2018

Examining Incidental Vocabulary Acquisition from Captioned Video: Does Test Modality Matter?

Frank Boers
fboers@uwo.ca

Nurul Aini Mohd Jelani
Western University

Examining Incidental Vocabulary Acquisition from Captioned Video:

Does Test Modality Matter?

Nurul Aini Mohd Jelani

&

Frank Boers

Abstract

Previous comparisons of vocabulary uptake from captioned and uncaptioned audio-visual materials have almost consistently furnished evidence in favour of captioned materials. However, it is possible that many such comparative studies gave an advantage to the captioned input conditions by virtue of their use of written word prompts in the tests. The present study therefore examines whether aurally presented test prompts yield equally compelling evidence for the superiority of captioned over uncaptioned video. Intermediate EFL learners watched a ten-minute TED Talks video either with or without captions and were subsequently given a word recognition and a word meaning test, with half of the test prompts presented in print and the other half presented aurally. While the results of the word recognition test were inconclusive, the word meaning test yielded significantly better scores by the group that watched the captioned video. However, this was due entirely to their superior scores on the printed word prompts, not the aural ones. This suggests that evaluations of the benefits of captions for vocabulary acquisitions should take input-modality – test-modality congruency into account.

Introduction

Studies on incidental vocabulary acquisition, that is, vocabulary acquisition occurring as a by-product of message-focused activities, have looked mostly at acquisition from reading (e.g., Horst, Cobb, & Meara, 1998; Pigada & Schmitt, 2006; Waring & Takaki, 2003), with only a comparatively small number of studies so far investigating vocabulary acquisition from listening (e.g., Van Zeeland & Schmitt, 2013; Vidal, 2003, 2011). Vocabulary acquisition through listening deserves more attention from applied linguists, however. For one thing, advances in technology have created easier access to authentic L2 listening materials, and so more opportunities for learning from such materials are now available. For another, word learning includes knowledge of the phonological aspects of the word, which is knowledge that is not always easily derived from reading only (van Zeeland, 2013). This is perhaps most obviously the case when the target language uses a logographic script (e.g., Mandarin Chinese), but it also holds true to a considerable degree in the case of some alphabetic languages, such as English, where the phonological form of many words has evolved such that their present written form—laid down as long ago as the 18th century (Freeborn, 2006, pp. 386–392)—is no longer a reliable cue for their present pronunciation.

Studies which investigate vocabulary uptake from listening generally show poorer gains than what has been observed in the context of reading (Brown, Waring, & Donkaewbua, 2008; Vidal, 2011). Explanations for this include the real-time nature of most authentic listening (where, unlike in the case of reading, one cannot usually return to earlier passages to verify or fine-tune one's interpretation) (Rost, 2002; Vandergrift, 2007) and the challenge of segmenting the stream of speech into distinct semantic units

such as words (Goh, 2000). Among the suggestions for assisting learners' listening comprehension is the use of audio-visual materials (e.g., Baltova, 1994; Neuman & Koskinen, 1992), because the visual component can offer clues to what is being talked about. With the growing availability of audio-visual materials such as television programmes, this medium for L2 learning has attracted increasing attention from applied linguists in recent years (e.g., Rodgers, 2013; Rodgers & Webb, 2011; Rodgers & Webb, 2017; Webb & Rodgers, 2009; also Peters and Rodgers, this volume). When using audio-visual materials, the addition of on-screen text corresponding to the aural input has also been recommended, because simultaneously seeing the words can help learners segment the speech stream they are being exposed to and thus to discern the spoken words therein (Vanderplank, 2010, for a review). On-screen text to accompany audio-visual input may be captions in the same language as the aural component (e.g., L2 captions accompanying L2 audio) or subtitles providing a translation into a different language (e.g., subtitles in the viewers' L1 accompanying L2 audio). The present article focuses exclusively on the first type (but see, e.g., Peters, Heynen and Puimege, 2016, for a recent comparison of the affordances of L2 captions and L1 subtitles for L2 vocabulary acquisition).

A substantial body of research in which the effects of captions are gauged in comparison with uncaptioned input conditions suggests that captions benefit learners' comprehension of audio-visual input and also their uptake of new words from this input (see Montero Perez, Van Den Noortgate and Desmet, 2013, for a meta-analysis of the available studies at the time). However, in the majority of the studies that have contributed to this strand of research on L2 vocabulary uptake, recognition and/or meaning-recall post-tests were used that presented the learner-participants with the

written form of the target words. This means that the participants who had been assigned to a comparison condition where they watched the uncaptioned version of the audio-visual material were presented in the post-test with printed word forms which they had not been exposed to before. This arguably put them at a disadvantage: Had they been presented with aural test prompts instead, they might have found it easier to recognize these as the words they heard in the videos. It is therefore at least theoretically possible that the positive effects of captioning on vocabulary uptake attested in these studies were overestimations owing to this input-modality – test-modality congruency. It is this possibility—pointed out earlier by Sydorenko (2010)—which the present study intends to explore.

