这样一个女儿 … (If you had such a daughter …)	如果你有这样的女儿 (If you had such a daughter …)
00:01:03,850 --> 00:01:07,430	her chi, the boundless energy of life itself …	她的气 生命本身的无限能量 … (her chi, the boundless energy of life itself …)	她天生精力充沛 充满着能量 (She was born with boundless energy …)
00:01:07,740 --> 00:01:10,030	speaking through her every motion …	通过她的每一个动作说话 … (speaking through her every motion …)	浑身散发着生命的活力 (and her body exudes the vitality of life …)
00:01:10,870 --> 00:01:15,810	could you tell her that only a son could wield *chi*?	你能告诉她 只有儿子才能掌握气 (you could tell her that only a son could wield *chi*?)	你会告诉她 只有男人才可以展示自己的力量 (you could tell her that only a male could wield *chi*?)

For example 6, it is important to introduce the concept of *chi*, which, in the film, refers to the qualification in martial arts that Mulan has held since she was a child. Without further explanation, this meaning may be very confusing to viewers. In Chinese, *chi* (气) refers to the ancient philosophical concept or relates to Chinese medicine and refers to a human being's internal dynamics that enable various organs to function. In Chinese works of martial arts, such an understanding of *chi* is widespread. The word *chi* could be literally translated into gas, air, steam, etc.; however, in this specific context, there are no corresponding English words to describe Mulan's talent in martial arts. To describe her talent as *chi* is a good fit, filling the gap created by the absence of the concept in English and aligning with the transformation of cultural dimensions. Whereas the MT engine translates the word *chi* literally, the officially published subtitles specifically translate the concept as "boundless energy or power". For those with an awareness of the concept of *chi* as used in the stated historical background, understanding may not be a problem; but for those without such background knowledge, the official translation may bridge most of the viewers' knowledge gaps.

Regarding errors of acceptability, the automatic dubbing function does not create errors of grammar and spelling, while idiomatic errors total 20. When examining errors of functional equivalence, idiomatic errors can be attributed to the inherent difficulty of translating culturally loaded words, as exemplified by the difference in the collocations in example 6. From the perspective of acceptability, one way to improve translation quality is to set up a corpus containing slang, idioms, common expressions, collocations, and so on. Another method may be to pre-edit the ST so as to rephrase the original message, thus allowing the MT engine to process the information more easily. However, this is a time-consuming process, and may conflict with motivations to save time and rely on the machine for efficiency.

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

2.3 Synchrony

In the case of the previous three parameters proposed by Pedersen (2017, p. 217), the quality assessment of subtitles was conducted before their automatic dubbing by the YSJ MT engine. Based on powerful MT technology, intelligent voice technology, and the AI algorithm, the automatic dubbing function of YSJ supports the option of voice-overs in multiple languages and tones, ensures that audio and vision are synchronized intelligently, eliminating overlap, and optimizes the raw results further by lip-synching and mixing with reverberation.

After automatic dubbing by YSJ, 38 automatically generated subtitles were found to overlap in time, which indicates errors in isochrony. To fix the time overlap, YSJ suggests three sequential processes to eliminate the overlap automatically: pan the subtitles,[1] adjust the speed of speech, and slow down the video clip to extend the timeline for subtitles. When the gap between subtitles is less than or equal to two seconds, it will move the subtitles. When the dubbing time is longer than one second and less than or equal to 1.2 seconds, the speech speed is adjusted. When the dubbing time is longer than 1.2 seconds or less than or equal to 1.5 seconds, the video display is extended, which means that the corresponding video clip is slowed down. The method of slowing down the video achieves synchronization between the machine-translated subtitles and the dubbed audio. However, this approach may also have negative impacts on the viewers' perception of the quality of the dubbing.

