

“So I Think دبّة is bear!” An Initial Data-Driven Explanation of How Arabic Students Use Captioned Video to Learn Vocabulary

Arab Journal of Applied Linguistics

e-ISSN 2490-4198

Vol. 07, No. 02, 2022, 1-34

© AJAL

<http://www.arjals.com>

Paula Winke, Elizabeth Huntley,¹ and Susan Gass, Michigan State University, USA

Abstract

We explored how L2 learners of Arabic make use of video-based captions to learn vocabulary. Thirteen students watched a four-minute captioned video twice on a computer equipped with eye-tracking technology. The study included 19 novel words, 2 familiar words with novel meanings, and 2 words with familiar morphological roots (23 total). In this paper, we focus on one word, *bear*, and plotted the learners' visual attention to that word as it appeared on screen. We triangulated the eye-tracking data with an L2 form recall task, a prior vocabulary knowledge scale task, a free-recall task, and a semi-structured interview with stimulated recall. Our conclusion is that Arabic programs should include more captioned-video content both inside and outside of class and at an appropriate proficiency level to help learners with L2 Arabic's complex form-meaning mapping.

Keywords: *incidental vocabulary learning, eye-tracking, second language acquisition*

¹ Corresponding author: Elizabeth Huntley
Email: huntle21@msu.edu

Introduction: Captions and how they help people learn second languages (L2s)

With a rise in the demand for authentic materials in foreign language curricula (Tomlinson, 2012) and the expanded use of educational technology (Levy, 2019), captioned videos have become increasingly popular as part of foreign or second language (L2) instruction (del Mar Suárez & Gesa, 2019; Mirzaei, Meshgi, Akita, & Kawahara, 2017). Captioned videos² are authentic and provide rich sources of foreign language learning input (Mirzaei *et al.*, 2017; Montero Perez, 2022; Montero Perez, Noortgate, & Desmet, 2013). Captions augment listening comprehension (Danan, 2004; Rodgers & Webb, 2017) and support vocabulary learning (Bird & Williams, 2002; Chai & Erlam, 2008; Koskinen, Wilson, Gambrell, & Neuman, 1993; Montero Perez *et al.*, 2013; Peters & Webb, 2018; Sydorenko, 2010). Captions enable learners to more effectively process input because they help learners segment and decode chunks of the aural stream into words and phrases (Bird & Williams, 2002; Montero Perez, Peters, & Desmet, 2014; Vanderplank, 1988), resulting in better intake (Winke, Gass, & Sydorenko, 2013), which is an integral part of learning (Gass, 1997). Captioned videos' multi-modality (i.e., visuals, aural, and textual) helps learners process input more deeply and strengthen form-meaning mapping, both of which help learners build up their L2 mental lexicon (Gass,

² Captions are used in this paper to describe text that is in the same language as the video (e.g., Arabic in video and Arabic in caption). They differ from subtitles which are translations of the video (e.g., Arabic in video and English in subtitles)

Winke, Isbell, & Ahn, 2019; Mayer, 2014; Mayer, Lee, & Peebles, 2014). In sum, there is a significant amount of research that demonstrates the significance of caption use in learning a foreign language.

Despite its widespread popularity in other foreign language curricula, captioned materials are not widely promoted in L2 Arabic classrooms. No major textbook series offers captioned video as a viewing option, either for constructed or authentic materials. Even language learning websites which do offer captions suggest (without offering evidence) that their use may be detrimental. For example, a post at the blog Transparent Language implores users to “not to focus on [captions] while answering the questions and try to rely solely on your listening skills” (Hanan, 2020). The website Playaling, which hosts authentic materials captioned in both dialect and Modern Standard Arabic, posts at the bottom of each page, “TIP: To train your ear, try listening without the captions first!” (Playaling, 2022). We argue that not only are captions beneficial, but they can furthermore provide instructors of L2 Arabic with great curricular flexibility by making authentic materials more accessible to learners (see Moufarrej & Salameh, 2019). In other words, rather than being seen as crutches, captions can provide students with useful information about the language being learned. However, more research is needed to understand how L2 Arabic learners make use of captions.

The current empirical findings on the effects of caption use in Arabic are not straightforward. Winke, Gass, and Sydorenko (2010) with evidence from English learners

of Arabic, Chinese, Russian, and Spanish found the results to be complex. Their study measured learning from a captioned video in three ways. There was an aural and written vocabulary test and a comprehension test. For learners of Russian and Spanish, it is the case that learning was better on all three measures when students watched a version of a video with captions before watching the same video without captions. However, for Arabic (and Chinese), the findings were less straightforward. In the case of Arabic, watching a video first with captions and second without benefited students in aural vocabulary gains. In a follow-up study Winke, Gass, and Sydorenko (2013) found that the time students spent on captions depended on many factors (e.g., vocabulary knowledge, type of content, and proficiency). What these and other research projects show is that captions are indeed beneficial, but one needs a detailed understanding of how best to use them to promote learning.