Literature review

Montero Perez, et al. (2013) conducted a meta-analysis of studies available at the time which investigated the effects of captioned video in comparison with uncaptioned versions (a) with regard to comprehension of the videos (15 studies) and (b) with regard to uptake of new words from the videos (10 studies). The conclusion was very favourable (with large pooled effect sizes) of the use of captions for both purposes. Since then, a few more studies on vocabulary acquisition have appeared that include a comparison of the effects of captioned and uncaptioned input conditions. Montero Perez, Peters, Clarebout and Desmet (2014) have furnished additional evidence that captions aid vocabulary uptake. By contrast, in an experiment conducted by Bisson, Van Heuven, Conklin and Tunney (2014), neither watching captioned or watching uncaptioned video were found to promote vocabulary acquisition. However, this null result is exceptional and may well have been due to the fact that the participants in that

particular study had no prior knowledge at all of the language used in the video material—a situation of L2 video use which is unlikely in L2 instructional settings.

As already mentioned, the majority of the studies so far have made use of vocabulary tests which presented participants with the written form of the target words, and this input-modality – test-modality congruency may have given an advantage to those participants who had seen those written forms in the captions while watching the videos. They would perhaps not have enjoyed this advantage had the prompts been presented aurally. Theoretical support for this proposition is lent by Transfer-appropriate Processing models, according to which the nature of a learning experience will to a considerable degree determine the nature of resulting knowledge and thus also what purposes this knowledge will most readily serve (Morris, Bransford, & Franks, 1977; also see Barcroft, 2015).

Only a handful of comparisons of vocabulary uptake from captioned and uncaptioned video have so far made use of aural test prompts, and not always with the express purpose of analysing effects (Rodgers, 2013). Markham (1999) used an aural multiple-choice test format to test advanced ESL students' ($N = 118$) recognition of aurally presented words they had encountered in two educational television programmes. The test scores were significantly higher when students had watched a captioned version of the programmes. This result suggests that these advanced L2 learners' uptake of the phonological form of the words benefited from seeing their orthographic representation (i.e., the captions) in conjunction with the aural input. However, Sydorenko (2010) argued that this might not hold true at lower levels of L2 proficiency, where learners may find it harder to divide their attention in the case of multi-modal L2 input (see Sweller, 1994, for theoretical support for this hypothesis). In

her study, beginning learners of Russian watched three short video clips from a popular Russian comedy either with or without captions. After this, two vocabulary tests were used. The first was a word recognition format, where the participants were required to decide if they had encountered the given words in the videos. The second—which was absent from Markham's (1999) study—invited the participants to explain the meaning of the same words. In both vocabulary tests, half of the target words were presented in print and the other half were presented aurally—a design feature we will replicate. Evidence of the role of input-modality – test-modality congruency emerged in the word recognition test: The group that had watched the clips with captions did better on the written test prompts than the group which had watched the uncaptioned clips, but not on the aural prompts, such that the two groups obtained similar overall scores. The word meaning test did not furnish evidence of a modality congruency effect, however: The group that watched the captioned version of the video clips outperformed the group that watched the uncaptioned version on both the written and aural sections of the test. One part of a more elaborate study by Winke, Gass, and Sydorenko (2010) used a similar design. Intermediate-level students of Spanish watched three short wildlife documentaries with or without captions, and were afterwards asked to provide the meaning of selected words, half of which were presented in print and the other half aurally. The best scores were again obtained after watching the captioned videos, without evidence of a mediating role of test modality. It needs to be mentioned that the number of participants was very small in both investigations just described, with treatment groups made up of only eight or nine learners. Given this limitation, it is probably still Markham's (1999) study which provides the most robust statistical evidence to suggest that captioning is beneficial for vocabulary uptake regardless of the

modality of the tests. The fact remains, however, that this was a study with highly advanced ESL learners, for whom processing aural and written L2 input simultaneously may not have posed a challenge. It may also be worth mentioning that the participants in Markham (1999) were tested after they had watched one video and then again after they had watched a second video. It is therefore likely that their engagement with the second video was influenced by their expectation of another (aural) test, resulting in greater attention to language proper in addition to content. Anticipation of a test is indeed known to alter learners' engagement with input (Montero Perez, Peters, & Desmet, 2015), which is also why Hulstijn (2001) considers surprise testing a useful criterion for distinguishing studies on incidental learning from ones on deliberate learning.

In sum, there appears to be insufficient evidence to confidently rule out the potential mediating role of input-modality – test-modality congruency in research on the benefits of captioning for incidental vocabulary acquisition. The study we report below is intended to complement what evidence is already available.

Research Questions

The research questions we address in this study are as follows:

1) Does captioning benefit intermediate learners' uptake of novel words from L2 video input, as gauged by means of a word recognition test (i.e., a test which requires learners to identify words encountered in that input)?

2) Does captioning benefit intermediate learners' uptake of novel words from L2 video input, as gauged by means of a word meaning test (i.e., a test which requires learners to translate or paraphrase new words encountered in that input)?

3) Are the attested benefits (if any) of captioning mediated by test modality, such that the benefits are more noticeable for written test prompts than for aural test prompts?

Method

Participants

Three groups of high school students were involved in this study. They were all 16-year-old students at a public high school in Penang, Malaysia. One group ($n = 15$) was used for the purpose of a 'norming' procedure, to determine which words in the video were highly likely to be unknown to learners at their level of proficiency. The two other groups were the actual treatment groups in the experiment. They were assigned to the captioned ($n = 32$) and the uncaptioned ($n = 34$) viewing conditions by the flip of a coin.

