Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions

Elke Peters[*], Eva Heynen, Eva Puimège

University of Leuven, Leuven, Belgium

ARTICLE INFO

ABSTRACT

Recent research has shown that learners can learn new words while watching TV programs. However, the number of words learned tends to be low. Several studies have demonstrated that first language (L1) subtitles as well as captions (= subtitles in the foreign language) have the potential to increase learning gains compared to when no on-screen text aids are provided. However, the evidence regarding the differential effect of both types of subtitles is still inconclusive. This paper reports on two exploratory studies investigating the effect of L1 subtitles and captions on different aspects of word knowledge among English-as-a-foreign language (EFL) learners in Flanders (Belgium). Data were collected in two different educational settings: intermediate EFL learners from a general school and low-proficiency EFL learners from a vocational school. Although learning gains were generally low, results indicated that captions have the potential to increase form learning. However, learners who were exposed to the audiovisual input with L1 subtitles did not perform better than the captions group in the tests focusing on the meaning of the target items. Additionally, findings also suggested that learners' vocabulary size and an item's frequency of occurrence in the video clip correlated positively with word learning.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Most research into incidental foreign language vocabulary acquisition has been conducted in the field of reading. Recently, more studies have started to explore the effects of audiovisual input on learning foreign language vocabulary (Lin, 2014; Montero Perez, Peters, Clarebout, & Desmet, 2014; Rodgers, 2013; Sydorenko, 2010; Winke, Gass, & Sydorenko, 2010, 2013). Audiovisual input provides foreign language learners with authentic input and rich contexts. That is why it has been argued that audiovisual input can have a similar, positive effect on vocabulary acquisition as reading (Lin, 2014; Lin & Siyanova-Chanturia, 2015; Webb, 2015). However, only a handful of studies have explored these benefits empirically with (young) adult learners (Montero Perez et al., 2014; Rodgers, 2013; Sydorenko, 2010; Winke et al., 2010, 2013). Generally, these studies have reported low learning gains (Montero Perez et al., 2014; Rodgers, 2013), which might be explained by the fact that vocabulary learning through audiovisual input is challenging because of online processing demands. Unlike in reading, learners cannot go back to a previous word or sentence making guessing more difficult.

There are however a number of factors that might foster learning gains through audiovisual input. Studies have shown that word learning through audiovisual input can be boosted by providing on-screen text such as captions (= L2 subtitles) (Montero Perez et al., 2014; Montero Perez, Van Den Noortgate, & Desmet, 2013) or L1 subtitles (Danan, 1992; d'Ydewalle &

---

* Corresponding author.
E-mail addresses: elke.peters@kuleuven.be (E. Peters), evaheynen@hotmail.com (E. Heynen), epuimege@gmail.com (E. Puimège).

Pavakanun, 1997, 1995; Koolstra & Beentjes, 1999). Yet, only a few studies have investigated the differential effect of L1 subtitles and captions on vocabulary learning (Bianchi & Ciabattoni, 2008; Bisson, Van Heuven, Conklin, & Tunney, 2014; Frumuselu, De Maeyer, Donche, & Colon Plana, 2015; Vulchanova, Aurstad, Kvitnes, & Eshuis, 2015). In addition, none of these looked at different aspects of word knowledge and they did not take learners' vocabulary size and an item's frequency of occurrence in the input into account; two factors that seem to play a facilitative role in learning words through audiovisual input (Montero Perez et al., 2014; Peters & Webb, submitted for publication; Rodgers, 2013). Therefore, this paper reports on two studies investigating the effect of L1 subtitles and captions on word learning by English-as-a-foreign language (EFL) learners. In addition, these studies also take the item's frequency of occurrence and learners' vocabulary size into account.

2. Literature review

2.1. L1 subtitles

L1 subtitles or interlingual subtitles provide viewers with authentic foreign language input and the input's (condensed) translation into the national language (Almeida & Costa, 2014; Koolstra, Peeters, & Spinhof, 2002). The use of L1 subtitles is the preferred way of making foreign language movies and TV programs available in most European countries, especially in smaller countries such as The Netherlands and Belgium (Flanders). Dubbing, on the other hand, is more common in larger countries such as Germany, France and Spain (Almeida & Costa, 2014; Koolstra et al., 2002). A recent study on the use of subtitling, requested by the European Commission (Safar et al., 2011), suggests that subtitles can foster language learning, awareness and motivation. Additionally, the European Survey on Language Competences (European Commission, 2012) also showed that there is a positive relationship between foreign language proficiency and learners' exposure to foreign language input (via TV and movies). These survey studies show the beneficial role of L1 subtitles for informal language learning.

Empirical evidence for the beneficial effect of L1 subtitles mainly stems from research carried out in the 1990s (Danan, 1992: pilot study; d'Ydewalle & Pavakanun, 1995, 1997; d'Ydewalle & Van de Poel, 1999; Koolstra & Beentjes, 1999). Overall, these studies suggest that L1 subtitles might be beneficial for vocabulary learning. The positive findings of these studies can (in part) be explained by the automatic reading of the subtitles, which was demonstrated in a number of eye-tracking studies (e.g. d'Ydewalle, Praet, Verfaillie, & Van Rensbergen, 1991). Consequently, both L1 subtitles and sound seem to be processed in parallel (Danan, 2004). In spite of the L1 subtitles' condensed form,[1] Koolstra et al. (2002) argue that this does not lead to information loss nor does it seem to distract from the picture.

2.2. Captions

Although Danan (2004) argues that L1 subtitles are a powerful pedagogical tool, recent research has tended to focus on captions rather than L1 subtitles (see e.g. Montero Perez et al., 2014; Montero Perez, Peters, & Desmet, 2015; Sydorenko, 2010; Winke et al., 2010, 2013). Unlike L1 subtitles, captions are intralingual subtitles, providing the viewers with aural as well as written input in the foreign language (L2 audio + L2 subtitles). Captions are primarily used for the deaf and hard-of-hearing.

Most studies focusing on the use of captions have explored its effects on listening comprehension, whereas fewer studies have looked at its effect on vocabulary acquisition (Montero Perez et al., 2013). A recent meta-analysis of the effectiveness of captioned video for listening comprehension and vocabulary learning (Montero Perez et al., 2013) showed that although the effect was sometimes moderated by test type, there was generally a clear and large effect of captions on both listening comprehension and vocabulary acquisition.

The positive effects of captions can be explained by the fact that they help learners segment the speech stream and distinguish separate words (Danan, 2004). Additionally, captions help learners pay (more) attention to words in the captions, as was also shown in two eye-tracking studies (Montero Perez et al., 2015; Winke et al., 2013). By providing learners with the correct word form, captions thus help to develop learners' word recognition.

Both Danan (2004) and Vanderplank (2010) stress that captions can indeed lead to vocabulary learning provided that the audiovisual input is not above the learners' proficiency. Webb and Rodgers (2009a) hypothesize that knowledge of the most frequent 3000 word families should be "the minimum vocabulary size necessary in order to watch movies for language learning" (p. 420). They demonstrated that knowledge of the most frequent 3000–4000 word families corresponds to 95% text coverage in American and British movies and knowledge of the most frequent 2000–4000 words corresponds to 95% text coverage in TV programs (Webb & Rodgers, 2009b).