Exploring the fine-grained nuances of caption use within the context of L2 Arabic

As noted above, we know from two decades of research on L2 learning that captions are beneficial. It is time for researchers to explore the nuances of how captions impact learning at a more fine-grained level. Much of this research has been conducted using eye trackers which are able to track eye movements as someone is looking at a video and/or reading a text. The assumption behind this research is the eye-mind link, which theorizes that eye gaze indexes focal attention (Godfroid, 2020; Rayner, 1998). In other words, eye movements provide a window on what someone is thinking and/or focusing

on while watching a video or reading a text. Thus, research in this area analyzes how long one focuses on a video or text and how much “jumping around” they do. That is, do they read something and then go back someplace earlier in the text? Do they fixate on new, unknown words, or do they just gloss right over them? At least four studies on L2 learners’ caption use have involved eye-tracking to explore the more fine-grained nuances of caption use (Bisson, Van Heuven, Conklin, & Tunney, 2014); Gass *et al.*, 2019; Lee & Révész, 2018; Montero Perez, Peters, & Desmet, 2015; Winke *et al.*, 2013). Out of these four eye-tracking and L2 captions studies, Montero Perez et al. (2015) specifically investigated vocabulary learning as an outcome variable. Montero Perez et al. investigated L2 learners’ processing of novel French vocabulary through French captioned video versus the same video (with aural French), but with English subtitles. Additional variables were full versus keyword captioning and whether the learners were informed that they would have vocabulary tests afterwards. The authors found a correlation between the learning of words and the total fixation durations and second pass durations (going back to look at words) with those participants who knew they would be tested on vocabulary after watching a video with captions. In other words, the longer a person fixed on a target word in the captions, the more likely it was that the person would correctly recognize the word on a form recognition test. We wondered if the same would be true of L2 Arabic.

One issue with Arabic vocabulary in particular is that the surface form can differ

drastically from the lexical root rather, or in addition, to the use of prefixes or suffixes to denote case or tense. This is unlike English and other European languages, where changes in the surface form do not as readily impact root access (Alhawary, 2009; Giaber, 2017). We speculated that because the meaning of Arabic words is highly dependent on the word's root, the frequent modifications between root and surface form may render word (meaning) identification difficult, especially in the aural mode. In writing, one can more clearly identify the root, and from there gain at least a sense of the word's meaning. Processing a word solely from aural input may be more difficult for learners. Learners have to (a) identify the root, (b) determine the general meaning of the root, and (c) determine how the morphological complexities have modified the root meaning. We suspect that captions may help learners overcome these challenges by overtly scaffolding form-meaning mapping.

Indeed, prior research has found that, for beginning learners of L2 Arabic, processing Arabic text is slow, while expert L2 readers can bundle features and rely on fewer overall features in understanding text (Wiley, Wilson, & Rapp, 2016). Prior research on captions (Winke *et al.*, 2013) has found that college-aged, L2 Arabic learners spent more time fixating on captions than learners of Spanish or Russian. This was likely due to the greater differences between their L1 (English) and L2 (Arabic) scripts. Learners of Spanish and Russian spent less time making form-meaning connections given the greater similarities between their L1 and L2 writing systems.

In this paper, we borrow an approach from Montero Perez et al. (2015) and outline areas of interest around individual captioned words in a fully L2 captioned video to try to explore a potential link between L2 vocabulary learning and one's attention to specific novel L2 words in the captions. We explore whether extra attention or certain patterns of attention may influence the learning of those words. In this study, we focus on L2 Arabic, and we address some methodological concerns from Montero Perez et al. (2015). In contrast to their study, we describe how interest areas around the targeted words were drawn, and we explain our corrections to the interest areas to combat drift. Montero Perez et al. wrote that they did not ask learners to comment on their learning processes related to the L2 French targeted words in their study because it seemed "questionable whether learners [would be] able to recall why they paid attention to a particular word," (p. 325), although other researchers have suggested that it is valuable to triangulate attentional data with qualitative data, such as first-person verbal reports, to fully understand learners' mental engagement with input (Godfroid & Schmidtke, 2013; Robinson, Mackey, Gass, & Schmidt, 2012). Furthermore, in Montero Perez et al.'s (2015) study, the number of times each word occurred in the video was not reported. Exposure frequency, however, plays a role in incidental vocabulary acquisition (Godfroid *et al.*, 2017; 2018; Malone, 2018; Mohamed, 2018). But what we do not know is what fluctuations there are in attending to each occurrence of a word. Thus, our research question for the current brief study was the following:

When Arabic learners learn vocabulary from captioned videos, can we find evidence in the eye-tracking metrics and through first-person verbal reports (interviews with stimulated recall) that the new knowledge came, at least in part, from the captions?

Method

Learners

We recruited 13 second- through fourth-year L2 Arabic learners from our university's Arabic program.³ Information about the individual learners is in Table 1. The age range of 12 of the participants was 19-24 (average is 21.5). The Arabic language program at our university was using the *Al-Kitaab* textbook series (Brustad, Al-Batal, & Tuñisī, 2007; 2010; 2011; 2013) at the time of data collection. Second-year learners have four contact hours per week, third-year learners have three, and fourth-year have two. Although six of the learners reported that they have relatives who speak the language, only one learner (P10) said that she spoke Levantine Arabic (as well as English) at home.

³ One student was in the first year of the program but he had studied Arabic in high school.