Although they had been learning English as a school subject for about nine years, these students' level of English language proficiency was estimated as only intermediate, which corresponds to the proficiency level addressed by their textbook. This textbook, prescribed by the Ministry of Education of Malaysia, was organised thematically, around ten different topics. At the request of the teacher, who agreed to incorporate the experiment into her course, a video was chosen for the experiment that was thematically related to one of those textbook topics. Informed consent to participate in the study was obtained from all the students and their parents. The students were informed the study was about the use of audio-visual materials in the English-subject classroom, but the focus on vocabulary acquisition was not disclosed to them until the data collection was completed.

To ascertain that the two treatment groups were comparable as far as their level of English proficiency was concerned, we compared the scores they obtained in their most recent final-year English examination, taken approximately two months prior to the experiment. The two treatment groups' mean scores (in %) were virtually identical: $M =$ 44.25 (SD =14.01) and $M =$ 42.67 (SD = 15.59), for the captioned and the uncaptioned input conditions, respectively. An independent-samples t-test confirms there was no significant difference between the treatment groups' exam scores: t = 0.44; $p =$ .66. For the norming procedure, we selected 15 students with considerably higher scores on the same examination ($M =$ 61.53; SD = 6.05). The rationale for this was that words unfamiliar to this more proficient group would in all likelihood be unfamiliar also to their less proficient peers in the actual treatment groups.

Materials

As mentioned, the experiment was conducted as part of the students' regular EFL course, and the choice of video was made in consultation with the students' teacher, who specifically requested for the video content to align with a theme from the course book. The teacher also expressed preference for a relatively short video so as to facilitate its integration in one of her regular (80-minute) class periods. One source of freely available videos on a wide range of topics from the realms of, for example, science, technology, business, and entertainment is *TED Talks* (https://www.ted.com). The duration of these talks ranges from four to 20 minutes, and so the bank of available TED Talks (which is continuing to grow) contains videos which are short enough to be incorporated in typical lesson periods. At first glance, one may question the suitability of these materials for use with high school EFL/ESL students, because the topics dealt

with in TED Talks are often of a specialized nature and the talks may consequently contain a fair amount of specialized or technical vocabulary. However, these are not talks by specialists for other specialists working in their field. Instead, it is the presenters' task to make the content of their talks accessible for a lay audience. An analysis of 80 TED Talks by Scotto di Carlo (2014) found that the presenters usually resort to analogies, comparisons, examples and visuals to clarify the technical concepts they introduce. Even though it has been suggested (Coxhead & Walls, 2012) that one needs knowledge of the 4,000 most frequent word families in English to understand 95% of the running words of the average TED Talk—95% being the 'coverage' often considered necessary for adequate listening comprehension (van Zeeland & Schmitt, 2012), the deliberate efforts on the part of TED Talk presenters to clarify terms and elucidate concepts may render these videos comprehensible also for non-advanced language learners. It is perhaps not surprising, then, that TED Talks are gaining popularity in L2 instructional settings (e.g., Takaesu, 2013).

After browsing the bank of available TED Talks in search of a relatively short one related to one of the themes of the students' textbook, it was decided (in consultation with the teacher) to use a ten-minute talk by Bart Knol on the subject of battling malaria, *three ways to kill mosquitoes*:

https://www.ted.com/talks/bart_knols_cheese_dogs_and_pills_to_end_malaria.

As mentioned, 15 students whose scores on the most recent English examination were higher than those of the actual treatment groups were used to estimate which words in the selected TED Talks video were likely to be as yet unfamiliar to the treatment groups. This procedure was preferred over administering a pre-test to the

treatment groups, because a pre-test experience can affect the way learners subsequently engage with the input materials.

 To select target words, the TED Talk was first screened by means of Nation and Heatley's (2002) *Range* program for words beyond the 1,000 most frequent word families of English. As a second step, words that have cognates in the participants' L1, such as *malaria*, *aroma, larvae, parasites, experiment* and *pill*, were excluded as potential target words. The remaining list of 35 potential target words was then presented to the aforementioned group of 15 comparatively proficient students, who were asked to try and explain the meaning of each word, either by giving an English synonym or by means of a translation into their L1. Seven of these 35 words—*repel, transmission, trigger, synthetic, remarkable, outbreak,* and *epidemic*— turned out to be unfamiliar to all of the students. This, however, was deemed too small a sample for the purpose of the study, especially given the need for a set of written and a set of aural prompts in the post-test, and so it was decided to select additional target words that at least 11 of the 15 students did not know. Eight more words were thus added to the list of target words—*fascinating, equivalent, volunteer, immobile, scattered, insecticide, convinced,* and *tablet*—, thus yielding a final list of 15 target words. Except for *remarkable* and *tablet*, which occurred seven and three times, respectively, all the target words occurred only once in the video.

Apart from the 15 target words, we included 13 additional words in the vocabulary test. These additional words did *not* occur in the video. Some of these were more common words (e.g., *violent*) that we deemed useful for affective motives—to avoid frustration on the part of the students in case they found the test too hard. Others were

words which showed phonological resemblance to the actual target words (e.g., *propel* to match the target word *repel*), a design feature we borrowed from Markham (1999).

