tesol QUARTERLY

The effect of imagery and on-screen text on foreign language vocabulary learning from audio-visual input

Elke Peters (KU Leuven)

In the past years, there has been a surge in the number of studies focusing on learning vocabulary from audio-visual input. These studies have shown that learners can pick up new words incidentally when watching TV (Peters & Webb, 2018; Rodgers & Webb, 2019). Research has also shown that the presence of on-screen text (L1 or L2 subtitles) might increase learning gains (Montero Perez et al., 2014; Winke et al., 2010). Learning is sometimes explained in terms of the beneficial role of on-screen imagery in audio-visual input (Rodgers, 2018). However, little is known about the effect of imagery on word learning and how it might interact with L1 subtitles and captions.

This study investigates the effect of imagery in three TV viewing conditions: (1) with L1 subtitles, (2) with captions, and (3) without subtitles. Data were collected with 142 Dutch-speaking EFL learners. A pretest-posttest design was adopted in which learners watched a 12-minute excerpt from a documentary. The findings show that the captions group made the most vocabulary learning gains. Secondly, imagery was positively related to word learning. This means that words that were shown in close proximity to the aural occurrence of the words were more likely to be learned.

KEYWORDS: TV, imagery, captions, subtitles, vocabulary learning, audio-visual input

Introduction

There has been a considerable amount of research into language learning from exposure to foreign language (FL) input. Most studies have focused on written input (e.g., Pellicer-Sánchez & Schmitt, 2010; Webb & Chang, 2015), but studies on the effects of audio-visual input have recently been gaining traction in the field of SLA (e.g., Peters & Webb, 2018; Rodgers & Webb, 2017). There is now growing evidence that FL vocabulary can be picked up incidentally through watching short video clips (Montero Perez, Peters, Clarebout, & Desmet, 2014; Neuman & Koskinen, 1992), full-length TV programs (Peters & Webb, 2018), and extensive TV viewing (Rodgers & Webb, 2019). Neuman and Koskinen (1992) were among the first to highlight the potential of audio-visual input for vocabulary learning. This claim has recently been reiterated by Webb (2015), who argued that watching TV extensively could fill the need for greater FL input, especially in learning contexts with little exposure to authentic FL input, and that TV viewing could thus be a valuable vocabulary learning method. Furthermore, FL learners seem to prefer TV viewing activities over reading activities when engaging with the FL outside of the classroom (Peters, 2018; Peters, Noreillie, Heylen, Bulté, & Desmet, 2019).

A number of studies have addressed how the learning gains from TV viewing can be enhanced by adding on-screen text, such as subtitles in the FL (= captions) or in the first language (L1) (e.g., Koolstra & Beentjes, 1999; Montero Perez et al., 2014; Peters, Heynen, & Puimège, 2016; Winke, Gass, & Sydorenko, 2010). However, most studies have compared either type of subtitles with no on-screen text or compared the effect of captions with the effect of L1 subtitles, whereas little research has focused on a comparison of the three viewing conditions (= audio-visual input with captions, L1 subtitles, or no subtitles). Further, little is known about the role of visual support or imagery

in learning new words, even though imagery has been claimed to be beneficial for word learning from audio-visual input (Rodgers, 2018; Sydorenko, 2010). The present study aims to address these gaps by investigating the effect of imagery on word learning from audio-visual input and by exploring the effect of on-screen text (captions, L1 subtitles) compared to no on-screen text.

Vocabulary learning from audio-visual input

Although most research into incidental vocabulary learning has focused on reading (e.g., Pellicer-Sánchez & Schmitt, 2010; Webb & Chang, 2015) and some studies have explored vocabulary learning through listening (van Zeeland & Schmitt, 2013; Vidal, 2003, 2011), recent studies have shown that learners can also learn new vocabulary (Peters & Webb, 2018; Rodgers & Webb, 2019) when they watch foreign-language TV. Peters and Webb (2018) found that EFL learners picked up four words on average while watching a full-length, one-hour TV program compared to 1.5 words in the control group. Additionally, they showed that words can be learned at the level of meaning recall as well as meaning recognition. Puimège and Peters (2019) demonstrated that in addition to learning the meaning, learners can also learn the form of unfamiliar lexical items (= single words as well as formulaic sequences) through TV viewing. Looking at extensive TV viewing (= watching 10 full-length episodes from one TV show), Rodgers and Webb (2019) found that adult EFL learners learned six new words on average and that the learning gains were comparable to those found in reading studies. However, the control group also learned five new words.