Table 1*Participant Background Information*

ID	Pseudonym	Gender	Year in Arabic program	Arabic spoken in home?
1	Ahmed	M	2nd	N
2	Barb	F	2nd	N
3	Cherine	F	2nd	N
4	Daim	M	2nd	N
5	Fara	F	4th	N
6	Gina	F	4th	N
7	Helen	F	3rd	N
8	Irene	F	4th	N
9	Jamila	F	3rd	N
10	Kalthum	F	3rd	Y
11	Larry	M	4th	N
12	Monica	F	3rd	N
13	Ned	M	3rd	N

Materials*Video*

The video used in the experiment was a short clip (4 minutes and 16 seconds, 234 total words) from a longer nature-themed video from National Geographic (2006) that tells the story of a mother bear and her cub. The video is commercially available in English. We translated, dubbed, and fully captioned the video clip into Modern Standard Arabic (Winke *et al.*, 2010; Winke *et al.*, 2013; Gass *et al.*, 2019).

Target Items

Target vocabulary items were selected by reviewing the video content with two Arabic instructors. We asked them to choose words in the video that they felt were novel, salient,

Table 2

Words Used in the Present Study.

Targeted Vocabulary	# of times in video	Ave. previewing knowledge [1-5] (SD)	Ave. post viewing translation results [0-3] (SD)
Novel form & meaning			
bear / دببة	4	1.85 (1.52)	1.38 (1.56)
forest / غابة	4	1.46 (0.52)	0.38 (0.87)
to appear / يظهر	4	2.23 (1.24)	0.31 (0.63)
mating season / موسم الجماع	2	1.62 (0.77)	0.46 (0.52)
nervousness / توتر	2	1.15 (0.38)	0.00 (0.00)
tragedy / مأساة	2	1.15 (0.55)	0.00 (0.00)
desire / رغبة	2	1.92 (1.19)	0.33 (0.48)
to loses / يفقد	1	1.38 (0.65)	0.23 (0.83)
chases / يطارد	3	1.00 (0.00)	0.00 (0.00)
grass / عشب	1	1.15 (0.38)	0.00 (0.00)
twins / توأمين	1	1.08 (0.28)	0.08 (0.28)
to be repeated / يتكرر	1	1.38 (0.65)	0.33 (0.63)
to nurse; to feed / إرضاع	1	1.38 (0.77)	0.15 (0.38)
to kill / يقتل	1	3.15 (1.46)	2.08 (1.44)
to clash / يتصادم	1	1.15 (0.55)	0.00 (0.00)
buy time / شراء الوقت	1	1.62 (0.77)	0.85 (0.80)
light colored / شقراء	1	1.15 (0.38)	0.23 (0.83)
dark colored / داكنة	1	1.15 (0.38)	0.00 (0.00)
Novel meaning			
cub / صغير	12	4.77 (.32)	2.00 (2.00)
territory / منطقة	3	4.69 (.60)	2.00 (2.00)
Familiar form (root)			
male / ذكر	7	3.77 (1.01)	1.23 (1.01)
female / أنثى	3	1.00 (0.00)	0.00 (0.00)
to eliminate / التخلص	1	1.62 (0.96)	0.69 (0.85)

and not directly covered in the students' textbook (*Al-Kitaab* textbook series, Brustad, Al-Batal, & Tuñisī, 2007; 2010; 2011; 2013). There were 23 target items in total: 13 nouns, 8 verbs, and 2 compound nouns. Of these target items, 18 were categorized as new (in the sense that students had not seen them in their textbook), 2 were familiar textbook words whose video-based meanings were novel, and 3 were derived from trilateral roots familiar to students (i.e., students had previously studied vocabulary derived from the same root). The list of vocabulary words under these distinctions, and the number of times each appeared in the video, are in Table 2.

Post-Tests

Comprehension Test

General comprehension was evaluated with a free-recall test where participants wrote down everything they remembered. A free-recall comprehension test was chosen over a multiple-choice test because free-recall can capture global understanding in finer detail, allowing us to better compare responses across participants (Winke & Gass, 2016). Participants responded in Arabic or English, and they were allowed to write or type their responses.

Vocabulary Tests

Vocabulary acquisition was measured with a pen-and-paper translation test. Participants saw the target item in Arabic and were asked to translate it into English. Target word items were generally ordered according to their first appearance in the video, with

semantically-related items grouped together when possible to mimic what a teacher might do in class. The translation test included a measure of prior vocabulary knowledge (Wesche & Paribakht, 1996). For each item, participants rated familiarity on a five-point Likert scale.

Background questionnaire

Participants filled out a brief background questionnaire in which information was gathered about past Arabic study, travel to an Arabic-speaking country, Arabic-speaking family members, and motivation for studying Arabic. De-identified background information with pseudonyms is in Table 1.

Interview

Participants took part in a short follow-up, semi-structured interview involving a customized stimulated recall. Questions in the interview template focused on participants' impressions of the captions and how they used them, as well as how they made use of previous familiarity with roots to ascertain target items. The second author, who collected the data in the lab and who speaks Arabic, individually selected video clip segments for the stimulated recall based on each participant's interview responses. There were two main criteria we used for selection: 1) the clip contained the target item(s) for which the participants had shown evidence of learning and/or 2) the clips contained items which were familiar textbook words, but whose video-based meaning was novel and appeared to elude participants. The goal was to see how participants were

interpreting the multimodal input, including when there was a disparity between the previously known meaning of the target item (such as "to remember") and the video-based meaning ("male"). Thus, we were able to tailor our follow-up probes in each interview with video-stimuli.

