

EFFECTS OF CAPTIONING ON VIDEO COMPREHENSION AND INCIDENTAL VOCABULARY LEARNING

Maribel Montero Perez, iMinds - ITEC KU Leuven Kulak

Elke Peters, KU Leuven

Geraldine Clarebout, KU Leuven

Piet Desmet, iMinds - ITEC KU Leuven Kulak

This study examines how three captioning types (i.e., on-screen text in the same language as the video) can assist L2 learners in the incidental acquisition of target vocabulary words and in the comprehension of L2 video. A sample of 133 Flemish undergraduate students watched three French clips twice. The control group ($n = 32$) watched the clips without captioning; the second group ($n = 30$) watched fully captioned clips; the third group ($n = 34$) watched keyword captioned clips; and the fourth group ($n = 37$) watched fully captioned clips with highlighted keywords. Prior to the learning session, participants completed a vocabulary size test. During the learning session, they completed three comprehension tests; four vocabulary tests measuring (a) form recognition, (b) meaning recognition, (c) meaning recall, and (d) clip association, which assessed whether participants associated words with the corresponding clip; and a final questionnaire.

Our findings reveal that the captioning groups scored equally well on form recognition and clip association and significantly outperformed the control group. Only the keyword captioning and full captioning with highlighted keywords groups outperformed the control group on meaning recognition. Captioning did not affect comprehension nor meaning recall. Participants' vocabulary size correlated significantly with their comprehension scores as well as with their vocabulary test scores.

Keywords: Video, Listening, Vocabulary, Multimedia, CALL

APA Citation: Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of Captioning on Video Comprehension and Incidental Vocabulary Learning. *Language Learning & Technology* 18(1), 118–141. Retrieved from <http://llt.msu.edu/issues/february2014/monteroperezetal.pdf>

Received: May 10, 2012; **Accepted:** April 8, 2013; **Published:** February 1, 2014

Copyright: © Maribel Montero Perez, Elke Peters, Geraldine Clarebout, Piet Desmet

INTRODUCTION

The emergence of multimedia learning environments (Brett, 1995) and the overall accessibility of video (DVD, YouTube, etc.) have created important platforms for enhancing second language (L2) listening development (Vandergrift, 2011). These platforms are increasingly being used in classroom practice (Grgurović & Hegelheimer, 2007; Vanderplank, 2010) and provide learners with a number of listening support options, mostly realized in the form of a “technological overlay” (Robin, 2007, p. 109) such as native language (L1) subtitles (L2 video, L1 on-screen text), reversed subtitles (L1 video, L2 text), and captioning (L2 video, L2 text). This study focuses on captioning, which maximally exposes learners to the L2, and which has been proven to be beneficial for augmenting comprehension (Baltova, 1999; Chung, 1999; Markham, 2001) and fostering vocabulary learning (Markham, 1999; Sydorenko, 2010).

While extensive research has addressed the potential of full captioning for video comprehension and vocabulary learning, research studies have not provided answers to several questions. Researchers have suggested simplifying full captions (Garza, 1991; Winke, Gass, & Sydorenko, 2010) but research that has

attempted to reduce the textual density by offering only keyword captions has yielded inconclusive results (Guillory, 1998; Park, 2004). Additionally, research on vocabulary learning has identified attention as being a crucial component for vocabulary learning (Hulstijn, 2001), yet no study has looked at the effects of salience in the captioning line for drawing L2 learners' attention to target vocabulary and enhancing incidental vocabulary learning, that is, vocabulary learning as a by-product of listening for meaning (Gass, 1999).

This study expands on previous captioning research by investigating not only video with full captions but also video with keyword captions and video with full captions and highlighted keywords in order to reveal the effectiveness for L2 learners' vocabulary learning and comprehension of video content in the context of L2 French.

Literature Review

Our literature review focuses on three elements: (a) the importance of lexical coverage for successful listening comprehension and vocabulary learning, (b) the effectiveness of captions for L2 video comprehension, and (c) the effectiveness of captioned video for L2 vocabulary acquisition.

Research has indicated that comprehension of written or aural input is crucial for acquiring new linguistic items incidentally (Gass, 1997; Lee & Van Patten, 2003). Although Rost (2002) identified listening as one of the most important sources for L2 acquisition, only a handful of researchers have investigated the relationship between listening comprehension and vocabulary learning and tried to discover which factors affect comprehension and chances for word learning. In a series of studies, Webb (2010, 2011) and Webb and Rodgers (2009) investigated the importance of the lexical coverage of television programs in relation to incidental vocabulary learning and text comprehension. They found that knowledge of the 2,000 to 4,000 most frequent word families provides 95% coverage of television programs, which, according to Nation (2006), is the coverage required for incidental learning and adequate text comprehension. Webb and Rodgers' results indicated that to achieve 98% coverage, 5,000 to 9,000 word families are necessary, depending on the television genre. van Zeeland and Schmitt (2012) suggested that learners with a vocabulary size between 2,000 and 3,000 word families have 95% coverage, which is sufficient for adequate understanding of storytelling passages (audio only). Although the findings of the aforementioned studies present slightly different figures on the number of words needed, they provide accumulative evidence for the claim that vocabulary size is highly correlated with listening success (Staehr, 2009) and vocabulary learning (Neuman & Koskinen, 1992). A general finding, however, is that most L2 learners have not yet developed a sufficiently large vocabulary to understand L2 video. In spite of the richness of video material (Baltova, 1999), it is not necessarily appropriate for adequate comprehension and learning (Danan, 2004). One solution to help learners is to provide them with L2 subtitles or captioning.

The bulk of captioning research has focused on the effectiveness of captioning for L2 learners' comprehension of video content (Baltova, 1999; Chung, 1999; Garza, 1991; Huang & Eskey, 1999-2000; Markham, 2001; Neuman & Koskinen, 1992; Park, 2004; Winke et al., 2010). In general, researchers have found that captioning enhances comprehension of L2 clips. Bird and Williams (2002) suggested that captions aid speech decoding and segmentation by helping listeners visualize the speech stream and clearly indicating word boundaries. Captioning has therefore been characterized as a "mediating device" (Vanderplank, 1988, p. 280), helping the learner when automated sound-script recognition falls short. Markham (1999) suggested that captions also help the development of word recognition skills. By doing so, it makes ambiguous speech clearer the next time it will be encountered (Bird & Williams, 2002; Garza, 1991) and enables learners to successfully cope with input that is slightly above their actual proficiency level (Danan, 2004; Neuman & Koskinen, 1992).

While it is generally acknowledged that captioning improves L2 listeners' understanding (e.g. Baltova, 1999; Huang & Eskey, 1999-2000), Pujolá (2002) and King (2002) have argued that the presence of

captions makes it difficult to conclude whether comprehension scores reflect participants' listening or reading skills. Although this claim has not been tested empirically, studies have shown that the availability of captions does not compromise successful auditory processing (Bird & Williams, 2002; Markham, 1999, as cited in Danan, 2004; and Vanderplank, 2010). Other researchers have tried to find out if the amount of textual density in the captioning line can be reduced by exploring the potential of keyword captioned video (Guillory, 1998; Park, 2004), as suggested by Garza (1991) and more recently by Winke et al., (2010). Keywords are expected to tackle the problem of textual density (Guillory, 1998) while still providing support. Guillory's study examined to what extent keywords can support understanding of L2 video for beginning learners of French. The results showed that learners in the keyword group (receiving only 14% of the full captioning text) performed as well as the full captioning group on the comprehension questions. Moreover, the keyword group significantly outperformed the no captioning group. Unlike Guillory's study, Park's research (2004) indicated that only the more advanced students were able to benefit from keyword captions; lower-level students did not significantly outperform the no captioning group. It is, however, unclear to what extent this finding can be attributed to the keyword captions or to the video input selected for his study. Since Park used the same input for both lower-level and higher-level learners, there might have been a mismatch between lower-level learners' proficiency and the level needed for the video (Vanderplank, 2010; Winke, et al., 2010).

