Effects of captioned videos on learners' comprehension

Sijia Guo
Macquarie University, Australia
sijia.guo@mq.edu.au

Helena Sit
University of Newcastle, Australia
helena.sit@newcastle.edu.au

Shen Chen
University of Newcastle, Australia
shen.chen@newcastle.edu.au

Abstract: The emergence of mass media and media literacy has impacted education, including second language programs. Language and communication in social media bring educators new ways of teaching and engaging students in multimodal literacy practices. In most Australian institutions, language studies other than Chinese are offered in a transmedia mode. However, few studies have examined the efficacy of Chinese language acquisition through transmedia in the format of captioned video at tertiary level. In this case study, the use of captioned video was investigated within the context of the use of captioned video within the context of a multimedia Chinese language-learning environment and its effects on learners' language comprehension at Macquarie University. Both quantitative and qualitive research methods were employed. The results provided both empirical evidence and practical strategies on

the feasibility, acceptability and efficacy of a captioned video transmedia intervention into Chinese language acquisition. The findings contribute to the literature on multimedia-based activities that bring digital humanities into language literacy.

Keywords: language and digital world, language and media literacy, language education, Chinese as a second or foreign language (CSL/CFL), captions

Introduction

Captioned videos were initially developed as an aid to hearing impaired individuals in the early 1970s (Taylor, 2005). Research into the value of captioned video as a tool for second language (L2) learning began in the early 1980s (Garza, 1991). Videos, which incorporate native speakers' presence and voices, have become an effective resource facilitating learners' comprehension and motivation (Weyers, 1999). A spectrum of studies has shown that it helps learners connect auditory to visual input, which can aid form-meaning mapping – a key process contributing to second language acquisition (SLA). Currently a number of studies have shown that captioned videos have a positive influence on learners' listening comprehension (Hosogoshi, 2016; Mirzaei et al., 2017; Montero Perez et al., 2013; Yang & Chang, 2014) and vocabulary acquisition (Montero Perez et al., 2014; Peters et al., 2016). For instance, a meta-analysis conducted by Montero Perez, Van Den Noortgate, and Desmet (2013) found that fully captioned videos had a significant effect on learners' listening comprehension (Hedges' g = 0.99) and vocabulary learning (g = 0.87). Despite the progress on research in this area, the existing literature shows that few cases have focused on Chinese language learning.

In Australia, the Chinese language, and subsequently Chinese heritage speakers, have spread far and wide. According to the 2016 census conducted by the Australian Bureau of Statistics (ABS), Mandarin was the second most common language spoken at home at 2.5% (n = 596,711), while Cantonese was listed fourth at 1.2% (n = 280,943) (ABS, 2017). Heritage speakers refer to speakers who learnt a minority language at home as a child. However, the language is never fully developed because the speaker grows up in another dominant language environment (Valdés, 2000). In this study, Chinese heritage learners refer to those who were born and grow up in Chinese families in Australia. They speak Mandarin or other dialects of Chinese language such as Cantonese, Shanghainese, and so on. On the other hand, non-heritage Chinese speakers include both Australian domestic and international students whose ethnic backgrounds are not Chinese.

This study compared the effects of captioned videos on the language comprehension of both cohorts of Chinese language learners. In particular, it investigated the effects of captioning during video-based listening activities among heritage and non-heritage Chinese learners in an Australian university. It further attempted to determine whether the beginner learners were listening to the videos or whether they were reading the texts (Vandergrift & Cross, 2014). In addition, the study aimed to identify the differences between non-heritage and heritage Chinese speakers in terms of captioned video comprehension.

Background

Application of captioned videos and L2 listening comprehension

In the last two decades, the application of captioned videos on learners' listening comprehension has been researched extensively (e.g., Danan, 2004; Diao et al., 2007; Hosogoshi, 2016; Mirzaei et al., 2017; Montero Perez et al., 2014; Montero Perez et al., 2013b; Taylor, 2005; Vandergrift, 2011; Winke et al., 2010; Yang, 2014; Yang & Chang, 2014; Yeldham, 2018). A significant number of studies have shown that compared with non-captioned videos, captioned videos can lead to better understanding in language learning (Mirzaei et al., 2017; Montero Perez et al., 2013; Rodgers & Webb, 2017). On the other hand, the findings in some studies have provided only limited support for the use of captions (Montero Perez et al., 2014).

