

The effect of key-words video captions on vocabulary learning through mobile-assisted language learning

Hassan Saleh Mahdi

University of Bisha, Saudi Arabia

Video captioning is a useful tool for vocabulary learning. In the literature, video captioning has been investigated by many studies, and the results indicated that video captioning is useful to foster vocabulary learning. However, most of the previous studies have investigated the effect of full captions on vocabulary learning. In addition, most of the previous studies have utilised the computer as a tool for the study. Very few studies have used mobile devices as a tool for the experiment. Therefore, this study aims at investigating the effect of keywords videos captioning on vocabulary acquisition using mobile devices. Thirty-four Arab EFL university learners participated in this study and were randomly assigned to two groups (keyword captioned video and full captioned video). The study is an experimental one in which pre- and post-tests were administered to both groups. The results indicated that keyword captioning is a useful mode to improve learner's pronunciation. However, there was no significant difference between the two modes of captioning.

Keywords: MALL, caption, keyword caption, full caption, vocabulary acquisition, technology

Introduction

Vocabulary learning is an essential part of mastering a language. Vocabulary is considered as the building blocks of a language (Schmitt, Schmitt, & Clapham, 2001, p. 53). Without knowledge of words and their meanings, it is impossible to convey the message in a language. Although listening, speaking, reading and writing are the basic skills that language learners need to master, vocabulary is essential to all these four skills. Therefore, it is important to consider vocabulary learning in any program devoted to language learning.

Vocabulary can be presented either traditionally or technologically. Using technology in language learning is the topic of several studies. These studies have revealed a number of encouraging results which indicated that vocabulary learning with the help of technology could be more effective than through the use of traditional ways of presenting vocabulary. Implementing mobile devices in vocabulary learning has received several studies to examine its effect on vocabulary learning. The results of these studies revealed that using mobile devices in vocabulary learning is an effective tool. With the help of these technologies, vocabulary can be presented in different modes (e.g. subtitling, annotation, and captioning). Video captioning is implemented to increase vocabulary learning. There are two ways of captioning: full captions and keywords captions. The impact of video captioning on vocabulary acquisition has

been investigated by many studies. The focus of the previous studies lies on full captions. In addition, almost all of the previous studies were conducted with the help of a computer. However, as far as the author knows, there is only one study which investigated the effect of keyword captions on vocabulary learning (i.e. Yang & Chang, 2013). They proposed three modes of captions: full, keyword-only, and annotated keyword captions and investigates their contribution to the learning of reduced forms and overall listening comprehension. The results revealed that all three groups exhibited improvement while the annotated keyword caption group exhibited the best performance with the highest mean score. Yang and Chang, (2013) focused on reduced forms. However, as far as the author knowledge, word pronunciation has not yet been investigated when mobile devices are used. Therefore, this study is an attempt to fill-in this gap and investigates the effect of keyword video captioning on word pronunciation in comparison to full video captioning using mobile devices.

The study is based on Mayer (2009) five ways to reduce extraneous processing. According to Mayer (2009, p.86), extraneous processing is cognitive processing during learning that does not serve the instructional goal, such as attending to irrelevant information or trying to make up for confusing layout of the lesson. The five principles for reducing extraneous processing are coherence, signalling, redundancy, spatial contiguity, and temporal contiguity.

Vocabulary learning

There are four basic skills required to master a language (i.e. listening, speaking, reading and writing). However, language learners need enough vocabulary to master these skills. Therefore, vocabulary learning is an essential component of language learning. Wikins (1972) as cited in Zhang, Song and Burston (2011) points out that "without grammar very little can be conveyed, without vocabulary nothing at all can be conveyed". Vocabulary learning is not just knowing the meaning of the word. It involves many aspects. Nations (2001, p26) lists the aspects that are necessary to know a word. They are a form (spoken, written and word parts); meaning (form and meaning, concept and referents, and associations); and use (grammatical functions, collocations, and constraints of use).

Vocabulary learning can occur in two different ways: intentional and incidental (Nation, 2001). Intentional vocabulary learning refers to any activity aiming at committing lexical information to memory (Robinson, 2001). On the other hand, incidental vocabulary learning refers to learning from context such as from reading or listening. It is a byproduct of something else (Gass & Seinker, 2008). Both ways are required to increase the size of vocabulary that language learners need to communicate in a language.

Moreover, vocabulary learning can occur in two environments (i.e. technology-based and traditional environment). Technology-based refers to the use of new technologies for vocabulary learning such as TV, computer, personal digital assistants (PDA) and mobile devices. The other environment is the traditional ways such as word cards, dictionaries, and word lists. With the help of technology-based environments, vocabulary can be presented in different modes such as video captioning, subtitling, and annotations.