Although far fewer studies have investigated the effects of captioned video on L2 learners' vocabulary learning (e.g. Baltova, 1999; Danan, 1992; Sydorenko, 2010), two general conclusions can be drawn from the results of previous research. First of all, it has been shown that captions significantly help learners on written form recognition (Neuman & Koskinen, 1992; Sydorenko, 2010) and aural form recognition tests (Markham, 1999). However, Sydorenko (2010) found that learners in the video only group outperformed the captioning group on an aural form recognition test and therefore concluded that captioning success may depend on test modality. Yet, previous research has tended to suggest that captions help L2 learners isolate word forms (Winke et al., 2010) and pay attention to them, which may subsequently stimulate noticing of these forms. With regard to noticing, it is well documented in the literature on vocabulary acquisition that noticing unknown words in the input is the first step in the acquisition process (Huckin & Coady, 1999; Hulstijn, 2001). The crucial role of attention is also at the basis of Vanderplank's "speculative model" (1990, p. 228) on language learning through captioned video. In his model, the "taking out" of language from captioned video consists of both attention and adaptation. Vanderplank defines attention as a conscious selection process that is based on systematically "noting and gathering" (p. 229) information and on a reflective component in which learners notice a gap while comparing their L2 knowledge with the captioned video input. Adaptation, the selection of linguistic elements that learners pay attention to "for [their] own purposes" (Vanderplank, p. 229), aligns with Gass' position that learners have "their own focus of attention" (1999, p. 321).

Second, captions also help learners make form-meaning connections in the mental lexicon, which constitutes a crucial process in the acquisition of lexical items (Van Patten, Williams, & Rott, 2004). With regard to those form-meaning connections, studies have found that captions help learners not only to recognize the meaning of target words (Huang & Eskey, 1999-2000; Neuman & Koskinen, 1992) but also to provide translations (Sydorenko, 2010; Winke et al., 2010) or to produce the target forms themselves (Baltova, 1999; Danan, 1992). While these studies provide valuable information on the potential of captions for vocabulary acquisition, further research is needed to investigate whether captions can be enhanced in order to stimulate vocabulary learning.

As shown in this section, previous research has lent support to the use of captioning in classroom contexts. Yet, it also reveals a number of issues that require further research. One of the recurring questions has been whether captions can be enhanced by providing L2 learners with salient keywords rather than full captions (Garza, 1991; Guillory, 1998; Park, 2004), or by visually enhancing keywords in the captioning line (Winke et al., 2010). While two studies have examined the effectiveness of keywords

for L2 learners' understanding of video, they produced conflicting results (Guillory, 1998; Park, 2004). Moreover, although attention has been identified as a crucial component of vocabulary learning (Hulstijn, 2001), we are not aware of studies that have focused on the effects of salience in the captioning line, either through keywords or highlighted items in full captions, in the context of vocabulary learning. A study comparing full captions to (highlighted) keywords may provide more information on (a) what learners notice in the input and (b) how full captions and keywords can enhance attention and processing quality in terms of vocabulary gains.

Research Questions

This study investigates the potential of three types of captioned video, varying in the amount of text and the salience of the lexical items in the captioning line, for content comprehension and incidental vocabulary learning. In particular, we have included three captioning groups (see [Table 1](#)): full captioning (FC), keyword captioning (KC), and full captioning with highlighted keywords (FCHK) and compared their results with those of a control group (NC).

Table 1. *Overview of Experimental Conditions*

Name of condition	No captioning (NC)	Full captioning (FC)	Keyword captioning (KC)	Full captioning with highlighted keywords (FCHK)
Participants per condition	32	30	34	37
Amount of textual support	None	Full	Partial	Full
Amount of salience	No salience	No salience	Salience	Salience of keywords

We were guided by two main research questions:

- 1) Does the type of captioning have a differential effect on L2 learners' understanding of video content, as measured by three comprehension tests consisting of global and detailed questions?

We hypothesize that learners in the captioning groups will outperform the NC group because of the availability of on-screen text. With regard to the differences between FCHK, FC, and KC, we adopt a null hypothesis as earlier studies have provided no or inconclusive results in this respect.

- 2) Does the type of captioning have a differential effect on L2 learners' incidental learning of target vocabulary words, as measured by a form recognition, clip association, meaning recognition, and meaning recall test?

We expect KC and FCHK to outperform FC and NC and for FC to outperform NC on the form recognition and clip association tests because the lexical items' salience may help the groups to notice words (Brett, 1998) and associate them with the correct clip more easily. With regard to meaning recognition, we hypothesize that FCHK will outperform KC and FC as the availability of highlighted keywords and full captioning may facilitate the inference of word meaning and initial form-meaning mapping. We also expect the differences between the captioning groups to disappear for meaning recall as the latter might be hard to achieve after the relatively short exposure to the target words. Yet, the captioning groups are still expected to outperform NC. [Table 2](#) summarizes our hypotheses.

Table 2. Research Hypotheses

Test	Hypothesis
Listening comprehension	FC, KC, and FCHK > NC
Form recognition and clip association	KC and FCHK > FC > NC
Meaning recognition	FCHK > FC and KC > NC
Meaning recall	FC, KC, and FCHK > NC

METHODS

Participants

The participants were 133 undergraduate law students (55 males, 78 females) at a Flemish university ($M_{\text{age}} = 17.98$ years, $SD = .51$). All were native speakers of Dutch, except for four French-speaking students whose data were excluded from the analysis. All students had an obligatory course in Legal French, which focuses on communicative competence and legal vocabulary. Classes were organized in six parallel groups of approximately 25 students. Participants were informed that attending the experiment would count as two of the 14 obligatory hours of online training. The participants could be considered as (high-) intermediate learners of French, as measured by the self-designed vocabulary size test (see [Instruments](#) and [Results](#) section).

The four conditions (NC, FC, KC, and FCHK) were randomly assigned to the six groups. Four groups were assigned to one of the four conditions, while the other two groups were divided equally over the four conditions in order to obtain a balanced composition.

Materials

Video Selection

Since our study was embedded in a formal classroom setting, we selected three relatively short,¹ but authentic French clips² from a Belgian and Swiss current affairs program for native speakers of French. The clips, which were available online, had a single narrator and included short interviews in which at least one interlocutor was shown. The first video (2'25'', 376 words) discussed the production and export strategy of a French brewery. The second and third clip presented the marketing strategy (4'24'', 772 words) and history (3'32'', 576 words) of the Lego © factory. The first and second clip were used without modifications. We removed a two-minute section of the third clip because it required too much economic knowledge. The three clips were manually transcribed and captions were added with MAGpie (http://ncam.wgbh.org/invent_build/web_multimedia/tools-guidelines/magpie).

Target Word Selection

An important criterion for selection was that the clips contained a number of words that were very likely to be unfamiliar to the participants, that is, the target words (TWs) of this study. We adopted two procedures to verify familiarity with the TWs:

First, we compiled a list of 140 possible TWs appearing in the clips and a set of pseudowords. A representative group of students ($N = 40$) of the same university, with a proficiency level similar to that of the actual participants of this study, indicated whether or not they knew each of the 140 words. We retained all the words that were not known by at least 70% of the students, which resulted in a set of 20 TWs (see [Appendix A](#)). We did not set the cognition level at 100% because students did not need to provide the actual translation and may have overestimated their knowledge.

Second, prior knowledge of the TWs in the present study was controlled for by means of a pretest. The prior knowledge test was administered to the participants four weeks before the learning session and

contained the 20 potential TWs and 18 distracters (easier words). The test format was similar to the Vocabulary Knowledge Scale (VKS) (Wesche & Paribakht, 1996) and measured participants' depth of vocabulary knowledge. For each item, participants indicated the corresponding level (see [Instruments](#) section, [Table 4](#), Test 3). We left out the fifth level (*I can use this word in a sentence, write a sentence*) because it was very unlikely that the learners would be familiar with the TWs. In addition, it is hard to determine objectively whether the sentence demonstrates knowledge of the word (Bruton, 2009; Nation & Webb, 2011). Results of the prior knowledge test showed that learners were familiar with three TWs, which were subsequently excluded from the TWs (see [Appendix A](#)).

