

AI vs. human translation: Shaping the future of language learning through AI-generated bilingual captions

Ailin Ma*

CW CHU College, Jiangsu Normal University, 221116 Xuzhou, China

Abstract. The impact of captions on language learning is one of the key research topics today. Researchers have found that the use of different types of captions will cause different learning effects on learners, and the order of using captions will also have different effects on language ability development. With the constant development of today's artificial intelligence technology, however, there is still a research gap in the relevant issues of AI-generated captions. Therefore, this paper summarizes and analyzes the research related to artificial intelligence and captions and explores the relationship between AI-generated bilingual captions and English learners' language learning. The study found that although AI-generated translation still has some shortcomings, its positive effect on English learners' language learning is very obvious. Compared with manual translation, AI translation may be able to better mobilize learning enthusiasm and help ESP learning. In addition, teachers and learners are advised to use AI translation more flexibly and pay attention to some words that are easily mistranslated, such as cultural vocabulary and emerging vocabulary.

1 Introduction

Now more and more platforms, such as YouTube and other websites, have been equipped with AI-powered auto subtitle generators and machine translation functions; only a few websites, such as the TED official website, still use manual translation with translator's signature for video translation. At the same time, with the development of diversified teaching, more and more teachers use instructional videos in teaching to train students' listening, speaking, reading, and writing skills. With the widespread use of instructional videos and the popularity of AI-generated bilingual captions, teachers are increasingly likely to use videos with bilingual captions generated by AI for teaching.

Many past research results have shown that captions have a positive effect on improving English proficiency for English as a foreign language (EFL) learner when watching videos. This type of research mainly focuses on two aspects: subtitle types and improving English proficiency. The concepts of L1 and L2 captions used here come from Montero Perez [1]. L2 captions refer to screen text transcribed in the target language, while L1 captions refer to

* Corresponding author: 3020212366@jsnu.edu.cn

captions that translate audio into the learner's native language. This study will also use relevant terms based on this definition.

Some studies have compared the effects of full captions, keyword captions, and no captions on learners [2-5]. There have also been studies comparing L2 captions with L1 captions and no captions [6, 7]. The benefits of providing two captions concurrently, namely bilingual captions, have also been proven by many studies. Recently, some scholars have studied the role of the order of using L1 and L2 captions in helping learners master words [8]. The study showed that watching videos in the order of L1 captions, L2 captions, and finally, no caption helps learners acquire a larger vocabulary, and this study also fills the gap in the research on the order of caption use.

Moreover, video captions promote the development of English proficiency in many aspects, and this direction has also been studied by many researchers. For example, Maribel Montero Perez et al. explored the effective role of videos with keyword captions and full captions on learners' listening comprehension [9]; and Andi Wang et al. used eye-tracking technology to show that captions such as L1 captions are beneficial to improving vocabulary level [2].

The research community has conducted comprehensive research on the field of artificially generated captions, but few studies have introduced the rapidly developing topic of artificial intelligence to explore the role and effectiveness of AI-generated bilingual captions in helping learners with language learning. This study aims to analyze and explore the impact of bilingual captions on English learners' language learning by analyzing previous studies, as well as the advantages and disadvantages of AI-generated translation compared with manual translation, and finally summarizes the possible impact of AI-generated bilingual captions on language learning for English learners and provides comments and suggestions for teachers who use videos for teaching. This study mainly explores the impact, problems, and possible improvements on Chinese native speakers in the process of using captions to learn English, aiming to provide reference value for research in the field of TESOL.

2 The impact of bilingual captions on English learners' language learning

2.1 Positive influences

Many studies have shown that students watching videos with bilingual captions have a very positive effect on their English language learning.

First, when students watch videos with bilingual captions, they will have a more thorough understanding of the video content. Qiao Yuanyuan, Qianyun Hu, and others have also conducted similar studies to prove this point [10, 11]. Understanding the content of the video is a very important part of using videos for teaching. Only by grasping and understanding the overall content of the video can learners better grasp new vocabulary and improve their English listening skills in this context. Otherwise, vocabulary acquisition and listening practice without context can easily cause ambiguity, and it is not conducive to students learning the usage of vocabulary in real contexts and it will also make learning more difficult. For students who do not have a high level of English proficiency, it is difficult to grasp the overall content of the video when watching videos without captions or with pure L2 captions; if only L1 captions are provided, it is not conducive to students learning Chinese and English in correspondence. Providing students with bilingual captions in Chinese and English means that students receive learning intake from three channels all together: English pronunciation in listening; L1 captions and L2 captions at the visual level. They can allocate their attention

freely according to their learning ability and ensure that they fully understand the video content.