The vocabulary test consisted of three components (see Appendix 1). One component, henceforth the word recognition task, presented the students with each of the words and the learners were asked (by means of a yes/no response) if these occurred in the video. This was intended to probe the learners' episodic memories of the words (Tulving, 1993), i.e., memories where the words are explicitly associated with the context or circumstance where they were encountered. The first 13 words (including seven of the target words), were printed on the answer sheet. The second 15 words (including the other eight target words) were presented aurally.

A second component in the vocabulary test concerned only the aurally presented words. The students were asked to try and write these down on the answer sheet (before deciding whether they occurred in the video). This was done with a view to assessing whether seeing the printed words in the captioned video would generate more accurate orthographic representations of the words than only hearing them.

The third component of the vocabulary test, henceforth the word meaning task, concerned both the written and the aural prompts again. The students were asked to explain the meaning of the words which they recognized from the video. They could do so either by providing a synonym in English or by providing a translation in their L1.

To ascertain that the students watched the video for its content, we also created ten multiple-choice comprehension questions (see appendix 2), which were administered before the vocabulary post-test. A few of these were general comprehension questions (e.g., concerning the presenter's principal motive for studying mosquitoes), but most concerned recall of more precise facts (e.g., which part of the human body African

malarial mosquitoes prefer biting). For each question, four answer options (one correct and three foils) were available, and the students were asked to select 'a', 'b', 'c', or 'd'. None of the words in the vocabulary test were used in the content questions. Analogous to the vocabulary test, the first five questions were printed on the worksheet, but the remaining five were presented aurally. The aural presentation comprised both the 'stem' of the multiple-choice question and the four response options.

Procedures

The same classroom procedures were used for both groups, with the only exception that one group watched the captioned TED Talks video and the other group the uncaptioned version. The students watched the video twice. This mimics what was done in virtually all previous studies on the effects of captioning, and it was also the teacher's habit to play audio recordings twice when she administered listening comprehension exercises. The students were told that questions about the content of the video would follow, but that a test on vocabulary would follow as well was not announced. This is in keeping with this study's purpose of examining incidental acquisition. After having watched the video twice, the students first tackled the comprehension test. Then followed the vocabulary test.

The multiple-choice comprehension questions were scored in a binary fashion, with one point awarded for every correct response and zero points for wrong responses. The yes/no responses in the word recognition task (i.e., where students were asked if the given words occurred in the video) were also scored in a binary fashion, as were the students' written reproductions of the aurally presented words—recall that the purpose of adding the writing task was to gauge whether seeing captions leaves accurate

memories of orthographic form. In scoring the responses in the word meaning task, however, it was deemed necessary (similar to Sydorenko, 2010) to give credit for partially correct responses. Half points were therefore awarded for responses that were semantically closely related to the target word. A scoring protocol was developed by the first author in collaboration with a second Malay-English bilingual EFL teacher. For example, for the target word *epidemic*, it was decided the participants would be awarded no point if they gave no response or gave a response unrelated to the meaning of the word, 0.5 if they wrote *disease* or its L1 translation *penyakit*, and one point if they wrote *plague* or the L1 translation *wabak.* These scoring guidelines were then applied by the first author in a first assessment round. Any cases of doubt (mostly about whether a response was close enough to the target meaning to be credited with half a point) were in a second round discussed with the aforementioned second assessor until agreement was reached.

Results and Discussion

Comprehension of the video content

Before turning to the results of the vocabulary test, it is worth ascertaining whether the students processed the video for its content, as per instruction. It is that purpose which the ten multiple-choice comprehension questions served. Given the generally high success rate (see below) on this test, it seems safe to say that the students indeed engaged with the content of the TED Talk. Somewhat against expectations, considering the substantial body of research which has demonstrated facilitative effects of captions for comprehension (Montero Perez, et al., 2013; but also see Rodgers & Webb, 2017), the group that watched the uncaptioned version of the video did as well as the group

that watched the captioned version. The average scores were 8.26 and 7.56, respectively. As the Shapiro-Wilk Test indicated that the scores in the uncaptioned condition were not normally distributed, the Mann-Whitney Test (i.e., the non-parametric counterpart of the t-test for independent samples) was used, and this revealed no significant difference between the two groups' scores ($z = 1.83$; $p = .067$).

Word recognition

One of the three components of the vocabulary test was the word recognition task. This was intended to probe the students' episodic memories of encountering the target words in the video. One of the proclaimed affordances of the use of captions is that it helps learners to segment speech and thus to discern lexical items in the speech flow. At first sight, this is corroborated by the test data. Altogether, the target words (i.e., the 15 words which occurred in the TED Talk) were identified 58.33% of the time by the students who had watched the captioned video, and this compares to a lower proportion, 52.16%, in the uncaptioned condition. The difference falls short of statistical significance, however, according to Yates Chi-Square: $\chi^2 = 3.57$; $p = .059$. It is worth mentioning that the students who watched the captioned video said 'yes' to more of the target words regardless of whether these were presented in written or in aural form (7.59% and 4.94% more, respectively), which suggests that test modality did not matter much in the word recognition task (unlike what was reported in Sydorenko's (2010) experiment with beginning L2 learners).