2.3. L1 subtitles versus captions

Both L1 subtitles and captions result in more lexical learning compared to no on-screen text. However, as far we know, only four studies have compared the effect of captions and L1 subtitles. Bianchi and Ciabattoni's (2008) study focused on L1 (= Italian) subtitles, captions and audio only (= control group). One week after the treatment, learners (18–45 years old) were tested on their knowledge of words that had occurred in the clips, which was either an excerpt from a movie with a strong

---

[1] L1 subtitles are characterized by space restrictions and omissions, as a result of which literal translations are not possible (Diaz Cintas & Remael, 2014).

relationship between imagery and audio (*Harry Potter*) or an excerpt from a movie without an imagery-audio relationship (*Fantasia*). Compared with the pretests, the descriptive statistics[2] showed that the increase was the highest in the two experimental groups. The best results were found in the L1 subtitles group. However, care should be taken when interpreting these results, because each condition consisted of participants from three proficiency levels: beginners, intermediate and advanced learners, with often a low number of participants in each subgroup (e.g. N = 5 in the beginners captions group and in the beginners control group). Although the authors argue that these three proficiency levels benefited differently from the three treatments, it might not be possible to come to any firm conclusions, as the authors pointed out themselves.

In a study focusing on Dutch as a foreign language and English L1 speakers, Bisson et al. (2014) examined three types of subtitles in a movie excerpt: L1 subtitles (L2 sound, L1 subtitles), reversed subtitles (L1 sound, L2 subtitles), and captions (L2 sound, L2 subtitles). No differences in learning gains were found between the three conditions and the control group, which had not been exposed to the audiovisual input. The authors attributed this lack of difference to the test format (recognition), which might not have been sensitive enough to measure any learning gains. However, another explanation might be found in the fact that Dutch was an unknown foreign language to the participants in this study. Webb and Rodgers (2009b) argue that incidental vocabulary acquisition through audiovisual input may occur when learners are familiar with the most frequent 3000 words, provided 95% text coverage is sufficient. In Bisson et al.'s study, the participants were not familiar with the most frequent 3000 words, as they had no knowledge of the target language, Dutch. In addition to vocabulary learning, this study also analyzed learners' processing of the three types of subtitles by means of eye-tracking. Their analysis revealed more regular reading with captions and L1 subtitles than in the case of reversed subtitles. Their findings also corroborated previous research that participants read most of the words in the subtitles (= automatic reading of subtitles).

Vulchanova et al. (2015) compared the effect of L1 (Norwegian) subtitles, captions (in English), and no subtitles (English audio only) in an animated cartoon among 16 and 17-year-old EFL learners. Vocabulary acquisition was measured four weeks after the video had been shown. The tests tapped into learners' knowledge of the meaning of the target words (definitions) and learners' recall of the words' occurrence in the video. They only found an effect of L1 subtitles, which was moderated by age, in the meaning recognition test. However, in the separate analyses for the two age groups (16 years old vs. 17 years old), the type of subtitles no longer affected word learning. It is not unlikely that the word recall test in which learners were asked whether the word had occurred in the video might have been too challenging, as a time interval of four weeks might be too long. Additionally, it is difficult to fully appreciate their findings, as no descriptive statistics were provided.

In a longitudinal study, Frumuselu et al. (2015) explored the effect of L1 (Spanish) subtitles and captions (English) on EFL university learners' acquisition of informal and conversational speech (phrasal verbs, idioms, informal words, expressions etc.). Learners of different proficiency levels (A2-C1 according to the Common European Framework of Reference (CEFR)) with L1 Spanish or Catalan (90% of the learners) or another L1 (10% speakers of other L1s, such as Dutch, German, Russian, …) watched 13 subtitled episodes from a popular TV series over a period of seven weeks. The results showed that there were learning gains in both groups, but the captions group performed better than the L1 subtitles group. Moreover, this effect was not mediated by learners' proficiency level, as measured by a lexical pretest. However, it could be argued that the lexical pretest might not have been sensitive enough to distinguish between different levels of proficiency, as it tested learners' preknowledge of 30 items from the TV series. Additionally, it is not unlikely that the effect found might also be attributed to the composition of the two groups, as the majority of learners in each group was reported to have a B2 or C1 level according to the CEFR. Vanderplank (2010) argues that L1 subtitles might be more beneficial for beginners, whereas captions might be better for higher-level learners.

Although the aforementioned studies on L1 and L2 subtitles add to our understanding of on-screen text aids, it is clear that the evidence regarding the differential effect of both types of subtitles is still inconclusive. This may not be surprising, as the studies reviewed in this subsection differed in a number of respects, e.g. type of audiovisual input (movie, TV series, cartoon), length of input (short intervention vs. longitudinal treatment), test formats, and time of test administration. These contradictory findings clearly highlight the need for more research into the effect of on-screen text aids.

2.4. Frequency of occurrence and learners' vocabulary size

Research into incidental vocabulary acquisition through reading has shown that repeated encounters with a new lexical item have a positive effect on learning single words (Chen & Truscott, 2010; Pellicer-Sánchez, 2016; Webb, 2007) as well as collocations (Pellicer-Sánchez & Siyanova-Chanturia, 2016; Webb, Newton, & Chang, 2013). Similar findings have been reported for the effect of repetition on explicit learning of single words and collocations (Laufer & Rozovski-Roitblat, 2011; Peters, 2012, 2014). Recently, studies have also looked at the effect of frequency of occurrence on incidental vocabulary acquisition through listening (van Zeeland & Schmitt, 2013; Vidal, 2003, 2011). Although frequency of occurrence had an effect on word learning, its effect seemed to be much smaller than in reading (Vidal, 2011). Additionally, its effect tended to be dependent on the aspect of word knowledge tested (van Zeeland & Schmitt, 2013). Although there is no agreement on the exact number of occurrences that is necessary, it is clear from the studies above that frequency of occurrence has the potential to affect word learning positively.

To the best of our knowledge, only two studies explored the role of frequency of occurrence in a viewing study. Rodgers (2013) found a medium (r = 0.30) correlation between frequency of occurrence and word learning in the more challenging test. Similarly, Peters and Webb (submitted for publication) found a positive effect of frequency of occurrence on word

---

[2] No inferential statistics were provided.

learning as measured in a meaning recall and a meaning recognition test after EFL learners had watched a one-hour documentary.

Few incidental vocabulary acquisition studies have taken learners' vocabulary size into account. From reading research, we know that there is a positive relationship between learners' vocabulary size and word learning (Horst, Cobb, & Meara, 1998). In spite of the scarcity of studies into vocabulary acquisition through audiovisual input, three studies have reported on the relationship between vocabulary size and word learning. Yet, findings seem to be mixed, as Rodgers (2013) did not find a significant correlation, whereas Montero Perez et al. (2014) and Peters and Webb (submitted for publication) did. It, thus, seems warranted to explore the role of learners' vocabulary size in learning new words, in more detail.

3. Rationale and research questions

Studies exploring the effects of L1 subtitles or captions have demonstrated that both might foster vocabulary acquisition. Nevertheless, the findings that have emerged from studies comparing the effect of L1 subtitles and captions seem to be contradictory. Additionally, the effect of mediating variables, such as frequency of occurrence or learners' vocabulary size, has hardly been addressed in previous studies.