Of the 17 retained TWs, there were seven nouns, four verbs, and six multiword units (verb and noun or pronoun and verb). None of the TWs were cognates. Since we selected authentic videos, it was not possible to control for the frequency of the TWs.³ Four TWs occurred more than once in the clip (see [Appendix A](#)). All TWs appeared as highlighted keywords in the FCHK condition and in isolation in the KC condition. The four TWs with a higher occurrence were keyword captioned for every occurrence in the clip.

Five experienced lecturers of French were asked to verify the appropriateness of the clips for the target audience and to rate the correlation between the audio and visual images in the clips. In their judgment, the clips were appropriate and visual images were considered supportive of the dialog but did not provide explicit information on the meaning of the TWs. There were, however, contextual clues available in the sentences containing the TWs. These clues are important to infer word meaning (Nation, 2001) and establish initial form-meaning connections. For example, in the sentence *Lego est revenue à la une après avoir frôlé le naufrage, il y a dix ans à cause des jeux électroniques* (“Lego made it back to the top, after being on the verge of disaster, ten years ago because of electronic games”), the context makes it possible to infer the meaning of the underlined TWs.

Keyword Determination Procedure

We define keywords as words that are important for the meaning of the sentence or paragraph. The keywords were presented like the TWs: either in isolation (KC) or as highlighted words (FCHK). The same five experienced lecturers of French were asked to highlight keywords for comprehension in the three video transcripts (as in Guillory, 1998, and Park, 2004). We processed their selections, compared them with our keywords and selected a final set of keywords, representing 17.11% (or 295 out of 1,724 words) of the total number of words in the three videos.

Instruments

Vocabulary Size Test

Previous research has revealed that vocabulary size is linked to listening success (Staehr, 2009; van Zeeland & Schmitt, 2012; Webb & Rodgers, 2009) and vocabulary learning (Baltova, 1999; Webb & Rodgers, 2009). Vocabulary size is also claimed to give a rough estimate of learners' language proficiency (Milton, Wade, & Hopkins, 2010). Therefore, we designed a 50-item⁴ multiple choice test that comprised three parts corresponding to the following word frequency bands: 2,001-4,000 (21 items), 4,001-5,000 (15 items), and 5,001-7,000 (14 items), based on the Routledge (Lonsdale & Le Bras, 2009), Verlinde (Selva, Verlinde, & Binon, 2002), and DPC corpus (Paulussen, Macken, Trushkina, Desmet, & Vandeweghe, 2006) frequency lists. Every test item was a written multiple choice question containing four Dutch translation options of the item (see [Figure 1](#)). Although a study on aural input should ideally include a spoken vocabulary size test, previous studies have shown that the use of a written vocabulary size test did not compromise the findings regarding the significant correlation between vocabulary size and listening comprehension (e.g., Staehr, 2009; van Zeeland & Schmitt, 2012).

Frequency band 2,001–4,000	<ul style="list-style-type: none"> • élu • slim (smart) • verkozen (<i>elected</i>) • stabiel (stable) • verplicht (obligatory)
Frequency band 4,001–5,000	<ul style="list-style-type: none"> • concertation • optreden (performance) • toegeving (concession) • overleg (<i>deliberation</i>) • opvatting (belief)
Frequency band 5,001–7,000	<ul style="list-style-type: none"> • prodigieux • nietsnut (good-for-nothing) • gevaarlijk (dangerous) • veilig (safe) • buitengewoon (<i>extraordinary</i>)

Figure 1. Sample of vocabulary size test where the first row contains the French word and its correct translation is in italics.

Three Comprehension Tests

We developed a comprehension test for each clip. Together, the three tests included 41 items: 19 short open-ended questions, 14 true-false items, and eight combination items (see Table 3). All participants were native speakers of Dutch and were asked to answer the open-ended questions in their L1 (Buck, 2001). The development of the test items was inspired by Buck’s “competency-based” default listening construct (p. 114), which measures general understanding, detailed content, and inferencing ability. Of the 41 items, 25 items focused on understanding the main ideas and 16 items targeted more detailed yet relevant elements. We did not include inferencing questions due to two limitations: (a) we used relatively short clips, providing very concrete and factual information; and (b) in order to prevent a task effect on vocabulary learning, the comprehension tests did not focus on the sections containing the TWs, posing extra limitations on the information that could be included in the tests.

The comprehension questions targeted the keywords, which means that questions only focused on important ideas. Moreover, if we had focused on ideas that were not represented in the keywords, we would not have been able to investigate differences between NC and KC (see Table 3).

Table 3. Sample of Comprehension Questions

Question Type	Comprehension Questions*
Short open-ended question	1. Explain why the French Craftworks association was established. (video 1) Answer: In order to <i>stimulate the export</i> of French beer.
	2. Explain why Lego is a classic marketing example. (video 2) Answers: Lego has an <i>online platform</i> where customers can present their creations. <i>By posting pictures of their creations</i> on the <i>online platform</i> , they provide inspiration for new Lego sets.
	3. According to the interviewee, what caused the crisis at Lego? (video 3) (with picture of the interviewee). Answer: Lego wanted to <i>grow so fast</i> that they started to <i>manufacture new</i>

	<i>products</i> . They did however <i>not have the expertise</i> to do so.
True/False questions	Indicate if the statement is true or false and, if false, correct the sentence. Annick states that the brewery's export rate is satisfactory since they export more than the national average. (video 1) Lego refunds purchases when customers put pictures or videos of their Lego constructions on the Internet. (video 2)
Combination task	Who says what? Combine each statement with the corresponding picture** (video 2). « Il y a <i>une révolution</i> dans la brique de mon enfance » → (picture of toy store communication manager) « On a un laboratoire de 20.000, 200.000 <i>utilisateurs</i> » → (picture of marketing specialist) « C'est <i>un jouet</i> qu'on garde longtemps » → (picture of Lego retailer)

Note: *We checked the internal consistency of the 41 items and found acceptable reliability ($N = 133$, Cronbach's $\alpha = .73$).

** The pictures are not reproduced due to copyright restriction. The words in italics were keyword captioned for KC and FHCK.

Four Vocabulary Tests

It is generally assumed that vocabulary acquisition is not an “all or nothing phenomenon” (Laufer, Elder, Hill, & Congdon, 2004, p. 209) but an incremental process with noticing as initial step (Hulstijn, 2001). Therefore, we assessed vocabulary learning by means of multiple tests, as suggested by Nation and Webb (2011). We first measured three aspects of word knowledge: form recognition, clip association, and meaning recall.⁵ The form recognition test assessed whether learners were able to recognize the TWs, a selection of keywords that were not TWs, and some distracters. Learners ticked off “yes” if they thought the word appeared in the clips, and “no” if they did not (see Table 4, Test 1). This test did not include non-words but we controlled for guessing by also presenting a clip association test. The clip association test contained the same TWs, keywords, and distracters as the form recognition test and built on learners' answers on the previous test. If learners had checked “yes” on the form recognition test, they were asked to indicate in which clip the word had occurred (Brewery or Lego). This allowed us to check if they could associate the word with a vague meaning (see Table 4, Test 2). The form recognition and clip association tests were combined with a VKS identical to the one used in the prior knowledge test (see Table 4, Test 3). The VKS enabled us to check whether learners were able to translate the TWs into their L1 (meaning recall).

Table 4. Target Word “amertume” in the Four Vocabulary Tests

Test	
1	Form recognition Word used in the clips? <input type="radio"/> Yes <input type="radio"/> No
2	Clip association If yes, in which clip? <input type="radio"/> Brewery <input type="radio"/> Lego
3	VKS (meaning recall) Amertume I don't remember having seen this word before. I have seen this word before but I don't know the meaning. I think I know the meaning:... (provide translation).