Research has shown (González Martínez, 2019, p. 6) that viewers prefer natural and fluent speech and that any noticeable discrepancies in timing or pacing may cause a distraction, negatively affecting their viewing experience. Following the acceptance of the suggestions to pan the lines, a total of 29 time overlaps remain, of which 21 may be adjusted for the speed of speech. Accordingly, eight time overlaps remain following the adjustment of speech speed and one overlap problem may be solved by slowing down the video clip. Because the purpose of this study was to test the ability of YSJ's automatic dubbing, all the revisions were accepted; finally, seven time overlaps remained that could not be resolved. These final seven overlaps could be classified as errors in isochrony, because the overlapping timing results in a mismatch between spoken lines, pauses, and the duration of the original actors' dialogues in comparison to the automatic dubbing.

---

[1]   In this context, the term "pan" refers to the horizontal movement of the subtitles across the screen. Panning may help to eliminate any overlap between utterances and the video and does not involve shortening the subtitles in any way.

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

Table 8

Example 7

Time in ST	Time in TT	Source text (ST)	Target text (TT)	Official theatrical version
00:13:52,093 --> 00:13:53,543	00:13:52,093 --> 00:13:53,543	Black Wind and I rode alongside	黑风和我骑在 (Black Wind and I rode on …)	刚才我骑着黑风 (Right now I rode Black Wind)
00:13:53,543 --> 00:13:55,303	00:13:53,543 --> 00:13:57,250	two rabbits running side by side	两只兔子 旁边 并排跑着 (two rabbits running side by side)	看见两只兔子在草丛里并排跑着 (I saw two rabbits running side by side in the grass.)
00:13:55,313 --> 00:13:57,943	00:13:55,313 --> 00:13:58,240	I think one was a male, one was a female.	我想一只是公的 一只是母的(I think one was a male, one was a female.)	我猜一只是公的 一只是母的(I think one was a male, one was a female.)
00:13:58,183 --> 00:13:59,743	00:13:58,183 --> 00:14:01,500	But you know, you can't really tell.	但你知道 你无法真正分辨出 (But you know, you can't really tell.)	不过其实我也分不清楚 (But actually I'm not clear.)

Following example 7, the MT results reveal that although the revisions generated by YSJ have been accepted, several dubbed audio clips were seen to overlap with one another and numerous dialogues were still audible while no lip movement was observed on screen. These resulted in serious synchrony errors. As seen in the example, the durations of the TT in the last three units of the subtitles do not match the time in the ST. To ensure the isochrony of automatically dubbed film, it is necessary to post-edit the raw output, that is, to speed up the speech of audio generated by the MT engine and adjust the TT by reducing the number of words.

In the most basic sense, the seven errors of synchrony are the result of the TT's long line length. As with the functions YSJ has recently introduced, there are three ways to ensure synchrony: panning the subtitles, adjusting the speed of speech, and slowing down the video clip to extend the timeline for subtitles. These measures could eliminate several errors to a certain extent, while their identification and adjustment still require refinement.

3. Conclusion

This study critically evaluated the application of automatic dubbing in China and pointed to gaps in translation quality. The author proposed the FAS model, based on the FAR model, to assess the quality of automatic dubbing and compared MT subtitles with human-translated subtitles, specifically official theatrical subtitles which are the same as the dubbing texts, to provide suggestions for the future improvement of MT engines. The findings of this study reveal semantic errors in functional equivalence and problems with idiomaticity and

synchrony that may be attributed to *Mulan*'s cultural context and the limited ability of current MT engines to factor in the context of original content. Although errors exist in automatically dubbed products, MT engines such as YSJ have introduced various functions to bridge gaps and increase efficiency. This study suggests that even when MT engines fail to redress certain errors, they still point to multiple ways of generating high-quality translations through post-editing. The study recommends the improvement of MT platforms and the viewing experience. These include creating a corpus containing culturally loaded words, character names, locations, titles, slang words, idioms, and collocations prior to MT. To aid the MT engine's processing capabilities, adjusting the ST is also crucial. Post-editing together with human fine-tuning is essential to ensuring the smooth flow and coherence of subtitles, as the MT engine cannot deal fully with the suspension of disbelief in films. Furthermore, the study suggests that human assessments should be conducted using the FAS model together with the display of both MT subtitles and official theatrical subtitles. On-site observation and questionnaires may be used to investigate viewers' experiences. In addition, a breakdown of viewers based on age, gender, and educational background may provide a more objective reference for assessing the quality of subtitles. Exploring the translation quality from different perspectives and analysing genres beyond feature films are areas worthy of future investigation.