Video captions and vocabulary learning

Video captioning is one of the modes that technology can provide for vocabulary learning. It is defined by Danan (2004, p.232) as "on-screen text in a given language combined with a soundtrack in the same language". Language learners have difficulty in decoding the speech of native speakers of the target language. The use of video captioning is an effective tool for language learners to decode the speech of native speakers presented in videos. Captions help learners link the written words to their actual speech. Several studies were conducted to investigate the effects of video captioning on language learning (e.g. Brett, 1995; Garza, 1991; Hsu, 1994; Huang & Eskey, 1999-2000). The results indicated that video captioning increases general comprehension and helps language learners understand better. It provides visual, contextual and non-verbal input for language learners (Brett, 1995).

Regarding vocabulary learning, several studies were conducted to find out whether video captioning is useful for vocabulary learning (e.g. Aldera & Mohsen, 2013; Hsu, Hwan, Chang & Chang, 2013; Stewart & Pertusa, 2004; Sydorenko, 2010; Yuksel & Tanriverdi, 2009). The results indicated that learners exposed to captioned videos outperformed non-captioned video students. Captioned videos foster vocabulary learning because it may contribute to a conscious focusing on the form and it encourages attention which is an essential for language learning (Vanderplank, 1990). Video captioning is also effective in word recognition and recall (Perez, Noortgate & Desmet, 2013). From the previous studies, there is a consensus that implementing captioned videos help language learning and lead to better comprehension and vocabulary learning.

Mobile devices and vocabulary learning

The advent of mobile devices has influenced the life of millions of people around the globe. Currently, there are more than 7 billion mobile phones subscriptions worldwide. Mobile devices can be integrated into education as a tool to facilitate language learning. They are equipped with different input modalities such as MP3 players, YouTube, etc. which can be used to present a target language in an effective way.

To find out the effectiveness of mobile devices on language learning, several studies were conducted (see Burston, 2013 for more information). Burston (2012, p.16) concludes that "the learning outcomes of MALL implementations are unquestionably positive in nearly 80% of the cases". Mobile devices can be used to facilitate vocabulary learning. Many studies were conducted to examine the benefits of mobile devices on vocabulary learning (Chen, Hsieh & Kinshuk, 2008; Browne & Gulligan, 2008; Kennedy & Levy, 2008; Lu, 2008; Saran, Seferoglu & Cagitay, 2012, Stockwell, 2007; Thornton & Houser, 2005; Wong & Looi, 2010; Zhang, Song & Burston, 2011;). The results of these studies revealed that using mobile devices is a useful tool for vocabulary learning.

Captioned videos through mobile devices and vocabulary learning

Mobile devices have great potential to provide additional activities for language learning. The current types of mobile devices can present the target language in different modes. In this regard, captioned videos can be displayed with the help of mobile devices to foster vocabulary learning. Mobile devices are equipped with many

programs that can present the videos clearly and in an interesting way. Researchers have investigated the effects captioned videos on vocabulary learning with the help of a computer (e.g. Aldera & Mohsen, 2013; Markham, 1999; Stewart & Pertusa, 2004; Sydorenko, 2010; Yuksel & Tanriverda, 2009). These studies have found that captioned videos are useful in fostering vocabulary learning. However, using mobile devices to present captioned videos for vocabulary learning has not fully investigated. Very few studies have dealt with this topic. For example, Hsu et al. (2013) investigated the effects of different display modes on vocabulary acquisition. They compared the effect of the full caption, target-word caption, and non-caption on vocabulary acquisition. Students used Personal Digital Assistants (PDAs) to play the videos. Pre- and post-tests were used. Results indicated that both modes (full caption and target-word captions) were good for vocabulary acquisition. Both groups outperformed non-captioned group. Therefore, this study is significant to find out the effect of mobile devices on vocabulary learning through keyword captioning. The study will attempt to answer the following question: Which mode of captioning is more useful for word pronunciation (keyword caption or full word caption)?

Reducing Extraneous Processing in Multimedia Learning

Using multimedia aids in language learning is a useful tool. However, there are some cautions to be considered. One of these cautions is the extraneous processing in learning. Mayer (2009, p.85) defines extraneous processing overload as "a situation in which the cognitive processing of extraneous material in the lesson is so demanding that there is little or no remaining cognitive capacity to engage in essential or generative processing. Extraneous processing overload is likely to occur when the lesson contains attention grabbing extraneous material or when the lesson is designed in a confusing way. This leads to an extraneous material which is defined by Mayer (2009, p.85) as "information from the lesson that is not needed to achieve the instructional goal."