The two experiments presented here seek to investigate the effect of L1 subtitles and captions on form recognition and on initial form-meaning mappings of unfamiliar words among EFL learners in an ecologically valid but controlled way. These experiments may shed new light on how audiovisual input and on-screen text aids can be used in the foreign language classroom.

The research questions that were addressed are:

1. Is there a relationship between the type of subtitles (L1 subtitles vs. captions) and EFL learners' initial word learning?
2. Is there a relationship between frequency of occurrence and initial word learning?
3. Is there a relationship between learners' vocabulary size and initial word learning?

It was hypothesized that captions would be more beneficial for learning an L2 item's form (recognition or recall) because they will help learners pay attention to the correct word form. L1 subtitles, on the other hand, were expected to be more helpful for learning an item's meaning, as learners are expected to read and process the translated subtitles providing them with the meaning of the target item. Additionally, it was expected that learners' vocabulary size as well as an item's frequency of occurrence would correlate positively with word learning.

To answer these questions, two experiments resembling real-life English learning experiences were set up. Both experiments adopted a between-subject, pretest-posttest design. In each experiment, two intact classes were assigned to one of two experimental conditions: an L1 subtitles condition or a captions condition. So learners either watched a video clip with L1 subtitles or with captions. Learners were tested one week prior to and immediately after having watched the clip to measure any learning gains.

4. Experiment 1

4.1. Materials and method

4.1.1. Participants

Participants were recruited from two classes in a general secondary school in Flanders (Belgium). Thirty-one Dutch-speaking EFL learners took part in experiment 1. However, data of three pupils was removed because they were not present during the second session of the experiment, resulting in 28 participants: 16 learners (8 males, 8 females) from year 5 (17 years old), who were assigned to the captions groups, and 12 learners from year 6 (18 years old), who were assigned to the L1 subtitles group. Year 5 learners would typically have had three and a half years of instructed English at the time of data collection; year 6 learners four and a half years. As L1 subtitles are the preferred way of making foreign language TV programs and movies available in Flanders, these learners can be expected to have been familiar with watching subtitled TV programs. They will, however, have been less familiar with watching captioned TV programs.

4.1.2. Learning materials

A documentary about "eating insects" was selected as a video clip. The documentary was 50 min in length, but only 13 min of the documentary were selected for reasons of feasibility. The clip consisted of 1888 words (tokens), of which 90.15% were 2K tokens as measured by Cobb's *Lextutor's VocabProfile* (http://www.lextutor.ca/vp/). Given the learners' proficiency level and vocabulary knowledge (see Results section), the documentary should not have been too difficult. *Aegisub*, a free online source tool (http://www.aegisub.org), was used to create the L1 Dutch subtitles and English captions. The tool conforms to subtitling principles, such as the 6-s rule and the maximum characters (41) one line can contain (Diaz Cintas & Remael, 2014).

4.1.3. Target items

A list of potential target items, which were all taken from the video clip, was discussed with the English teacher of the two classes. Thirty-nine target items in total were selected (see Table 1). Some easy, high-frequency items (e.g. *girl*, *fish*, *street*,

Table 1

Target items with PoS (part of speech), frequency of occurrence, frequency rank in COCA.

Item	PoS	Frequency of occurrence	Frequency rank in COCA	Item	PoS	Frequency of occurrence	Frequency rank in COCA
dawn	noun 1		4044	girl	noun	3	364
pests	noun 1		6942	street	noun	2	555
throat	noun 2		2475	crawling	adjective 1		3874
prejudice	noun 4		20,062	untapped	adjective 1		18,955
water bug	noun 5			revolting	adjective 1		22,491
vendor	noun 1		4680	savvy	adjective 1		9876
gimmick	noun 1		13,197	inedible	adjective 1		28,268
resources	noun 2		770	delicious	adjective 1		5036
cricket	noun 12		8537	lucky	adjective 3		2145
diet	noun 2		2158	nice	adjective 1		900
carbon footprint	noun 1		/	cater for	verb	1	7050
greenhouse gases	noun 1		30,788	pride oneself	verb	1	10,731
snapshot	noun 1		7839	thrive	verb	1	4681
demand	noun 1		1293	head to	verb	4	937
creature	noun 1		2375	regress	verb	1	17,700
risk	noun 1		652	discover	verb	2	959
eggs	noun 7		1383	increase	verb	1	660
lunch	noun 6		1595	think	verb	3	56
fish	noun 1		950	swallow	verb	1	3569
market	noun 1		403				

market, *nice*, *eggs*, *to think*), which were expected to be known to most participants, were also chosen for reasons of test motivation (see also Table 1 for frequency rank of target items). The target items consisted of nouns, verbs and adjectives. Frequency of occurrence was not controlled for in the video, but was taken into account in the analyses (see also Scoring and analyses). However, other word-related variables that could have a played a role were not considered in the analyses simply because there were not enough instances of those parameters (e.g. compounds [*water bug*, *greenhouse gases*] or phrasal and reflexive verbs [*cater for*, *pride oneself*]).

4.1.4. Tests

4.1.4.1. Vocabulary size test. Learners' prior vocabulary knowledge was tested in a frequency-based (COCA/Davies, 2008) multiple choice vocabulary test (Peters, Van Rompaey, & Velghe, submitted for publication). The test measures learners' sight vocabulary[3] (Laufer & Ravenhorst-Kalovski, 2010) and gives an estimate of learners' vocabulary knowledge at different frequency levels up to the most frequent 5000 words. The test consists of four parts corresponding to four frequency bands: 0–2000 (2K), 2001–3000 (3K), 3001–4000 (4K), and 4001–5000 (5K). Each part contains 30 items. An item is presented in isolation; five options are provided: 1 correct answer, 3 distracters, and an *I don't know* option. The Cronbach's alpha of the test was 0.96 (N = 30).

Example of test item:
talk:
0 to speak
0 to remove something
0 to give in return for money
0 to start an important activity
0 I don't know the answer

4.1.4.2. Pretests and posttests. Learning gains were measured in a test consisting of two parts to take into account partial learning (Nation & Webb, 2011): a spoken form recognition and a spoken meaning recall test part. The same test was used as pretest and immediate posttest.

The test consisted of 50 test items, the 39 target items and 11 nonwords. Nonwords were added to the test to control for guessing and any pretest-posttest learning effect (Nation & Webb, 2011, p. 265). The nonwords were retrieved from http://lextutor.ca/freq/lists_download/pnwords.html. The aural presentation of the target items was recorded to control for differences in pronunciation, pace, and intonation in the different test administrations. A native speaker was asked to clearly pronounce each word once with a pause of eight seconds between each word.

---

[3] Sight vocabulary is defined as those words whose meaning is so familiar to a learner that they can be understood out of context.

In the form recognition test part, learners were asked to tick off "yes" or "no" when they had heard or seen the word before. Immediately after answering an item in the form recognition test, learners were also asked if they could provide the meaning of that target item (= meaning recall part). Learners could supply an L1 translation, a synonym, or an example. All items were presented in their aural form, not in their written form. The order and the pronunciation of the items were identical in the pretest as well as the posttest, because the aural presentation of the target items was recorded. Reliability of the tests was acceptable (Cronbach's alpha ranged from 0.71 to 0.77).