Interestingly, the students in the captioned condition were more inclined than their peers in the uncaptioned condition to also mark the words which did *not* occur in the video as having been encountered in the video. The ratios of such mistaken 'yes'

responses were 37.73% and 29.62%, respectively, and this difference was found to be significantly significant ($\chi^2 = 6.38$; $p = .012$). This casts some doubt on the reliability of the responses concerning the target words, especially in the group which had watched the captioned version of the video, who apparently believed a greater number of the test prompts overall corresponded to words used in the video. An ingenious feature of the research design in Montero Perez, et al. (2015) was to use two videos and to ask participants who claimed to remember encountering a given word to indicate in which of the two videos they had encountered it. This then served as supporting evidence that the participants were truly reporting episodic memories. As the present study used only one video, a question of this kind was unfortunately not included.

We can only speculate about why the students who watched the captioned video were more inclined than their peers who watched the uncaptioned version to mistakenly consider non-target words as ones which occurred in the video. According to the teacher who assisted with the data collection, the use of captioned video was uncommon in the English courses taken by her students so far. The Malaysian television culture does not habitually use captions either; instead, the use of L1 subtitles is customary. Given their lack of familiarity with captioned video, it is conceivable that the students in this study focused on the captions without assuming that these were exhaustive renderings of what the presenter of the TED Talk was saying. This may then have led them to make guesses at which of the words in the test were likely to have occurred in the video but not included in the captions. If so, their higher number of 'yes' responses to non-target words does not necessarily imply that their 'yes' responses to the actual target words were less reliable than those of the group that watched the uncaptioned version of the video. However, as said, this must remain speculation.

Orthographic representations

Another potential bonus of captioning is that learners may learn the orthographic form of words alongside their phonological form (e.g., Borras & Lafayette, 1994). That is why we asked the students not only to decide whether the aurally presented words occurred in the video but also to write them down. In both groups, students' written reproductions of the words exhibited various spelling mistakes—demonstrating the imperfect alignment of phonology and orthography in English—, such as *outbrake instead of *outbreak*, *appydemic instead of *epidemic*, *skatted instead of *scattered*, and *insectiside instead of *insecticide*. Rather surprisingly, the overall rates of spelling mistakes did not differ between the two groups: 67.03% and 69.63% for the captioned and uncaptioned conditions, respectively ($\chi^2 = 0.27$; $p = .603$). Having seen the written form of the words in the captions was apparently not sufficient to give the group that watched the captioned video a noticeable advantage over the comparison group in their development of accurate orthographic knowledge of these new words.

It is worth recalling in this regard that seven of these eight aurally presented target words occurred only once in the video (and so the students saw them only twice, as they watched the video twice). It is well known that repeated encounters matter for measurable incidental vocabulary acquisition (Webb, 2014a), and this includes spelling (Pigada & Schmitt, 2006). Also, unlike most other reading conditions, reading captions involves time pressure and therefore leaves little opportunity to contemplate the form of words. Besides, the task instruction encouraged the students to focus on the content of the TED Talk; it did not stimulate a focus on the linguistic packaging of that content. Given all of these conditions, one should not be over-optimistic about swift and

accurate uptake of the orthographic features of recently encountered words (Laufer, 2003).

Word meaning

When students indicated they remembered encountering a target word in the video, they were invited to interpret its meaning. The group that had watched the captioned video made such attempts in 57.14% of the instances where they had indicated recognition of the target words. This compares to 48.12% in the uncaptioned condition. The former group's greater inclination to propose word meanings was significant: $\chi^2 =$ 4.10; $p = .043$. So, watching the captioned video not only led students to indicate recognition of more of the target words; it also led them to make more interpretation attempts concerning these words.

That does not mean, of course, that these interpretations were always accurate. Interestingly, some of the proposed meanings appeared totally unrelated to the content of the video but instead seemed based on the students' analysis of the test prompt as such. For example, some students mistook *outbreak* for 'going out for a break' and some interpreted *remarkable* as 'something that can be marked again'. Some thought *immobile* referred a mobile telephone, a misinterpretation reminiscent of Laufer's (1997) observations about 'synforms' or deceptive look-alikes.

Nonetheless, a fair proportion of the meaning-guessing attempts were at least partially successful, and this demonstrates the considerable potential of authentic audio-visual input for incidental vocabulary acquisition (Webb, 2014b). Of a total of 160 meaning-guesses made by the group that watched the captioned video, 82 (i.e., 51.25%) received at least partial credit. This compares to 50 out of 128 attempts (i.e., 39.06%) in

the uncaptioned condition. The difference between these 'success' ratios is borderline significant: Yates $\chi^2 = 3.78$; $p < .052$.

While the above analyses provide some insight into the two groups' response behaviours, a more straightforward way of estimating the affordances of captioned video for uptake of word meaning is by comparing the two groups' test scores at the participant level. Table 1 sums up the descriptive statistics concerning the word meaning task (recall that the scores include half points given for partially correct responses).