In addition, research also proved that watching bilingual captions is also very helpful for vocabulary acquisition [10-12]. When using pure L1 captions or L2 captions, students lack visual intake of the target language or native language, and it is difficult to connect the auditory and visual input content during the learning stage. When using bilingual captions, students can compare Chinese and English content simultaneously and have the most intuitive grasp of English words and meaning context. For target words whose meanings are unknown, bilingual captions also help students find the corresponding Chinese interpretation in time. At the same time, the role of bilingual captions in improving the listening ability of English learners has also been proven by some studies [9].

Some studies have pointed out that there is little difference in the results of bilingual captions and English-only captions on content comprehension and vocabulary acquisition [11]. However, the research subjects are students in Bilingual Programs, whose English level is more likely to be higher than that of normal English learners and have a better grasp of English, which may have an impact on the experimental results. Nonetheless, in the interviews after the experiment, both teachers and students expressed their willingness to use bilingual captions for English learning and were more motivated to use them.

2.2 Possible problems

The effectiveness of bilingual captions on English learners' language learning has been discussed and confirmed more in terms of content comprehension and vocabulary acquisition, while bilingual captions do not have a clear advantage in other aspects of English learning, such as listening. When learners watch bilingual captions, there are too many elements presented in the video, and too much redundant information is not conducive to students' language learning. In particular, the appearance of L1 captions will make students more dependent on the text, because the native language is the language that learners are more familiar with. Meanwhile, visual learning is more intuitive and simpler than auditory learning, which also reduces the focus on audio. Compared with bilingual captions, no captions and keyword captions may have a more obvious improvement in the listening ability of English learners.

3 AI-generated translation versus manual translation

3.1 Advantages of AI-generated translation

With the continuous development of artificial intelligence technology, many related technologies have made corresponding progress, and AI-generated translation is no exception. Compared with manual translation, the cost of AI-generated translation is significantly lower. The time required to complete the translation for AI is very short, and AI can even translate almost simultaneously while watching a video; at the same time, the monetary cost of using AI-generated translation is very small compared to hiring professional translators.

Furthermore, AI is supported by a large database, and its memory capacity is much greater than that of humans. This means that once a similar situation occurs, AI can quickly extract and link to the words that may be used in that situation, and eliminate some incorrect translations in advance to make the best solution; such an advantage is reflected in the translation of technical terms in some specialized fields. Furthermore, compared with ordinary machine translation, AI has functions such as predictive analysis and can

supplement missing and unrecognized words according to the context. Zhang Xiaoyan's related research also proves these advantages of AI-generated translation.

3.2 Existing shortcomings of AI-generated translation

Although AI translation has many pleasing advantages, it also has many areas that need improvements. The biggest problem is that AI translation is still less flexible than translation by professionals, which is mainly reflected in the following points. First, AI translation will not always naturally bring in context to translate specific content. For example, the translation of "take a poll" will naturally translate into a public opinion poll without context by AI, but it is a survey of student learning satisfaction, aiming to collect student opinions [13]. These translations without context can easily cause misunderstandings. Meanwhile, the existence of some emerging vocabulary phrases, such as new product names and advertising slogans, will also significantly reduce the translation quality of AI. Once AI cannot extract the direct corresponding translation from the large database, it needs to translate new words based on association. However, the association function of AI translation still has a lot of room for improvement, so the accuracy of emerging phrases translated by AI is not high. There are also words with similar pronunciations that are difficult for AI to recognize and translate correctly, while professional translators can understand and translate accurately based on the context; the translation of culturally related content is also difficult for AI, while people can handle this type of translation based on cultural background and their own experience.

On the top of flexibility, there are also problems with sentence segmentation in AI translation. Generally, AI translation segments sentences and forms subtitles based on sound pauses, which can easily lead to unreasonable sentence segmentation: subtitles are too long, too short, or the meaning is cut off.

The above shortcomings are most likely to be reflected in speech videos because speech videos place higher demands on flexibility. The speaker's accent, background noise, grammatical errors during impromptu speeches and many others need to be taken into account during translation. In addition, if there are auxiliary speech tools during the speech, such as pictures and texts on the screen, general AI translation will not translate them, which also hinders understanding to a certain extent.