Example of item tested in form recognition and meaning recall test.

Target item	Form Recognition	Meaning recall
Aural presentation of item	yes — no	… … … … … … … … … … … ….

We did not administer any delayed posttests because our focus was on initial learning during which target items were encountered for the first time. We agree with Hulstijn (2003) that long-term knowledge of word meanings or word forms requires repeated encounters and retrieval opportunities. Given the use of two short video clips in which most target items occurred only once, it is probably not reasonable to expect (m)any long-term learning gains without additional practice or encounters. Additionally, deliberate learning in the time period between immediate and delayed posttests could not be controlled for. As pointed out by Nation and Webb (2011), this is particularly problematic in the case of low scores, as only a few learners engaging in deliberate learning could inflate delayed posttest scores considerably.

4.1.5. Questionnaire

A short questionnaire was developed to verify whether the learners had concentrated on the content of the video clip and to explore what they had learned in terms of content. One question specifically tapped into learners' recall of (other) words they might have learned while watching the video clip as learners might provide words that were not tested in the posttest.

4.2. Procedure

The data collection took place in two sessions in two consecutive weeks. In session 1, all learners took the vocabulary size test, then the form recognition and meaning recall test. Learners were told that they were taking part in a study on vocabulary knowledge. It was not mentioned that they would be tested on the same words again one week later. In the second session, learners watched the video clip containing the target items. They were told that they would have to answer comprehension questions afterwards. Participants were at this point not informed that they would be tested on the words used in the video. Having watched the video clip twice, they first filled in the questionnaire before taking the posttest. All learners were then debriefed about the aim and methodology of the study.

4.3. Scoring and analyses

All tests were scored dichotomously with 0 for an incorrect answer and 1 for a correct answer. The meaning recall tests were scored by two raters; the interrater reliability was r = 0.99 in the pretest and in the posttest. To answer our research questions, a repeated measures logistic regression in SPSS was carried out (= Generalized Estimating Equation or GEE analysis in SPSS) for each of the two tests (one GEE for the form recognition test part, one for the meaning recall test part) because it is appropriate for the analysis of dichotomous response data (correct or incorrect response in the posttest). A GEE analysis allows for the analysis of treatment variables, participant variables, and item variables in one model. This means that we could include type of subtitles (captions or L1 subtitles), learners' vocabulary size, and frequency of occurrence in one model. The analysis is based on the number of observations and not on total test scores per participant. This means that the combination "participant, item, response" defines for each observation a particular score (correct/incorrect) on a particular item for a particular participant. The odds ratio (= exp^b or exponential parameter estimate) predicts the odds of a correct response with a parameter.

A GEE can also be used when data is not normally distributed, as was the case in our tests. Unlike AN(C)OVAs, a GEE or repeated measures logistic regression does not require the assumptions of homogeneity of variance or linearity to be met either. All other assumptions were met: a dichotomous response variable, no multicollinearity, and an acceptable case-parameter ratio (10 > 1). The following parameters were always entered into the model: treatment (type of subtitles), frequency of occurrence, learners' vocabulary size, and the interactions condition × frequency of occurrence, condition × vocabulary size, and frequency of occurrence × vocabulary size. Parameters that did not contribute significantly to the regression model were removed and then the model was refit again.

4.4. Results

4.4.1. Vocabulary size test

As the L1 subtitles group had received one year more of formal instruction, it was important to verify whether the two groups differed significantly in their prior vocabulary knowledge. As can be seen in Table 2, both groups obtained similar

Table 2
Mean scores in percentages on vocabulary test: total score and score per part (SD in brackets).

Condition	Year	N	Total	2K	3K	4K	5K
Captions	5th year	16	77.50% (14.44)	92.92% (7.49)	83.96% (16.73)	69.79% (14.48)	63.33% (21.74)
L1 subtitles	6th year	12	79.38% (12.95)	91.67% (7.04)	86.11% (11.27)	74.72% (16.42)	65.00% (19.92)

Table 3
Number and percentage of correct/incorrect responses in posttests.

Condition	Form recognition		Meaning recall	
	Incorrect response	Correct response	Incorrect response	Correct response
Captions	114 51.8%	106 48.2%	285 80.7%	68 19.3%
L1 Subtitles	94 67.6%	45 32.4%	198 79.2%	52 20.8%
Total	208 57.9%	151 42.1%	483 80.3%	120 19.9%

Table 4
GEE analysis of form recognition test: test of model effects.

Parameter	Wald Chi-square	df	Sig	b	Exp^b	CI for Exp^b	
						Lower	Upper
Intercept	15.871	1	<0.0001	−5.42	0.004	0.000	0.052
Condition	16.516	1	<0.0001	5.47	236.71	16.95	3305.58
Frequency	5.149	1	0.023	0.09	1.10	1.01	1.19
VocSize	8.616	1	0.003	0.05	1.05	1.02	1.08
Condition × VocSize	11.772	1	0.001	−0.05	0.95	9.21	0.98

Note: Frequency = frequency of occurrence; VocSize = Vocabulary size; CI = Confidence interval.

scores and did not differ significantly in terms of vocabulary knowledge ($t = -0.36$; $df = 26$; $p = 0.73$). The two groups mastered the most frequent 2000 words (>90% score on 2K section).

4.4.2. Form recognition test

Table 3 provides the results for the form recognition test part and also for the meaning recall test part (correct and incorrect responses). The results are based on words for which learners did not have any preknowledge (= 0 score in pretest).

As there was hardly any guessing and no pretest effect given the lack of learning gains for nonwords in the form recognition test (pretest = 3.21; posttest = 1.98 out of 11), the form recognition posttest results can be considered reliable.

As can be seen in Table 3, learners in the captions groups were able to recognize more target items than the L1 subtitles group, which corresponds to recognizing 14% words more in relative gains or 3 words more in absolute gains. The GEE analysis, based on 359 observations, revealed four significant parameters: type of subtitles, frequency of occurrence, learners' vocabulary size, and an interaction effect between type of subtitles and vocabulary size (see also Table 4). The interaction effect revealed that the effect of the type of subtitles was conditional on learners' vocabulary size. The captions group's odds of correct response increased, when their vocabulary size was larger. There was also a positive relationship between frequency of occurrence and form recognition. The odds of a correct response increased by 10% when the frequency of occurrence increased. So we found evidence that captions are more beneficial for learning formal aspects of word knowledge, although its effect was conditional on learners' vocabulary size.

4.4.3. Meaning recall test

The GEE of the meaning recall test was based on 603 cases or observations (= items not known in the meaning recall pretest). As can be seen in Table 3, learners in both groups provided a correct answer in 20% of the cases.[4] The GEE analysis showed that two parameters contributed significantly to the regression model: an item's frequency of occurrence in the video and learners' vocabulary size (see Table 5). The type of subtitles was not a significant predictor. There was a positive relationship between an item's frequency of occurrence and learners' odds of a correct response in the meaning recall test (see Table 5). When frequency of occurrence increased by 1, the odds in favor of a correct response increased by 11%. Similarly, there was a positive relationship between learners' vocabulary size and vocabulary learning (see Table 5). With 10 words known more in the vocabulary test ($\exp^{10 \times b} = \exp^{10 \times 0.03} = 1.30$), the odds of a correct response increased 30%.