Montero Perez et al. (2013a) investigated 226 university students' comprehension tests which consisted of global and detailed questions after watching (1) non captioned, (2) fully captioned, and (3) keyword captioned videos. The findings revealed that the fully captioned video group outperformed the groups in the other two video settings. According to the survey analysis, compared with keyword captions, which were considered highly distracting, the learners seemed to have a strong desire for full captions. They believed that captions could facilitate learners' speech decoding and the meaning-making process. In the same vein, Winke et al. (2010) investigated the influence of using captioned video on second- and fourth-year learners of Arabic, Chinese, Spanish, and Russian listening comprehension. The results indicated that captioned videos, which aided overall comprehension, were more effective than videos that were not captioned. This finding echoed Bird and Williams's (2002) study, which claimed that bimodal modalities could lead to better recognition in memory. Furthermore, they examined the effect of the order of using captions and found that for learners of Arabic and Chinese, showing captions a second time tended to be more effective. They suggested that for languages whose orthography is not close to the native language, the aural modality is preferable when

the written symbols are not familiar to the learners. Mirzaei et al. (2017) noticed that the introduction of a novel captioning method, partial and synchronised captioning (PSC), reduced the captions shown on the screen, which in turn helped learners decrease their dependence on captions. The PSC was evaluated against non-captioned and fully captioned videos. The findings illustrated that when shown less than 30% of the captions, the participants were still able to achieve the same level of comprehension as those in the full caption group.

On the contrary, there are some studies that have shown no significant difference between captions and no captions. Montero Perez et al. (2014) examined the effects of three caption types (non-captioned, fully captioned, and keyword captioned) on learners' comprehension of videos. In their study, 133 Flemish university students who were considered as high-intermediate learners of French watched three videos twice and completed three comprehension tests. The results showed that even though the captioned group significantly outperformed the non-captioned groups on form recognition and clip association, the participants in all conditions achieved similar scores on the comprehension tests. Moreover, compared to the other two groups, the keyword captioned group reported lower scores on the usefulness of captions, which contrasted with Yang and Chang's (2014) study. These studies have not reached consensus about hence further exploration is necessary.

Learners' language proficiency level and the caption reading

The current literature has shown that learners' focus on captioned videos largely depends on their language proficiency level. According to Yeldham (2018), less-proficient learners tend to read texts more than listening to videos, while more proficient learners who can utilize a wider range of cues such as captions, visual, and audio, may benefit more during video watching. Similarly, in Sydorenko's (2010) study, learners' survey results revealed that the high beginners of Russian learners preferred to read captions rather than use visuals or listen to videos. Winke, Gass, and Sydorenko (2013), who investigated the caption-reading behaviour of four foreign language learners, including Arabic, Chinese, Spanish, and Russian, found that lower intermediate and intermediate learners of the languages mentioned above tended to read the captions. The foreign language learners fixated on the captions area on the screen 68% of the time. In their study, Chinese learners spent more time reading captions in order to derive meaning when the video content was not familiar. They believed that a lower vocabulary breadth and depth was the main reason for this (Stæhr, 2009). Caimi's (2006) study on lower-intermediate EFL learners' reactions to subtitles showed that students were overloaded with visual, aural, and captions, which led to caption reading. Numerous studies have echoed the same finding; that is,

less-proficient learners tend to read captions as a crutch to help them understand the videos (Pujolà, 2002; Taylor, 2005).

On the other hand, studies have shown that more proficient learners tend to tune in to both the spoken message and the captions. In Taylor's (2005) study, more third-year students (50%) than first-year students (23%) reported that they were able to utilise all three channels – image, sound, and captions – while watching videos. Similarly, the more proficient learners in Pujolà's (2002) study reported that they only considered captions when they experienced comprehension difficulties. All these studies indicate that the effect of using captioned video is dependent on the proficiency level of learners. The caption is not the only source for learners to reach comprehension.

Theoretical explanation of how learners process captioned videos

To explain how the captions were processed by learners, the following theoretical frameworks were discussed in the current literature: working memory, dual coding theory, and cognitive load theory (Winke et al., 2013; Yang, 2014; Yeldham, 2018). Working memory (WM) theory posits that there are a central executive and two subsystems in the human brain: the phonological loop and visual-spatial sketchpad (Baddeley, 2007). In other words, visual and aural information is processed through separate channels and each channel has limited WM store (Baddeley, 2000). According to the cognitive load theory, due to the limited capacity of WM, if information input is simultaneously transmitted via multiple channels, it will lead to cognitive overload and have a negative influence on comprehension (Mayer, 2009; Sweller, 1994). When the overload takes place – for example, when watching videos with captions – it will narrow learners' cognitive focus to help them better comprehend. Therefore, less-proficient learners tend to read captions rather than attending to aural cues (Hulstijn, 2003). In Winke et al.'s (2013) study, the authors posited that the Chinese learners experienced a split-attention effect when the videos contained unfamiliar scripts. The dual coding theory depicts that verbal codes and non-verbal codes are received through different channels (Clark & Paivio, 1991). The ears can process sensory input such as verbal and non-verbal cues. Graphical information is processed via visual channels. In practice, better learning takes place via binary codes than by a single channel (Mayer, 2009).

