

Exploring second language viewers' use of cognitive strategies in learning Chinese through multimedia learning resources with captions and social annotations

Shuzhou Lin¹, Ting Wang¹, Wei Wei², Jiahui Yang³, Sara Cañizal Sardón¹

¹University International College, Macau University of Science and Technology, Macau, China, ²Faculty of Applied Sciences, Macao Polytechnic University, Macau, China,

³Departamento de Filología Inglesa (Lengua Inglesa), Universidad de Sevilla, Sevilla, Spain.

Abstract

Although linguistic captions are perceived as helpful in developing L2 vocabulary knowledge, the use of social annotations in video learning remains underinvestigated. This study investigated L2 learners' use of cognitive strategies in learning through multimedia resources in a MOOC context when both caption and social annotations modes were available. Triangulated data were collected from thirteen L2 MOOC learners from Africa, including think-aloud data, post-video interviews, and their notes. Findings suggest that 1) captions and social annotations lead to different cognitive strategies and videos with social annotations are found more engaging, 2) captions facilitate L2 viewers to conduct more bottom-up listening strategies at sentence level, while social annotations facilitate more top-down listening strategies, and 3) social annotations are used not only to enhance and expand the understanding of the video, but also to create a sense of belonging and further motivate viewers to achieve a higher level of engagement.

Keywords: *Social annotation; caption; multimedia learning; cognitive strategies; Cognitive Theory of Multimedia Learning (CTML).*

1. Introduction

This study aims to investigate the cognitive strategies adopted by L2 learners in multimedia learning with captions and social annotations. Social annotations refer to the comments created collaboratively by users on the same online document (Sun, Hwang, Yin, Wang, & Wang, 2022), and they focus on the content of the document. In a digital reading and writing environment, social annotations are perceived as beneficial to learners in terms of 1) increasing motivation and engagement levels (Yang, Zhang, Su, & Tsai., 2011), 2) better conditions for improving reading comprehension and writing quality (Li & Lai, 2022; Yang *et al.*, 2011), 3) better opportunities for peer feedback (Chang & Windeatt, 2016), and 4) promotion of knowledge sharing (Yang *et al.*, 2011). However, linguistic-oriented captions, also known as subtitles, are also widely promoted as pedagogical and scaffolding tools for second language viewers in multimedia learning environments, including L2 subtitles (Mahdi, 2017), hyperlinks (Garrett-Rucks, Howles, & Lake, 2015), and pictorial information (Aldera & Mohsen, 2013). Contributions of this study include 1) a better understanding of how linguistic-related captions and content-related social annotations affect L2 viewers' use of cognitive strategies in video learning, and 2) the development of more effective, engaging, and beneficial multimedia language learning resources equipped with captions and social annotations based on L2 learners' cognitive learning processes.

2. Literature review

Annotations are often used as visual aids to facilitate learning in multimedia learning environments (Unal, 2021; Mohsen, 2016). In the context of applied linguistics or educational research, social annotations refer to 1) personal comments on a specific item and/or peers' annotations (Li & Lai, 2022; Sun *et al.*, 2022; Yang *et al.*, 2011), 2) questions and answers to learning materials (Yang *et al.*, 2011), and 3) spontaneous thoughts and reflections (Li & Lai, 2022). The taxonomy of social annotations includes views (Li & Lai, 2022), judgments and evaluations (Pham, 2021), and emotions and feelings (Chang & Windeatt, 2016). Conventionally, social annotations are examined in the context of digital reading and collaborative writing (Pham, 2021; Chang & Windeatt, 2016). In the fields of computer science and electronic engineering, social annotations are sometimes referred to as 'barrage' or 'bullet screen' (Chen, Zhou, & Zhi, 2019).

There are three main discussion threads in previous studies on annotations in multimedia learning. Apart from the first thread which investigates the link between the incidental learning of vocabulary knowledge and the availability of annotations in a multimedia learning environment, the second one targets annotations based on topic and content, and concludes that 1) topic-level annotations are not as effective as lexical annotations for promoting vocabulary learning (Unal, 2021), and 2) content-related social annotations significantly

improve learners' writing quality and peer commenting skills (Pham, 2021). For example, Unal (2021) investigated how lexical and topic-level annotations and working memory capacity affect incidental vocabulary learning, and concluded that lexical annotations result in better performance in vocabulary meaning recall. Finally, the third thread investigates the impact of social annotations on learners' emotions, motivation, and engagement. Social annotations are found to 1) increase learners' motivation and engagement (Li & Lai, 2022; Yang *et al.*, 2011); 2) develop learners' confidence, a sense of community, and mutual trust (Sun *et al.*, 2022; Chang & Windeatt, 2016). For instance, Li and Lai (2022) compared the effects of social annotations and online forums on L2 learners' online collaborative writing, and found that social annotations not only led to better learning outcomes but also enhanced learners' engagement levels.