A large number of the studies examining vocabulary learning from audio-visual input have investigated how on-screen text and captions in particular can fuel vocabulary learning through TV viewing (for a recent overview of research into captions, see Vanderplank, 2016a, 2016b). Previous research has established that captions have the potential to boost

vocabulary learning (see e.g., the meta-analysis of 10 studies by Montero Perez, Van Den Noortgate, and Desmet, 2013). Further, foreign language learners also perceive captions as helpful for making form-meaning connections of unfamiliar words (Montero Perez, Peters, & Desmet, 2013; Winke, Gass, & Sydorenko, 2013). Studies on captions have also explored the effectiveness of different types of captions (Montero Perez et al., 2014; Montero Perez, Peters, & Desmet, 2018) and the order effects of viewing condition (viewing with captions before viewing without captions or vice versa) when learners watch a clip twice (Winke et al., 2010). Taken together, these studies have all found beneficial effects of captioned TV viewing for word learning, even though its effect might depend on the word knowledge aspect tested (Montero Perez et al., 2014, 2018) or the orthography of the foreign language (Winke et al., 2010). Although the beneficial effects of captions have often been explained in terms of their potential to help learners segment the speech stream (Montero Perez et al., 2013a; Winke et al., 2010), it should at the same time be noted that some of the positive effects found might also be attributed to test-modality congruency whereby the captioned viewing group is favored because of the written presentation of the target items in the posttests (Jelani & Boers, 2018; Sydorenko, 2010). For instance, Jelani and Boers (2018) found that learners who had watched TV with captions scored better on a word meaning test than learners who had watched TV without captions, but the difference between the two groups was mainly due to the written word prompts in the test and not the aural ones.

Research has also addressed the question whether the use of L1 subtitles might foster vocabulary learning. Empirical evidence for L1 subtitles mainly stems from studies conducted with children. One well-known study is that by Koolstra and Beentjes (1999) who found that young learners picked up more words when they watched a 15 minute clip with L1 subtitles compared to no L1 subtitles. Further, their findings showed that learners who frequently watched English-language TV outside the classroom benefited more from

watching TV than learners who did so less often. Similar findings have been found in other studies with children (e.g., d'Ydewalle & Van de Poel, 1999).

To date, there has been little agreement on which type of subtitles (captions or L1 subtitles) is more effective for vocabulary learning. The studies comparing captions with L1 subtitles have produced contradictory findings (e.g., Bianchi & Ciabattoni, 2008; Bisson, Van Heuven, Conklin, & Tunney, 2012; Frumuselu, De Maeyer, Donche, & Colon Plana, 2015; Peters et al., 2016; Pujadas & Muñoz, 2019), probably as a result of different types of input (movies, TV shows, short clips), different methodologies (one-off interventions, longitudinal studies, different word knowledge aspects tested) and different participant profiles. Nevertheless, captions have been argued to be more beneficial for learning the form of unfamiliar words (Peters et al., 2016) and more beneficial for intermediate and advanced language learners (Danan, 2004), whereas L1 subtitles are said to be more useful for "less skilled learners" (Danan, 2004, p.75).

The role of imagery in vocabulary learning from audio-visual input

The key characteristic of audio-visual input, compared to other types of input, is its combination of aural text and visual images. In his theory of multimedia learning, Mayer (2014) states that "people learn more deeply from words and pictures than from words alone" (p.1). In this line of reasoning, learning will improve when learners have access to visual as well as aural information because they will make mental connections between both types of information provided there is temporal proximity. It is, thus, not surprising that several studies have referred to on-screen imagery or visual support as an explanation for the learning gains from audio-visual input (Peters et al., 2016; Peters & Webb, 2018; Rodgers & Webb, 2019).