---

[4] The L1 subtitles group had relative learning gains of 4%, which corresponds to absolute gains of 0.5 word learned more, but this was not significantly more than in the captions group.

Table 5
GEE analysis of meaning recall test: test of model effects.

Parameter	Wald Chi-square	df	Sig	b	Exp^b	CI for Exp^b	
						Lower	Upper
Intercept	48.738	1	<0.0001	4.01	0.02	0.006	0.056
Frequency	8.547	1	0.003	0.10	1.11	1.03	1.18
VocSize	17.975	1	<0.0001	0.03	1.03	1.01	1.04

Note: Frequency = frequency of occurrence; VocSize = Vocabulary size; CI = Confidence interval.

5. Experiment 2

5.1. Materials and method

The second experiment was similar in design to the first experiment. It also adopted a between-subject (L1 subtitles vs. captions), pretest-posttest design. It differs from experiment 1, though, in the profile of its participants and in the tests used. Unlike experiment 1, in which participants were recruited from a general secondary school, the participants in the second experiment were recruited from a vocational school, because such language learners tend to be undersampled in SLA research (Plonsky, 2015). In addition to form recognition (= recognizing the form of the target item), this experiment also explored the effect of L1 subtitles and captions on two other aspects of word knowledge: form recall (= providing the form of the target item) and meaning recognition (= recognizing the meaning of a target item in a multiple choice test). This allowed us to measure learning gains at a more difficult level (form recall) and at an easier level (meaning recognition).

5.1.1. Participants
The initial sample consisted of 30 learners from a vocational school, i.e. a school which provides pupils with a hands-on, job-specific training. However, due to students being absent during one of the data collection sessions or students not completing all tests, data of 12 learners had to be removed, bringing the total number of participants to 18. Learners' age ranged from 17 to 20 (mean = 17.94). Most students had Dutch as their L1; other L1's were Berber (3), Russian (1), Dari (2), Kurdish (1). One participant was bilingual (Dutch-Berber). Participants were assigned to either the L1 subtitles (N = 10) or the captions (N = 8) condition. These learners had one hour of English per week and were low-proficiency to pre-intermediate learners of English ranging considerably in their vocabulary knowledge (see Table 7 in Vocabulary Size Test). Participants were familiar with L1 subtitles; some also watched captioned TV-programs on a regular basis, as was revealed in a questionnaire.

The participants in this study being low-proficiency learners tend to be undersampled in SLA research, as most studies focus on highly literate (university language) learners (Plonsky, 2015). It should be mentioned that data collection and participant cooperation were a serious challenge, as these participants, who are being trained for very specific jobs, are not used to taking tests or exams. This is also illustrated in the considerable data loss.

5.1.2. Learning materials
An episode from the series *The Simpsons* was selected because learners were already familiar with this series and their characters. The episode contained few twists and cultural references. Additionally, there were few references to other episodes or other American TV series. The episode containing 2226 words (tokens) (86.86% 2K tokens) was about 20 min in length. The software Jubler (www.Jubler.org) was used to create captions and L1 subtitles.

5.1.3. Target items
Eighteen target items taken from the *Simpsons* episode were selected for this experiment. The items could differ in part of speech and frequency of occurrence. Two items, *pen pal* and *honor code*, were compounds (see also Table 6).

The selection of potential target words was based on a number of steps. First, the participants completed a general pretest, which tested the learners' meaning recognition of 115 vocabulary items. Based on the results of this test, a list of potential target items occurring in the video clip was compiled. The participants' teacher was, then, asked to highlight words from the list, which she considered unknown to the group of participants. The remaining 35 words were tested through three pretests: a form recall test, a form recognition test and a meaning recognition test. Words that were partially unknown to at least 70% of the participants were selected for the posttests, resulting in a list of 18 target words.

5.1.4. Tests

5.1.4.1. Vocabulary size test. Participants' prior vocabulary knowledge was tested with the frequency-based vocabulary test (Cronbach's alpha = 0.98; N = 18) that was also used in experiment 1.

5.1.4.2. Pretests and posttests. Learning gains were measured from pretest to posttest at three levels of sensitivity: written form recall, written form recognition, and written meaning recognition. The order of the items differed in each test. The

Table 6
Target items with PoS, frequency of occurrence and frequency rank in COCA.

Item	PoS	Frequency of occurrence	Frequency rank in COCA
to initiate	verb	1	3548
to pledge	verb	1	5253
faint	verb	1	11,170
carve	verb	1	3833
infallible	adjective	1	20,678
curb	noun	1	8166
quilt	noun	1	6206
installment	noun	1	10,477
inappropriate	adjective	1	5099
pen pal	noun	3	/
detention	noun	1	6772
baldness	noun	1	23,269
foreigner	noun	2	4935
disobey	verb	1	16,097
squirrel	noun	2	7008
achieve	verb	1	1153
appointment	noun	2	3040
honor code	noun	5	/

instructions were always given in Dutch. In the form recognition and meaning recognition pretest, nonwords were also used in order not to draw too much attention to the target items and to control for guessing. In the three pretests, 35 words in total were tested (17 more than the 18 target items; see also Target items). The nonwords and the 17 words that were already known were no longer tested in the posttests to avoid test fatigue. In the posttest, only the aforementioned 18 target items were tested. To avoid a test effect, the tests were administered in the following order: first form recall, then form recognition, and finally meaning recognition. The reliability for all tests (N = 18) was high and ranged from 0.90 to 0.94.

In the form recall test, learners were provided with the target item's Dutch translation and asked to supply the target item's form.

Kaalheid: ……… … … … … … … … … …
Bereiken: … … … … … … … … … … ….
Erecode: … … … … … … … … … … … …

In the form recognition test, learners were asked whether they recognized the written form of the items. They had to tick off those words that they had seen or heard before.

Example of form recognition test:
    0 inappropriate
    0 pen pal
    0 detention

Finally, in the third test, learners had to tick off the correct meaning of the target item. Each item was presented in isolation. Each test item contained the correct answer, two distracters, and one I don't know option. Learners were explicitly asked not to guess.

Inappropriate
0 ongewenst (*unwanted/unwelcome*)
0 ongepast (*inappropriate*)
0 onverantwoordelijk (*irresponsible*)
0 ik weet het niet (*I don't know*)

5.1.4.3. Comprehension task. The comprehension task was used to verify whether learners had understood the gist of the story. The task consisted of two open-ended questions in Dutch that measured general understanding.

5.1.5. Questionnaire
A short questionnaire was designed to determine whether learners had focused on the content of the episode. The questionnaire also contained two questions on the participants' previous experiences with captions and L1 subtitles. The answers to these questions were used to help interpret the results of the vocabulary tests.

5.1.6. Procedure

The pretest and posttest data were collected in three sessions. In the first session, organized two weeks prior to the experimental treatment, learners took the vocabulary size test. One week later (in session 2), they took the pretests. In session 3, learners first watched the episode. They were informed that afterwards they would have to answer comprehension questions. Having completed the questionnaire, they were administered the posttests (first the form recall test, then the form recognition test, finally the meaning recognition test).

