

The effects of captions in teenagers' multimedia L2 learning

LAURENCE LWO

*College of Humanities & Social Sciences, National Taiwan Ocean University,
2 Pei Ning Road, Keelung, Taiwan, 20224
(email: lolnotdog@gmail.com)*

MICHELLE CHIA-TZU LIN

*Guei-Shan Junior High School, Taiwan, 6F, No.6, Chen-Goa E.R.,
Taoyuan City, Taiwan
(email: miche0927@yahoo.com.tw¹)*

Abstract

This study aims to explore the impact of different captions on second language (L2) learning in a computer-assisted multimedia context. A quasi-experimental design was adopted, and a total of thirty-two eighth graders selected from a junior high school joined the study. They were systematically assigned into four groups based on their proficiency in English; these groups were shown animations with English narration and one of the following types of caption: no captions (M1), Chinese captions (M2), English captions (M3), and Chinese plus English captions (M4). A multimedia English learning program was conducted; the learning content involved two scientific articles presented on a computer. To track the learning process, data on oral repetition were collected after each sentence or scene was played. A post-test evaluation and a semi-structured interview were conducted immediately after viewing. The results show that the effect of different captions in multimedia L2 learning with respect to vocabulary acquisition and reading comprehension depend on students' L2 proficiency. With English and Chinese + English captions, learners with low proficiency performed better in learning English relative to those who did not have such captions. Students relied on graphics and animation as an important tool for understanding English sentences.

Keywords: animation, captions, multimedia, second language learning

1 Introduction

Using computer technology as a multimedia learning system and applying it to L2 learning has become an important trend. Since Paivio put forward his Dual-Coding Theory in 1971, there has been much important follow-up research (Moreno & Mayer, 1999, 2000, 2002; Paivio, 1990). Mayer's (2001) generative theory is one of

¹ Corresponding author

the most important. In terms of L2 learning, it is worth investigating three major processes – selecting, organising, and integrating – when learners are presented with visual and verbal information, such as illustrations and text. Especially important is investigating in detail whether the captions in animations play a supporting or hindering role in L2 learning. Some research shows that using captions can improve a learner's acquisition of an L2 language through the combination of visual, audio, and written information, allowing the viewer to select input from images, voice-overs, and captions simultaneously to further organise and integrate them. This increases the learner's knowledge, recall, and handling of the language material and, hence, effectively increases his acquisition of the foreign language and has a positive influence on the learner's intake of language material, interest in learning, and confidence (Koolstra & Beentjes, 1999; Stewart, 2004; Taylor, 2005). Even as multimedia learning resources are continually being updated in the hope of enhancing learning effectiveness, research on cognitive burdens raises doubts about the effectiveness of presenting information in a multimedia format. Multimedia-related research has not led to any unifying conclusion (Sweller, 1989, 1994; Sweller & Chandler, 1994; Mayer & Moreno, 1998). Research on the topic of whether captions support or hinder learning, in particular, is still inconclusive.

Although different individuals have different preferences when it comes to sensory input (Yuen, 1991), multimedia materials containing both visual and audio input may fit different learners' needs. Some concerns include the following. Will the learner focus only on the captions and forget to listen? Is the use of native (e.g., Chinese) captions helpful for L2 learning, or is it distracting? Furthermore, when being presented with an animation, which includes captions and voice-overs, the learner has to focus on two types of visual information simultaneously (i.e., captions and animations). Will learners try to pay attention to everything all the time and thus be distracted? Or will the concurrent input of visual (texts, images, and animations) and audio (voice-overs) information result in an excessive cognitive burden?

The existing research on using foreign language videos or animations with captions in the foreign language mainly examines the impact that different ways of presenting the captions has on the learner's listening, vocabulary acquisition, and reading comprehension. The following issues related to viewing different captions during animation scenes are worth testing and require close observation to verify their effects on students' understanding: (1) a sentence-by-sentence, scene-by-scene investigation of the L2 learning process and (2) investigating the method of sentence-by-sentence learning in combination with learners observing different types of scenes and recording. Hence, the purpose of this study is to examine the impact of computer animation with different forms of captions (no captions, Chinese captions, English captions, and Chinese + English captions) on students' vocabulary acquisition and sentence comprehension.

2 Related works in the field

2.1 Generative theory of multimedia learning

Mayer and Sims (1994) state that one of the most important functions of instructional materials is to assist students in constructing referential connections between

two forms of mental representation: the verbal representational system and the visual representational system. These referential connections are most easily built when both verbal and visual materials are presented concurrently. Their experiments found that predominantly spatial learners could devote more cognitive resources to building referential connections between visual and verbal representations of the presented material (*op. cit.*).

Mayer's (2001) generative theory of multimedia learning posits that learners engage in three major processes – selecting, organising, and integrating – when they are presented with visual and verbal information, such as illustrations and text. When first presented with a text, the learner must select the relevant words to be retained as a text base in the verbal working memory; when first presented with illustrations, the learner must select relevant images to be retained as an image base in the visual working memory. The learner must then organise the text base into a coherent verbal representation and the image base into a coherent visual representation. The last phase of the generative theory is that the learner must integrate the verbal and visual representations by making one-to-one connections between features of the two representations. According to this theory, meaningful learning is enhanced when a learner can construct and coordinate visual and verbal representations of the same material (Plass, Chun, Mayer, & Leutner, 1998).