Extraneous processing in multimedia learning can be overcome. Mayer (2009, p.86) explored five principles for reducing extraneous processing in multimedia learning. He summarised them as follows:

1. Coherence techniques involve deleting extraneous words, sounds, and pictures from a multimedia lesson.
2. Signalling involves highlighting the essential words and pictures in a multimedia lesson.
3. Redundancy techniques involve removing redundant captions from narrated animations.
4. Spatial contiguity involves placing words next to corresponding graphics on the screen or page.
5. Temporal contiguity involves presenting corresponding narration and graphics simultaneously.

Based on Mayer's principles for reducing extraneous processing in multimedia learning, one can claim that keyword captioning is a useful tool for vocabulary learning because it fulfils the five principles proposed by Mayer (2009). First, coherence techniques involve deleting extraneous words. Only keywords are shown to help students capture their pronunciation as they occur. Second, signalling involves

highlighting the essential words. Third, redundancy techniques involve removing redundant other words that are not intended for the current lesson. Fourth, spatial contiguity involves placing written words next to corresponding sound as shown by the movement of the speaker's mouth. Fifth, temporal contiguity involves presenting keywords and their written form simultaneously. To address these issues, this study seeks to answer the following questions:

1. Does keyword captioning foster vocabulary acquisition among EFL learners?
2. What is the best mode for learning word pronunciation (keyword caption or full caption)?

Method

Previous research has shed light on how various modes and modalities may contribute to vocabulary learning (e.g. subtitling, captioning, and annotation). To investigate the effect of captions on L2 pronunciations, two levels of captions were established: 1) full-text captions and 2) keyword captions. The study was designed to explore the following hypothesis: captioning of the authentic video has a positive effect on L2 pronunciation. To examine this hypothesis, an empirical study was conducted based on an experimental design using a quantitative approach.

Two video clips for all input conditions—full-text captions, and keyword captions—were chosen from YouTube. The captions appeared in the centre of the video clips. In the case of full captions, the captions were divided into phrases and shown simultaneously with the spoken utterances. The keyword captions were set at a default of two seconds, which is enough for the participants to read the word and listen to its pronunciation. The study was carried out in the first semester in the academic year 2017 at University of Bisha, Saudi Arabia. The participants (N = 34) were randomly distributed among no captions, full-text captions, and keyword captions groups.

Participants

The participants in this study were 34 native speakers of Arabic enrolled in the English Department at Bisha University, Saudi Arabia. The students' ages ranged between 19 and 21. Due to University rules and the cultural values that support gender segregation in classes, all of the participants were male students. The participants were randomly assigned to one of the two groups: 17 participants in each group.

Study design

The study employed a between-subjects design in which participant viewed videos under one of two modes: keyword captions and full captions. The groups were asked to play the videos and then complete the vocabulary test. In other words, the independent variable was the L2 learning mode, which consisted of two modes (i.e., keyword captions and full captions). The dependent variable (word pronunciation) was measured by a pronunciation test. Word pronunciation was measured prior to the intervention (i.e., pretest), and immediately after the intervention (i.e., posttest).

A pre-test and a post-test were administered, comprising the same content. The test consisted of 40 words selected from the videos sent to the mobiles of the participants. Participants were required to pronounce the selected words. Each word pronounced correctly was given one point. The total number of correct words for each group was analysed.

Materials Used and Target Vocabulary

The two groups of participants all watched the same video clips, and only the modes of captions were different. It was assumed that these clips were appropriate to the participants in terms of both language level and interest. The treatment involved students in all groups watching the videos three times. The duration of each clip is about 6 minutes which is considered an appropriate time span for learning. Forty target words from the clips were chosen based on the proficiency level and background of the students. The main concern in selecting the target words for this study was that they should be unknown to the students in order to ensure that the effectiveness was the result of the mode (full or keyword captions).

Procedures

The researcher first administered the pre-test prior to the treatments. The students were asked to read the 40 words aloud. Each student was asked to read aloud the 40 words. Then, the experiment was conducted. During the experiment, the students were assigned into two groups. Each group had a separate room. The videos were sent to their mobiles via Whatsapp. The learners were instructed to play the video using their own mobile devices and completed their task over a 30-min period. After viewing the video, the students were asked to do the post-test. They were asked to produce the words orally. Each student was asked to read aloud the 40 words and record them, then send them to the instructor. Each correct word had got 1 point. This study consisted of four steps. The steps were as follows.