5.2. Results

The scoring procedure and analyses were identical to the ones in experiment 1. The tests were scored dichotomously. A GEE in SPSS (repeated measures logistic regression) was run for each of the three vocabulary tests. Only items with which learners were not familiar in the pretest were analyzed.

5.2.1. Vocabulary size test

The participants ranged considerably in their vocabulary knowledge (see Table 7). Only three learners seemed to master the most frequent 2000 words in English. Although the subtitles group had a larger vocabulary size on average, the t-test did not reveal a significant difference between the two groups ($t = -1.24$; $df = 16$; $p = 0.23$).

Table 7
Vocabulary size test scores in percentages (standard deviation in brackets).

	Total	2K	3K	4K	5K
Captions	20.00%	35%	15.42%	15.83%	13.75%
	(21.99)	(25.70)	(21.00)	(23.01)	(19.96)
L1 Subtitles	32%	59%	27.67%	20.33%	21.00%
	(19.10)	(25.49)	15.40	(20.45)	(24.55)

5.2.2. Form recall

The analysis was based on 247 observations (items not known in the pretest) (see Table 8). There were more correct responses (21.5% responses) in the captions group than in the L1 subtitles group (11.1% responses) (the captions group learned about 2.4 words more (= 18.5%)). The GEE revealed a model with three parameters (see Table 9): type of subtitles, frequency of occurrence, and vocabulary size. The analysis showed that captions were more beneficial for learning the form of an unfamiliar word, although it must be emphasized that there were very few correct responses in both conditions. In addition to the use of captions, frequency of occurrence was also positively correlated with learning gains (see Table 9). Learners' odds of a correct response increased three times for each additional occurrence. Finally, learners' vocabulary size was also positively correlated with their learning (see Table 9). In the form recognition test, we thus found evidence for the beneficial effect of captions for learning the form of unknown words.

5.2.3. Form recognition

Table 8 shows that there were slightly more correct responses in the captions group (29.1%) than in the L1 subtitles group (25.5%). However, the GEE, based on 212 observations, did not reveal a relationship between the type of subtitles and form recognition (see Table 10). Only one parameter seemed to affect learning gains in this test, viz. the interaction between frequency of occurrence and vocabulary size, which was positively correlated with learning gains in the form recognition test. (see Table 10).

Table 8
Number and percentage of incorrect and correct responses in Form recall, form recognition and meaning recognition test.

	Form recall (247 observations)		Form recognition (212 observations)		Meaning recognition (219 observations)	
	Incorrect	Correct	Incorrect	Correct	Incorrect	Correct
Captions	95	26	78	32	93	19
	78.5%	21.5%	70.9%	29.1%	83%	17%
L1 Subtitles	112	14	76	26	73	34
	88.9%	11.1%	74.5%	25.5%	68.3%	31.8%
Total	207	40	154	58	166	53
	83.8%	16.2%	72.6%	27.4%	75.8	24.2

Table 9
GEE analysis of form recall test.

Parameter	Wald Chi-square	df	Sig	b	Exp[b]	CI for Exp[b]	
						Lower	Upper
Intercept	51.70	1	<0.0001	−4.25	0.01	0.00	0.05
Condition	5.79	1	0.016	1.25	3.49	1.26	9.64
Frequency	16.94	1	<0.0001	0.76	2.17	1.50	3.14
VocSize	9.73	1	0.003	0.03	1.03	1.01	1.05

Note: Frequency = frequency of occurrence; VocSize = Vocabulary size; CI = Confidence interval.

Table 10
GEE analysis of form recognition test.

Parameter	Wald Chi-square	df	Sig	b	Exp[b]	CI for Exp[b]	
						Lower	Upper
Intercept	8.71	1	0.003	−2.46	0.14	0.02	0.44
Frequency	1.23	1	0.81	0.36	1.43	0.76	2.68
VocSize	0.06	1	0.27	0.01	1.01	1.00	1.05
Frequency × VocSize	4.38	1	0.04	0.024	1.024	1.00	1.05

Note: Frequency = frequency of occurrence; VocSize = Vocabulary size; CI = Confidence Interval.

Table 11
GEE analysis of meaning recognition test.

Parameter	Wald Chi-square	df	Sig	b	Exp[b]	CI for Exp[b]	
						Lower	Upper
Intercept	27.03	1	<0.0001	−2.31	0.10	0.00	0.06
Frequency	0.70	1	0.41	−0.24	0.79	0.45	1.38
VocSize	2.25	1	0.13	−0.03	0.97	0.94	1.01
Frequency × VocSize	14.33	1	<0.0001	0.07	1.07	1.03	1.10

Note: Frequency = frequency of occurrence; VocSize = Vocabulary size; CI = Confidence interval.

5.2.4. Meaning recognition

The GEE analysis of the meaning recognition test was based on 219 observations (items not known in the pretest). As can be seen in Table 8, there were more correct responses in the L1 subtitles condition (31.8%) than in the captions condition (17%). Surprisingly, the type of subtitles did not affect learning gains in the meaning recognition test. A GEE analysis revealed that one parameter was positively correlated with word learning, viz. the interaction between frequency of occurrence and learners' vocabulary size (see also Table 11). The odds of a correct response in the meaning recognition test increased by 3% when the frequency of occurrence and the vocabulary size test score increased by 1 (see Table 11).

6. Discussion

This paper discusses two exploratory experiments that investigated the differential effect of L1 subtitles and captions on word learning. Data were collected in two different educational settings. In spite of differences in the design and procedure (e.g. documentary vs. cartoon, one viewing vs. two viewings; different test types and different test administrations; different participant profiles and proficiency levels), the results of experiment 2 largely match those of experiment 1. The findings showed that incidental vocabulary acquisition while watching a short video is possible. The findings also suggest that learners benefited more from captions than from L1 subtitles if the aim is learning new word forms. Additionally, it was found that both frequency of occurrence and learners' vocabulary size were positively correlated with learning.

6.1. The type of subtitles: L1 subtitles vs. captions

The findings of the form recall test (experiment 2) suggest that captions have the potential to result in more word learning than L1 subtitles. The evidence for the beneficial effects of captions on form recognition remains inconclusive, because we did not find an effect in experiment 2 and the effect in experiment 1 was conditional on learners' vocabulary size. Although it is not immediately clear why the form recognition findings are not in line in the two studies, the different test item presentation might account for some of the differences. In experiment 1, the items were provided in their aural form; in experiment 2 in their written form. It is not unlikely that the type of input (documentary vs. cartoon) might have played a role as well. Thus,

we found only partial support for our hypothesis that captions are more beneficial for form learning. However, we did not find any evidence that L1 subtitles are more effective than captions for learning the meaning of new words.

The positive effect of captions on form recognition (experiment 1) and form recall (experiment 2) might be explained by the fact that captions do not only draw learners' attention to the word forms, but they provide them with the correct written word form and help learners distinguish between separate words (Danan, 2004). From eye-tracking research we know that captions are read automatically (Montero Perez et al., 2015; Winke et al., 2013). Our findings seem to corroborate Frumuselu et al.'s (2015) results, even though their study focused on extensive viewing (13 episodes of a TV series) and used different test measures. Bianchi and Ciabattoni (2008) found a positive effect of captions but only in their beginners group. In spite of these methodological differences, findings seem to point to the beneficial effect of captions for word learning.