To fill the gaps in the research of multimedia learning, this study asked: what cognitive strategies are used by L2 viewers when captions and social annotations are provided in multimedia learning? The Cognitive Theory of Multimedia Learning (CTML), proposed by Mayer (2009), was adopted as the theoretical framework of this study. According to CTML, the cognitive process of multimedia learning includes: 1) selecting, which means selecting relevant verbal and pictorial information from multimedia resources for processing in working memory; 2) organizing, which means organizing the selected verbal and pictorial information into the coherent verbal and pictorial models, thus leading to the creation of internal connections; 3) integrating, which means building external connections by integrating the verbal and pictorial models with prior knowledge.

3. Research design

3.1. Research context

Thirteen (6 females and 7 males) African students from a three-year learning Chinese as a second language program were recruited for this study. The participants were second-year international students from a vocational college, aged between 18 and 24. Their Chinese proficiency level was around HSK level 3 (equivalent to The Common European Framework of Reference for Languages (CEFR) level B1). Based on the MOOC platform, this program is called Improve Chinese Communication through Multimedia Learning. Several carefully chosen Chinese language videos are included in this 8-week program, with the goal of enhancing the communication and proficiency of international students in Chinese. Both linguistic-based captions and social annotations (comments generated by native Chinese viewers) are provided along with the videos. Due to the pandemic, participants had been learning online (on the MOOC platform) for at least 12 months, and they were familiar with the operating system.

3.2. Data collection

The experiment was conducted on a MOOC platform where students voluntarily video-recorded the whole learning and data collection process with an extra phone. Think-aloud interviews, observations, and post-video interviews were used to gather the data for this study. Participants were informed about the purpose of the study and the procedures of data collection, and then demonstrated how to do think-aloud and control the operating system on the MOOC platform. They were required to watch a two and half-minute long Chinese video about spicy Chinese cuisine. With no time limit, they were allowed to pause, rewind, fast-forward, and restart the video, take photos or notes, and use online dictionaries on their cellphones. Moreover, they were given control over access to the type of on-screen text, which means that they could manually switch between two modes: caption mode and social annotation mode. In the caption mode, viewers can see the transcripts of the video, while in the annotation mode, they can access comments made by L1 viewers about the video's subject matter and their feelings toward it.

Post-video interviews were conducted with the same predetermined questions following each participant's completion of the video they had just watched in order to gather information on the participants' cognitive processes. The interview questions were designed based on the cognitive processes from CTML (Mayer, 2009).

3.3. Data analysis

Participants watched the video about four times on average, thus about four short post-video interviews were conducted for each participant. Participants reportedly selected and deselected various types of information in the video while in caption mode and social annotation mode. Based on the three cognitive processes from CTML (Mayer, 2009), think-aloud and post-video interview data were coded deductively first. Observation data such as time spent on each video, the number of pauses, and notes taken were collected and analyzed, as these learning behaviors in two modes indicate different cognitive processes and strategies in multimedia learning. For example, the duration of video learning may suggest learners' motivation and engagement levels under different learning modes.

4. Findings

4.1. Quantitative data

Table 1 presents the results of the descriptive data analysis. Overall, to learn the two-and-a-half-minute-long instructional video, the duration of video learning in social annotation mode was longer than in the caption mode. Post hoc analyses were conducted using Bonferroni's post hoc test. Participants took significantly longer time to learn videos with social annotations ($M = 305$ seconds, $SD = 62.798$) than with linguistic-oriented captions ($M =$

232.32 seconds, SD = 101.584). Moreover, they also paused the video with social annotations more frequently ($M = 7$, $SD = 4.619$) than captions ($M = 1.62$, $SD = 1.387$). No statistically significant difference was found in the number of notes taken under two different modes. Since participants were encouraged to watch the videos as many times as they wanted, a longer time spent on videos with social annotations may suggest a higher engagement level. Therefore, the descriptive data suggest that captions and annotations lead to very different cognitive strategies, levels of motivation and engagement.