		I am sure that this word means:... (provide translation).
4	Meaning recognition	Amertume 1. evenwicht (balance) 2. ontgoocheling (disappointment) 3. bitterheid (bitterness) 4. deskundigheid (expertise)

In addition to the first test, we also administered a multiple choice meaning recognition test that included only the 17 TWs in order to check if learners were able to recognize the translation of the TWs among four Dutch translation options (see Table 4, Test 4). The tests in Table 4 were all administered immediately after the treatment.⁶

Questionnaire

Participants completed a questionnaire, consisting of a five-point Likert scale, including statements on video comprehension (research question 1), vocabulary learning (research question 2), and the usefulness of captions for both tasks. The questionnaire data were used to help us clarify the findings of our quantitative research.

Procedure

Three months before the experiment, the materials and procedures were pilot tested with a group of 22 students in their last year of secondary school who were enrolled in a French summer course. The participants completed all the tests used in the present study and evaluated the materials by means of a questionnaire. Our main finding was that the comprehension tests were too easy because many participants achieved maximum scores. However, the questionnaire results showed that they considered the clips challenging. As a result, some questions were rewritten for the present study and extra open-ended questions were added. The results of the pilot test did not reveal problems with the vocabulary tests or the TWs.

One month prior to the learning session, participants in the present study completed a vocabulary size and prior knowledge test (see Figure 2). They were told that such tests are typically administered at the beginning of the academic year. However, they were not informed about the aim of the experiment.

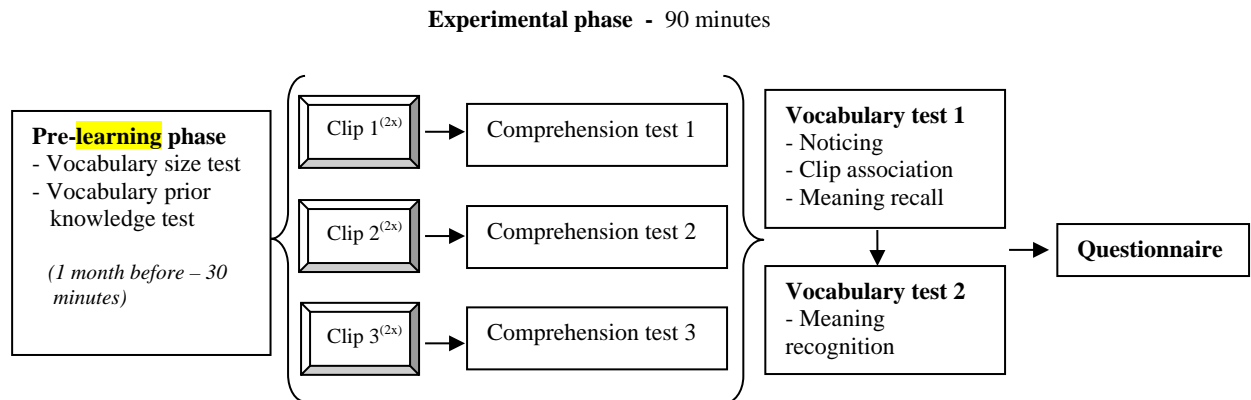


Figure 2. Overview of procedures.

The learning session took place in three computer rooms with a PC and headset for each participant. Students were told that the exercises were part of a research study on the use of video in the L2 classroom. We explained that they were going to watch three short clips. Learners completed a

comprehension test after viewing each clip twice (see [Figure 2](#)). Students were prompted to focus on the meaning of the clips and were forewarned of the comprehension tests. Participants were informed that each comprehension test contained short, open-ended questions, true-false questions, and combination tasks. They were not allowed to make notes, either during the clips or in between viewings. In order to be able to gauge incidental learning, we did not inform participants about the vocabulary tests, as suggested by Hulstijn (2001), who states that the incidental character is ensured at the level of test announcement. At the end of the learning session, all participants were debriefed about the aim of the experiment

We tested all the students in two consecutive sessions of 90 minutes. The supervisors made sure that participants involved in the first session did not see participants of the second session and could thus not inform them about the procedures, as this might have compromised the incidental nature of the vocabulary acquisition this experiment aimed to measure. All the tests were pencil-and-paper tests.

Scoring

Comprehension Tests

Full credit (1 point) was given for exact answers. Partial credit (0.5) was given for partially correct answers to open-ended questions. For example, partial credit was given when learners provided only one of the two parts of the answer to the question *Explain why Lego is a classic marketing example*.

[Lego has an online platform where customers present their creations.]/[The customers provide inspiration for new Lego sets.]

Each true-false item consisted of two tasks - (a) indicate whether the sentence is true or false and (b) if false, correct the sentence - and had a maximum score of 2 points.

Vocabulary Tests

One point was given for each correct answer on the form recognition, clip association, and meaning recognition tests. The VKS could have been scored in two ways (Bruton, 2009, p. 294): we could have either used the level number or recoded answers into known words (i.e., correctly translated words, 1 point) and unknown words (0 points). Because we used the VKS to measure meaning recall rather than progress in the scale (Nation & Webb, 2011), we chose to recode into binomial scoring (see [Table 5](#)).⁷

Table 5. Scoring of VKS

VKS	Level	Binomial scoring
I don't remember having seen this word before.	1	0
I have seen this word before but I don't know what it means. Or: wrong answer at level 3 or 4.	2	0
I think I know the meaning of the word and correct translation.	3	1
I am sure that this word means ... and correct translation.	4	1

Although most TWs were unknown to students, we wanted to account for their minimal prior knowledge. We did so by calculating retention scores (Horst, Cobb & Meara, 1998; Peters, 2006, p. 77) for the meaning recognition and meaning recall test (see Equation 1):

$$(1) \quad \frac{(\text{testscore} - \text{priorknowledge})}{(\text{max test} - \text{priorknowledge})}$$

As shown by the equation, we calculated the difference between the number of correct responses on the

test (meaning recognition and meaning recall) and the number of known words, as measured by the prior knowledge test, and then divided this number by the total number of TWs (max test) minus the prior knowledge score. Since known words were dismissed from the analysis, the result is a retention score representing the percentage of the actual learning gains.

Analyses

We conducted a multivariate analysis of covariance (Tacq, 1997) for each of the two research questions. The independent variable was type of captioning (NC, KC, FC, and FCHK); the covariate was the score on the vocabulary size test. There were two sets of dependent variables: the comprehension tests (research question 1) and the vocabulary tests (research question 2). We set the significance level of the p -value at .05 in all statistical analyses.

As an effect size measure, we used partial eta squared (η_p^2), which refers to the proportion of total variance explained by an effect “in which the effects of other independent variables and interactions are partialled out” (Richardson, 2011, p. 135). Effect size values have the advantage of being independent of sample size. We used Cohen’s rules of thumb (1988) for interpretation: small, $\eta_p^2 > .0099$, medium, $\eta_p^2 > .0588$, and large, $\eta_p^2 > .1379$.

RESULTS

Vocabulary Size Test

Table 6 summarizes the test scores for each of the three frequency bands and shows that participants were most proficient on the first part, which assessed knowledge of the 2,001 to 4,000 most frequent words. Scores decreased as the word frequency level increased (see Table 6). Since vocabulary size is an estimate of language proficiency, scores indicate that participants’ proficiency level in the current study ranged from intermediate to high intermediate. This is also the required level at the end of high school education according to the criteria of the Common European Framework of Reference (Council of Europe, 2011). A one-way ANOVA revealed that the participants in the four conditions did not differ significantly in terms of vocabulary size, $F(3, 129) = 0.01, p = .998, \eta_p^2 < .01$. The 50-item size test had an acceptable reliability index ($N = 133$, Cronbach’s alpha = .78) and was used as a covariate in further analyses in order to control for learner variables.

Table 6. Mean Scores and Standard Deviations on Vocabulary Size Test

		All students ($N=133$)		NC ($n = 32$)		FC ($n = 30$)		KC ($n = 34$)		FCHK ($n = 37$)	
Word Frequenc	Items	M	SD	M	SD	M	SD	M	SD	M	SD
2,001–4,000	21	16.47	2.63	16.81	2.09	16.17	3.12	16.41	2.35	16.46	2.91
4,001–5,000	15	9.87	2.20	10.03	2.07	9.83	2.51	9.97	2.10	9.68	2.21
5,001–7,000	14	7.83	2.56	7.44	2.99	8.27	2.63	7.68	2.54	7.97	2.12
Total score	50	34.17	6.13	34.28	6.20	34.27	7.12	34.06	5.65	34.11	5.89

Research Question 1

We checked the correlations between the three comprehension tests and found that the scores of comprehension test 1 correlated significantly with test 2 ($r = .39, p < .001$) and test 3 ($r = .30, p = .001$); test 2 correlated significantly with test 3 ($r = .35, p < .001$). In order to take into account the relationship between the dependent variables, we performed a MANCOVA (Tacq, 1997).