With a few exceptions (Garza, 1991; Winke et al., 2010; 2013), notably absent are studies that examine the acquisition of less-commonly taught languages such as Chinese using captioned video. Further, no study has been conducted to date comparing the effects of captioned video on the language acquisition of heritage and non-heritage learners. Winke et al. (2013) point out that "[i]t is, therefore, worthwhile

to investigate whether languages with different writing systems, phonological features and levels of cognitive demand differentially affect language learners' use of captions" (p. 257). For this reason, our study was timely conducted to address this current research gap, aiming to better understand the influence of captioned videos on learners' Chinese language learning and to identify the different strategies to guide non-heritage and heritage Chinese speakers during captioned video watching.

The current study was designed to answer the following specific research questions:

1) When a video was viewed twice with the second viewing including annotated keyword captions, would the captioned video be more effective than the non-captioned?

2) What kinds of benefits do captioned videos bring to learners of Chinese with a Chinese heritage background and others with a non-heritage background?

3) To what extent do captioned videos benefit Chinese learners' listening comprehension?

Research method

Participants

Altogether 66 participants (28 male; 38 female) in the first year of a Chinese language course at an Australian university participated in this study. Among them were 50 non-heritage Chinese learners and 16 heritage Chinese learners. Their level of Chinese proficiency was classified as beginners (the non-heritage learners) and advanced beginners (the heritage learners) within the context of the Australian university where the study was conducted. The non-heritage learners were from a range of language backgrounds: English, Korean, Hindi, Khmer, Vietnamese, Italian, Thai, Spanish, German, Russian, Japanese, and Samoan. The heritage group consisted of Tagalog, Korean, Shanghainese, and Cantonese speakers. None of the non-heritage participants had previous knowledge of Chinese. Both the non-heritage and heritage cohorts were taught by the first author using the same teaching materials at a similar pace. By the end of the experiment, both the heritage and non-heritage learners were able to read and write 300 Chinese characters.

Materials for experiments

The video clips. All the captioned videos were designed and developed on the basis that the content covered the expressions that the students learnt in a teaching week. The primary purpose of the captioned videos was to foster learners' listening

comprehension skill, while the accompanying texts explained the meaning and pronunciation of new words. Each video clip lasted for around two to three minutes. The content of the video clips was consistent with the language syllabus and textbook which covered topics such as family story, transportation enquiry, lifestyles and hobbies, shopping, and so on. In addition, the context of the clips included locations surrounding the university, places with which the students were already acquainted (e.g., the library, classrooms, train/bus stations, shopping centres, student hub/cafeteria, restaurants etc.).

Figure 1. A screenshot of an annotated keyword captioned video

Comprehension tests. The comprehension tests evaluated the effectiveness of each caption method – fully Captioned (FC first), annotated keyword captioned (AKC first), and non-captioned (NC first) – on learners' listening comprehension and provided data for research questions 1 and 2 by comparing the results between non-heritage and heritage learners. Every time that the participants finished watching a certain version of a video twice, they were asked to take a comprehension test, which consisted of multiple choice, fill in the blank, and closed questions. All the questions were designed to examine learners' understanding of the content of the videos in terms of general and specific information. A sample comprehension test can be found in Appendix A.

Procedure

A quasi-experimental design was adopted for the study. At phase 1, three modes of captions including full, annotated keyword, and not captioned were proposed to investigate their impact on beginning learners of Chinese in two elementary Chinese units. The non-heritage learners were assigned into three treatment groups according to the tutorial classes they enrolled in at the beginning of the semester. The three

groups watched different versions of five video clips twice in weeks 4, 5, 8, 9, and 10. Due to higher proficiency in listening and speaking, the setting of the heritage group was the same as the third group of non-heritage speakers. After each video viewing, all the participants completed a comprehension test. Table 1 illustrates the experimental procedure.

Table 1. Students' video watching procedure

	Non-heritage Group 1 (n = 21)	Non-heritage Group 2 (n = 14)	Non-heritage Group 3 (n = 15)	Heritage Group (n = 16)
1st time	Full captioned (FC)	Annotated keyword captioned (AKC)	No captions (NC)	No captions (NC)
2nd time	Annotated keyword captioned (AKC)	Annotated keyword captioned (AKC)	Annotated keyword captioned (AKC)	Annotated keyword captioned (AKC)

At phase 2, the participants were asked to conduct an anonymous post-session questionnaire to solicit feedback in terms of listening comprehension and general learning experiences regarding the use of captioning video following the last video view. During phase 3, those students who expressed interest in volunteering for the focus groups were then interviewed, thus providing their in-depth viewpoints on the effectiveness of captioning videos on their language learning. The data was intended to elucidate the statistic results further.