Furthermore, the study has identified several drawbacks inherent in the FAS model. First, the model deducts points based on errors, whereas the MT does not award extra points for well-executed translations. Secondly, AVT includes multimodal content: consequently, automatic dubbing does not intelligently and completely identify all the information on the screen, which results in an absence of information. Furthermore, regarding synchrony, various aspects such as accent, tone, emotion, and speech speed remain major problems, with post-translation human editing remaining a necessary step. The YSJ platform provides practical solutions for character tones, tunes suitable for a multitude of particular scenes, speech speed, etc. At the same time, there are polyphonic characters in Chinese, that is, characters which are capable of multiple pronunciations or readings. For instance, these are characters that may be pronounced differently depending on the meaning or the context in which they appear. For example, the character "行 (xing)" may have several meanings and pronunciations – such as "to walk", "to go", "behaviour", and "a row" – depending on the context in which it is used. YSJ conveniently allows translators to preselect the specific polyphony of Chinese characters.

Although many errors may be found in automatically dubbed products, MT engines such as YSJ have introduced multiple technologies to bridge the gaps and increase efficiency. Even while the MT engine is unable to process certain errors, it may suggest several methods for generating high-quality translation through the process of post-editing.

In conclusion, this article has highlighted the challenges and possibilities facing automatic dubbing in China, emphasizing the importance of MT engines being improved continuously. It has also considered various factors that are able to enhance the translation quality of

Jin, H., & Yuan, Z. (2023). The application of machine translation in automatic dubbing in China: A case study of the feature film *Mulan*. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *22*, 80–94.

synchrony. However, the present study remains narrow in focus and pertains only to feature films. A greater number of film genres are certainly worthy of being subjected to similar analysis.

References

Bellés-Calvera, L., & Quintana, R. C. (2021). Audiovisual translation through NMT and subtitling in the Netflix series 'Cable Girls'. *Proceedings of the Translation and Interpreting Technology Online Conference, Online,* 142–148. https://doi.org/10.26615/978-954-452-071-7_015

Burchardt, A., Lommel, A., Bywood, L., Harris, K., & Popović, M. (2016). Machine translation quality in an audiovisual context. *Target*, *28*(2), 206–221. https://doi.org/10.1075/target.28.2.03bur

Bywood, L. (2020). Technology and audiovisual translation. In Ł. Bogucki & Deckert, M. (Eds.), *The Palgrave handbook of audiovisual translation and media accessibility* (pp. 503–517). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-42105-2_25

Caro, N. (Director). (2020). *Mulan* [Film]. Walt Disney Pictures.

Chaume, F. (2004). Synchronization in dubbing: A translational approach. In P. Orero (Ed.), *Topics in audiovisual translation* (pp. 35–52). John Benjamins. https://doi.org/10.1075/btl.56.07cha

González Martínez, R. (2019). *Audiovisual translation: A contrastive analysis of The Lord of the Rings: The Two Towers* [Bachelor's thesis, Universidad de Valladolid]. UvA Campus Repository. https://uvadoc.uva.es/handle/10324/39485

Matamala, A. (2016). Terminological challenges in the translation of science documentaries: A case-study. *Across Languages and Cultures, 11*(2), 255–272. https://doi.org/10.1556/Acr.11.2010.2.7

Matamala, A., & Ortiz-Boix, C. (2016). Accessibility and multilingualism: An exploratory study on the machine translation of audio descriptions. https://doi.org/10.24310/TRANS.2016.v0i20.2059

Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *The Journal of Specialised Translation*, *28*, 210–229.

Xiao, W., & Gao, J. (2020). 机器翻译字幕质量评估研究——以"网易见外"英译中字幕为例 [Assessing machine translation quality in interlingual subtitling: A case study of ArcTime. *Artificial Intelligence and Robotics Research*, 2021, *10*(2), 206–213. https://doi.org/10.12677/AIRR.2021.102020