Chun & Plass (1996b) tested how reading comprehension can be facilitated through the application of multimedia tools in a language learning program. In their study, 160 university students were enrolled in a second-year German language program, which used the multimedia application CyberBuch. The results indicated that a dynamic, visually advanced multimedia tool did facilitate overall comprehension and that annotating individual vocabulary items with both visual and verbal information was more useful than providing only verbal information. Their findings support the dual-coding theory and its extension to multimedia learning; they also emphasise the importance of both visual and verbal information as reinforcement in both top-down (knowledge-based strategy) and bottom-up (text-based strategy) processing in reading in a foreign language.

Mayer & Moreno (1998) tried to test a straightforward prediction of a dual-processing theory of working memory. Their assumptions of dual-processing theory are as follows: (1) working memory includes an auditory working memory and a visual working memory, which are analogous to the phonological loop and visual-spatial sketch pad, respectively; (2) each working memory has a limited capacity; (3) meaningful learning occurs when a learner retains relevant information in each memory store, organises the information in each memory store into a coherent representation, and makes connections between corresponding representations in each memory store; and (4) connections can be made only if corresponding pictorial and verbal information are in working memory at the same time. Mayer and Moreno conducted two experiments to determine whether there was evidence for dual processing systems in working memory. They found that, after viewing these two animations, when describing the major steps in the process of lightning forming or the operation of a car's braking system, learners who had had concurrent narration outperformed those who had had concurrent on-screen text. They concluded that multimedia learners can integrate words and pictures more easily when the words are presented auditorily, rather than visually.

Plass *et al.* (1998) extend the generative theory of multimedia second-language learning. They claim that two aspects of language learning are targeted, namely the learning of individual vocabulary items and the overall comprehension of text reading. The application of the generative theory of multimedia learning to vocabulary learning suggests, first of all, that learners of a second language have two separate verbal systems and a common imagery system. Second, in addition to translations of words linking the two verbal systems, storage in the second verbal system has an additive effect on learning. Words that are coded dually in two modes (verbally and with pictures) would be learned better than those coded only verbally. In contrast to vocabulary learning, which involves rote learning on a low level, reading comprehension is a constructive process that involves the construction of meaning on a higher level. According to the generative theories of comprehension, when reading a text, learners must build referential connections in working memory between the mental representations of ideas or propositions that have been presented in different modes. Comprehension occurs when these connections are stored in long-term memory, but storage may be hindered if learners are not able to build long-term memories. We conducted an experiment to determine the visual and verbal preferences in a second-language multimedia learning environment. The participants were 103 English-speaking college students who were enrolled in second-year German language courses. The students were asked to read a 762-word German language story presented by a computer program. As a tool to assist learning key words in the story, students could choose to view a translation on the screen in English (i.e., verbal annotation) or view a picture or video clip representing the word (i.e., visual annotation), or both. The researchers observed that students were able to remember word translations better when they had selected both visual and verbal annotations during learning than when they selected only one or no annotation; in addition, students comprehended the story better when they had the opportunity to receive their preferred mode of annotation. These findings are consistent with the generative theory of multimedia learning, which assumes that learners actively select relevant verbal and visual information, organise the information into coherent mental representations, and integrate these newly constructed visual and verbal representations with one another.

2.2 Literature review on captions

As discussed by Markham (1999), subtitles refer to “on-screen text in the native language combined with the second language soundtrack”, and captions refer to “on-screen text in the second-language combined with the second-language soundtrack” (*op. cit.*: 321). The literature review here is presented in two parts, which discuss captions and subtitles separately. Reviewing the studies of subtitles and captions can provide us with an understanding of their independent effects on the first language (L1) and the second language (L2).

Some subtitle studies provide evidence that subtitles are a useful tool for learning a second language (Borràs & Lafayette, 1994; Koolstra & Beentjes, 1999). Borràs and Lafayette (1994) found that multimedia presentations including English subtitles facilitated better general comprehension of material by native English speaking university students who studied French as a foreign language. Investigating subtitled

television programs viewed by Dutch children in grades 4 and 6, Koolstra and Beentjes (1999) designed three experimental conditions: (1) watching an English television program with Dutch subtitles; (2) watching the same English program without Dutch subtitles; (3) watching a Dutch television program (as a control group). Their results showed that the subtitled group significantly outperformed the other two groups in vocabulary acquisition. Furthermore, the results showed that there was no interaction effect found between conditions. Koolstra and Beentjes (1999) conclude that both sixth and fourth grade Dutch children learned English words from a subtitled English television program. The results of these studies showed that L1 on-screen text might contribute to L2 learning when subtitles are presented in the L1.

The positive effects of captions on L2 learning have been documented in earlier research. Captioned video material improves L2 learners' reading vocabulary knowledge (Neuman & Koskinen, 1992), listening word recognition (Markham, 1999) and reading and listening comprehension (Garza, 1991; Markham, 1989). In these studies, video materials were presented with on-screen text in the second language. In general, these studies provide evidence that L2 captions promote L2 learning. However, captioning may not be useful for enhancing all L2 learning. Taylor (2005) indicates that the length of time a learner has been studying the L2 probably has an impact on the effects of captions. By testing 85 beginner students of Spanish, Taylor (2005) found that there were no significant differences between the captioning and no-captioning groups on either free recall of, or multiple-choice tests on, the information from the video. The conclusion from Taylor's study is that "captioning might not be as effective for enhancing beginning learners' comprehension as it [was] for more experienced learners" (*op. cit.*: 426). From Taylor's study, we see that the learner's L2 proficiency may play a role in determining the effects of captions. In this study, the learner's proficiency in English was treated as an independent variable.