Table 1. Descriptive data

	Mode	N	Mean	SD	Sig.
Duration	Caption	13	232.31	101.584	0.000
	Social annotation	13	305.38	62.798	
Pauses	Caption	13	1.62	1.387	0.000
	Social annotation	13	7.00	4.619	
Notes	Caption	13	1.77	1.878	0.134
	Social annotation	13	3.08	4.132	

4.2. Cognitive strategies to cope with linguistic-oriented captions

The different engagement levels between the two modes can be explained by the different cognitive strategies reported by the L2 viewers during three periods: selecting, organizing and integrating.

As for the caption mode, the viewers reported that in the selecting period, they paid attention to three main resources. First of all, pictorial information from images was used to help them 1) understand the unfamiliar vocabulary and phrases, 2) make sense of the sentence plots, and 3) facilitate their understanding of the details of the video content. Secondly, audio information was selected from the narration to 1) learn the pronunciation of unknown words, and 2) assist in the understanding of video content. Thirdly, textural information from captions was selected to 1) learn the characters of unfamiliar words and phrases, and 2) support their understanding of the details.

In the organizing period, the cognitive strategies mainly involved: 1) organizing textual information from a target word and its surrounding words into a verbal model to make sense of the less familiar words or phrases; and 2) organizing pictorial and audio information with linguistic captions to learn, check, or enhance their understanding of less familiar words or phrases, and content details. For example, one participant reported that:

(S2) When I heard a new word, I paid attention to the images, and focused on what he meant, but I also needed the characters in the captions, otherwise I couldn't understand the new word.

In the final period, what they had newly learned from the video was reported to be integrated with 1) background knowledge, such as already known words and prior knowledge of Chinese cuisine, and 2) personal experiences. The integration helped them to further turn working memory into long-term memory. One participant explained:

(S12) To learn the new word “蒸” (to steam), the images helped me learn the word but I need to listen to it ... When I learned this new word in steamed tofu, I thought about other delicious foods like dumplings and buns, which were all cooked in the same way.

4.3. Cognitive strategies to cope with social annotations

Under the social annotation mode, participants showed great interest in the comments generated by native Chinese viewers. In the selecting period, they paid attention to 1) textual information from social annotations to enhance and expand their understanding of the video, and 2) pictorial information from comment-related images to make sense of the comments. As there were numerous scrolling lines of comments, it was discovered that they selectively read the social annotations and ignored the audio commentary.

In the organizing period, participants were found to organize the verbal information and pictorial information in three ways: 1) organizing textual information from social annotations to understand the content of comments, and learning how to write social annotations properly to communicate with L1 viewers; 2) organizing textual information from contradictory social annotations to evaluate the comments and video content; 3) organizing relevant pictorial information from the images to assist in the learning of social annotations. The quote below demonstrates how the verbal model and pictorial model were organized:

(S9) The comment said: “Tofu sweats a lot.” In the video, I can't see tofu sweating a lot, so it seems that this viewer has tried cooking tofu before and this comment tells you the possible problem.

After the verbal and pictorial models based on social annotations were organized, they were integrated with 1) video content learned earlier, 2) similar or contradictory experiences in L1 and L2 contexts, and 3) cultural and learning experiences. Moreover, more than half of the participants tried to generate comments to interact with the chef and L1 viewers, and reported that they were happy to see others holding the same opinions as their own, which gave them a sense of belonging. The quote below suggests how this external connection was built:

(S7) This comment also questioned the use of sugar. In my country, if we want a sweet dish we only add sugar, and if we want a sour dish we only use vinegar. We don't mix sugar with chilli or vinegar. When I see others having the same question, I feel that I am not alone.

5. Discussion and implications

To comprehend the captions and social annotations, different cognitive strategies were used by L2 learners to select, organize and integrate the textual, video and audio information.

In caption mode, participants selected verbal information and pictorial information, and then organized them into verbal models and pictorial models to learn the unfamiliar words and some details about the content. This indicates that detailed internal connections between the verbal and pictorial models were built at both the linguistic and content levels (Mayer & Moreno, 2003). In relation to the learning of new words, participants were found to select the pronunciation from the audio narration, characters from captions, and relevant images. They then organized this information and integrated it with their background knowledge and personal experience to enhance their understanding of the words. In return, a better comprehension of the details of the captions at the sentence level added more detail to viewers' understanding of the video, such as the ingredients and cooking procedures. In other words, a bottom-up model can be observed when L2 viewers interplay with linguistic captions in instructional videos at the sentence level (Vandergrift, 2004).