In terms of data analysis, the survey data and the comprehension tests scores were calculated and compared between the different groups of participants. The focus group transcripts were loaded into NVivo9 (N9). The analysis assisted in identifying connections between specific themes as well as gaps and discrepancies.

Data analysis and results

Comprehension tests results

To answer the first research question, an analysis of the one-way ANOVA test was used to examine the effect of conditions (the type of captions) on listening comprehension among the three non-heritage groups, with the significance level set to 0.05. The effect size ($\eta p2$) was reported and interpreted based on Cohen's rules of thumb: small ($\eta p2 > .01$), medium ($\eta p2 > .06$), and large ($\eta p2 > .14$). To answer the second research question, a t-test was conducted to check the differences of listening

comprehension test scores between the third non-heritage group and the heritage group.

Research question 1: The first research question investigated non-heritage learners' listening comprehension after watching different types of captioning videos. Specifically, the question explored whether captions were more effective than non-captions when a video was viewed twice with the second viewing using annotated keyword captions. To examine whether a significant difference existed among the three non-heritage groups, a one-way ANOVA was employed. Table 2 presents brief information of the three groups of non-heritage learners' comprehension test scores, including the number of tests and group mean scores.

Table 2. Non-heritage groups comprehension results analysis

	N	Mean	Std. Deviation
FC first group	21	7.2310	1.10005
AKC first group	14	6.9821	1.18218
NC first group	15	7.6222	1.23357
Total	50	7.2787	1.16692

The results revealed that the participants' listening comprehension test scores in the three conditions did not differ significantly: $F_{(2, 47)} = 1.125$, $p = 0.333 > 0.05$, $\eta p2 = .046$ which in Cohen's (1988, pp. 284-7) terms would be considered a small effect size.

When comparing the mean scores among the three groups, learners who viewed NC first obtained slightly higher scores in the comprehension test than those who saw captions (FC and AKC) first, but the difference was not statistically significant.

Research question 2: To answer the second research question, which investigated the effect of captioning between non-heritage and heritage Chinese learners, learners from the third non-heritage group and the heritage group watched the five videos twice (NC first, AKC second).

The t-test results showed that the heritage group performed significantly higher on the comprehension tests than the non-heritage group, $t_{(29)} = -2.168$, $p = 0.038 < 0.05$. The heritage group's comprehension tests scores (M = 8.54, SD = 1.12) were higher than the non-heritage group (M = 7.62, SD = 1.23). This finding revealed a positive answer to the second research question. The effect size, Cohen's d = $(8.5375 - 7.6222)\diagup 1.176478 = 0.778 > 0.5$ can be considered a medium effect size.

Survey results

Fifty-six (out of 66) students completed the online survey immediately after watching the last video, making the overall response rate 84.8%. Forty-two of the students were non-heritage Chinese learners while 14 were heritage learners; 65% were Australian domestic students, with the remaining 35% from the UK, New Zealand, Russia, China, Korea, Japan, Indonesia, Vietnam, Thailand, and Cambodia.

Figure 2. Learners' understanding of the videos
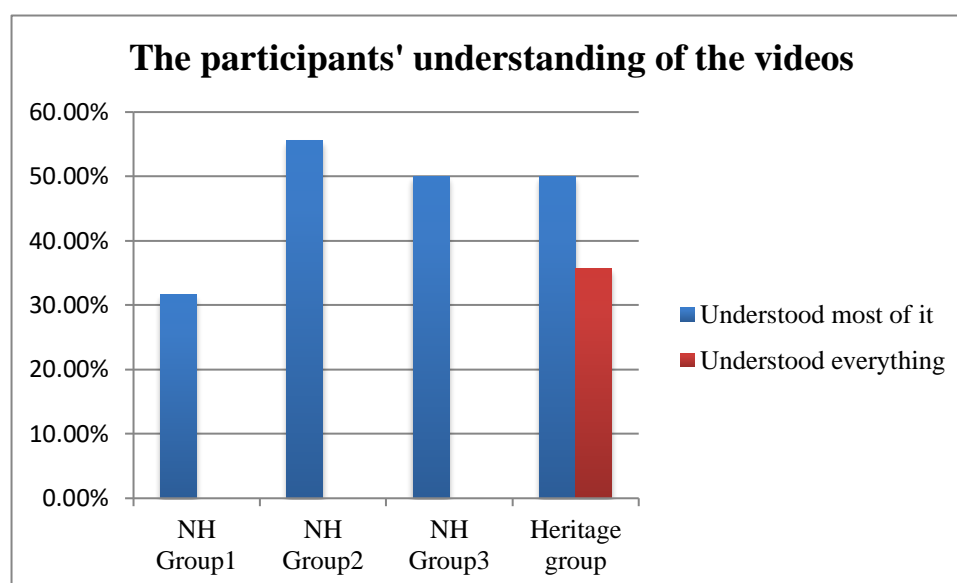
As Figure 2 shows, when asked about their understanding of the captioned videos, 35.7% of the heritage learners (n = 5) expressed that they fully understood the videos, in contrast to the non-heritage learners (n = 0). More than half of the students (55.6%) from the non-heritage group 2 indicated that they understood most of the videos, which was the highest among the three non-heritage groups. For more details of the survey results see Authors (2018).