Another issue that needs to be addressed here is the potentially mediating role of imagery. Although the relationship between imagery and the aural presentation of the lexical items was not the focus in our study, it is not unlikely that such visual clues may have helped the learning of some items in the captions group, especially in experiment 2, as there was more visual support in *The Simpsons* episode than in the documentary about eating insects (experiment 1). Such a semantic match between on-screen imagery and the aural presentation of the lexical items might have helped learners establish initial form-meaning links in the mental lexicon, which would be in line with Bianchi and Ciabattoni's (2008) findings, as their study also revealed that the combination of captions and visual support was particularly beneficial, especially for beginner learners. Not only were these learners in those instances provided with the written and aural form but also with access to the meaning of some of the target items by means of visual clues. When provided with the meaning via L1 subtitles, learners will also see the visual clue or image but it might be more difficult to link the meaning of the target item to the L2 aural form in the speech stream.

Finally, it should also be mentioned that the test format in experiment 2 (written test items) might have favored the captions group to some extent. Unlike the participants in experiment 1 who took aural posttests, they were presented with the written form of the items in the captions as well as in the test.

6.2. Frequency of occurrence

A second finding is the beneficial effect of frequency of occurrence on incidental word learning, although it was dependent on learners' vocabulary size in two tests in experiment 2. It seems that repeated encounters with unfamiliar words increase the likelihood that the item will be noticed and retained. This finding supports Rodgers' (2013) and Peters and Webb's (submitted for publication) findings and is also in line with findings that have emerged from reading and listening research (Chen & Truscott, 2010; Rott, 1999; Vidal, 2011; Webb, 2007; Webb et al., 2013; van Zeeland & Schmitt, 2013).

Our aim was not to compare the exact frequency of occurrence (e.g. 1–3–5 occurrences), but we wanted to take this parameter into account, even though most items occurred only once and less than half of the items in experiment 1 and less than one third of the items in experiment 2 occurred more than once. Consequently, our two studies do not allow us to pinpoint the exact number of encounters necessary. They only showed that the odds of learning a word increased when the frequency of occurrence increased. To put it more concretely, the odds of learning a word occurring 12 times, such as *cricket* (experiment 1), were 2.6 higher than learning a word occurring once. Obviously, this does not mean that words with one occurrence were not learned, as some of the best-learned items (*water bug*, *snapshot*, or *crawling* in experiment 1) were only encountered once in the input. This highlights that there is not a one-to-one relationship between frequency of occurrence and word learning, but that other parameters might play a role as well (Pellicer-Sánchez & Siyanova-Chanturia, 2016; Peters & Webb, submitted for publication). As mentioned in the previous subsection, one such factor could be imagery or visual support.

The effect of frequency of occurrence was about the same in the form recognition and meaning recall test in experiment 1 ($Exp^b = 1.10$ and 1.11) but differed in the form recall ($Exp^b = 2.17$), and form recognition and meaning recognition test in experiment 2 because in the latter two, its effect was dependent on learners' vocabulary size. Nevertheless, our findings are different from van Zeeland and Schmitt's (2013) study who found that the effect was smaller in the meaning recall test than in their form recognition and grammar recognition test. Although the results cannot be directly compared because of different test formats, our findings also seem to suggest that the effect of frequency of occurrence was larger in experiment 2 (*The Simpsons* episode) than in experiment 1 (documentary), which might be explained by some of the target items' relevance to understanding the input. In experiment 2, *pen pal* and *honor code* were two of the best-learned items. These two items occurred three and five times in the input, but they were at the same time relevant to understanding the gist of the episode. It might thus not always be possible to disentangle frequency of occurrence and relevance, as relevant words probably occur more frequently as well.

6.3. Learners' vocabulary size

A final result from these two studies is the positive relationship between vocabulary size and learning in all test types. The more words a learner knows, the more words they will learn. The effect of this parameter seems to be quite robust, as the effect of vocabulary size was about the same in all tests, viz. the odds of a correct response increased 2%–5% for one word known more in the vocabulary size test.

The second experiment also showed that even learners with small vocabulary sizes might be able to pick up new words when watching TV with either L1 subtitles or captions. Webb and Rodgers (2009a,b) argued that learners probably need to know 3000 word families in order to benefit from audiovisual input. The results of experiment 2 show that on-screen text and

visual clues might compensate for learners' insufficient vocabulary knowledge, as none of those learners knew 3000 word families. Another interpretation could be that the threshold of vocabulary knowledge put forward by Webb and Rodgers might be lower, as this has not been empirically investigated yet.

At the same time, our findings also seem to support previous claims (Danan, 2004; Vanderplank, 2010) that the audio-visual learning materials should not be above the learners' proficiency level. When learners' prior vocabulary knowledge was very low, as was sometimes the case in experiment 2, learning did not, or hardly, occur,[5] as there were no learning gains for knowledge of form-meaning connections in such cases. Additionally, these learners were unable to understand the gist of the episode, as became clear in the two comprehension questions.

Vanderplank (2010) argues that captions might be better suited for higher-level learners. Although it was not the aim to investigate the interaction between proficiency level and type of subtitles (L1 subtitles vs. captions), our findings seem to indicate that this was not the case in the present study, as not only the (high-)intermediate learners in experiment 1 but also the low-proficiency learners in experiment 2 benefited from the captions. Positive findings of captions for different proficiency levels have been reported in Frumuselu et al. (2015) (high-intermediate learners) and in Bianchi and Ciabattoni (2008) (beginner learners). In the latter, the beginners performed even better when there was a strong relationship between the aural input and the imagery. Also in the *Simpsons* episode (experiment 2), the imagery was probably more supportive than in the documentary (experiment 1). It could, thus, be that it is the combination of captions and imagery that is particularly conducive to learning, irrespective of proficiency level. Obviously, this interpretation remains speculative, but it does highlight the need for more research to clarify the role of proficiency and other variables (imagery) in learning words through audiovisual input.

Only few viewing studies so far (Frumuselu et al., 2015; Montero Perez et al., 2014, 2015) have taken learners' vocabulary knowledge into account in their analyses. Given the strong relationship between reading as well as listening and vocabulary (Schmitt, Jiang, & Grabe, 2011; Stæhr, 2009) on the one hand and the facilitative role vocabulary size seems to play in learning new words on the other (Horst, Cobb, & Meara, 1998; Montero Perez et al., 2014; Peters & Webb, submitted for publication), we would argue that a general vocabulary measure should be used to control for individual differences between participants.

7. Conclusion

Previous research has demonstrated that both L1 subtitles and captions boost word learning compared to not offering subtitles. The findings of the two studies presented here show that EFL learners can indeed learn new words when watching a TV program with L1 subtitles or captions in class. Although we only found an effect of captions on form learning, other factors such as frequency of occurrence and a learner's vocabulary size might have a bigger impact on the learning process. Nevertheless, it should be emphasized that the number of words learned in both studies was small. Most learning gains will have been partial learning gains, as learners' moved from no knowledge of an item to recognition of the form or from form recognition to meaning recognition or recall.