Given the positive effects of L1 and L2 on-screen text, another research path focuses on investigating which type of on-screen text better promoted L2 learning. Using L1 and L2 as on-screen text, Markham, Peter, and McCarthy (2001) examined the effects of using L1 (English) captions, L2 (Spanish) captions, and no captions with an L2 (Spanish) soundtrack on university-level Spanish as a Foreign Language students' comprehension of DVD material. The results of a written summary and a multiple-choice vocabulary test showed that the group with English captions significantly outperformed the group with Spanish captions, which in turn, significantly outperformed the group without any captions. As shown in Markham *et al.* (2001), L1 captions yielded better outcomes than L2 captions.

There are some empirical gaps in what has been tested. First, the languages investigated were in the same language family. Second, in the studies reviewed, only one language with captions was used in an experimental condition. Third, except in the studies by Taylor (2005) and Markham (1989), few studies examined the effects of captions by considering learners' target language proficiency. In contrast to earlier studies, the present study uses two languages from two different language families with captions and presents them simultaneously on the screen to examine whether proficient and less proficient learners were significantly different in their L2 learning.

3 Method

To compare the effects of different captions in L2 multimedia learning, we adopted a quasi-experimental method. A multimedia reading program for English learning was designed entirely by the first author.² Qualitative and quantitative data were collected as the basis of the analysis.

3.1 *The participants*

A total of 32 fourteen-year-old students from five classes were selected from a cohort of eight classes at a public junior high school in Taiwan. At the junior high school stage in Taiwan, students have experience in learning English from elementary school level. The participants were divided into two proficiency groups: sixteen students with high proficiency in English and sixteen students with low proficiency in English. Their proficiency in English was measured by scores from achievement tests, which had been conducted in the previous semester. The selection criteria for the two groups were as follows:

- 1) Students with high English competence had average marks that were above 80 in the mid-term and final examinations in grade seven.
- 2) Students with low English competence had average marks of approximately 60 in the same tests in grade seven.

Students differing in their proficiency in English were selected and assigned to each of the four captioning groups based on their performance level and the male-to-female ratio. The four groups included no captions (M1), Chinese captions (M2), English captions (M3), and Chinese + English captions (M4). There were eight students in each group, four with high marks and four with low marks; four males and four females.

3.2 *Study tools*

A multimedia reading program for testing and observations was designed by the first author and programmed in Action Script.

3.2.1 The multimedia reading program. This multimedia system has two lessons aimed at the two different student levels described in section 3.1: the first is “Sleeping animals” (Fig. 1), and the second is “Snails” (Fig. 2). Four types of learning modalities with different captions were made from the two articles using Flash software.

The article in the first lesson is called “Sleeping animals” (O’Neil, 1999b). The main scenes are divided into four learning units, which are about a lizard, a snake, a bat, and people, respectively. Descriptions of the habitat and mode of sleeping for the first three learning units are provided. The sentence pattern is simple, and there is a high degree of similarity.

² Funding for this work was provided by the National Science Council in 2005–2006. Project no. 94-2413-H-019-001, Utilizing dual-coding theory in computer assisted English learning.



Figure 1. scene from “Sleeping animals”



Figure 2. scene from “Snails”

The first scene is about a lizard and contains three sentences, which are designed as follows:

- 1) *This is a lizard.*
- 2) *It can sleep on a branch.*
- 3) *It can sleep hanging on.*

The second scene is about a snake and proceeds similarly:

This is a snake...

The article in the second unit is called “Snails” (O’Neil, 1999a). The main scenes can be divided into two acts about the snail’s organs and its way of life. Compared to the first unit, this unit has more complex sentence structures, and there is no repetition in the sentence patterns.

The first scene is about the snail's organs and contains three sentences that are designed as follows:

- 1) *Snails are animals with shells on their backs.*
- 2) *Their eyes are on the end of their feelers.*
- 3) *The breathing hole is under their shell.*

The second scene is about the snail's way of life and includes sentences such as the following:

- 1) *This snail makes slimy mucus...*

3.2.2 Instruments of measurement. The research team designed a checklist for recording observations (Appendix 1) by asking questions such as "Did you understand the sentence just then?"; "Did you hear the sentence or did you read it?"; "Can you say the sentence in Chinese?" "What did you see in the image just then?" A post-test evaluation included eight questions on vocabulary recognition, four questions on vocabulary use, and five questions on sentence comprehension (Appendix 2). Lastly, an interview was conducted with students. These three research tools form the basis of the analysis for the experiment and have been evaluated by five professionals.

3.3 Study procedure

3.3.1 Pilot study. After the research tools were developed, the researcher (the second author) who served in a junior high school selected four grade 8 students (two with high and two with low proficiency in English) to conduct the pilot study. This was to ensure that the experiment would run smoothly and to increase the validity of the test tools.

3.3.2 Formal survey. Before the formal survey, the researcher gathered the 32 study participants to explain the procedures of this study. They were told how to repeat data and interact with the researcher orally during the viewing of the animations. This was done so that they would have some conceptual knowledge of the retelling method. After this, the researcher allowed the students to watch sample animations in the multimedia reading program for English learning, letting them become familiar with the system and ensuring the study participants had the ability to repeat data and comprehend during viewing.