Even though the effect of imagery has not been the main focus in empirical studies on TV viewing, some research evidence tentatively points to the beneficial effects of imagery on learning. For instance, it was shown that input with imagery tends to result in more learning than input without imagery (Neuman & Koskinen, 1992). Further, when comparing three input modalities (audio + video + captions, audio + video, and audio + captions), Sydorenko (2010) found that the translation test results of words supported by visual images were higher (36% correct) compared to words not clearly supported by images (6% correct). Additionally, learners themselves also reported to have used images in the video to find the meaning of unknown words (Sydorenko, 2010). From eye-tracking research, we know that learners spend time on the images in the input and even more so when no on-screen text is provided (Bisson et al., 2012). Further, it has been argued that on-screen imagery might be more beneficial when captions are provided because of the semantic match between the visual image, the written form in the captions, and the spoken form of the lexical item in the audio, which might help learners create initial form-meaning links in the mental lexicon (Bianchi & Ciabattoni, 2008; Peters et al., 2016). However, in line with the dual-processing theory of working memory (Baddeley, 2007), one could argue that learners might experience processing difficulties when engaging with both pictorial and written information. As a result of cognitive overload, learners might split their attention between the imagery and the captions (Ayres & Sweller, 2014; Mayer & Moreno, 1998). Yet, eye-tracking research has shown that FL learners are capable of adequately processing both images and on-screen text (Bisson et al., 2012), provided they are familiar with the script of the foreign language (Winke et al., 2013).

The role of imagery in audio-visual input was explicitly addressed in a recent corpus study by Rodgers (2018), who showed that TV genres might differ in the amount of visual support they provide. He found that documentaries contain more imagery related to target

words and with temporal proximity compared to narrative TV genres, which might indicate that there is more potential for vocabulary learning in documentaries compared to narrative genres. In the documentary analyzed (*Planet Earth*), 65% of the words occurred simultaneously with the image, while 7% of the words were still shown within a timeframe of five seconds before or after the aural occurrence. However, research has not systematically investigated the effect of on-screen imagery on vocabulary learning in an intervention study. The present study aims to fill this gap.

Factors affecting vocabulary learning from audio-visual input

Previous research into vocabulary learning from audio-visual input has explored a number of learner-related and word-related variables that might affect the learning gains from audio-visual input. For instance, several studies (e.g., Montero Perez et al., 2014, 2018; Peters & Webb, 2018) have shown that learners' prior vocabulary knowledge, often measured by means of a vocabulary levels test or vocabulary size test, is positively related to word learning. The more words a learner knows, the more words they will pick up while watching TV.

Researchers have also addressed the role of a number of word-related factors in vocabulary learning through TV viewing. Frequency of occurrence has been found to affect word learning positively (Peters & Webb, 2018), although its effect seems to be smaller in longitudinal studies that investigated extensive viewing (Rodgers & Webb, 2019). In a recent study, Puimège and Peters (2019) found that a word's corpus frequency was positively correlated to word learning at the level of form recall. They argue that a word's corpus frequency could to some extent reflect learners' familiarity with a target item. Finally, there seems to be strong evidence that cognates are more likely to be picked up incidentally than non-cognates when learners watch TV (Peters & Webb, 2018). Moreover, in a study which

compared vocabulary learning through reading with vocabulary learning through listening, Vidal (2011) found that the effect of cognates was larger in spoken input compared to written input.

Rationale and research questions

Despite the growing evidence for the benefits of audio-visual input for vocabulary learning, little is known about the effect of imagery on word learning. Although several studies have explored the effects of on-screen text, it remains unclear which type of subtitles is most beneficial and how different types of subtitles might interact with imagery. The aim of the present study is to explore the role of imagery in word learning from audio-visual input and to compare the differential effect of captions, L1 subtitles, and no subtitles. Such a study is needed if we want to understand which factors have an impact on incidental vocabulary learning through TV viewing. The following research questions were addressed:

1. Does imagery have an effect on word learning from audio-visual input?
2. Does input modality (captions, L1 subtitles, no subtitles) have an effect on word learning from audio-visual input?
3. Is there an interaction effect of imagery and input modality on word learning from audio-visual input?