4 The impact of AI-generated bilingual captions on English learners' language learning

4.1 Positive impact on learning motivation

AI-generated bilingual captions are likely to motivate students to learn. With the rapid development of the Internet, the types of videos uploaded are becoming more and more diverse, and the number of videos is increasing. At the same time, videos are updated very quickly, especially videos that pay great attention to timeliness, such as news, video blogs, etc., some of which can be updated multiple times a day. However, due to the limited number of professionals in manual translation and limited attention, professional translators usually make captions for certain videos instead of translating every English video. Yet manual translation takes a long time, and for these fast-appearing videos, the speed of manual translation cannot keep up with the speed of video updates. Therefore, although current learners can access a variety of original English videos to help them learn English, there are very few videos that have been manually translated. In most cases, people can find instructional videos that they are interested in, but without captions, which can make learning steeply more difficult, and thus may also significantly reduce motivation.

AI-generated bilingual captions may be able to solve this problem and greatly increase learners' motivation for learning English. AI translation is efficient and fast and can translate multiple videos in parallel and ensure the basic quality of translation. As long as AI translation is used, bilingual captions can be quickly generated for each video without time or quantity barriers. Therefore, when learners find an instructional video that they are interested in, AI translation can quickly generate bilingual captions for learners to learn.

4.2 Positive impact on English for Specific Purposes (ESP)

According to Omnia Ibrahim Mohamed, ESP is a significant branch of English language teaching and is a kind of teaching method used by teachers to guide students in learning English relevant to their major or profession [14]. The videos used for ESP learning will inevitably involve many professional words, which is a huge challenge for manual translation. It is very difficult to translate a large number of professional words without relevant background. To successfully complete the translation, professionals must be familiar with the relevant professional background many days in advance. This not only takes a long time but also cannot guarantee the final high-quality translation. Therefore, there are very few professional videos on the Internet with high-quality manual translation captions.

As for AI translation, it has a powerful information base and a vocabulary base for various professional fields. Translation in specific fields is a kind of translation by selecting words in a restricted scope, which is simpler for AI translation. Many studies have also proved that AI translation has a very high accuracy in specific fields, such as medicine, so AI-generated bilingual captions might be able to provide reliable and efficient translation for learners who want to learn English in specific fields.

4.3 Possible negative effects

First, AI-generated bilingual captions may make learners dependent. The original intention of AI-generated bilingual captions is to help learners learn English more effectively, but the convenience of AI-generated bilingual captions may also make some learners forget their original intention of learning. People tend to pay more attention to content they are familiar with. When every video is equipped with bilingual captions, students with weak willpower may tend to focus on the L1 captions and ignore the English ability they need to practice and improve.

Moreover, the quality of bilingual captions generated by AI is not 100% accurate yet, and AI often makes mistakes. Especially when it comes to cultural translations, such as slang, or emerging words, such as advertising slogans, the quality of AI translation will be greatly reduced. If these words are the core content of the video, it will seriously hinder learners' understanding of the video, and learning with such captions can also affect the effectiveness of English learning.

5 Suggestions

5.1 Teachers

When teachers select videos with bilingual captions as teaching materials, they are advised to first choose videos that have been translated by professionals and strictly reviewed, because this is the best choice to ensure the accuracy of the captions and will not cause learning problems for students. When teachers select new video materials and there is no professional manual translation, using AI to generate bilingual captions is a good choice.

According to current research, the reliability of bilingual captions generated by AI is also very strong. AI-generated bilingual captions can be translated quickly and are very practical. However, it may be better for teachers to check the captions after the generation of AI, focusing on some culturally relevant phrases, emerging words, and words with similar pronunciations.

Additionally, when teachers select videos with bilingual captions generated by AI as instructional videos, it is very necessary to think about which aspect of students' English ability they want to train. If teachers want to improve students' vocabulary and let students have an overall grasp of the video content, it is recommended to choose videos with AI-generated bilingual captions. If teachers want to train students' English abilities in other aspects, such as listening, it is recommended to choose videos with other caption types.

5.2 Learners

Learners often watch original English videos to improve their English. However, without captions, it may be difficult to confirm whether what they hear and think is correct. In this case, it is recommended to use AI-generated bilingual captions. Such captions can be turned off or on by learners, which means they can freely choose to turn captions on or off according to their own learning needs.