The experiment consisted of one-on-one learning on the computer by the student as the researcher observed the student's learning process, using the oral repeating method to obtain an idea of how the student was faring with the viewing and listening. After observing the student, an interview was conducted to collect information about the subject's vocabulary learning, lesson comprehension, focus and attitude toward learning while viewing the animated multimedia learning system. The experiment lasted for five weeks.

During the experiment, students were shown the animations in the multimedia reading system for English learning. Once the learning system was started, the screen was paused and shut off after each scene and sentence so that the students could answer the researcher's questions orally. The researcher then recorded each answer

in a table. If the students were able to answer the question fully, then the number of sentences was counted; if the students were not able to answer the question fully, then the correct vocabulary items used were ticked to make a tally of the total number of words correctly answered. Then, the next scene in the learning system was played. After viewing the first lesson, the participants completed a post-test evaluation. After viewing both lessons, an interview was conducted to obtain the students' understanding of the two lessons. Each day, two students completed the experiment. It lasted for about two hours in total, with thirty minutes for each lesson.

3.4 Data analysis

Quantitative and qualitative data were collected in this study. For the quantitative data, SPSS for Windows was used to carry out the statistical analysis and processing. This included descriptions of the sample, frequency distribution, and percentage analysis for each variable, means, standard deviations, and three-way ANOVA. For the qualitative data, the researcher wrote out the interview verbatim to complement the parts that were unclear or needed further exploration in the quantitative data about the students' understanding and attitudes. Then an analysis was performed using the three qualitative analytic phases developed by Humberman and Miles (1994): data reduction, conclusion drawing, and verification.

4 Results

In total, 32 students completed the post-test evaluation after watching the two lessons, and 32 interviews and checklists were collected. Below is a descriptive analysis of the learning achievements and attitudes after viewing the animations with different captions.

4.1 Post-test results

The post-test scores for the first lesson are shown in Table 1.

The more advanced students fared better when the animation was not accompanied with captions. They had a mean score of 38.75 ($SD = 2.5$) out of a possible 40 in vocabulary recognition; a mean of 19.5 ($SD = 9$) out of 30 in vocabulary use; a mean of 22.5 ($SD = 7.55$) out of 30 in sentence comprehension. In total, the more advanced students scored a mean score of 80.75 ($SD = 16.32$) for the animation with no captions. The scores in the other three groups are presented in the same format. The post-test scores for the second lesson are shown in Table 2.

A three-way analysis on variance was performed (Table 3) based on (1) student proficiency in English, (2) gender, and (3) animation caption types. There were significant differences (under 0.05) in the post-test scores for students of different proficiencies ($F = 19.30$, p -value of 0.00), but no significant differences among different caption types and between genders.

4.2 Analysis of retelling records

The results were totalled from the first four questions from the two animation lessons for a total of 64 oral repetitions. The statistical results of the first four responses to the four types of captions are shown in Tables 4 and 5.

Table 1 Post-test results of animation one

Student proficiency	Captions	Vocabulary recognition (S = 40)			Vocabulary application (S = 30)		
		students	mean	sd	students	mean	sd
High	None	4	38.75	2.50	4	19.50	9.00
	Chinese	4	38.75	2.50	4	16.50	3.00
	English	4	38.75	2.50	4	21.00	14.28
	C + E	4	35.00	0.00	4	15.00	12.37
Low	None	4	32.50	9.57	4	15.00	12.49
	Chinese	4	32.50	9.57	4	12.75	1.50
	English	4	36.25	2.50	4	15.00	10.39
	C + E	4	36.25	2.50	4	15.00	14.28

Student proficiency	Captions	Sentence comprehension (S = 30)			Total (S = 100)		
		students	mean	sd	students	mean	sd
High	None	4	22.50	7.55	4	80.75	16.32
	Chinese	4	18.00	12.96	4	73.25	12.58
	English	4	20.25	8.26	4	75.50	25.85
	C + E	4	21.00	3.46	4	75.50	15.78
Low	None	4	16.50	7.55	4	64.00	12.44
	Chinese	4	13.50	5.75	4	58.00	10.71
	English	4	22.50	3.00	4	74.00	10.99
	C + E	4	22.50	3.00	4	71.25	17.33

Then, based on what the students paid attention to and saw on the screen during the viewing, a percentage of the number of correct repeating answers were calculated, and the statistical result using ANOVA is as shown in Table 6. The results tallying the objects the students noticed in the two lessons are shown in Tables 7 and 8.

To sum up the statistical results, the learning experiment showed the following: First, students of low proficiency showed, in the learning of certain sentence structures, a better level of performance with both English captions and Chinese + English captions than without any captions. Second, the participants showed the highest level of comprehension of sentences and vocabulary with Chinese captions; the next level of comprehension was with no captions. The highest level of comprehension in reading was with English captions; the highest level of comprehension in both reading and listening was with English captions. Third, in the test asking students to “translate into Chinese”, the group with Chinese captions performed the best, followed by the group with Chinese + English captions; the lowest-scoring group was the one with English captions. Fourth, in the test asking students to “repeat in English”, the results of the low proficiency group showed that the score of those with Chinese + English captions was better than that of those with no captions in the