Table 7 summarizes mean scores on the comprehension tests. The results of the MANCOVA contradicted

our initial hypotheses and showed type of captioning had no significant effect of on comprehension scores, Wilks's lambda, $F(9, 301.93) = 0.40, p = .935, \eta_p^2 = .01$.

Table 7. Mean Scores and Standard Deviations on the Comprehension Tests

	All students (<i>N</i> = 131)		NC (<i>n</i> = 32)		FC (<i>n</i> = 28)		KC (<i>n</i> = 34)		FCHK (<i>n</i> = 37)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Comprehension 1	11.21	2.63	10.95	2.76	11.70	2.32	10.79	2.28	11.43	3.02
Comprehension 2	4.81	1.33	4.78	1.30	5.11	1.46	4.63	1.43	4.78	1.19
Comprehension 3	12.11	2.88	12.19	2.75	12.50	3.68	11.91	2.48	11.92	2.73

Note. The maximum scores of comprehension tests 1, 2, and 3, were 16, 7, and 18, respectively.

The covariate vocabulary size was significantly related to the comprehension scores, Wilks's lambda, $F(3, 124) = 8.81, p < .001$, and had a large effect: $\eta_p^2 = .18$. The positive *b*-values for each comprehension test (see Table 8), indicated a positive relationship between vocabulary size and comprehension: the larger one's vocabulary, the better the comprehension score.

Table 8. *F*-statistics for Vocabulary Size

Test	Source	df	F	p	η_p^2	b
Comprehension 1	Voc size (covariate)	1	10.11	.002*	.07	.12
Comprehension 2	Voc size (covariate)	1	5.93	.016*	.05	.05
Comprehension 3	Voc size (covariate)	1	21.59	<.001**	.15	.18
	Error	126				

Note. * $p < .05$. ** $p < .001$.

Research Question 2

We checked the correlations between the four dependent variables measuring vocabulary learning (cf. procedure research question 1). Table 9 displays significant correlations between all tests.

Table 9. Correlations Between the Four Dependent Variables Measuring Vocabulary Learning

	Form recognition	Clip association	Meaning recognition	Meaning recall
Form recognition	1.00	.85*	.46*	.40*
Clip association	.85*	1.00	.52*	.43*
Meaning recognition	.46*	.52*	1.00	.45*
Meaning recall	.40*	.43*	.45*	1.00

Note. * $p < .001$.

We found the highest mean scores on the form recognition test (see Table 10). The retention scores were highest for the meaning recognition test. Results of the MANCOVA in Table 11 indicate that type of captioning significantly affected vocabulary learning, Wilks's lambda, $F(12, 301.91) = 4.63, p < .001, \eta_p^2$

= .14. Vocabulary size was significantly related to vocabulary learning, Wilks's lambda, $F(4, 114) = 16.64, p < .001, \eta_p^2 = .37$.

Table 10. Mean Scores on the Vocabulary Tests

	All students ($N = 122$)		NC ($n = 30$)		FC ($n = 27$)		KC ($n = 31$)		FCHK ($n = 34$)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Form recognition	9.87	3.17	7.13	3.26	11.07	2.54	11.16	2.33	10.15	2.82
Clip association	7.63	2.90	5.73	3.17	8.78	2.52	8.23	2.17	7.85	2.84
Meaning recognition	.59	.14	.53	.14	.60	.14	.62	.09	.63	.17
Meaning recall	.14	.20	.13	.22	.16	.22	.13	.17	.14	.19

Note. The smaller number of participants on the vocabulary tests is due to the fact that some participants accidentally skipped one or two pages due to double-sided printing of the test, which consisted of 9 pages. The maximum score on the form recognition and clip association test was 17.

The analysis of the between-subjects effects (see Table 11) showed that type of captioning significantly affected the first three components of vocabulary knowledge tested, that is, form recognition, clip association, and meaning recognition. Vocabulary size scores were significantly related to all four vocabulary components, and b -values reported in Table 11 indicate a positive relationship: the larger one's vocabulary, the more vocabulary gains.

Table 11. Results of MANCOVA on Vocabulary Acquisition

Test	Source	df	F	p	η_p^2	b
Form recognition	Type of captioning	3	14.46	<.001**	.27	
	Voc size (covariate)	1	7.91	.006*	.06	.12
Clip association	Type of captioning	3	7.51	<.001**	.16	
	Voc size (covariate)	1	14.92	<.001**	.11	.16
Meaning recognition	Type of captioning	3	4.85	.003*	.11	
	Voc size (covariate)	1	46.90	<.001**	.29	.01
Meaning recall	Type of captioning	3	.03	.993	.00	
	Voc size (covariate)	1	37.27	<.001**	.24	.02
	Error	117				

Note. * $p < .05$. ** $p < .001$.

The results of the post hoc analysis (Bonferroni correction) partially confirmed our hypotheses: captioned groups significantly outperformed the control group on form recognition and clip association. However, no differences between the captioning groups were found. For meaning recognition, only KC and FCHK significantly outperformed NC. No significant differences between KC, FCHK, and FC were found. Contrary to our hypothesis, we found no differences between the captioned groups and the control group

on the meaning recall test.

The effect sizes showed a decreasing effect of type of captioning (FC, KC, FHCK, or NC) as the vocabulary component tested increased in difficulty. The effect size of type of captioning was largest for form recognition ($\eta_p^2 = .27$) and clip association ($\eta_p^2 = .16$) and smaller for meaning recognition ($\eta_p^2 = .11$). The opposite pattern was found for the covariate (vocabulary size): the deeper the knowledge component tested, the larger the effect size of the covariate (see Table 11).

Questionnaire

The descriptive statistics for each questionnaire statement are listed in Table 12. Results of the comprehension statements revealed that all conditions gave similar scores regarding the understanding and difficulty level of the questions.

Table 12. Questionnaire Results

	All students (<i>N</i> = 133)		NC (<i>n</i> = 32)		FC (<i>n</i> = 30)		KC (<i>n</i> = 34)		FCHK (<i>n</i> = 37)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Comprehension (RQ1)										
1. I didn't understand the French video clips and I had no idea what to answer to the comprehension questions.	2.00	.86	2.03	.82	1.83	.87	1.91	.87	2.19	.88
2. I think viewing the video twice was important to help me understand it.	3.38	1.26	3.53	1.22	2.97	1.33	3.44	1.24	3.51	1.24
3. I would have preferred seeing the video three times.	2.77	1.30	3.12	1.31	2.33	1.24	2.71	1.22	2.86	1.36
4. The comprehension questions were easy (1); average (2); difficult (3); very difficult (4).										
Video 1	1.86	.76	1.97	.82	1.70	.79	1.79	.85	1.95	.58
Video 2	1.79	.75	1.75	.72	1.83	.75	1.65	.85	1.92	.68
Video 3	2.22	.87	2.13	.91	2.13	.86	2.12	.98	2.46	.69
Vocabulary (RQ 2)										
5. I was able to infer word meaning from the textual context.	3.49	.88	3.66	.97	3.37	.96	3.50	.83	3.43	.80
6. The video images made it possible for me to infer the meaning of unknown words.	3.13	.99	3.06	1.16	3.27	.94	3.06	.92	3.14	.95
7. If I had known that vocabulary exercises were to follow, I would have paid more attention to the words.	3.82	1.24	3.84	1.08	3.53	1.36	3.94	1.35	3.92	1.16
Usefulness of (keyword) captioning (RQs 1 & 2)										

8. I did not need extra support to understand the video.	2.96	1.16	3.66	.90	2.60	1.10	2.38	.99	2.22	1.08
9. If I had been able to activate captioning, I would definitely have done that.			4.25	1.05						
10. I managed to understand the videos because (keyword) captioning was available.	3.60	1.05			3.90	.89	3.00	1.10	3.92	.89
11. I was distracted by the presence of captioning. Consequently, I focused less on the audio.	2.23	1.15			1.73	1.05	2.62	1.02	2.27	1.22
12. Captioning helped me to notice unknown words.	3.51	1.19			3.60	1.28	3.68	1.04	3.30	1.24
13. I learned new words thanks to the availability of (keyword) captioning.	3.30	1.10			3.52	1.15	3.18	1.14	3.24	1.01

Note. 1 = I do not agree at all, 5 = I completely agree.