Procedure

We ran the study in an Eye-Link 1000 eye-tracking lab. After signing a consent form, each learner sat in a chair in front of the Eye-Link 1000 and placed their foreheads on a headrest, approximately 50 centimeters away from the computer screen, which was connected to the eye-tracking apparatus. We then ran participants through a 9-point eye-tracking calibration and validation procedure. Following the calibration, participants watched the video (Trial 1), took a short break, then watched the video a second time (Trial 2). Participants took the post-tests following the video viewing: first, the free-recall general comprehension test, and second, a simultaneous test of translation and prior familiarity. Participants then filled out a background questionnaire while the researchers evaluated each participant's comprehension and translation test. The results from the comprehension and translation tests informed the semi-structured interview with stimulated recall.

During the interview, the second author replayed parts of the video that had targeted words that the learner learned. The students were asked what they were thinking about when they originally saw the video and were also asked general questions

about their use of the captions.

Analysis

Eye-movement recording and metrics

Participants' eye movements were recorded with the EyeLink 1000 eye-tracking system from SR Research. We sampled at a rate that captured participant eye movement every 4 milliseconds. We set three predefined areas of interest for our learner eye tracking: (a) the full video image area, (b) the captions area (both as in Winke *et al.*, 2013), and (c) all instances of individual target items within the captions area (as in Montero Perez *et al.*, 2015). The first two areas of interest were static and were on screen the entire length of the video encompassing large areas of the screen. The third type of interest areas (over the targeted vocabulary words) were smaller and dynamic. There was one for each occurrence for each individual vocabulary word. The second author drew each one over the word on the original video, but elongated the area of interest above and below the word to about an inch in size to accommodate normal upward and downward eye-fixation drift, as well as off-calibration. Each interest area only appeared when the vocabulary word was on screen. Each was time stamped and appeared and disappeared on cue in relation to when the word appeared and disappeared as part of its caption in the video.

To measure visual attention, we calculated for each learner the eye-tracking metric of *dwell time*, which is the total amount of time looking in milliseconds. For further details

about conducting eye-tracking research, see (Godfroid, 2020).

Measurements of vocabulary learning

We assessed the responses on the vocabulary translation test as follows: 0 (no response given); 1 (incorrect), 2 (partially correct), and 3 (correct). By combining the translation responses with the 5-point prior-knowledge Likert responses, we coded each person into four levels of learning in regard to each targeted vocabulary word, with the lowest level (Level 0) rare, as expected:

- **Level 3:** Had limited prior knowledge (1-2 on prior knowledge test), and knew the word (2-3 on translation)
- **Level 2:** Had prior knowledge (3-5 on prior knowledge test), and knew word (2-3 on translation). This represents a steady state.
- **Level 1:** Had limited prior knowledge (1-2 on prior knowledge test), did not know word (0-1 on translation)
- **Level 0:** Had prior knowledge (3-5 on prior knowledge test), did not know word (0-1 on translation)

The only two points of interest to us are Levels 3 and 1. Level 2 represents a steady state, i.e., they knew the word before the study. Level 0 is a bit of an anomaly because it appears that what they knew at the beginning did not carry through to the end. In what follows, we will focus on one word, دبة *'bear,'* because that was the one word that showed the greatest amount of learning (most likely because it represented the subject matter of the

video) and hence provides the richest data.

Findings

1. Who learned which words?

There were five words that had high scores in the post viewing translation test, namely the Arabic equivalents of 1) bear, 2) to kill, 3) cub, 4) territory, and 5) male. As can be seen in Table 2, four of these words had relatively high scores on their assessment of prior knowledge. Given that so many students were familiar with these words prior to the study, we focused on the word for 'bear' because there was evidence of learning given relatively low average previewing knowledge. To summarize, the Arabic word for 'bear' was learned (Level 3) by three students (IDs 6, 7, and 8), and it not learned (Level 1) by 7 students (1, 2, 4, 5, 9, 11, and 12). Our comparison in this study involves the two extremes (those who learned vs. those who did not).

2. Is there evidence in the eye-tracking metrics that the captioning of the words that were learned were attended to more (or differently) by those who learned the words, as compared to those who did not?

To answer this question, we visually inspected the data on the learning (Level 3) and non-learning (Level 1) of 'bear' for attentional patterns. We plotted the two groups' dwell times (Figure 1 and Figure 2) on the vocabulary word for 'bear.' What we see in Figure 1 is that all three learners in Group 1 appear to peak in terms of dwell time at about the fifth or

sixth appearance, after which there is a slight decline in dwell time across the 8 occurrences of the word in the video. We suggest that this may be indicating a moment of recognition. The continued attention to the word following the “peak” moment, or spike in attention, may be thought of as reinforcement or “checking” time. In other words, learners are looking at the word to verify form-meaning connections (a form of meaning activation). In fact, there is still dwell time that is higher than it was on the earliest appearances. This reflects the need to have multiple exposures for learners for learning and time to solidify new information.