Table 2 *Post-test results of animation two*

Student proficiency	Captions	Vocabulary recognition (S = 40)			Vocabulary application (S = 30)		
		students	mean	sd	students	mean	sd
High	None	4	37.50	5.00	4	20.75	7.23
	Chinese	4	37.50	2.89	4	21.00	9.27
	English	4	37.50	2.89	4	20.75	7.23
	C + E	4	33.75	6.29	4	22.75	6.13
Low	None	4	28.75	9.47	4	9.75	9.61
	Chinese	4	32.50	5.00	4	15.25	10.84
	English	4	30.00	9.13	4	4.00	4.62
	C + E	4	31.25	11.09	4	19.00	12.70

Student proficiency	Captions	Sentence comprehension (S = 30)			Total (S = 100)		
		students	mean	sd	students	mean	sd
High	None	4	21.00	3.46	4	79.25	11.03
	Chinese	4	25.50	5.75	4	84.00	17.03
	English	4	19.50	3.00	4	70.25	6.95
	C + E	4	27.00	3.46	4	83.50	9.40
Low	None	4	18.00	6.93	4	56.50	23.62
	Chinese	4	24.00	4.90	4	71.75	20.01
	English	4	16.5	5.75	4	50.50	14.20
	C + E	4	21.00	6.00	4	75.75	19.17

Table 3 *Three-way analysis on variance of post-test*

Type	ANOVA				
	Σx^2	MS	df	F value	p-value
English proficiency	9418.78	9418.78	1	19.30	.000***
Gender	657.03	657.03	1	1.34	.263
Caption	762.844	254.28	3	.52	.674

*** $P = .000$

second lesson. Fifth, in the test asking students “what they noticed” using the oral repeating method, the results for the high proficiency group showed that Chinese + English captions were more helpful than the other three forms of captions.

Additionally, the number of times images were remembered after viewing was 903 (71.46%) in the second lesson, higher than the 679 (43.42%) in the first lesson. This shows that the images in the second lesson more easily left an impression on the students. Turning to the content of the animations, the first one, “Sleeping animals”,

Table 4 Tally of the number of sentence ($S_n = 152$) and words ($W_n = 936$) comprehended and correctly repeated in English in thinking out loud test

	No captions		Chinese captions		English captions		Chinese + English captions	
	S_n	W_n	S_n	W_n	S_n	W_n	S_n	W_n
	Comprehension	47	567	60	369	79	273	87
Repeated	53	496	60	330	71	245	68	268

$S_n = 152$ (total for 19 sentences in two animations viewed by 8 people).

$W_n = 936$ (19 sentences in two animations).

Table 5 Tally of the mode of sentence comprehension and number of sentences correctly repeated in Chinese in thinking out loud test

Captions	Mode of comprehension			Repeat in Chinese	
	Hearing	Reading	H + R	Able	Unable
None	96	0	56	85	67
Chinese	104	0	48	141	11
English	17	52	83	80	72
C + E	23	18	111	137	15

$S_n = 52$ (total for 19 sentences in two animations viewed by 8 people).

had more scenes with complex images. Therefore, the students were able to retain only 43.42% of the total images, and these were mostly conspicuous objects in the background that drew the students' attention. The main characters, such as the snake and the lizard, did not leave a stronger impression than these background objects.

The "Snail" animation used the snail as the main character and had only two scenes. Throughout the animation the scenes were simpler. The moon in the background of the last scene was the most conspicuous, but the students were able to recall most of the appearances of the snail.

4.3 Interviews and analysis of observation results

Qualitative data such as interview and observation results can be summarised with the following key points.

4.3.1 While viewing animations, the animated images draw the most attention, followed by captions, and lastly voice-overs. The researcher asked the study participants in the groups with captions the following question: "What did you notice as you watched the animations?" All of the students said they saw the animations; 90% of the participants paid attention to the captions; only 75% paid attention to the voice-overs.

Table 6 Summary of level of comprehension in thinking out loud test

Content	Results		
1. Did you understand the question?	Animation one	Between high proficiency group (not significant) Between low proficiency group (F = 6.005, p < .05) (M3 > M1) (M4 > 1)	Between high proficiency group (not significant) Between low proficiency group (not significant)
2. Did you hear it or read it?	Hearing Reading H + R	Mean (M2 > M1 > M4 > M3) Mean (M3 > M4; M3, M4 > M2, M1) Mean (M4 > M3 > M1 > M2)	
3. Can you say the sentence in Chinese?	Mean (M2 > M4 > M1 > M3)		
4. Can you repeat the sentence in English?	Animation one	Between high proficiency group (not significant)	Between high proficiency group (not significant)
	Animation two	Between low proficiency group (not significant)	Between low proficiency group (not significant) (F = 3.480, p < .05) (M4 > M1)
5. What images did you see?	Animation one	High proficiency group (not significant) (F = 11.232, p < .01) (M4 > M1) (M4 > M2) (M4 > M3)	Between high proficiency group (not significant)
	Animation two	Between low proficiency group (not significant)	Between low proficiency group (not significant)

(M1: no captions, M2: Chinese captions, M3: English captions, M4: Chinese + English).