1. A pre-test: Participants completed a pronunciation test of 40 words to assess their baseline knowledge regarding word pronunciation.
2. Intervention: Participants were divided into two groups, who watched the Video clips with full and keyword-only. Except for different modes of captions, the sequence and the number of the clips were all the same to ensure participants received equal input.
3. A post-test: After watching the video clips, participants took the same test again.

Data analysis

To answer the research questions, a paired sample t-test was performed. The paired sample t-test was employed to compare the pre-test and posttest scores of the two groups. The independent variable is the different groups, including the full caption, and keyword-only caption groups. The dependent variable is the post-test score of the pronunciation test.

Results and discussion

One of the objectives of this study was to examine the effectiveness of the keyword caption mode with respect to the students’ pronunciation. It was therefore first necessary to ensure that the two groups had comparable pronunciation level before beginning the treatment. The results were shown in Figure 4.1. This was determined by way of the pre-test for which the mean values and standard deviations of the scores were 13.94 and 6.28 for the first group (keyword), 13.94 and 8.97 for the second group (full caption). The results of the pre-test among the two groups do not show a significant difference ($t = 0.41 > .05$); that is, it was ascertained that the two groups of students had equivalent prior knowledge before the experiment.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 prefull	14.9412	17	8.97546	2.17687
prekey	13.9412	17	6.28958	1.52545

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 prefull - prekey	1.00000	9.86154	2.39178	-4.07034	6.07034	.418	16	.681

Fig. 4. 1 Pre-test of full and keyword captioned words

To compare the pre-test and post-test scores of the two groups, the paired-sample t-test was applied. Figures 4.2 and 4.3 show the results in terms of mean scores (M), number of participants per group (N), standard deviation (SD), and t value.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 prefull	14.9412	17	8.97546	2.17687
postfull	18.0588	17	9.49652	2.30324

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 prefull - postfull	-3.11765	3.38900	.82195	-4.86011	-1.37519	-3.793	16	.002

Fig. 4.2 Pre- and post-test of full captioned words

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 prekey	13.9412	17	6.28958	1.52545
postkey	21.3529	17	10.37992	2.51750

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 prekey - postkey	-7.41176	8.33711	2.02205	-11.69831	-3.12522	-3.665	16	.002

Fig. 4.3. Pre- and post-test of keyword captioned words

Overall, the results indicated that all the participants in the two groups showed significant improvement after the experiment. In particular, the keyword caption group demonstrated most significant improvement in the mean score by 21.35 while the full caption group showed the least increase by 18.05. However, the results of post-test for both groups indicated that there was no significant difference among both groups ($sig.355 > .05$).

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 postfull	18.0588	17	9.49652	2.30324
postkey	21.3529	17	10.37992	2.51750

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 postfull - postkey	-3.29412	14.25116	3.45641	-10.62139	4.03315	-.953	16	.355

Fig. 4.4. Post-test full and keyword captions

Generally, the improvement made on word pronunciation made by captioning (either full or keyword) was found to positively impact pronunciation. The effect size of post-test of both groups is .33 which is medium. Specifically, keyword captions were found to enhance learners' ability to pronounce words better than full captions, potentially resulting from the addition of the written word along with its pronunciation which contributed to students' pronunciation.

While past research has investigated the effect of full captions on vocabulary acquisition, few have examined keyword captions to ascertain the maximum effect of captions in pronunciation. Overall, this study has found that captions (either full or keyword) contributed to better pronunciation.

Conclusion

The current study explores how different types of captions (i.e. full and keyword captions, impact EFL students' pronunciation. Inspired by past studies, which proposed that word captioning is useful to improving vocabulary acquisition, keyword captioning was utilized to make EFL learners aware to the pronunciation of English words. Previous studies have demonstrated a positive outcome when introducing captions or subtitles to enhance vocabulary acquisition. However, few investigated the effect of keyword captioning on word pronunciation in comparison to full captioning. This study revealed that keyword captioning is a useful mode to improve learner's pronunciation. The results indicated that learners' pronunciation improved when caption is used. However, no significant difference between the two modes (keyword and full captioning). The results suggest that keyword captioning has the potential to help EFL learners to improve word pronunciation. Though the difference between the two modes is not significant, the study suggests that keyword captions may be more helpful than full captions. Further studies are needed to explore this issue in detail.

There are some limitations to the present study. First, given that this study employed a relatively small sample of participants, generalization is unlikely regarding how the two types of captions may be distinguished in effect. For future research, larger sample and a long-term treatment may provide additional evidence and expand understanding in pronunciation. Second, while participants in the study shared a homogeneous academic background in terms of the university they came from and their English proficiency level, and all are male students, for future research, participants could be grouped by alternative variables such as gender, proficiency levels, and so forth before the intervention. The study has found that keyword captions may be beneficial in terms of assisting with the learning of pronunciation. For future pedagogical design, keyword caption is recommended.