In the mode of social annotations, participants selected, organized and integrated information for evaluating the content of social annotations, learning how to make comments in L2 properly, and communicating with L1 viewers. They integrated the verbal and pictorial presentations of social annotations with knowledge learned earlier in the video, personal experiences, and other background knowledge, to further confirm and evaluate the social annotations selected and content details. It suggests that both internal and external connections were made for social annotation learning (Mayer & Moreno, 2003). The overall understanding of the video and social annotations from other viewers, including personal and cultural interpretations of the video, was used to evaluate and confirm some other details in the video, which fits in the top-down model (Vandergrift, 2004). Both quantitative and qualitative data suggest that social annotations lead to higher engagement and motivation levels, which is in accordance with the conclusions of previous studies (Li & Lai, 2022; Yang *et al.*, 2011). For example, longer learning time and more pauses were recorded under the mode of social annotation, serving as evidence of higher learning engagement and motivation. Social annotations can also create a sense of belonging for L2 learners (Chang & Windeatt, 2016) and encourage them to interact with L1 viewers.

At least two pedagogical implications can be summarized. To begin with, giving learners the autonomy to select captions or social annotations based on their needs may accommodate their individualized learning goals. In addition, meaningful, comprehensible, and topic-relevant social annotations can enhance their understanding of video and lead to higher learning engagement and motivation levels.

References

- Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education*, *68*, 60-75. doi: 10.1016/j.compedu.2013.04.018
- Chang, H., & Windeatt, S. (2016). Developing collaborative learning practices in an online language course. *Computer Assisted Language Learning*, *29*(8), 1271–1286. doi: 10.1080/09588221.2016.1274331
- Chen, G., Zhou, S., & Zhi, T. (2019). Viewing mechanism of lonely audience: Evidence from an eye movement experiment on barrage video, *Computers in Human Behavior*, *101*, 327-333. doi: 10.1016/j.chb.2019.07.025
- Garrett-Rucks, P., Howles, L., & Lake, W. M. (2015). Enhancing L2 reading comprehension with hypermedia texts: Student perceptions. *CALICO Journal*, *32*(1), 26-51. doi: 10.1558/calico.v32i1.26131
- Li, S. C., & Lai, T. K. (2022). Unfolding knowledge co-construction processes through social annotation and online collaborative writing with text mining techniques. *Australasian Journal of Educational Technology*, *38*(1), 148-163. doi: 10.14742/ajet.6834
- Mahdi, H. S. (2017). The use of keyword video captioning on vocabulary learning through mobile-assisted language learning. *International Journal of English Linguistics*, *7*(4), 1-7. doi: 10.5539/ijel.v7n4p1
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511811678
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning, *Educational Psychologist*, *38*(1), 43-52. doi: 10.1207/S15326985EP3801_6
- Mohsen, M. A. (2016). Effects of help options in a multimedia listening environment on L2 vocabulary acquisition. *Computer Assisted Language Learning*, *29*(7), 1220-1237. doi: 10.1080/09588221.2016.1210645
- Pham, V. P. H. (2021). The effects of lecturer's model e-comments on graduate students' peer e-comments and writing revision. *Computer Assisted Language Learning*, *34*(3), 324-357. doi: 10.1080/09588221.2019.1609521
- Sun, C., Hwang, G. J., Yin, Z., Wang, Z., & Wang, Z. (2022). Trends and issues of social annotation in education: A systematic review from 2000 to 2020. *Journal of Computer Assisted Learning*, 1-22. doi: 10.1111/jcal.12764
- Unal, B. (2021). Glossing and incidental vocabulary learning in L2 reading: a cognitive load perspective. *International Review of Applied Linguistics in Language Teaching. Advance online publication*, 000010151520200164. doi: 10.1515/iral-2020-0164

- Vandergrift, L. (2004). 1. Listening to learn or learning to listen?. *Annual Review of Applied Linguistics*, 24, 3-25. doi: 10.1017/S0267190504000017
- Yang, S., Zhang, J., Su, A., & Tsai, J. (2011). A collaborative multimedia annotation tool for enhancing knowledge sharing in CSCL, *Interactive Learning Environments*, 19, 45-62. doi: 10.1080/10494820.2011.528881