With regard to the results of the vocabulary statements, we observed that all groups considered the textual context and images helpful for word meaning inferences. Moreover, participants would have focused more on the words if they had been forewarned of a vocabulary test. Learners in the NC group were most confident about not needing extra support, but would activate captions if available.

When asked about the usefulness of captions, the FC and FCHK groups considered captioning more useful than the KC group. The latter also provided the highest, yet still average, mean score for the distraction induced by the support. The captioning groups were equally positive about the usefulness of captions for recognizing and learning new words.

DISCUSSION

This study was motivated by the need to explore variations on captioning, either by providing the learner with more salience and/or less textual density. The results suggest that (a) the type of captioning did not affect comprehension scores; (b) learners' vocabulary size positively correlated with comprehension and incidental vocabulary learning; (c) the presence of on-screen text supported receptive vocabulary learning; and (d) salience induced by KC or FCHK did not result in greater vocabulary gains than captioning without salience (FC).

Research Question 1

The analyses revealed that participants in all conditions achieved similar scores on the comprehension tests, which is at variance with our hypothesis. Thus, our findings only provide limited support for the use of captions and keyword captions and therefore contradict previous research (e.g. Baltova, 1999; Chung, 1999; Markham, 2001). An important explanation for the similar test scores might lie in the information targeted by the comprehension questions. First of all, the comprehension questions did not focus on the TWs, in order to prevent an effect of the comprehension tests on vocabulary retention. Because learners did not need the TWs to answer to the comprehension questions, the differences between the four conditions might have been minimized. Perhaps the parts containing the TWs did cause problems for the NC group, but this was not reflected in the comprehension test scores as the questions did not require the TWs occurring in the captions. Second, because of the nature of the clips, which presented concrete factual information, we could only ask literal comprehension questions, rather than a combination of literal and inferencing questions (Buck, 2001). Third, it is possible that captioning groups decoded the

clips better than the NC group (Bird & Williams, 2002), a hypothesis that is supported by the higher scores of the FC, KC, and FCHK groups on the form recognition test (cf. research question 2). Yet, our findings suggest that a better decoding does not necessarily result in a better understanding of the message.

It could be hypothesized that the clips were too easy for the participants and therefore minimized the comprehension differences between the groups. Yet, what needs to be stressed is that the groups achieved only intermediate scores on the comprehension tests, which shows that content understanding was challenging for the participants. Although the usefulness of captions was not reflected in higher comprehension scores, questionnaire results indicated that participants in the FC and FCHK groups considered captions useful for understanding the clips (see Table 12, question 10). While the NC group indicated that they did not need extra support to understand the videos, they unambiguously indicated that they would use full captions if they had been available (see Table 12, question 9).

With respect to the keyword captions, we found that salience of the keywords for the KC and FCHK groups did not help learners to outperform FC and NC on the comprehension test. Our results are in line with Park's findings for lower intermediate and intermediate learners (2004) and therefore contradict research on reading comprehension in which it has been demonstrated that salience and attention allocation are important for successful text comprehension (Gaddy, van den Broeck, Sung, 2001). Moreover, learners in the KC group reported lower mean scores on the usefulness of keyword captions when compared to the FC and FCHK group (see Table 12, question 10). In order to reveal the potential of keyword captioning, further research is necessary.

Finally, learners' scores on the vocabulary size test were significantly related to their comprehension scores, and had a large effect size. Our result provides additional evidence in support of previous research on this relationship (e.g., Staehr, 2009; van Zeeland & Schmitt, 2012).

Research Question 2

Form Recognition and Clip Association

The results of our analyses partially confirm our hypothesis and reveal that the captioning groups (FC, KC, and FCHK) significantly outperformed the control group on the form recognition test. Although mean scores were generally lower for clip association, we found a similar pattern. Overall, our results confirm the usefulness of captions at the decoding level and their capacity to help learners to isolate words (Winke et al., 2010) and pay sufficient attention to these items. Our results are consistent with Vanderplank's model (1990), in which the role of subtitles was considered crucial to taking out words from captioned input. This ties in with previous research that has found positive effects of captions for enhancing written form recognition (e.g. Neuman & Koskinen, 1992; Sydorenko, 2010). On the other hand, we did not find evidence for the hypothesis that the salient KC and FCHK groups would outperform the FC group on the form recognition and clip association test. Interestingly, the availability of on-screen text, rather than salient input, had an overall positive effect on form recognition and clip association. This finding was confirmed by our questionnaire results (see Table 12, questions 12 and 13) which showed that the FC, KC, and FCHK groups found the availability of captioning equally useful for recognition and word learning.

How can we explain that salience did not enhance noticing? In his model, Vanderplank (1990) claimed that the "taking out" of language consists of a combination of attention and adaptation. More particularly, he stated that learners might pay attention to words and select "language attended to for own purposes" (p. 229). Although keywords were not made salient for the FC group, learners might have noticed these words for different reasons.

One possibility is that unfamiliarity with the TWs drew learners' attention and produced a "conscious focus on form" which occurs "particularly when new or striking expressions are used" (Vanderplank,

1988, p. 276). By comparing the language used in the captioned video with their own lexis, learners might have noticed gaps in their knowledge (Gass, 1997).

Another reason may be found at the level of test announcement, that is, the comprehension test might have directed the learners' attention towards particular words even though they were not textually salient. Hulstijn (2001, p. 268) suggested that, on the one hand, learners can "serve under an intentional condition" when trying to understand the text with the subsequent comprehension test in mind. On the other hand, they may "serve under an incidental condition in that they are being exposed to unfamiliar words" and are not aware that a vocabulary test will follow (p.268). Because learners were prompted to focus on the meaning of the clips and were only forewarned of the comprehension test, learners in the FC group might have recognized words as well as the salient groups because they considered them important for the test and the meaning of the clips. This explanation seems plausible since the majority of the TWs were considered keywords for comprehension by our experienced lecturers. Moreover, this finding confirms the claim that learners can simultaneously attend to form and meaning when the form is important for the meaning (Baltova, 1999; Van Patten, 1990).

Meaning Recognition and Recall

In order to determine the quality of processing, we included a meaning recognition and recall test. Although we hypothesized that participants in the FCHK group would outperform participants in the other conditions on the meaning recognition test, the group scored only slightly, but not significantly, higher than the KC and FC groups. The results of the meaning recognition test differed from the previous two vocabulary tests in that KC and FCHK significantly outscored only the NC group and therefore seem to have processed the TWs more elaborately than the NC group. The FC group did not differ significantly from the NC group. This finding might be explained by the fact that the TWs were better isolated for the salient keyword groups. Prince (1996, p. 489) underlined the importance of "isolating the word from the context, so that context provides the means to identify the meaning of the new word." Yet, since the salient keyword groups did not significantly outperform the FC group, this result should be interpreted cautiously.

In order to find out how captioning might have helped word meaning inferences, learners were asked about the usefulness of contextual and visual clues. Results revealed that learners in all conditions considered the textual context most helpful and reported very similar scores (see [Table 12](#), questions 5 and 6). A similar appreciation was found for the support of visual clues, although scores were slightly lower. Given the rather abstract nature of most TWs, the audio-video correlation could indeed only provide scarce visual clues.