Table 1: Scores on the word meaning task

	Written prompts (max = 7)			Aural prompts (max = 8)			Total (max = 15)		
	mean	median	SD	**mean**	median	SD	**mean**	median	SD
Captioned (n = 32)	**2.87**	2.75	1.38	**1.21**	1	1.07	**4.08**	4	1.71
Uncaptioned (n = 34)	**2.05**	1.5	1.62	**1.20**	1	1.5	**3.25**	2.5	2.91

The Shapiro-Wilk Test indicated that the scores on this word meaning test were not normally distributed in the uncaptioned condition, and so again the Mann-Whitney Test was used to further compare the two groups' scores. Taking all 15 target words together, the group that watched the captioned video was indeed found to perform significantly better than the group that watched the uncaptioned video: $z = 2.48$; $p = .013$; $d = .35$. However, this is due exclusively to their superior performance on the

section of the test where the words were presented in print: $z = 2.78$; $p = .005$; $d = .55$.

When the words were presented aurally, the two groups' average scores were virtually

identical ($z = 0.79$; $p = .43$). Compared to the data for the word recognition task, these

results thus indicate more clearly that having seen the written form of a word while

watching captioned video is helpful in a post-test which also presents the word prompt

in written form, but not necessarily so in a test where the word prompt is presented

aurally.

    Table 1 also shows that the test performance by both groups was better in the case

of the set of prompts presented in print as compared to those presented aurally. This

suggests that the aural test mode made the task more challenging overall (e.g., Chang &

Read, 2013). However, this needs to remain a tentative observation in the present study,

because the target words in the written and aural sections of the test were not

necessarily well matched in terms of learning difficulty. Meticulous efforts to try and

match the two sets for item difficulty were not deemed vital in the present study, which

addresses the question of whether test modality influences performance differently in

the two treatment groups rather than whether one test modality is inherently more

challenging than another.

    Regardless of modality, rates of successful interpretations varied widely across the

set of target words, from 0% to 50%. The target item attracting the highest number of

correct responses was *tablet*. This is not surprising, because in the video the TED Talk

presenter actually showed the audience a tablet while talking about it. It is a crucial

piece of information in the video. *Tablet* is also one of two words which occurred more

than once in the video. On the other hand, the meaning of the other word, *remarkable*,

which occurred as many as seven times, was supplied by only one of the 66 participants (and was thus one of the words whose meaning was least likely to be picked up). This illustrates that repeated encounters with a word is but one of several influential factors in incidental word learning (e.g., Webb, 2008).

Comparing the two treatment groups' performance on the word meaning task, it is striking that the standard deviation was considerably larger in the uncaptioned than the captioned condition. Some of the students in the uncaptioned condition appeared much better able than most of their peers to cope with the aural input and to derive word knowledge from it. In a similar vein, it appears that the mediating effect of input–test mode congruency varied considerably in the captioned condition, such that the correlation between the scores on the written and aural parts of the test was negligible in the group that watched the captioned video: $r^2 = .003$. By contrast, in the group that watched the uncaptioned condition, a significant positive correlation indicates that students who did relatively well on the aural part also tended to do relatively well on the written part of the test: $r^2 = .529$ ($p < .0001$).

Conclusion and Limitations

The first two research questions we sought to address in this study concerned the benefits of captioning for intermediate L2 learners' uptake of novel words as gauged by (a) a word recognition task and (b) a word meaning task. While the data of the word recognition task are to be interpreted cautiously because of the issues with guessing, the results generated by the word meaning task support the proposition that captions are beneficial for vocabulary acquisition. This result is consistent with the findings of the majority of previous studies (Montero Perez, et al., 2013; Montero Perez, et al., 2014).

The third research question, which constituted the main impetus for this study, concerned the role of test–modality congruency in gauging the merits of captions for vocabulary uptake. More precisely, like Sydorenko (2010), we explored the possibility that the use of written test prompts gives an advantage to treatment groups assigned to a captioned video condition. We found evidence for this role of test modality in the word meaning task, where the better scores achieved by those who watched the captioned video were indeed confined to the written prompts. This finding calls for a more cautious interpretation of the results reported in many previous studies on the merits of captioning, because these did not take the role of test modality into consideration.

It stands to reason that learning styles and learner profiles (e.g., L2 proficiency, L2 listening experience, familiarity with the L2 script) will play a part in evaluations of the usefulness of captioning (e.g., Winke, Gass, & Sydorenko, 2013). In addition, the purpose of captioning needs to be taken into account. For example, captions may be considered temporary scaffolding, part of preparing students for the ultimate challenge of unassisted L2 listening comprehension (Rodgers & Webb, 2017). From that perspective, weaning students' off captions and gradually introducing uncaptioned video when they are 'ready' for it appears well-justified (King, 2002).