All participants were asked two open-ended questions regarding the advantages and limitations of captioned video. In their answers, several students (n = 45) believed that captions helped them understand the videos. However, nearly one-third of the students (n = 20) found the videos moved too quickly, which prevented them from understanding.

> Excerpt 1. *Watching the same video twice with the second time having subtitles and highlighted keywords or unknown words helped me understand how the words fit in context. (Non-heritage group 3, NC*

first)

Excerpt 2. *Keyword captions with highlighted colour. (Non-heritage group 2, AKC first)*

Excerpt 3. *Subtitles and videos help learners in understanding the context in which words should be used. (Heritage group)*

Excerpt 4. *Because the keywords/phrases go quickly, while intending to write it down, you miss some of the other parts following that. Likewise, without the subtitle, while trying to figure some parts of it, you miss the rest, maybe playing it again would have helped a bit. (Non-heritage group 1, FC first)*

Excerpt 5, *The pace of the video, sometimes it's too fast, and add of Pinyin$_2$ would be helpful. (Non-heritage group 3, NC first)*

Focus group interview results

After watching the last video, volunteers (n = 15) participated in a focus group interview with the purpose of exploring students' preference of captions and their perceptions of listening comprehension development. The interview data were analysed using an inductive approach. There were five themes that emerged from the participants' comments:

1. Annotated keyword caption aided vocabulary deciphering and understanding.
2. Full caption was overwhelming for beginners since it distracted their attention.
3. Heritage and non-heritage learners held different opinions regarding non-caption.
4. Captions helped to improve learners' listening comprehension.
5. Compared to less-proficient learners, more-proficient learners tended to utilise multiple cues (aural, visual, and captions) while watching videos. However, less-proficient learners considered captions as crutches.

The first three themes answered the first and second research questions. The fourth theme confirmed that captions did help heritage and non-heritage learners with their listening comprehension. The last finding which emerged from the interviews of both non-heritage and heritage learners helped us understand the reasons why these two groups of learners had different perceptions about caption. Participants' interview excerpts are presented according to the themes.

Since all the participants watched the annotated keyword captioned videos for

---

2 Pinyin is a romanization writing system of Chinese.

the second time, their feedback regarding caption types was designed as the first question in the interview. The majority of the interviewees (n = 14) expressed that annotated keyword captions were useful in terms of listening comprehension and vocabulary learning. Only one heritage learner commented that having no captions would not affect his video comprehension.

Excerpt 1. *I like the subtitles, especially with the new words being subtitled in the videos, makes it easier to understand the new words. The good thing is being able to watch the video twice because what you don't pick up on the first one, you can focus on the second. (Non-heritage group 1, FC first)*

Excerpt 2. *I thought keyword captions was useful because I find when you learn a new language when people speak at normal speed, all the words slide together and you can't really decipher all the individual words. But with the captioning, when the character says the word, and then I read it at the same time, I think, "Oh yeah, that's that word." If I didn't have the captions, it will all just come out in one big jumble. (Non-heritage group 3, NC first)*

Excerpt 3. *I prefer it with keywords, definitely. Especially when it's words that we had just learnt and so, you need to really think about them. (Non-heritage group 3, NC first)*

Excerpt 4. *[Having] the keyword is definitely better than without. (Non-heritage group 3, NC first)*

In contrast, full caption was not preferred by the non-heritage group since it led to attention splitting and made them focus on reading rather than watching. Due to their relatively weak listening proficiency, multitasking such as combining listening, watching, and reading could be too challenging for beginners, especially non-heritage learners. However, one heritage learner believed that full captioning could help him better understand the use of specific vocabulary and sentence structures.