Generally speaking, we found that participants made initial form-meaning connections for a series of TWs. But which conditions fostered the most qualitative vocabulary processing? The results of the meaning recall test were low and, contrary to our hypothesis, no differences were found between the groups. These findings do not support previous research (e.g. Danan, 1992; Sydorenko, 2010; Winke et al., 2010), which showed beneficial effects of captioning on meaning recall. Our study differs, however, from the cited articles in that they focused on low intermediate and beginning students respectively.

Plausible explanations for the low gains might be the following: (a) Since captioning as such does not provide concrete information on word meaning, learners are required to construct meaning based on inferring processes. Yet, the meaning of difficult words may remain unknown even after two or more viewings. (b) Because of the "real-time nature" (Buck, 2001, p. 6) of listening, learners are left with very little time to infer word meaning from context, which might prevent them from hearing the rest of the video (Goh, 2000). (c) It has been shown that inferring word meaning is a difficult, slow, and often an unsuccessful process (Liu & Nation, 1985). Because we wanted to make sure that words were unfamiliar, we were forced to focus on low frequency words, which might have a higher "learning burden" (Laufer, 2005, p. 234). Thus, the meaning recall test was a very demanding test because participants were expected

to watch the clips, read the captions, remember the content, derive the meaning of unknown words successfully, and finally remember the meaning of a set of TWs after only two viewings of the clips.

Vocabulary Size

Our study corroborated the findings of previous research on the importance of vocabulary size for successful vocabulary learning (Webb, 2010; Webb & Rodgers, 2009). We found medium and significant effect sizes of vocabulary size for form recognition and clip association. The larger one's vocabulary size, the less decoding load a video presents and the more time learners can spend on specific lexical items (Goh, 2000; Pulido, 2007). Yet, the effect sizes for vocabulary size were considerably lower than the effect size for type of captioning, which suggests that captioning is crucial to form recognition and clip association.

For meaning recognition and recall, we found very large effect sizes for vocabulary size; these effects were considerably higher than the effect sizes for type of captioning. Indeed, it is well documented that greater lexical knowledge seems to facilitate guessing from context (Liu & Nation, 1985) and it has been characterized as "a critical factor" (Nation, 2001, p. 233) for successfully inferring word meaning.

Pedagogical Implications

When teachers intend to use video for stimulating vocabulary acquisition, they should be encouraged to use captioning because it might facilitate students' recognition of unknown words and their making initial form-meaning connections. Our results provide limited support for the use of on-screen text for improving content comprehension. We therefore recommend that teachers adopt a "staged video approach" (Danan, 1992) in which they gradually decrease the amount of text. Teachers could show the same video first with FCHK or FC, then with KC and finally in a NC mode. By doing so, they might help learners to progressively decrease the amount of support while at the same time optimizing word recognition and chances for word learning.

Teachers should also be encouraged to address the importance of vocabulary size as it has a positive effect not only on successful comprehension but also on vocabulary learning. Teachers could provide opportunities for learners to enlarge their vocabulary size through exposure to captioned video or by means of other incidental (e.g., extensive multimedia reading) or intentional (explicit vocabulary learning) tasks.

CONCLUSION AND LIMITATIONS

Although this study has tried to respond to the need to explore variations on standard captioning, it also presents a number of limitations. First of all, we could not assess understanding of the complete clips because our comprehension tests had to focus on the parts that did not contain the TWs. Moreover, the clips were short, which presents limitations in terms of the type of information targeted by the questions. Our results therefore provide only limited insight into how different types of captioning can improve comprehension. Second, this study has focused on a set of predefined TWs. Yet, learners might have noticed and established form-meaning connections of other words in the videos they were not familiar with. Because our vocabulary tests only included predefined TWs, we are unaware of any other learning gains (Pulido, 2007). Third, we assessed noticing using an indirect measure (i.e., the form recognition test). More accurate data on what exactly induced form recognition, taking into account potentially moderating variables such as frequency of occurrence and the part of speech of the TWs (Webb, 2007), might reveal when and why learners notice certain items. Finally, the written format of the vocabulary posttests may have favored the captioning groups to some extent. Unlike the control group, they also encountered the target words in their written form.

As a conclusion to this paper, we would like to indicate three future research directions. First of all, future studies might investigate the extent to which captioning, and keyword captioning in particular, can be

enhanced by adding access to the target words' meaning, such as a gloss containing the L1 translation (Sydorenko, 2010; Webb, 2010). Research questions could focus on the use of such glosses and their effectiveness for comprehension and vocabulary learning. Second, current research on captioning has almost exclusively measured the value of short clips, by means of isolated experiments. As one reviewer suggested, it may be interesting to look at the potential of captioning when used with full-length TV programs (Rodgers & Webb, 2011). We should also encourage longitudinal research in order to investigate whether systematic and long-term exposure to captioned video affects learners' use of captions and the effectiveness of these captions in terms of improved listening comprehension. Third, the availability of on-line measures such as eye-tracking might shed more light on the time learners spend on the captioning line (Winke, Gass, & Sydorenko, 2013) and provide an objectified measure to study the role of attention in vocabulary learning (Godfroid, Housen, & Boers, 2010). Further research on the topics mentioned above will undoubtedly allow us to gain a greater insight into how and when captioning can help language learning.

APPENDIX A. List of 20 Target Words

Target word	Type of word	Clip	# encounters in clip
Larguer	verb	Lego (2)	1
faire un tabac	multiword	Lego (2)	1
Assaut	noun	Lego (1)	1
fermentation	noun	Bénifontaine	1
ça cartonne	multiword	Lego (1)	1
*berceau	noun	Lego (2)	1
amertume	noun	Bénifontaine	1
Houblon	noun	Bénifontaine	3
*solidifier	verb	Lego (2)	1
gravir les échelons	multiword	Lego (2)	1
Mûrir	verb	Bénifontaine	1
Malt	noun	Bénifontaine	3
être à fond	multiword	Lego (1)	1
se disperser	verb	Lego (2)	1
*récompenser	verb	Lego (1)	1
doper les ventes	multiword	Lego (2)	1
frôler le naufrage	multiword	Lego (1)	1
Brasser	verb	Bénifontaine	2
divertissement	noun	Lego (1)	1
Levure	noun	Bénifontaine	4

Note. Words marked with * were not included in the analysis because too many participants were already familiar with the word before the experiment (as measured by the vocabulary prior knowledge test).

NOTES

1. Because a formal learning context differs from normal viewing situations, a number of practical limitations such as time constraints inevitably played a role in the selection of clips. In Flemish high schools for example, one learning session lasts only between 50 and 60 minutes, which would make it difficult to use long clips and exploit them in a meaningful way. In this context, short clips present ecologically valid L2 viewing and practicing opportunities. The length of the clips in this study was also similar to the clips used in previous studies on captioning effects (e.g. Park, 2004; Sydorenko, 2010; Winke et al., 2010).
2. “Video” and “clip” are used interchangeably.
3. The differences in occurrences of the TWs may be considered a limitation of the clips used. Although frequency of the TWs may have played a role for vocabulary learning in this study, as suggested by previous research (Webb, 2007), results show that learners were able to recognize more than the four TWs with higher occurrence (see [Table 10](#)). Nonetheless, results do not imply that they necessarily noticed the four TWs with higher occurrence correctly. A detailed analysis of the aspects that induced form recognition and vocabulary learning might clarify this aspect but was beyond the scope of the present study.
4. The initial vocabulary size test consisted of 53 items. As the reliability test indicated that three items correlated negatively, these items were left out of the analysis.
5. The concepts of meaning recognition and meaning recall are not clearly defined. We adopt the distinction made by Laufer et al. (2004) between active and passive recognition and recall, whereby passive recognition consists in “choosing the meaning of the target word from the four options provided” (p. 207) and passive recall consists in translating the target form.
6. This study did not include delayed post-testing. Hulstijn (2003) argued that a study measuring the effectiveness “during a learning session in which words are presented for the first time, requires only an immediate posttest” (p. 372) because it is difficult to ascertain whether learners’ scores on the delayed posttest should be ascribed to the experimental treatment or to other learning opportunities which occurred in the period between the learning session and the delayed test (Hulstijn, 2003; Nation & Webb, 2011). Nation and Webb pointed out that this problem could be avoided by using nonsense words instead of real TWs. Yet, the use of nonsense words in a study based on authentic video seems inconvenient as it would considerably affect the original speech stream.
7. We used the VKS to measure meaning recall rather than the progress made. Using the VKS as described in Nation and Webb (2011) might have offered more precise data concerning learners’ progress.