Given the great diversity in learner profiles as well as in pedagogic uses of audio-visual materials, further investigations of the affordances of captions would certainly be welcome. The study reported here is just a modest addition to the available body of research on the subject, with several limitations in addition to ones already mentioned. One is the absence of a delayed post-test, which could have helped to determine to what extent the memories left by encounters with the words in the two video conditions were durable. Another is that the norming exercise with the more proficient student group to

select target words offers no absolute guarantee that these target words were unfamiliar to all the participants in the actual treatment groups. Like us, Sydorenko (2010) and Winke et al., (2010) avoided pre-testing, but they asked their participants post facto about their prior knowledge of the target words. In the present study, there was insufficient time at the end of the class to elicit such confirmation from the students whether given target words were indeed new to them—and questions could be raised about the reliability of such self-reported data, anyhow. It would nevertheless have been useful to establish the two treatment groups' equivalence in ways supplementary to the comparison of their most recent English examination scores. Given the focus on vocabulary acquisition, administering a version of the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001; Webb, Sasao, & Ballance, 2017) could have been particularly useful. It also needs to be borne in mind that the study reported here involved just one video (and only one particular genre—TED Talks) and that the use of captions was relatively novel to the student-participants. It is hard to predict whether the influence of captioning might evolve with prolonged exposure to this type of multimodal input. Studies have confirmed that captions definitely attract attention (e.g., Bisson, et al., 2014; Chai & Erlam, 2008; Montero Perez, et al. 2015; Sydorenko, 2010; Taylor, 2005), but how much attention may be variable (Winke, et al., 2013).

In conclusion, the findings of this study suggest that test modality matters in comparing the benefits of captioned and uncaptioned videos. Still, it is important to note that, altogether, the results were favourable of the use of captions. After all, while the captioned condition generated better test performance on the written word prompts (suggesting that captions helped the learners to focus on these words while watching the video), the test performance on the aurally presented prompts was the same in both

treatment conditions. This suggests that—with the intermediate-level learners who participated in this experiment—the use of captions did not compromise uptake of the phonological form of target lexis (which is also what Markam (1999) found in his experiment with advanced learners). Overall then, the findings presented here are still in agreement with the recommendation (e.g., Vanderplank, 2010) to use captioned audio-visual materials, at least if the aim is to foster incidental vocabulary acquisition.

References

Baltova, I. (1994). The impact of video on the comprehension skills of core French students. *Canadian Modern Language Review, 50*, 507–31.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning.* Amsterdam: John Benjamins.

Bisson, M.-J., Van Heuven, W.J.B., Conklin, K, & Tunney, R. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 31*, 693–713.

Borras, I., & Lafayette, R.C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal, 78*, 61–75.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading a Foreign Language, 20*, 136–163.

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.

Chang, A.-C., & Read, J. (2013). Investigating the effects of multiple-choice listening

    test items in the oral versus written mode on L2 listeners' performance and

    perceptions. *System, 41*, 575–586.

Coxhead, A.J., & Walls, R. (2012). TED Talks, vocabulary, and listening for EAP. *The*

    TESOLANZ Journal, 20, 55–58.

Freeborn, D. (2006). *From Old English to Standard English* (3rd ed.). Basingstoke, UK:

    Palgrave MacMillan.

Goh, C.C. (2000). A cognitive perspective on language learners' listening

    comprehension problems. *System, 28*, 55–75.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring

    Second Language Vocabulary through Reading. *Reading in a foreign language*, *11*,

    207–223.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A

    reappraisal of elaboration, rehearsal and automacity. In P. Robinson (Ed.),

    Cognition and second language instruction (pp. 258–286). Cambridge: Cambridge

    University Press.

King, J. (2002). Using DVD feature films in the EFL classroom. *Computer Assisted*

    Language Learning, 15, 509–523.

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical

    factors affecting the difficulty of vocabulary acquisition. In M. McCarthy & N.

    Schmitt (Eds.), *Vocabulary Description, Acquisition and Pedagogy* (pp. 140–155).

    Cambridge: Cambridge University Press.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review, 59*, 567–587.

Markham, P. (1999). Captioned, videotapes and second-language listening word recognition. *Foreign Language Annals, 32*, 321–327.

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

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

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*, 308–328.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Nation, I.S.P., & Heatley, A. (2002). *Range: A Program for the Analysis of Vocabulary in Texts*. (Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation)

Neuman B., & Koshinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*, 96–106.

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

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: a case study. *Reading in a Foreign Language, 18*, 1–28.

Rodgers, M. (2013). *English Language Learning through Viewing Television: An Investigation of Comprehension, Incidental Vocabulary Acquisition, Lexical Coverage, Attitudes, and Captions*. Unpublished PhD thesis, Victoria University of Wellington, New Zealand.

Rodgers, M.P.H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESLO Quarterly, 45*, 689–717.

Rodgers, M.P.H., & Webb, S. (2017). The effects of captions of EFL learners' comprehension of English-language television programs. *Calico Journal, 34*,

Rost, M. (2002). *Teaching and researching listening*. London: Pearson Education.

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

Schmitt, N., Schmitt, D., & Clapman, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55–88.

Scotto di Carlo, G. (2014). The role of proximity in online popularizations: The case of TED talks, *Discourse Studies, 16*, 591–606.

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

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

Takaesu, A. (2013). TED Talks as an extensive listening resource for EAP students. *Language Education in Asia, 4*, 150–162.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*, 422–427.

Tulving, E. (1993). What is episodic memory? *Current Directions in Psychological Science, 2*, 67–70.

van Zeeland, H. (2013). Vocabulary and listening. In C.A. Chapelle (Ed.), *The Encyclopaedia of Applied Linguistics*. New York: John Wiley and Sons.

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 1–24.

van Zeeland, H., & Schmitt, N. (2013) Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*, 609–624.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*, 191–210.

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language teaching*, *43*, 1–37.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, 24, 56–89.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*, 219–258.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading a Foreign Language, 15*, 130–163.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*, 232–245.