Excerpt 5. *If it's full captions, then everybody just reads everything, and they don't pay attention to what's going on. With captions, you look at it but then there are not enough captions, so you still have to pay attention to the video. (Non-heritage group 2, FC first)*

Excerpt 6. *I think whole sentences distract you from actually listening to what they're saying. (Non-heritage group 1, FC first)*

Excerpt 7.   *If I had the full captions, then I could see certain words in the sentence. I can figure out the sentence structure. Whereas if it's just the keywords, I know what they're talking about, but I don't know which order I should put it in. (Heritage group)*

Compared to the heritage learners, five students from the non-heritage group 3 (NC first) commented that they did not fully understand the videos after the first time viewing without captions. However, two heritage learners believed that having captions or no captions would not substantially affect their understanding of the content or the context of the videos.

Excerpt 8.   *I would say obviously with subtitles was better. When I watched the first time, I didn't really understand much at all and then the second time is very much easier with the keywords and stuff. (Non-heritage group 3, NC first)*

Excerpt 9.   *Even without subtitles, you can still guess what they're talking about even if you don't know the exact words. It's just the subtitles let you know what the actual word is. (Heritage group)*

Excerpt 10.   *I think reading with subtitles is pointless, because if we're trying to do comprehension and listening, then only the keywords are really what we need to understand the full context. (Heritage group)*

As far as listening comprehension was concerned, most of the participants (n = 10), particularly non-heritage speakers, confirmed that the captions contributed to better understanding. However, heritage speakers mentioned that understanding the contents of the videos was not challenging for them. They tended to read the captions to get more information.

Excerpt 11.   *I believe that it has improved my listening comprehension. In class, we're not used to speaking in a fast pace. By being able to have the captions, listening-wise, I reckon I understood more. Over each week, I've been able to pick up more words quicker. (Non-heritage group 1, FC first)*

Excerpt 12.   *I would most likely read what they're saying rather than looking at what they're doing and how they're speaking. (Heritage group)*

The last theme revealed the reason why captions have different effects on

learners. More-proficient learners were able to multitask and get more information from various cues. And for less-proficient learners, they rely on reading keyword captions to get critical information from the videos.

Excerpt 13.   *I prefer watching with subtitles. Me, personally, at the moment I'm not a great speaker, not a great listener. When I listen to a conversation, I wouldn't be able to pick up on it. (Non-heritage group1, FC first)*

Excerpt 14.   *I found that when the keywords came up, the caption, I was looking more at the Pinyin, than the character because I'm not familiar with the characters. (Non-heritage group, 2 AKC first)*

Discussion

This study was set out to investigate the use of captions by beginner learners of Chinese while watching videos in the target language. To answer the first research question, which focuses on non-heritage learners' comprehension watching different settings of captioned videos, we found no significant differences among the three non-heritage groups, echoing the findings from Taylor (2005) and Montero Perez et al. (2014). Statistical results showed that captions neither helped nor hindered the non-heritage learners' listening comprehension. Supported by comments from our participants, beginners such as the non-heritage learners in this study, who had little knowledge in reading and writing in the target language, had difficulties in utilising three channels of image, sound, and captions simultaneously. Presenting captions in videos might make them confused and even lead to distraction.

Another explanation for the result is the orthographic differences between participants' native language and Chinese. As Winke et al. (2010) points out, "[languages] with scripts that are different from the native language script [are] easier to be altered to something unknown through hearing it" (p. 80). In this study, all the captions were written in Chinese characters except for a few that were presented in Pinyin in the annotated keyword captions; as such, participants were overwhelmed with new information through aural and visual channels and it may have been difficult to split their attention to read captions written in Chinese characters which are totally different to their first language. As shown in the study by Winke et al. (2013), English speakers who watched captions in a language with non-Roman scripts such as Arabic and Chinese, experienced an overburden of visual working memory capacity (Chandler & Sweller, 1991). Even though they placed more cognitive resources on

reading captions, it still led to less information intake. The current research further confirmed this interpretation.

The second research question compared the effects of captions on learners' listening comprehension between non-heritage and heritage groups. The t-test results showed that, in terms of video understanding, the heritage groups performed significantly better than their non-heritage counterparts. As shown in Figure 2, nearly one-third of heritage learners understood everything in the videos, while this was not reported by any of the learners in the non-heritage groups.

It was not surprising that the heritage group with relatively higher listening and speaking competence had higher scores than the non-heritage groups. Similar findings were reported by Yeldham (2018) who examined the use and effect of captioned videos on learners' listening competence in a number of previous studies. He also found that, compared with less-proficient learners, more-proficient learners show a higher capability to generally utilise a wider range of cues (caption, speakers, and visual) during video watching. As noted in this study, the heritage learners did not rely on captions to understand the videos, but with the help of these, they were encouraged to learn how to use the vocabulary and to construct sentences. Echoing Winke et al.'s (2013) findings, the interview data from this study also depicted that captions did help non-heritage students draw their attention to the language form. According to the noticing hypothesis put forward by Schmidt (1990, 1995, 2001), learners' attention to language form is necessary for language learning. He claimed that understanding the rules or principles of language learning can facilitate the production of output more directly.