References

- Aldera, A. S. & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills, *Computers & Education, 68*, 60-75.
- Brett, P., 1995. Multimedia for listening comprehension: the design of a multimedia-based resource for developing listening skills. *System, 23(1)*, 77-85.
- Burston, J. (2012). *Mobile language learning: Getting IT to work*. In J. Burston, F. Doa, & D. Tsagari(Eds.), *Foreign language instructional technology* (pp. 81_99). Nicosia: University of Nicosia Press.
- Burston, J. (2013). Mobile-Assisted Language Learning: A selected annotated bibliography of implementation studies 1994-2012. *Language Learning & Technology, 17(3)*, 157–224.
- Chen, N. Hsieh, W.& Kinshuk, (2008). Effects of short-term memory and content representation type on mobile language learning, *Language Learning & Technology, 12(3)*, 93-113.
- Danan, M. (2004). Captioning and subtitling: undervalued language learning strategies. *Meta, 49(1)*, 67–77.
- Garza, T. J. (1991). Evaluating the use of video materials in advanced foreign language learning. *Foreign Language Annals, 24(3)*, 239–258.
- Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course*. Taylor & Francis.
- Browne, C., & Culligan, B. (2008). Combining technology and IRT testing to build student knowledge of high frequency vocabulary. *The JALT CALL Journal, 4(2)*, 3–16.
- Hsu, J. (1994). *Computer assisted language learning (CALL): The effect of ESL students' use of interactional modifications on listening comprehension*, (Unpublished doctoral dissertation). Department of Curriculum and Instruction, College of Education, Iowa State University, Ames, IA.

Hsu, C. K., Hwang, G. J., Chang, Y. T., & Chang, C. K. (2013). Effects of video caption modes on English listening comprehension and vocabulary acquisition using handheld devices. *Educational Technology & Society, 16(1)*.

Huang, H. & Eskey, D. (1999-2000). The Effects of Closed-Captioned Television on the Listening Comprehension of Intermediate English as a Second Language (ESL) Students, *Journal of Educational Technology Systems, 28(1)*, 75-96.

Kennedy, C., & Levy, M. (2008). L'italiano al telefonino: Using SMS to support beginners' language learning. *ReCALL Journal, 20(3)*: 315–330.

Lu, M. (2008). Effectiveness of vocabulary learning via mobile phone. *Journal of Computer Assisted Learning, 24(6)*: 515-525.

Markham, P., (1989). The effects of captioned television videotapes on the listening comprehension of beginning, intermediate and advanced ESL students. *Educational Technology, 29*, 38-41.

Mayer, R. E. (2009). *Multimedia Learning*. Cambridge University Press

Nation, I.S.P., (2001). *Learning Vocabulary in Another Language*. Cambridge University Press, Cambridge.

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

Robinson, P. (2001). *Cognition and second language instruction*. Cambridge, UK: Cambridge University Press.

Saran, M., Seferoglu, G., & Cagiltay, K. (2012) Mobile language learning: Contribution of multimedia messages via mobile devices in consolidating vocabulary, *The Asia-Pacific Education Researchers, 21(1)*, 181-190.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing, 18(1)*, 55-88.

Stewart, M. A., & Pertusa, I. (2004). Gains to language learners from viewing target language closed captioned films. *Foreign Language Annals, 37(1)*, 438–447.

Stockwell, G. (2007). Vocabulary on the move: Investigating an intelligent mobile phone-based vocabulary tutor, *Computer Assisted Language Learning*, 20(4), 365-383, DOI: 10.1080/09588220701745817

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

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning*, 21(3): 217-228

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with Uni-Lingual (CEEFAQ) subtitles. *System*, 18(2), 221-234

Wong, L-H & Looi, C-K.(2010). Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: two case studies, *Journal of Computer Assisted Learning*, 26, 421–433

Yang, J. C. & Chang, P. (2013). Captions and reduced forms instruction: The impact on EFL students' listening Comprehension, *ReCALL*, 26(1): 44–61, doi:10.1017/S0958344013000219

Yuksel, D. & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *The Turkish On-Line Journal of Educational Technology*, 8, 48-54.

Zhang, H., Song, W., & Burston, J. (2011). Reexamining the effectiveness of vocabulary learning via mobile phones. *The Turkish Online Journal of Educational Technology*, 10(3), 203–214