Method

Participants

One-hundred forty-two EFL learners (L1 Dutch) from four fifth and two sixth grade classes of one secondary school in Flanders (the Dutch-speaking part in Belgium) participated in this study. Data of 24 participants were not included in the analyses because these participants were either absent in one of the data collection sessions or they had a much lower score on

the posttest than on the pretest. The latter was the case for one participant. This brought the total number of participants to 118 (M_{age} = 16.4). Learners in the fifth grade were in their fourth year of formal English instruction; learners in the sixth grade were in their fifth year of formal English instruction. Classes were randomly assigned to one of the experimental conditions: (1) captions, (2) L1 subtitles, or (3) no subtitles (= control group). Learners had an intermediate proficiency level (B1 according to the Common European Framework of Reference). Their score on the Vocabulary Size Test (Nation & Beglar, 2007) ranged from 66 to 123 out of 140, with an average of 91.57 (*SD* = 9.77) (see also Results section). Finally, it should be noted that Flanders uses L1 subtitles (and not dubbing) to make foreign language TV programs and movies accessible, so the participants in the present study were used to watching subtitled English-language TV (see also Peters, 2018). In addition to L1 subtitles, Flemish EFL learners also frequently watch English-language TV with captions or without subtitles (Peters, 2018; Peters et al., 2019).

Input

The audio-visual input used was an excerpt (11 minutes + 25 seconds) of the documentary *Planet Earth* (2006), which is the same documentary as the one analyzed in Rodgers (2018). Documentaries have been shown to contain more imagery in close proximity to target words than narrative TV genres, making this excerpt appropriate for our research purposes. As can be seen in Table 1, an analysis of the lexical profile of the documentary excerpt showed that 90% lexical coverage was reached with the 3000 most frequent word families in English, as indicated by the VocabProfile tool in Lextutor (Cobb, n.d.). A lexical coverage of 95% was reached with the 5,000 most frequent word families. Given learners' relatively high scores on the VST and their estimated vocabulary size of 9,157 word families, it can be assumed that the input was comprehensible for the learners (average score on 1K words was 9.03/10, on

2K words 8.97/10 and on 3K words 7.81/10), even though it should be pointed out that the VST is not a levels test, which can estimate coverage of frequency bands.

Table 1

Lexical profile of the input

Frequency Level	**Tokens**	**Cumulative tokens**
1K level	82.53%	82.53%
2K level	6.52%	89.05%
3K level	3.00%	92.05%
4K level	2.09%	94.14%
5K level	2.48%	96.62%
6K level	0.78%	97.40%
7K level	0.86%	97.66%
8K level	0.65%	98.31%
>8K level + off-list words	1.69%	100%

Target items

Thirty-six target items were selected from the audio-visual input. A large number of target items that differed in a number of ways (see Table 2) was selected in order to capture as much incidental learning as possible. Such an approach "may provide a more accurate representation of vocabulary learning gains than frequency-based selection of items" (Webb & Chang, 2015, p. 658). Fifteen items had imagery-support, 21 items were not supported by on-screen imagery. Ten of the 36 items were cognates[1], 26 were non-cognates. Frequency of occurrence ranged from 1 to 9. Some items were high-frequency items (e.g. *egg*, *food*), which were included for reasons of test motivation, while others were low-frequency words

according to Brysbaert and New's (2009) SubtLexUS frequency list. Finally, the items' concreteness, determined by means of Brysbaert, Warriner, and Kuperman's (2014) concreteness ratings, ranged from 1.71 to 4.97, which means that some items were concrete (*egg*), while others were abstract (*barely*).

Table 2

List of target items with their values*

Item	**Imagery**	**Cognate**	**Frequency of occurrence**	**Frequency (LogTrans)**	**Concreteness**	**PoS**
egg	0	0 (1/5)	1	3.1235	4.97	Noun
barely	0	0 (0/5)	1	3.1711	1.71	Adverb
emergence	1	0 (0/5)	1	1.0414	2.56	Noun
chaos	0	1 (5/5)	1	2.6812	2.79	Noun
Arctic	0	1 (5/5)	3	2.0043	n/a	Noun
ploy	0	0 (0/5)	1	1.8976	2.21	Noun
blizzard	0	0 (0/5)	1	2.0000	4.68	Noun
to outrun	1	0 (0/5)	1	2.0000	2.88	Verb
billion	0	1 (3/5)	2	2.7551	3.79	Noun
to stir	1	0 (0/5)	1	2.4800	3.76	Verb
confinement	0	0 (0/5)	1	1.6128	3.61	Noun
to depart	0	0 (0/5)	1	2.0414	2.86	Verb
ordeal	0	0 (0/5)	1	2.0864	2.04	Noun