Figure 2 contains the dwell times of those for whom learning did not take place. We do not see the same pattern of a clear peak followed by reinforcement as seen in Figure 1. For example, student 12 does appear to peak her visual attention to the word for ‘bear’ at the 6th appearance, but then does not dwell on the word at all after that. Student 11 peaks at time 8, but there is no time for further form-meaning linkage. Students 5 and 9 peak at time 5, but then do not attend to any great extent at later appearances of the word. Student 12 peaks at time 6 with little dwell time at later occurrences. Student 1 shows no dwell time throughout. Student 4 appears to be the exception and patterns more like those in Figure 1 than those in Figure 2. Thus, there are two elements that appear to be important and that differentiate the students in these two figures: (1) a peak, and (2) further dwell time, which we speculate is time for solidification.

Figure 1

Level 3 Learners' Dwell Time on Each Occurrence of the Word for 'Bear'

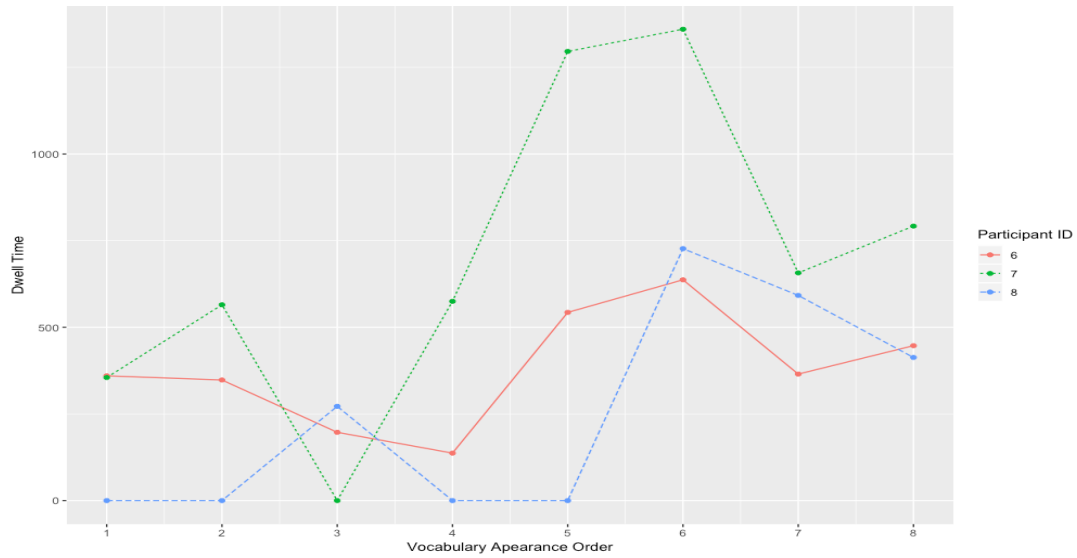
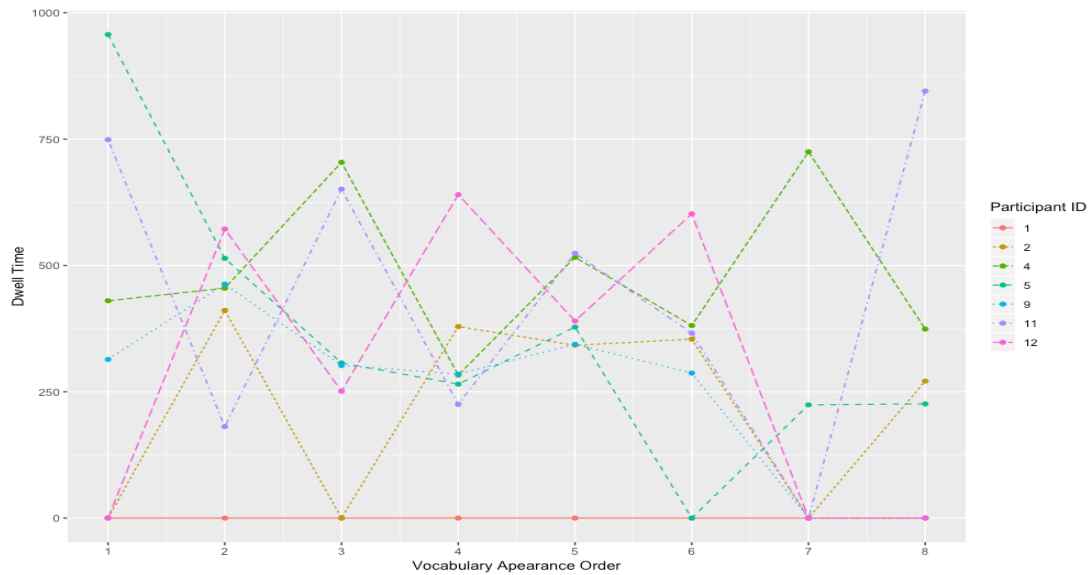


Figure 2

Level 1 Learners' Dwell Time on Each Occurrence of the Word for 'Bear'



3. Is there evidence in the first-person, verbal reports that new vocabulary knowledge came, at least in part, from the captions?