ACKNOWLEDGEMENTS

We are grateful to the LLT reviewers for their insightful comments and suggestions. All remaining errors are our own. We would also like to thank Dr. Carmen Eggermont for allowing us to conduct this study in her classes (KU Leuven Kulak) and the team of Roeland (Saint-Dizier 4, 2011) for their help in organizing the pilot study for this experiment. Finally, we would also like to thank the Media and Learning Unit (KU Leuven) and Dr. Hans Paulussen for developing the video player.

ABOUT THE AUTHORS

Maribel Montero Perez obtained her PhD in 2013 and works within the iMinds - ITEC research group at the KU Leuven Kulak. Her research interests concern the effectiveness of video for L2 listening and vocabulary acquisition and the use of eye-tracking for investigating vocabulary learning from watching captioned video.

E-Mail: maribel.monteroperez@kuleuven-kulak.be

Elke Peters is an assistant professor of English at the Department of Language and Communication at the KU Leuven. Her research interests include educational language policy and instructed second language acquisition, particularly vocabulary acquisition and learning of collocations.

E-Mail: elke.peters@arts.kuleuven.be

Geraldine Clarebout is head of the Center for Medical Education of the faculty of medicine, KU Leuven. Her PhD research focused on the use of tools in learning environments and the optimization of tool use. She is currently involved in quality assurance processes in education.

E-Mail: geraldine.clarebout@med.kuleuven.be

Piet Desmet is full professor of French and Applied Linguistics and Foreign Language Methodology at KU Leuven Kulak and coordinator of the iMinds - ITEC (Interactive Technologies) research group. His research takes a particular interest in adaptive learning environments, mobile language learning, serious gaming, parallel corpora for CALL and writing aids.

E-Mail: piet.desmet@kuleuven-kulak.be

REFERENCES

- Baltova, I. (1999). *The effect of subtitled and staged video input on the learning and retention of content and vocabulary in a second language*. Unpublished doctoral dissertation. University of Toronto, Toronto, Canada.
- Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, 23(4), 509–533.
- Brett, P. (1995). Multimedia for listening comprehension: The design of a multimedia-based resource for developing listening skills. *System*, 23(1), 77–85.
- Brett, P. (1998). Using multimedia: A descriptive investigation of incidental language learning. *Computer Assisted Language Learning*, 11(2), 179–200.
- Bruton, A. (2009). The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly*, 6(4), 37–41.
- Buck, G. (2001). *Assessing Listening*. Cambridge, UK: Cambridge University Press.
- Chung, J. (1999). The effects of using video texts supported with advance organizers and captions on Chinese college students' listening comprehension: An empirical study. *Foreign Language Annals*, 32(3), 295–308.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Council of Europe. (2011). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. Retrieved from http://www.coe.int/t/dg4/linguistic/Source/Framework_FR.pdf

- Danan, M. (1992). Reversed subtitling and Dual Coding Theory: New directions for foreign language instruction. *Language Learning*, 42(4), 497–527.
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *META*, 49(1), 67–77.
- Gaddy, M. L., van den Broeck, P., & Sung, Y.C. (2001). The influence of text cues on the allocation of attention during reading. In T. Sanders, J. Schimperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 89–110). Amsterdam: John Benjamins.
- Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, 24(3), 239–258.
- Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gass, S. M. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, 21, 319–333.
- Godfroid, A., Housen, A., & Boers, F. (2010). A procedure for testing the Noticing Hypothesis in the context of vocabulary acquisition. In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition* (pp. 169–197). Amsterdam: John Benjamins.
- Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55–75.
- Grgurović, M., & Hegelheimer, V. (2007). Help options and multimedia listening: Students' use of subtitles and the transcript. *Language Learning & Technology*, 11(1), 45–66. Retrieved from <http://llt.msu.edu/vol11num1/pdf/grgurovic.pdf>
- Guillory, H. G. (1998). The effects of keyword captions to authentic French video on learner comprehension. *CALICO Journal*, 15(1–3), 89–108.
- 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.
- Huang, H.-C., & Eskey, D. E. (1999–2000). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems*, 28(1), 75–96.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 21(2), 181–193.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.
- Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty, & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.
- King, J. (2002). Using DVD feature films in the EFL classroom. *Computer Assisted Language Learning*, 15(5), 509–523.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge. *Language Testing*, 21(2), 202–226.
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*, 5(1), 223–250.

- Lee, J. F., & Van Patten, B. (2003). *Making communicative language teaching happen*. New York, NY: McGraw Hill.
- Liu, N., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16(1), 33–42.
- Lonsdale, D., & Le Bras, Y. (2009). *A frequency dictionary of French core vocabulary for learners*. London, UK: Routledge.
- Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, 32(3), 321–328.
- Markham, P. (2001). The influence of culture-specific background knowledge and captions on second language comprehension. *Journal of Educational Technology Systems*, 29(4), 331–343.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in English as a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. del Mar Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83–98). Clevedon, UK: Multilingual Matters.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59–82.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.
- Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, 27(1), 95–106.
- Park, M. (2004). *The effects of partial captions on Korean EFL learners' listening comprehension*. Unpublished doctoral dissertation. University of Texas at Austin, Austin, TX.
- Paulussen, H., Macken, L., Trushkina, J., Desmet, P., & Vandeweghe, W. (2006). Dutch parallel corpus: A multifunctional and multilingual corpus. *Cahiers de l'Institut de Linguistique de Louvain*, 32(1–4), 269–285.
- Peters, E. (2006). Learning second language vocabulary through reading: Three empirical studies into the effect of enhancement techniques. Unpublished doctoral dissertation. University of Leuven, Leuven, Belgium.
- Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80, 478–493.
- Pujolá, J.T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, 14(2), 235–262.
- Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, 57(1), 155–199.
- Richardson, J. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6, 135–147.
- Robin, R. (2007). Commentary: Learner-based listening and technological authenticity. *Language Learning & Technology*, 11(1), 109–115. Retrieved from <http://llt.msu.edu/vol11num1/pdf/robin.pdf>
- Rodgers, M.P.H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs.

TESOL Quarterly, 45(4), 689–717.

Rost, M. (2002). *Teaching and researching listening*. London, UK: Longman.

Selva, T., Verlinde, S., & Binon, J. (2002). Le Dafles, un nouveau dictionnaire électronique pour apprenants du français. In A. Braasch (Ed.), *Proceedings of the Tenth EURALEX International Congress*, (pp. 199–208). Copenhagen, Denmark: CST.

Staehr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31, 577–607.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, 14(2), 50–73. Retrieved from <http://lt.msu.edu/vol14num2/sydorenko.pdf>

Tacq, J. (1997). *Multivariate analysis techniques in social science research: From problem to analysis*. London, UK: Sage.

Van Patten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287–301.

Van Patten, B., Williams, J., & Rott, S. (2004). Form-meaning connections in second language acquisition. In B. Van Patten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 1–26). Mahwah, NJ: Lawrence Erlbaum.

van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 33(5), 1–24.

Vandergrift, L. (2011). Second language listening: Presage, process, product, and pedagogy. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 455–471). New York, NY: Routledge.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *English Language Teaching Journal*, 42(4), 272–281.

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with unilingual (CEEFAX) sub-titles. *System*, 18(2), 221–234.

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

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65.

Webb, S. (2010). Using glossaries to increase the lexical coverage of television programs. *Reading in a Foreign Language*, 22(1), 201–221.

Webb, S. (2011). Selecting television programs for language learning: Investigating television programs from the same genre. *International Journal of English Studies*, 11(1), 117–136.

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning*, 59(2), 335–366.

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth vs. breadth. *The Canadian Modern Language Review*, 53, 13–40.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65–86. Retrieved from <http://lt.msu.edu/vol14num1/winkegasssydorenko.pdf>

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.