thawing	0	0 (0/5)	1	1.1139	3.8	Adjective
penguin	1	1 (5/5)	2	2.1703	5	Noun
downy	1	0 (0/5)	1	1.3424	n/a	Adjective
calf	1	1 (5/5)	4	2.1818	4.48	Noun
den	1	0 (1/5)	4	2.4955	4.57	Noun
food	0	0 (1/5)	3	3.8964	4.8	Noun
fur	1	0 (0/5)	1	2.6263	4.69	Noun
slope	1	0 (0/5)	3	2.1790	4.07	Noun
to toboggan	1	0 (0/5)	1	0.6990	4.76	Verb
tundra	1	1 (5/5)	2	1.1761	4.21	Noun
cub	1	0 (0/5)	9	2.0334	4.67	Noun
wilderness	0	1 (5/5)	1	2.3054	3.79	Noun
seal	0	0 (0/5)	2	2.8768	4.63	Noun
shelter	0	0 (0/5)	1	2.7752	4.64	Noun
to convert	0	1 (3/5)	1	2.2041	2.73	Verb
fragile	0	1 (4/5)	1	2.4200	2.86	Adjective
herd	1	0 (2/5)	5	2.5575	4.11	Noun
immensity	1	1 (5/5)	1	0.6021	2	Noun
sheer	0	0 (0/5)	1	2.3692	3.35	Adjective
steep	1	0 (0/5)	1	2.1004	3.76	Adjective
to lure	0	0 (0/5)	1	2.3385	3.59	Verb
pastures	0	0 (0/5)	1	1.8633	4.78	Noun
confidence	0	0 (1/5)	1	2.9974	2.17	Noun

Note: PoS = part-of-speech; *TAALES (Kyle & Crossley, 2015) was used to generate the corpus frequency and concreteness values.

Drawing on Rodgers (2018), imagery was operationalized as the visual occurrence of the target item in the timeframe of five seconds before or after the aural occurrence of the item. However, in contrast to Rodgers (2018), who only analyzed nouns, different PoS were included in the present study. Imagery was determined by two raters. In case of disagreement, a third rater was asked. This resulted in 15 items with imagery or visual support and 21 items without imagery. Secondly, because words with visual support might be more concrete and imageable, it was verified whether concreteness differed between the words with and without imagery by using Brysbaert et al.'s (2014) concreteness ratings. No statistically significant difference was found between the imagery-supported words and the words without imagery support (M_{imagery}= 3.97, $M_{\text{no imagery}}$= 3.47, U=175, p = .23).

Data collection instruments

Vocabulary tests.

Learning gains were measured at the level of form recognition and meaning recall in a paper-and-pencil test that consisted of two parts. One part focused on form recognition, while the other focused on meaning recall. The same test was used as pretest and posttest. Because the L1 subtitles and control group were exposed to spoken input only and the captions group to both spoken and written input, the test items were provided in their written and spoken form in order not to favor any group. Previous research has shown that only providing the written form in tests might favor learners assigned to the captions group as a result of test-modality congruency (Jelani & Boers, 2018; Sydorenko, 2010).

The form recognition part had a checklist format (yes – no), in which learners had to indicate whether they had heard or seen the word before. The test items consisted of the 36 target items and eight distractor items (= non-words[2]) to control for guessing. If learners indicated they recognized three or more non-words, their data were not included in the analysis of the form recognition test. This was the case for 23 students. In the meaning recall part of the test, learners were asked to provide all known meaning senses (translation, synonym, definition) of the test items. The procedure was as follows: first, learners were provided with the spoken and written form of the test item, e.g. *a cub*; next, they ticked off whether they had heard or seen the word *cub* before and then they gave the meaning of the test item *cub*. This test format was used to minimize the test duration, as two separate tests would take longer (see also Peters & Webb, 2018). Both the pretest and posttests had an acceptable reliability, Cronbach's alpha for the form pretest was .70, for the form posttest .74; Cronbach's alpha for the meaning pretest was .72 and for the meaning posttest .74.