To address this question, we looked at the interview transcript from learners 6, 7, and 8, who, as indicated by the translation test, appeared to have learned to recognize the written Arabic form of the word for 'bear' by watching the video clip with captions. This is Level 3 of learning. These three individuals indicated they did not know the word for 'bear' prior to watching the video. Like most of the 13 learners in the study, those who learned the written form for 'bear' from the captions described that the video was "very fast," and that it was hard, overall, to associate the words with the captions. For example, when asked what she typically did while watching the video the first time, Gina, participant 6, indicated the following:

Example 1

So, I was trying to read the captions, but then they were moving really fast, so I was like, "oh maybe I should be watching the video a little bit more," so I was trying to go back and forth, and that wasn't really working, because then if I wasn't reading it, I wasn't really understanding. So, then the second time I watched it, I tried to just listen more and watch it more, and read the subtitles [captions] less. (P06, Gina)

Similarly, Helen, participant 7, discussed her difficulties in reading along with the captions when asked what she did when viewing the first playing of the video.

Importantly, she used the captions to try to figure out the meanings of the words.

Example 2

I was just trying to read along. And I would try to figure out the words I know, and if I knew some of the roots... but then it would move too fast, so I would kind of, like, read the first couple, then watch the video. Sometimes I just tried listening, but then usually I would go back to the words.

(P07, Helen)

Participant 8, Irene, who also appeared to have learned the written form of the word for '*bear*' from the video and did not know the word before viewing, seemed to have an easier time with the video overall. She used the video strategically. When asked if she was able to learn new words from the video, she responded:

Example 3

Not really. I was looking at the video itself. I tried to listen instead of reading when I hear. Well, I mean, I was reading it, but I was trying to listen to it the first time and then read it the second time.

(P08, Irene)

The researcher then asked:

So, the first time the video played, you were trying to not read the captions? (Researcher)

To which she responded:

Yeah, I mean it's hard not to when it is in your face, but I was trying to and then the second time I was trying to read it more. I don't think I learned anything though. (P08, Irene)

She continues by talking about what she did while watching the video for the first time:

Example 4

Well, if you can read it and then hear it, and then see something that corresponds to it, that helps to create the bigger picture... Because if you can read it, then that helps with your comprehension, because you can hear [a] native person say, and that can help with you being able to say the word. And then also, having the actual video itself in the background, that reinforces the audio because then you can see what you're hearing. (P08, Irene)

She, too, used the captions as part of the information she used to learn words. Later in the interview, when directly asked if the captioning was helpful, participant 8, Irene, went on to explain that a drawback for her in comprehending the video was that her listening skills in Arabic are low. She pointed out that even if she knew the words, she may not have recognized them in the current video because her listening was not as good as it needed to be for the video.

Example 5

Because my listening is not as good, I think that I wouldn't have recognized a lot of words that I otherwise know. (P08, Irene)

Toward the end of the interview, participants were asked if the captions helped.

Participant 8, Irene, further noted that the full captioning both hindered and helped:

Example 6

It [captioning] made it kind of hard to understand, because I don't understand all the words. But, like I said, it kind of helps in a way too, because it helps more so with speaking and actually comprehending, because I can hear the word and I can speak it, because I can read it. Does it make sense? Like, I heard it, I saw it, and I can read it now. (P08, Irene)

The participants also focused on the word for 'bear' itself. Irene talked about the benefits of frequency.

Example 7

Irene: Was *dibba* right, was it right?

The conversation continued with the researcher asking Irene how she got to the word *dibba*:

Researcher: So, you said you learned *dibba* ["bear" #1], which...

Irene: That was a guess, but, I mean...

Researcher: Can you tell me how you think, how do you think you got *dibba*?

Irene: They said it a lot.

Participant 6 also recognized that the captioned video had served as a venue for learning. In her discussion with the researcher, he specifically talked about *dibba* and one other word.

Example 8

Researcher: Were you able to learn any new vocabulary from the video?

Gina: I'm not sure, possibly?

Researcher: So, here's everything that we had selected (hands participant a blank target word test). Is there anything from here where you were like, "oh I learned a new word!"

Gina: So I think that *dibba* [#1 bear] might be "bear" (note: student is correct here)? And then I think the only other one that I really might have been able to pick up on, and it was also kind of based on the root, it was *mawsim al-jimaaʿ*. [#12 "mating season;" participant

guessed “pack/group”]

Expanding beyond reports from those who learned the word for ‘bear,’ a common theme discussed across all participants was that the sheer speed of the video was difficult because the learners are not used to watching authentic videos in Arabic, or that when authentic videos are played in class, the videos are manipulated pedagogically to help the learners follow along. Illustrative comments from the learners in regard to this are as follows:

Examples 9-11

9. Most of the time I didn't really even finish the [captioned] sentence before it changed [went away].
(P01, Ahmed)
10. In class we don't have captions. (P05, Fara)
11. We don't use a lot of videos in class. We use some videos from a website *aswaat arabiyya* [an online collection of listening texts for students of Arabic; <https://www.laits.utexas.edu/aswaat/>]. Those don't have captions, at least if they do, I don't know how to turn them on! (P06, Helen)

Discussion

In this study, the eye-tracking metrics showed that the three students who learned the word for ‘bear’ were visually attending to the vocabulary word in the captions with a rather consistent pattern. Our small-scale study indicated the Arabic learners’ vocabulary was expanded through exposure to this short video clip, a relatively good outcome, given that the learners were not asked to learn *a priori*, and given that the video clip itself was

only approximately 4 minutes long. This study demonstrated eye movement patterns that differ between those who learn from the video and those who do not. We can conclude, based on our eye-tracking metrics, that L2 Arabic captions are used for novel vocabulary learning, especially when the words are highly salient. Based on our interview data, we found that the benefits of captions may, in part, be dependent on prior experience using L2 captions, as seen in Examples 6-8.