Addressing the third research question, survey and interview data also indicated that captioned videos led to better listening comprehension among both groups of learners. The heritage learners appeared to be slightly more competent than the non-heritage groups in terms of listening and speaking at the elementary level but were lacking in reading and writing skills. Despite the apparent difficulties experienced by the non-heritage learners, they overwhelmingly expressed their positive opinions towards captioning. Both non-heritage and heritage learners considered captions as useful and necessary for them to understand the context of conversations, decipher unfamiliar vocabulary, and occasionally acquire certain language expressions. Taylor (2005) suggests that the use of captions could help beginners reduce the pressure that they face when watching target language videos. As found in other studies, beginners perceived captions as a crutch (Syodorenko, 2010; Winke et al., 2010), which contribute to making videos more comprehensible. According to Krashen's (1981) input hypothesis, if the input is made comprehensible to learners, acquisition should take place. With the assistance of captions, information

at the "i+1" level – that is, beyond what the learner can fully understand – can be more accessible for learners to understand and acquire.

Lastly, another noteworthy point is that the research findings confirm the importance of being literate in the 21st century. Not only do students need to develop multiple literacies in regard to multimodality and the various information and communication technologies, but they also need to master multiple literacies in a range of transmedia modes. It is also recommended that "each subject area has its own discourse, genres, registers and textual resources that jointly convey meaning" (Queensland School Curriculum Council, 2001, p. 11) and therefore "developing curriculum that draws upon students' lives should also include thinking about the resources and knowledge that families use to navigate their daily lives" (Seely-Flint et al., 2017, p. 73).

The current research findings provided confirming evidence to support the notion that through specific teaching and learning materials (a set of various video clips created to align with the Chinese language syllabus), both groups of students were more capable of strengthening their abilities in Chinese word recognition, reading, and listening comprehension skills. The developed teaching activities which included audio-visual lesson materials helped the students gain experience in different types of texts and enhance their Chinese language acquisition. As the current research study further shows, the effect of captioning videos on students' Chinese language acquisition was evaluated as effective and beneficial for both groups of language learners.

Conclusion

This study has yielded a number of quantitative and qualitative data reflecting the effects of captioned videos on listening comprehension of learners of Chinese. It supports the use of captioned videos with non-heritage learners of Chinese, who found them to be conducive for making information more comprehensible, deciphering unfamiliar vocabulary, and understanding the context of the videos. For Chinese heritage learners, who do not rely on captions to follow videos, it helps them better acquire vocabulary and draw their attention to notice language forms, which in turn can lead to active language learning. As far as the setting of video watching was concerned, there was no significant difference among the three non-heritage groups (FC, AKC, and NC first and AKC second). However, it was also noted that beginners were more likely to experience cognitive overload if the orthographies of their native and target languages were vastly different, which in turn resulted in less information intake. Chinese heritage speakers, on the other hand, who are more proficient in

listening and speaking are more capable of utilising multiple channels to comprehend captioned videos.

In the current literature, there is a serious lack of studies on the use of captions in a language with non-Roman scripts such as Chinese. The present study focused only on heritage and non-heritage learners of Chinese at the beginner level. More studies are needed to explore the potential of using captions at intermediate or advanced levels. Moreover, it would be worth investigating how learners of Chinese process captions and what strategies they employ when doing so. In spite of this limitation, the study clearly demonstrated that both groups of Chinese learners benefited from learning Chinese language via captioned video. This study was significant in exploring diverse learners' performance and perceptions of their Chinese language learning experiences with caption-designed teaching and learning activities. The findings from this investigation have both wider and global implications for academics teaching the Chinese language who must reflect on how to design, develop, and implement effective approaches to meet the needs of diverse students in the 21st century.

About authors

Dr. Sijia Guo is a lecturer in Chinese Studies, Department of International Studies at Macquarie University. Her research focuses on task-based language teaching, online Chinese teaching via web-conferencing systems, second language acquisition and captioned videos in foreign language education.

Dr. Helena Sit is a Senior Lecturer and PhD supervisor in the School of Education at the University of Newcastle. She holds a PhD in Education and an MEd in TESOL from the University of Newcastle. Her research expertise includes second language education, international education, higher education and teacher education.