Vocabulary size test.

To take into account individual differences between the learners and the three experimental groups, learners had to take the Vocabulary Size Test (VST, Nation & Beglar, 2007). The VST is a frequency-based vocabulary size test, which samples 10 items from each of the first 14 frequency bands of 1,000 word families. The test contains 140 test items in total. The VST, which has a multiple choice format, gives a rough estimate of a learners' vocabulary size. Reliability of the VST was acceptable with a Cronbach's alpha of .80 (n = 118)

Questionnaire.

A short questionnaire (in Dutch) was administered to verify whether the participants had enjoyed watching the documentary and whether they had understood the video. The following four questions were asked:

1. What did you think of the video?
2. What did you (not) like?
3. What did you learn in terms of content? What did you (not) know before watching the video?
4. Are there any new words that you learned from the video excerpt?

The last question allowed us tap into learners' recall of words that were not targeted in the tests. The questionnaire was used to help us interpret the data.

Procedure

All data were collected in February 2017. The procedure consisted of two sessions. In the first session (50 minutes), learners took the Vocabulary Size Test and pretest. In the pretest, the items were provided in their spoken and written form. Learners were at this stage not informed that they would be tested again on the same words. In the second session (50 minutes), one week later, all learners watched the video excerpt twice. Depending on the group, they watched the video with captions, L1 subtitles, or no subtitles. Learners were told that they would have to answer comprehension questions afterwards. Having watched the video, they first completed the questionnaire before taking the unannounced vocabulary posttest. As in the pretest, learners could hear and see the test items.

Scoring and Analyses

The vocabulary pretest and posttest were scored dichotomously. A word recognized in the form recognition part of the test received one point, a word not recognized zero points. As

mentioned before, data of learners who ticked off three or more non-words were not analyzed. Data of 95 participants in total were analyzed in the form recognition test.

Some target items were polysemous, such as *calf* and *seal*. In the meaning recall test, a lenient scoring procedure was adopted, in which any correct meaning sense was considered correct in order not to overestimate the learning gains. As a result, more items were regarded as known in the pretest than when only the context-specific meaning had been scored as correct. For instance, it was assumed that when learners provided the meaning "part of a leg" for *calf*, they would be familiar with the meaning "baby of a cow" too, also because *calf* is cognate with Dutch in the meaning "baby of a cow". The meaning recall test was scored by two raters. Interrater reliability for both the pretest and the posttest was very high: Cohen's kappa (κ) = .93, $p < .0001$ for the pretest and Cohen's kappa (κ) = .94, $p < .0001$ for the posttest. Disagreements between the two raters were resolved through discussion.

To answer our research questions, a repeated measures logistic regression per test part (form recognition and meaning recall) was run in SPSS (version 25), i.e. the Generalized Estimating Equations (GEE) procedure in SPSS. The GEE analysis is appropriate for the analysis of a binary outcome variable (correct or incorrect score in the posttest). Further, it allows for the analysis of clustered data (items clustered within a participant). Learners' knowledge of the target items in the posttest was the dependent variable. Target items that were known in the pretest were not taken into account in the analyses. This means that the analyses were run on those items that could potentially be learned. The predictors in our regression model were: input modality (dummy coded), imagery (binary), the interaction between input modality and imagery, cognateness (binary), corpus frequency (logtransformed), frequency of occurrence, and learners' score on the VST (= prior vocabulary knowledge).

The variables learners' prior vocabulary knowledge (score on VST), cognateness, frequency of occurrence, and corpus frequency were also entered into the regression models because previous research has shown that these factors might affect vocabulary learning through TV viewing (Puimège & Peters, 2019; Peters & Webb, 2018). The assumptions for a repeated measures logistic regression were met: there was no multicollinearity[3] and the observations per predictor ratio was good (rule of thumb is 10:1 ratio, based on the smaller of two counts; 316 observations or correct responses in the form recognition test and 433 observations or correct responses in the meaning recall test). We started with a model including all predictors. When a predictor was not significant, it was removed from the model, but it was always checked if the removal resulted in a lower QIC (= better model fit). If not, the parameter was retained.