Table 7 *Animation one – number of images viewed*

Code	English name	# times shown	# time observed	Percentage observed
1. F	flower	256	176	68.8%
2. G	girl	64	34	53.1%
3. USB	upside-down bat	64	32	50.0%
4. YM	yellow moon	160	78	48.8%
5. P	people	64	30	46.9%
6. ST	stuck-out tongue	64	26	40.6%
7. AB	animation bat	160	62	38.8%
8. BS	blue sky	288	109	37.8%
9. B	branch	352	97	27.5%
10. AL	animation lizard	160	35	21.9%
Total: 43.42%				

Table 8 *Animation two – number of images viewed*

Code	English name	# times shown	# times observed	Percentage observed
1.YM	yellow moon	32	32	100%
2.AS	animation snail	224	179	79.9%
3.SM	slimy mucus	32	23	71.9%
4.BS	blue sky	160	112	70.0%
5.RC	red circle	64	44	68.8%
5.ST	silver trail	32	22	68.8%
7.F	flower	224	152	67.9%
8.RS	real snail	128	83	64.8%
9.T	tree	224	139	62.1%
10.C	cloud	192	116	60.4%
Total: 71.46%				

The students also expressed the degree of attention they paid to the animation images in the interview. For example:

“Initially, I couldn’t really understand, and then I saw the images, which made it easier to understand.” (G2-2, interviewed 2008/12/09)

“Because it is easier to understand with pictures, I had a stronger impression of the images than of the sounds.” (G1-4, interviewed 2008/12/10)

“It is easier to understand with pictures. I can maintain recall of the images for very long.” (G3-5, interviewed 2008/12/11)

4.3.2 More than half of the students in the no captions group wanted captions to go with the animations. The researcher asked the study participants in the no captions

group the following question: “When I played the first learning style (no captions), did you wish there were captions?” 66% of students said they wanted English captions, and 34% of the students wanted Chinese + English captions together. For example, student G1 said, “*The animations I saw didn’t have captions. I understood the first one a little, but I was completely lost during the second one. If you had shown me one with English and Chinese caption, then I would have understood.*” (G1-3, interviewed 2008/12/12).

4.3.3 *Regarding the processing of information by students in the English captions group, participants readily had access to corresponding Chinese words after reading the English words.* The researcher asked the participants in the English captions group the following question: “*When I played the third learning style (English captions), did the Chinese meanings of the English words emerge come to my mind?*” Almost all replied that they did. Some of them even wished there had been Chinese captions.

4.3.4 *The participants in the Chinese + English captions group thought that the Chinese captions helped them the most to understand the content of the stories, followed by the animations, the voice-overs, and lastly the English captions.* The researcher asked the study participants in the Chinese or Chinese + English captions groups the following question: “*With voice-over listening, Chinese captions, English captions and animations, which helped you the most to understand the content of the stories?*” Almost all said that the Chinese captions helped the most in terms of understanding the content of the stories. For example:

“The animations I saw had Chinese + English captions and English voice-overs. Listening can reinforce my impression of the vocabulary or sentences. The Chinese captions allowed me to quickly understand the meaning; the English captions allowed me to catch up and corrected my pronunciations.” (G4-1, interviewed 2008/12/12)

4.3.5 *Chinese and English captions shown simultaneously with other media, such as animations and voice-overs, resulted in a cognitive burden for some students.* A total of 62.5% of the participants said that the Chinese + English captions accompanied with animations and voice-overs were overwhelming. Because both the capacity and duration of working memory are limited, some students thought that, “*The caption type (Chinese + English captions) presented was too much information at one time, and it was only shown once. In such a short time, I could only remember the important parts. I can’t remember so much...*” (G4-1, interviewed 2008/12/16).

4.3.6 *In terms of viewing the contents, most students said that they liked the animations and the content of the stories.* In the interviews, up to 93.8% of participants thought that the animations allowed for forming a deeper impression of the material; all participants thought that animations were beneficial for learning English. For example, one student said the following: “*Because the main characters in the animations were obvious, I was able to link the words to the images, which made it easy for me to remember the vocabulary, and I do not have to memorise it!*” (G1-5, interviewed 2008/12/15).

5 Discussion and conclusions

The present study was conducted to examine how captions and animation influence L2 learning at the word and sentence comprehension levels.

5.1 In L2 multimedia programs, animated images and voice-overs had more impact than captions

This study found from the post-testing scores that the students' vocabulary recognition and application did not show a significant difference between the four caption types after viewing the two animations.

Neither the presence or absence of captions nor the inclusion of Chinese and English captions together or separately resulted in any difference in cognitive understanding. The group that had voice-overs but no captions produced the same effect described by Mayer and Moreno (1998). That is, the combination of visual animation and audio voice-overs produces enough learning inputs for L2 programs.

There was no significant difference, therefore, with respect to learning across the four caption types. This is different from the results of most related research, which has found captions to be more useful and that voice-overs are more helpful than texts (Chun & Plass, 1996a; Plass *et al.*, 1998).

We postulate that the possible reasons for this difference include the following. (1) The present research adopted a sentence-by-sentence and scene-by-scene approach, with viewing, recording, further viewing, and so on. The process was slow and full of interruptions, which might have affected the students' overall understanding and thinking. (2) As Kost, Foss, and Lenzini (1999) discovered in their research on foreign language reading, "visual aids are most beneficial in terms of reading comprehension of a foreign language". We postulate, therefore, that animated images will have a more significant impact on the level of comprehension than the presence or absence of captions and caption types. (3) The participants might have selected the information they needed at the right time, though there was an abundance of information presented at the same time.