Dr. Shen Chen is an Associate Professor of Language Teacher Education at the University of Newcastle. His research interests include second language curriculum design, pedagogical innovation, Computer Assisted Language Learning (CALL), and culture in foreign language education. He has published widely in the field of second language education.

References

Australian Bureau of Statistics. (2017). Census reveals a fast changing, culturally diverse nation. Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/lookup/Media%20Release3

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. (2007). *Working memory, thought, and action.* Oxford, UK: Oxford University Press.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509-533.

Caimi, A. (2006). Audiovisual translation and language learning: The promotion of intralingual subtitles. *The Journal of Specialised Translation, 6*, 85-98.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Cohen, J. W. (1988). *Statistical power anaylsis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Danan, M. (2004). Captioning and subtitling: undervalued language learning strategies. *META, 49*(1), 67-77.

Diao, Y., et al. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237-261.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24*(3), 239-258.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75.

Hosogoshi, K. (2016). Effects of Captions and Subtitles on the Listening Process: Insights from EFL Learners' Listening Strategies. *JALTCALL journal, 12*(3), 153-178.

Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning, 16*(5), 413-425.

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford, England: Pergamon.

Long, D. R. (1990). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition, 12*(1), 65-80.

Mayer, J. (2009). *Multimedia learning* (2nd ed.). Cambridge: Cambridge University Press.

Mirzaei, S. M., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. *ReCALL, 29*(2), 178-199.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology, 18*(1), 118-141.

Montero Perez, M., Peters, E., & Desmet, P. (2013). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21-43. doi:10.1017/S0958344013000256

Montero Perez, M., Noortgate, W. V. D., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720-739.

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134-148.

Pujolà, J.-T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL: The Journal of EUROCALL, 14*(2), 235-262. doi:10.1017/S0958344002000423

Rodgers, M. P. H., & Webb, S. (2017). The Effects of captions on EFL learners' comprehension of English-language television programs. *CALICO, 34*(1), 20-38.

Seely-Flint, A., Kitson, L., Lowe, K., Shaw, K., Humphrey, S., Vicars, M., & Rogers, J. (2017). *Literacy in Australia: Pedagogies for engagement.* (3rd ed.). John Wiley & Sons, Milton, QLD.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorail on the orale of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and*

awareness in foreign langauge learning (pp. 1-64). Honolulu, Hawai'i: University of Hawai'i.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-33). Cambridge: Cambridge University Press.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign langauge. *Stud Second Lang Acquis, 31*(4), 577-607.

Sweller, J. (1994). Cognitive load theory, learning difficutly, and instructional design. *Learning and Instruction, 4*(4), 295-312.

Syodorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology, 14*(2), 50-73.

Taylor, G. (2005). Perceived Processing Strategies of Students Watching Captioned Video. *Foreign Language Annals, 38*(3), 422-427.

Valdés, G. (2000). The teaching of heritage languages: An introduction for Slavic-teaching professionals. In O. Kagan & B. Rifkin (Eds.), *The learning and teaching of Slavic languages and cultures* (pp. 375-403): Slavica Pub.

Vandergrift, L. (2011). Second language listening. Presage, process, product and pedagogy. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 455-471). New York/ London: Routledge.

Vandergrift, L., & Cross, J. (2014). Captioned video: How much listening is really going on. *Contact, 40*(3), 31-33.

Weyers, J. R. (1999). The effect of authentic video on comunicative competence. *Mordern Language Journal, 83*(339-349).

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65-86.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal, 97*(1), 254-275.

Yang, H.-Y. (2014). Effects of advance organizers and subtitles on EFL learners' listening comprehension Sskills. *CALICO, 31*(3), 345-373.

Yang, J. C., & Chang, P. (2014). Captions and reduced forms instruction: The impact on EFL students' listening comprehension. *ReCALL, 26*(1), 44-61.

Yeldham, M. (2018). Viewing L2 captioned videos: What's in it for the listener? *Computer Assisted Language Learning, 31*(4), 367-389.

Appendix A.　A sample of comprehension test

Part 1: Multiple-choice

1) A 今年多大？（　　　　　）

   A. 20 岁　　　　B. 21 岁　　　　C. 22 岁

2) A 哪儿不舒服？（　　　　　）

   A. 嗓子疼　　　B. 头疼，全身都不舒服　　C. 肚子不舒服

3) 医生说他＿＿＿＿＿＿（　　　　　）

   A. 发烧了，是感冒　　　B. 嗓子发炎了　　　　C. 很好

4) What's the reason why A called Guo laoshi?　　（　　　　　）

   A.　他要来晚了　　　B. 他有问题　　　C. 他不能来上课

Part 2: Answer the following questions.

5) 医生说 A 应该做什么？

6) 他为什么不能来上课？