When doing listening practice, learners can use captions as a learning tool. If learners are doing intensive listening, they can pause and segment each sentence, and then turn on the captions after listening to check whether the sentences they heard are correct and whether their understanding is appropriate; if learners are doing extensive listening, they can check in each section.

Learners can also use videos with AI-generated bilingual captions to learn vocabulary. If learners do not need to learn words with a high degree of precision, they can roughly determine the meaning and approximate usage context of unfamiliar words by watching the bilingual captions. However, when using AI-generated bilingual captions, learners should also pay attention to some problems that may arise in AI translation, such as cultural vocabulary and emerging vocabulary. When encountering such vocabulary, it is recommended that learners check the dictionary or the Internet for further confirmation.

6 Conclusion

This study explores the impact of AI-generated bilingual captions on English learners' language learning. The result shows that AI-generated bilingual captions have far more advantages than disadvantages. Although AI translation cannot completely ensure the accuracy of translation, especially when translating cultural phrases, emerging vocabulary, and words with similar pronunciations, its overall accuracy and translation efficiency are sufficient to support effective learning for English learners. In addition, AI-generated bilingual captions are highly timely and applicable to almost all videos. Such captions may be able to better mobilize learning enthusiasm and help ESP learning, which may be beyond the reach of manually translated bilingual captions.

In order to better use AI-generated bilingual captions, teachers are advised to use such captions to train students' overall comprehension and vocabulary acquisition. It is recommended to check the captions before formal use and pay attention to error-prone points. Meanwhile, when students use AI-generated bilingual captions on their own, they are advised to freely choose to turn on and off captions according to their own learning needs. Students should also pay attention to some easy-to-mistake translations to avoid misunderstandings of key content.

This study uses the application of AI in education as the background to explore the impact of bilingual captions on English learners' language learning, providing new thinking directions for other researchers. However, this paper does not use specific experiments to conduct research. It is recommended that future research can use experimental methods to further verify the specific impact of AI-generated bilingual captions on English learners' language learning.

References

1. M. Montero Perez, Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Lang. Teach.* **55**(2), 163-192 (2022)
2. D. Aldukhayel, The effects of captions on L2 learners' comprehension of vlogs. *Lang. Learn. Technol.* **25**(2), 178-191 (2021)
3. P. Winke, S. Gass, T. Syodorenko, Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Mod. Lang. J.* **97**(1), 254-275 (2013)
4. M.P.H. Rodgers, S. Webb, The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO J.* **34**(1), 20-38 (2017)
5. F. Teng, Maximizing the potential of captions for primary school ESL students' comprehension of English-language videos. *Comput. Assist. Lang. Learn.* **32**(7), 665-691 (2019)
6. P. Markham, L.A. Peter, The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *J. Educ. Technol. Syst.* **31**(3), 331-341 (2003)
7. A. Hayati, F. Mohmedi, The effect of films with and without subtitles on listening comprehension of EFL learners. *Br. J. Educ. Technol.* **42**(1), 181-192 (2011)
8. I. Wi, F. Boers, Sequential Use of L1 and L2 Captions: Exploring the Benefits for Vocabulary Acquisition. *TESOL Q.* **58**(1), 511-521 (2023)
9. M. Montero Perez, E. Peters, P. Desmet, Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL* **26**(1), 21-43 (2014)
10. Y. Qiao, The influence of subtitle type on high school students' English video comprehension and incidental vocabulary acquisition, Master Thesis, Henan University (2021)
11. Q. Hu, M. Deocampo, The influence of English movies bilingual subtitles to Chinese students' English comprehension and vocabulary acquisition. *EURASEANs: J. Glob. Socio-Econ. Dyn.* **1**(44), 345-353 (2024)
12. L. Lyu, J. Zhu, X. Yv, A study on the effect of bilingual subtitles on original film and television works and incidental acquisition of college English vocabulary. *J. Kaifeng Inst. Educ.* **3**, 103-104 (2016)
13. X. Zhang, Assessing Real-time Bilingual Subtitling of AI-powered Simultaneous Interpretation: A Research Report on Experiment Conducted with Youdao SI, Master Thesis, Nankai University (2021)
14. O.I. Mohamed, N.N. Alani, English for Specific Purposes: An Overview: Definitions, Characteristics and Development. *Eng. Lang. Teach.* **2**, 510-20. (2022)