5.2 The effect of captions in L2 multimedia on reading comprehension varies, depending on the student's L2 proficiency

During the retelling process for the first lesson, with easier sentence patterns, we found that having either English captions or Chinese + English captions was more helpful than not having any captions for less proficient learners. This was also the case for the second lesson with more complex sentence structures, as the less proficient group also did better with Chinese + English captions than with no captions, when asked to repeat in English the sentences that were in the animations. This result is consistent with Koolstra and Beentjes' (1999) finding that young children could acquire elements of a foreign language by watching subtitles in video programs; the result is partially similar to Markham, Peter and McCarthy's (2001) finding that the group with English (L1) captions performed better than the no-captions or Spanish (L2) caption groups. Lin's (2005) experiment showed that the combination

of voice-overs in the L1 and captions in the target language can alleviate the cognitive burden of learners. It provided evidence of an effect with either L1 captions or narration in L2 multimedia.

This result shows that students of different proficiency levels show different responses to different caption types; this is especially evident in the less proficient group. With simple sentence structures, having Chinese + English captions or English captions paired with animations was sufficient to give the support needed for less proficient students. As for the more complex sentence structures, the less proficient group had their best rate of correct responses in recalling the sentences in the animations when having Chinese + English captions.

Seen in the light of the modality effect and generative theory (Mayer, 2001), we then postulate that a less proficient student is more likely to selectively take in material that they notice. They do not seem to be affected by the presence of too much information; therefore, with simple sentence structures, they can select the information they need when viewing animations with Chinese + English captions or English captions. Both of these caption styles allow them to understand foreign sentences. In the more complex sentence structures, there is a better outcome in repeating a sentence they have just heard only when they are given Chinese + English captions. These differences do not occur for more proficient students. As Katchen (1996) concluded, for the more advanced student, native (Chinese) captions might become a hindrance because it is easy to become dependent on them, and they can even make the student slower by distracting him.

In analysing the above two findings, it seems that there was no obvious effect on reading comprehension. There was only a small difference between students of different levels of proficiency. For more proficient students, there was no significant difference in their learning, regardless of what type of captions they saw. For the less proficient students, providing English or Chinese + English captions was helpful for comprehending simple sentence structures; with complex sentence structures, only Chinese + English captions had a positive effect on the correct repetition of the sentences.

5.3 The effect of a multimedia L2 program with captions on vocabulary acquisition is very slight

Based on post-test scores, students' vocabulary recognition and vocabulary use did not show a significant difference across the four captions types after viewing the two animations. In this study, the group with voice-overs but no captions did not perform better or worse than the groups with both voice-overs and captions. That is, the presence or absence of captions in the target or the native language have no significant effects on the outcome of vocabulary learning when learning a foreign language.

The concurrent presence of both Chinese and English captions had no effect on students' vocabulary learning. Based on the generative theory, combining Chinese captions (as supplemental visuals in Chinese) with English ones can alleviate the cognitive burden on the students. However, this was not the case in this study. In this study, all of the four learning styles had voice-overs and animations; therefore, the difference between each group was not very significant. The most important

information had already been transmitted to the working memory area through the visual animations and voice-overs, resulting in the difference between different caption types not being significant.

5.4 Animations were most beneficial when headings affected choice of imagery

When the students were asked what they noticed as they watched the animations, all of them said they saw the animation images. Up to 93.8% of participants thought that animations formed a deep impression; all participants thought that animations were beneficial in learning English.

At the beginning of each L2 lesson, headings were presented to show the title of the lesson. In the animation "Sleeping animals", the background (showing flowers and a girl) attracted the viewer's attention the most, whereas in the animation "Snails", apart from the moon in the last scene, which was the most conspicuous, the main character, the snail, was most able to attract the viewer's attention. This shows that headings, which are a key visual element in multimedia, not only deliver messages about the content, but also influence the viewer's choice of image memory.

The feedback compiled following student interviews demonstrated that the highest proportion of students liked the animations and the content of the stories, followed by those who liked the voice-overs and captions.

6 Suggestions for future research

The research has contributed in three main aspects. First, it was found teenage learners paid more attention to animations, including pictorial and auditory, than to captions in L2 multimedia learning. Second, it was found some students liked captions, either in native or foreign language, in L2 multimedia learning, despite the slight difference in effects shown by caption types. Third, it was concluded that the effect of different captions in multimedia L2 learning with respect to vocabulary acquisition and reading comprehension depends on students' L2 proficiency.

Since this study adopted the method of pausing after each scene and sentence to obtain students' responses, the whole process was time-consuming, and as a result, only 32 students were used as participants. Because the number of participants was low, this must be taken into consideration in the discussion. We used the retelling method and a subsequent interview to collect data separately. Although questions during the retelling process were avoided as much as possible, there were probably factors that interfered with the students' thought processes.

By proceeding sentence-by-sentence through the passage, a more fine-grained approach was adopted to examine the students' learning and problems during learning. Unfortunately, as a result, vocabulary, sentences, and visual elements became the main targets of the study, and information about the whole picture is missing.

Future research could adopt the use of sections of two or three meaningful sentences to conduct the retelling experiment, thereby allowing space for students to more fully comprehend the content of the articles and for more meaningful cognitive understanding. A more in-depth study could be done with more focus on

the component involving audio voice-overs and aural comprehension. In addition, the data collected via verbal reporting could be expanded so as to include the length of time spent on reading, frequency of eye movements, and other factors.

Acknowledgement

The authors would like to acknowledge the contribution of Tsai Hsin Yuan in programming the multimedia reading program used in this study.