Webb, S. (2014a). Repetition in Incidental Vocabulary Learning. In C.A. Chapelle

(Ed.), *The Encyclopaedia of Applied Linguistics*. New York: John Wiley and Sons.

Webb, S. (2014b). Extensive viewing: Language learning through watching television.

In D. Nunan & J.C. Richards (Eds.), *Language learning beyond the classroom* (pp.

159–168). London: Routledge.

Webb, S., & Rodgers M.P.H. (2009). Vocabulary demands of television programs.

Language Learning, 59, 335–366.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test:

Developing and validating two new forms of the VLT. *ITL - International Journal

of Applied Linguistics, 168*(1), 34–70.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for

foreign language listening activities. *Language Learning & Technology*, *14*, 65–86.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by

foreign language learners: An eye-tracking study. *The Modern Language Journal,

97*, 254–275.

Appendix 1: Vocabulary Test

Instructions:
Please indicate whether you remember seeing/hearing the following words in the video
you have just watched. Then, try to explain the meaning of the words. You can choose
to give a synonym of the words or you can choose to translate the words into your
mother tongue. You do not need to give a synonym or translate the words you have *not*
seen/heard in the video. Put a dash (-) instead.
There are two parts to this quiz. The first part presents the words in writing and the
second presents them aurally.

No.	Word	Did this word appear in the video? Yes/No	Give a synonym/Translate this word
1.	fascinating	(yes)	

2.	propel		
3.	repel	(yes)	
4.	violent		
5.	awesome		
6.	equivalent	(yes)	
7.	dissemination		
8.	transmission	(yes)	
9.	volunteer	(yes)	
10.	trigger	(yes)	
11.	tiger		
12.	remarkable	(yes)	
13.	refundable		

This is the second part of the quiz. Listen to the words and write them down in the first column. Indicate in the second column whether you have heard/seen these words in the video. Then try to guess the meaning by giving a synonym or a translation.

14.	(immobile)	(yes)	
15.	(textile)		
16.	(outbreak)	(yes)	
17.	(uptake)		
18.	(pandemic)		
19.	(scattered)	(yes)	
20.	(epidemic)	(yes)	
21.	(pesticide)		
22.	(insecticide)	(yes)	
23.	(infection)		
24.	(synthetic)	(yes)	
25.	(authentic)		

26.	(convinced)	(yes)	
27.	(tablet)	(yes)	
28.	(goblet)		

Appendix 2: Comprehension Test

Instructions:
The following questions are about the TED Talks video that you have just watched.
There are 10 multiple-choice questions. The first five will be in written form; while the
remaining five will be read aloud by your teacher.
For the first 5 questions, please circle the correct answer (A/B/C/D). For the remaining
5 questions, write down the correct option (A/B/C/D) in the space provided.

1. According to the speaker, why do some people get bitten more often than others?
        A. Because of the smell of one's body
        B. Because of a particular skin type
        C. Because of the health status of a person
        D. Because of the gender of a person

2. According to the speaker, how do mosquitoes choose their "victims"?
        A. By choosing female over male "victims"
        B. By randomly choosing their "victims"
        C. By sniffing the body smell of the "victims"
        D. By choosing "victims" who have malaria

3. Which time of the day are mosquitoes most active?
        A. In broad daylight
        B. At dusk
        C. At dawn
        D. At night

4. What was the speaker's main purpose for conducting various experiments on mosquitoes?
        A. Because he wants to know whether mosquitoes prefer him or his wife
        B. Because he wants to catch the mosquitoes and use them as bait
        C. Because he wants to find ways to kill the mosquitoes
        D. Because he is interested in how mosquitoes breed

5. What kind of malarial mosquitoes is mentioned by the speaker?
        A. African malarial mosquitoes
        B. American malarial mosquitoes
        C. Asian malarial mosquitoes
        D. Australian malarial mosquitoes

6. Your answer: _____
7. Your answer: _____
8. Your answer: _____
9. Your answer: _____
10. Your answer: _____

The aurally presented questions:

6. According to the speaker, which part of the body do Dutch mosquitoes prefer to bite?
   A. Arm
   B. Foot
   C. Face
   D. Stomach

7. What was the name of the cheese mentioned by the speaker which attracts mosquitoes?
   A. Feta cheese
   B. Cheddar cheese
   C. Parmesan cheese
   D. Limburger cheese

8. According to the speaker, how can dogs help fight malaria?
   A. By training the dogs to destroy mosquito breeding sites
   B. By training the dogs to identify people who are infected with malaria
   C. By barking at the mosquitoes that approach their owners
   D. Dogs can't help fight against malaria

9. What was the third way of fighting against malarial mosquitoes, as mentioned in the video?
   A. By taking a pill which makes the mosquitoes sick and die.
   B. By catching the mosquitoes and putting them in a cage.
   C. By releasing the mosquitoes somewhere else.
   D. By applying a special cream to your skin.

10. Which of the following statements is TRUE according to the speaker's research findings?
   A. Not all mosquito species bite on the same part of the body.
   B. The fight against malaria is almost won.
   C. Mosquitoes like eating certain types of cheese.
   D. Mosquito larvae cannot be killed.