In the beginning of this paper, we explained that captioned videos have become increasingly popular as part of foreign or second language (L2) instruction (del Mar Suárez & Gesa, 2019; Mirzaei *et al.*, 2017). Our results might allow for an expansion on what it means to “take advantage” of captions to learn vocabulary. In Arabic as a second language, this may be indeed quite complicated. Native speakers of English take a relatively long time to learn Arabic to advanced levels of proficiency relative to other languages, such as German, Greek, Turkish, or Spanish (Stevens, 2006). Arabic reading and writing in particular have been described as “difficult” for native-speakers of English to acquire. In addition to its unique orthographic features, many of which are novel to adult L1 English learners of Arabic, Arabic has a unique morphological system which can be challenging for learners (Wiley, *et al.*, 2016), and learning to parse the language through visual representation may be particularly important (Showalter & Hayes-Harb, 2015).

In sum, what we have shown in this brief report is that eye-tracking data can

provide direct evidence of L2 learning behavior. We are left, however, with a number of unanswered questions for future research. For example, why were so few words learned from the video? Students reported not having used captioned videos in their curriculum. Is training required to use captions appropriately?

It does appear that experience using captions or subtitles may be related to what students can gain.¹ For example, Irene, a student who was successful in learning, discussed her prior experience (Example 12). Even though, her experience was with subtitles and not captions, she did recognize a “eureka” moment of learning.

Example 12.

Researcher: Have you ever used captioned videos before? Or did you guys ever watch Arabic videos that also had the words below in class?

Irene: When I studied abroad in Jordan... when I was there, my roommate and I, we would always watch like Disney movies with Arabic and Arabic subtitles or English, either way. But that that helped us when we were in Jordan. That was two years ago. But that was really helpful. I just had to watch a video two days ago of an Arabic movie, and it had English subtitles. So, it would like there be a word that they would say and then I read it, and thought, “oh that's what that word is in Arabic!”

Researcher: Oh, but what about Arabic with Arabic subtitles?

Irene: Oh no, not really. It's normally Arabic speaking with English or English with Arabic

Researcher: Okay, and with the Disney videos in Jordan, was it ever Arabic-Arabic?

Irene: Hmm... it was Arabic-English.

Researcher: Okay, so you are used to reading the subtitles and needing to take all these in, but
 YOU never had Arabic on Arabic

A further issue relates to the kind of captioning we used (full captioning). Could we have captioned only the words that the teachers believed were meaningful for the learners to learn, as described by Montero Perez et al. (2014)? If we had done so, and told the learners there would be a test of vocabulary learning afterwards (Montero Perez *et al.*, 2018), would the words have been learned at a higher rate? We suspect that they would have been, in particular because it would have given the learners more time to look at the targeted words. Finally, as noted above, learners commented on the difficulty level of the video. The mismatch between their proficiency level and the level of the video (e.g., vocabulary and speed of delivery) may also have been a factor in how little was learned.

Even though captioned videos are authentic and rich sources of language-learning input (del Mar Suárez & Gesa, 2019; Mirzaei *et al.*, 2017; Montero Perez, Noortgate, & Desmet, 2013), captioning may not be used as much in certain instructional contexts due to various reasons, one of which may be students' difficulties in parsing and aurally attending to the authentic video, as we found in this study. In other words, the speed of authentic and pre-captioned video that could be easily used in the classroom may not be used because the speed of the authentic audio is considered inaccessible to the students. One could say the students' listening is not "up to speed," or one could say that the

authentic video is “too fast.” The result is the same: Less use of authentic, captioned video. We believe that Arabic instructors should push past that notion, and instead use both video and captioned video more to help students with listening comprehension. While it is true that captions enable learners to more effectively process input by helping them segment and decode chunks of the aural stream into words and phrases (Bird & Williams, 2002; Montero Perez, Peters, & Desmet, 2014; Vanderplank, 1988), resulting in better intake (Winke, Gass, & Sydorenko, 2013), according to this study, caption use may take practice. Learners need to learn through practice on how to segment and decode. Doing so over time through captioned videos with increasing presentational speed and increasing content difficulty may allow captions to be maximally facilitative of those processes.

References

- Alhawary, M. T. (2009). *Arabic second language acquisition of morphosyntax*. Yale 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, 509–533. <https://doi.org/10.1017/S0142716402004022>
- Bisson, M. J., Van Heuven, W. J., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 35(2), 399–418.
- Brustad, K., Al-Batal, M., & Tuñisī, A. (2013). *Al-kitaab fī ta'allum al-'Arabiyya: A textbook for intermediate Arabic* (3rd ed.). Georgetown University Press.
- Brustad, K., Al-Batal, M., & Tuñisī, A. (2007). *al-Kitāb fī ta'allum al-'Arabīyah, al-Juz' al-thālith*. Georgetown University Press.
- Brustad, K., Al-Batal, M., & Tuñisī, A. (2010). *Alif Baa: Introduction to Arabic letters and sounds*. Georgetown University Press.
- Brustad, K., Al-Batal, M., & Tuñisī, A. (2011). *al-Kitāb fī ta'allum al-'Arabīyah: Al-Juz' al-awwal*. Georgetown University Press
- Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics*, 14, 25–44.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies.