Although this paper reports on two experiments, care should still be taken when interpreting the results given the small sample size in both experiments. Additionally, the effect of captions on the form recall test (experiment 2) might to some extent have been dependent on the type of audiovisual input used, viz. a cartoon with visual clues. Future studies could investigate the role of imagery in audiovisual input in more detail. Additionally, more research is probably needed to study which type of audiovisual input (cartoon, documentary, movie, TV series, …) is better suited to vocabulary learning. Another limitation is that both experiments only used one short intervention and consequently they did not explore long-term learning effects. Finally, more research on the relationship between learners' vocabulary knowledge and vocabulary acquisition through audiovisual input, taking different genres and different levels of visual support into account, should be undertaken.

Our expectations of what and how much can be learned through audiovisual input in a short intervention should be modest given the challenging processing demands. Additionally, vocabulary acquisition is an incremental process in which words should be encountered and retrieved repeatedly before they can be firmly entrenched in the mental lexicon. The real value of audiovisual input and probably also of L1 subtitles and captions is probably to be found in the long run, which is in line with Webb's (2015) plea for extensive viewing outside the classroom. Large learning gains might only be observable in longitudinal studies. Nevertheless, the two experiments presented here show that even short video clips have the potential to result in vocabulary learning among intermediate and low-proficiency EFL learners.

References

Almeida, P. A., & Costa, P. D. (2014). Foreign language acquisition: The role of subtitling. *Procedia − Social and Behavioral Sciences, 141,* 1234–1238. http://dx.doi.org/10.1016/j.sbspro.2014.05.212.

---

[5] Learning occurred only at the level of form recognition.

Bianchi, F., & Ciabattoni, T. (2008). Captions and subtitles in EFL learning: An investigative study in a comprehensive computer environment. In A. Baldry, M. Pavesi, C. Taylor Torsello, & C. Taylor (Eds.), *From didactas to ecolingua: An ongoing research project on translation and corpus linguistics* (pp. 69–90). Trieste: EUT.

Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics, 35*(2), 1–20. http://dx.doi.org/10.1017/S0142716412000434.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. http://dx.doi.org/10.1093/applin/amq031.

Cobb, T.. VocabProfile Home. Retrieved from: http://www.lextutor.ca/vp/.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning, 42*(4), 497–527. http://dx.doi.org/10.1111/j.1467-1770.1992.tb01042.x.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Journal Des Traducteurs, 49*(1), 67–77. http://dx.doi.org/10.7202/009021ar.

Davies, M. (2008). *Corpus of contemporary American English*. Retrieved from: http://corpus.byu.edu/coca/.

Diaz Cintas, J., & Remael, A. (2014). *Audiovisual translation: Subtitling*. Manchester: St. Jerome Pub.

D'Ydewalle, G., & Pavakanun, U. (1995). Acquisition of a second/foreign language by viewing a television program. In P. Winterhoff-Spurk (Ed.), *Psychology of media in Europe: The state of the art—perspectives for the future* (pp. 51–64). Opladen, Germany: Westdeutscher Verlag GmbH.

D'Ydewalle, G., & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition? In P. Winterhoff-Spurk, & T. Van der Voort (Eds.), *New horizons in media psychology* (pp. 145–155). Opladen, Germany: Westdeutscher Verlag GmbH.

D'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28*(3), 227–245. http://dx.doi.org/10.1023/A:1023202130625.

D'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television. Automatic reading behaviour. *Communication Research, 18*(5), 650–666.

European Commission. (2012). *First European survey on language competences*. Retrieved from: http://ec.europa.eu/languages/library/studies/executive-summary-eslc_en.pdf.

Frumuselu, A. D., De Maeyer, S., Donche, V., & Colon Plana, M.del M. G. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education, 32*, 107–117. http://dx.doi.org/10.1016/j.linged.2015.10.001.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207–223.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty, & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 349–381). Malden, MA: Blackwell.

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51–60. http://dx.doi.org/10.1007/BF02299476.

Koolstra, C. M., Peeters, A., & Spinhof, H. (2002). The pros and cons of subtitling and dubbing. *European Journal of Communication, 17*, 325–354. http://dx.doi.org/10.1017/CBO9781107415324.004.

Laufer, B., & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in Foreign Language, 22*(1), 15–30.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4), 391–411.

Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System, 42*, 164–176. http://dx.doi.org/10.1016/j.system.2013.11.010.

Lin, P. M. S., & Siyanova-Chanturia, A. (2015). Internet television for L2 vocabulary learning. In D. Nunan, & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 149–158). New York/London: Routledge.

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.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal, 99*(2), 308–328. http://dx.doi.org/10.1111/modl.12215.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739. http://dx.doi.org/10.1016/j.system.2013.07.013.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition, 38*, 97–130. http://dx.doi.org/10.1017/S0272263115000224.

Pellicer-Sánchez, A., & Siyanova-Chanturia, A. (April 2016). An eye-tracking investigation of incidental collocation learning during reading. In *Paper presented at AAAL 2016*. Orlando.

Peters, E. (2012). The differential effects of two vocabulary instruction methods on EFL word retention: A study into task-effectiveness. *IRAL: International Review of Applied Linguistics in Language Teaching, 50*(3), 213–238.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research, 18*(1), 75–94. http://dx.doi.org/10.1177/1362168813505384.

Peters, E., & Webb S., Incidental vocabulary acquisition through watching a single episode of L2 television (submitted for publication).

Peters, E., Van Rompaey, T., & Velghe, T., The VocabLab tests: The development of an English and French vocabulary test (submitted for publication).

Plonsky, L. (June 2015). Sample demographics and generalizability in SLA: A systematic review. In Paper presented at the reliability and validity in SLA research workshop. Barcelona.

Rodgers, M. P. H. (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* (Unpublished Ph.D. thesis). Wellington: Victoria University.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*, 589–619.

Safar, H., Modot, A., Angrisani, S., Gambier, Y., Eugeni, C., Fonatenel, H., et al. (2011). *Study on the use of subtitling. The potential of subtitling to encourage foreign language learning and improve the mastery of foreign languages*. Retrieved from: http://eacea.ec.europa.eu/llp/studies/documents/study_on_the_use_of_subtitling/rapport_final-en.pdf.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. http://dx.doi.org/10.1111/j.1540-4781.2011.01146.x.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577–607. http://dx.doi.org/10.1017/S0272263109990039.

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

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching, 43*(1), 1–37. http://dx.doi.org/10.1017/S0261444809990267.

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56–89. http://dx.doi.org/10.1093/applin/24.1.56.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*(1), 219–258. http://dx.doi.org/10.1111/j.1467-9922.2010.00593.x.

Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. *Frontiers in Psychology, 5*, 1–10. http://dx.doi.org/10.3389/fpsyg.2014.01510.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65. http://dx.doi.org/10.1093/applin/aml048.

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan, & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). New York/London: Routledge.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*(1), 91–120. http://dx.doi.org/10.1111/j.1467-9922.2012.00729.x.

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427. http://dx.doi.org/10.1093/applin/amp010.

Webb, S., & Rodgers, M. (2009b). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366. http://dx.doi.org/10.1111/j.1467-9922.2009.00509.x/full.

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://llt.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. http://dx.doi.org/10.1111/j.1540-4781.2013.01432.x.

van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), 609–624. http://dx.doi.org/10.1016/j.system.2013.07.012.