References

- Borràs, L. and Lafayette, R. (1994) Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal*, **78**(4): 61–75.
- Chun, D. M. and Plass, J. L. (1996a) Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal*, **80**(2): 182–198.
- Chun, D. M. and Plass, J. L. (1996b) Facilitating reading comprehension with multimedia. *System*, **24**(4): 503–519.
- Garza, T. J. (1991) Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, **24**(3): 239–258.
- Humberman, A. M. and Miles, M. B. (1994) Data management and analysis methods. In: Denzin, N. K. and Lincoln, Y. S. (eds.), *Handbook of qualitative research*. London: Sage.
- Katchen, J. E. (1996) First language subtitle: help or hindrance? Paper presented at the *Annual Meeting of the Japan Association of Language Teachers*. ERIC Document Reproduction Service No. 421873.
- Koolstra, C. M. and Beentjes, J. W. J. (1999) Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology, Research and Development*, **47**(1): 51–60.
- Kost, C. R., Foss, P. and Lenzini, J. J. (1999) Textual and pictorial glosses: effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, **32**(1): 89–113.
- Lin, H. I. (2005) Developing a New Model for Captioned Video Teaching in Taiwanese Elementary Schools. *English Learning & Teaching*, **29**(4): 21–38.
- Markham, P. (1989) The effects of captioned television videotapes on the listening comprehension of beginning, intermediate, advanced ESL students. *Educational Technology*, **29**(10): 38–41.
- Markham, P. (1999) Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, **32**(3): 321–327.
- Markham, P. L., Peter, L. A. and McCarthy, T. J. (2001) The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, **34**(5): 439–445.
- Mayer, R. E. and Sims, V. K. (1994) For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, **86**(3): 389–401.
- Mayer, R. E. and Moreno, R. (1998) A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology*, **90**(2): 312–320.

- Mayer, R. E. (2001) *Multimedia Learning*. New York: Cambridge University Press.
- Moreno, R. and Mayer, R. E. (1999) Cognitive principles of multimedia learning: the role of modality and contiguity. *Journal of Educational Psychology*, **91**(2): 358–368.
- Moreno, R. and Mayer, R. E. (2000) A coherence effect in multimedia learning: the case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, **92**(1): 117–125.
- Moreno, R. and Mayer, R. E. (2002) Learning science in virtual reality environments: role of methods and media. *Journal of Educational Psychology*, **94**(3): 598–610.
- Neuman, S. B. and Koskinen, P. (1992) Captioned television as comprehensible input: effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, **27**: 95–106.
- O'Neil, S. (1999a) *Snail*. St. Leonards, Australia: Horwitz Martin Education.
- O'Neil, S. (1999b) *Sleeping Animals*. St. Leonards, Australia: Horwitz Martin Education.
- Paivio, A. (1971) *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.
- Paivio, A. (1990) *Mental representations: a dual coding approach*. New York: Oxford University Press.
- Plass, J. L., Chun, D. M., Mayer, R. E. and Leutner, D. (1998) Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, **90**(1): 25–36.
- Stewart, M. A. (2004) Gains to language learners from viewing target language closed-captioned films. *Foreign Language Annals*, **37**(3): 438–447.
- Sweller, J. (1989) Cognitive technology: some procedures for facilitating learning and problem solving in Mathematics and Science. *Journal of Educational Psychology*, **81**(4): 457–466.
- Sweller, J. (1994) Cognitive load theory learning difficulty, and instructional design. *Learning and Instruction*, **4**: 295–312.
- Sweller, J. and Chandler, P. (1994) Why some material is difficult to learn. *Cognition and Instruction*, **12**(3): 185–233.
- Taylor, G. (2005) Perceived processing strategies of students watching captioned video. *Foreign Language Annals*, **38**(1): 422–427.
- Yuen, C. Y. (1991) Multimedia: how it changes the way we teach and learn. *Electronic Learning*, **11**(1): 22–26.

Appendix I: Checklist for thinking out loud method – Unit 2

Sentence	Understanding sentence or words	Listening or reading	Speaking in Chinese	Repeating in English	Viewing the animation graphics
(1) Snails are animals with shells on their backs.	() sentence () word () none	() L () R	Y/N	Y/N	() C () BS () AS () RS () RC () BS () L () F*

*Each abbreviation represents an object in the scene. (C: cloud, BS: blue sky, AS: animation snail, RS: real snail, RC: red circle, BS: black square, L: leaf, F: food).

Appendix II: Post-viewing test 2

Class: _____ Name: _____ No.: _____

I. Vocabulary Recognition

1. (3) snail
 (1) 烏龜
 (2) 水蛇
 (3) 蝸牛
 (4) 鯊魚
 (5) 不知道

2. ...

II. Vocabulary selection and filling in the bank: (Fill in one answer in a blank)

- A. mucus B. breath C. shells D. a silver trail
 E. feed F. feelers G. move

Snails move slowly and make D. Children say that snails move slowly because of their heavy C. Maybe it is true. But one thing is sure – that their shells sometimes keep them from danger.

On the end of snails' F are two eyes. It makes them look like aliens (外星人) somewhat. You usually can find these “aliens” at night when they come out to E at that time.

III. Reading Comprehension: (Choose the best answer.)

- (1) 1. When are you more likely to find snails (比較可能)?
 (1) At night.
 (2) In the morning.
 (3) In the afternoon.
 (4) At noon.