Meta, 49, 67–77.

del Mar Suárez, M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: Do proficiency and aptitude make a difference? *Language Learning Journal*, 1736(May).
<https://doi.org/10.1080/09571736.2019.1617768>

EyeLink 1000 [Apparatus and software]. (2017). Mississauga, Ontario, Canada: SR Research.

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Erlbaum.

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. I. (2019). How captions help people learn languages: A working-memory, eye-tracking study. *Language Learning and Technology*, 23(2), 84–104. Retrieved from <https://www.lltjournal.org/item/3108>

Giaber, J. M. (2017). Differences in word formation between Arabic and English: Implications for concision in terminology translation. *Al-‘Arabiyya: Journal of the American Association of Teachers of Arabic*, 50, 53–79.

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports, and vocabulary learning scores. In J. M. Bergsleithner, S. Frota, & J. K. Yoshioka (Eds.), *Noticing: L2 studies and essays in honor of Dick Schmidt* (pp. 183–205). University of Hawai‘i at Manoa,

National Foreign Language Resource Center.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584.
<https://doi.org/http://dx.doi.org/10.1017/S1366728917000219>

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Yoon, H.-J. (2017). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584.
<https://doi.org/10.1017/S1366728917000219>

Hanan. (2020, April 10). Listening practice: Covid-19 benefits. *Arabic Language Blog*.
<https://blogs.transparent.com/arabic/the-only-beneficiary-of-coronavirus/>

Koskinen, P. S., Wilson, R. M., Gambrell, L. B., & Neuman, S. B. (1993). Captioned video and vocabulary learning: An innovative practice in literacy instruction. *The Reading Teacher*, 47, 36–43.

Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *Modern Language Journal*, 102(3), 557–577. <https://doi.org/10.1111/modl.12503>

Levy, M. (2019). World CALL: Are we connected? *Annual Review of Applied Linguistics*, 39, 59–73. <https://doi.org/10.1017/S0267190519000059>

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural

- enhancement, and working memory. *Studies in Second Language Acquisition*, 40(3), 651–675. <https://doi.org/http://dx.doi.org/10.1017/S0272263117000341>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press.
- Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, 28(5), 653–660. <https://doi.org/10.1002/acp.3050>
- Mirzaei, M. S., 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. <https://doi.org/10.1017/S0958344017000039>
- Mohamed, A. A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, 40(2), 269–293. <https://doi.org/http://dx.doi.org/10.1017/S0272263117000092>
- Montero Perez, M. (2022). Second or foreign language learning through watching audiovisual input and the role of on-screen text. *Language Teaching*, 1–30. <https://doi.org/10.1017/S0261444821000501>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>

- Montero Perez, M., Peters, E., & Desmet, P. (2014). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL*, 26(1), 21–43. <https://doi.org/10.1017/S0958344013000256>
- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, 99(2), 308–328. <https://doi.org/10.1111/modl.12215>
- Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, 31(1–2), 1–26. <https://doi.org/10.1080/09588221.2017.1375960>
- Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning and Technology*, 18(1), 118–141. Retrieved from <http://llt.msu.edu/issues/february2014/monteroperezetal.pdf>
- Moufarrej, G., & Salameh, C. (2019). The effects of songs on vocabulary retention in foreign language acquisition: The case of Arabic. *Al-'Arabiyya: Journal of the American Association of Teachers of Arabic*, 52, 101–123.
- National Geographic. (2006). Last stand of the great bear. 20th Century Fox. *Playaling*. Retrieved January 4, 2022, from <https://playaling.com>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*,

40(3), 551–577. <https://doi.org/http://dx.doi.org/10.1017/S0272263117000407>

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>

Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. W. (2012). Attention and awareness in second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247–267). Routledge.

Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English-language television programs. *CALICO Journal*, 34(1), 20–38. <https://doi.org/10.1558/cj.29522>

Showalter, C. E., & Hayes-Harb, R. (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, 36(1), 23–42. <https://doi.org/10.1017/s0142716414000411>

Stevens, P. B. (2006). Is Spanish really so easy? Is Arabic really so hard?: Perceived difficulty in learning Arabic as a second language. In K. M. Wahba, Z. A. Taha, & L. England (Eds.), *Handbook for Arabic language teaching professionals in the 21st century* (pp. 35–63). Lawrence Erlbaum Associates.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology*, 14(2), 50–73. Retrieved from <https://www.lltjournal.org/item/2688>

- Tomlinson, B. (2012). Materials development. In A. Burns & J. C. Richards (Eds.), *Pedagogy and practice in second language teaching* (pp. 269–278). Cambridge University Press.
- Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *English Language Teaching Journal*, 42(4), 272–281.
- Wesche, M. B., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13–40.
- Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance*, 42(8), 1186–1203. <https://doi.org/10.1037/xhp0000213>
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 66–87. <https://www.lltjournal.org/item/2711>
- 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. <https://doi.org/10.1111/j.1540-4781.2013.01432.x>
- Winke, P., & Gass, S. (2016, November). Using free recall and idea units for evaluating second language comprehension: Methodological choices and issues. *AL Forum: The Newsletter of the TESOL International Association Applied Linguistics Interest Section*. Retrieved from <http://newsmanager.commpartners.com/tesolalis/issues/2016-11-04/5.html>