Results

Vocabulary Size Test

As can be seen in Table 3, the captions group obtained higher scores than the L1 subtitles group and the no subtitles group. An ANOVA was run to determine whether the differences were statistically significant. All assumptions for an ANOVA (normality, homogeneity) were met. The ANOVA indicated that the groups differed significantly from each other, $F(2, 115) = 4.73$, $p = .01$, $n_p^2 = .08$. A Bonferroni post-hoc test showed that only the captions group differed significantly from the no subtitles group ($p = .008$); no other differences were found. Consequently, learners' VST scores were taken into account as a potentially mediating factor in the regression analyses.

Table 3

Descriptive statistics for the VST, per group

	N	Mean (SD) (Max=140)	[95% Confidence Interval]
Captions group	36	94.89 (10.53)	[91.33, 98.45]
L1 subtitles group	41	91.95 (9.74)	[88.88, 95.02]
No subtitles group	41	88.27 (8.15)	[85.7, 90.84]

Form recognition test

The descriptive statistics of the form recognition pretest and posttest as well as the meaning recall pretest and posttest are provided in Table 4.

Table 4

Descriptive statistics form and meaning pretest/posttest, per group

	Form Pretest Mean (SD) [95% CI]	Form Posttest Mean (SD) [95% CI]	Meaning Pretest Mean (SD) [95% CI]	Meaning Posttest Mean (SD) [95% CI]
Captions	26.19 (3.70)	29.00 (3.12)	17.08 (4.35)	21.06 (4.29)
	[24.70, 27.69]	[27.74, 30.26]	[15.61, 18.56]	[19.60, 22.51]
L1 subtitles	25.53 (3.88)	26.38 (3.93)	15.76 (3.57)	18.61 (3.62)
	[24.13, 26.93]	[24.96, 27.79]	[14.66, 16.85]	[17.47, 19.75]
No subtitles	24.11 (3.84)	24.51 (3.06)	14.24 (2.70)	16.93 (3.33)
	[22.83, 25.39]	[23.49, 25.53]	[13.39, 15.10]	[15.88, 17.98]

Note: the Form test is based on data of 95 participants, the Meaning test on data of 118 participants. CI = Confidence Interval; Max score possible = 36

The GEE procedure was run for 1030 observations (= items not known in form pretest), of which 316 were correct responses and 714 incorrect in the posttest.

The analysis revealed that the odds of learning a word were lower for words without on-screen imagery (see Table 5). Or to put it differently, when a word had visual support, the odds of a correct response were almost three times higher (1/Exp(B) = 1/0.355 = 2.82). Secondly, the analysis showed that the type of subtitles was related to word learning. The captions group performed significantly better than the no subtitles group (p = .001). The odds of learning a word were 2.5 times higher in the captions group compared to the no subtitles group (Exp(B)=2.45). There was no difference between the L1 subtitles group and the no subtitles group (p = .91). Further, pairwise comparisons with Bonferroni correction showed that the captions groups also learned more words than the L1 subtitles group (p < .0001). No significant interaction between group and imagery was found, which means that the effect of imagery was not dependent on the type of subtitles.

Table 5

Parameter estimates of GEE for form recognition test

Parameter	B	SE	95% Wald CI		Wald Chi-Square	*p*	Exp(B)	95% Wald CI for Exp(B)	
(Intercept)	-2.55	.92	-4.36	-.75	7.71	.006	.08	.01	.47
Group= Captions	.90	.27	.38	1.42	11.35	.001	2.45	1.46	4.13
Group= L1Subtitles	-.03	.27	-.55	.49	.01	.914	.97	.58	1.64

Group= NoSubtitles	0								
Imagery=0	-1.04	.25	-1.52	-.56	17.88	.000	.36	.22	.57
Imagery=1	0								
Cognate=0	-1.91	.23	-2.36	-1.46	69.43	.000	.15	.09	.23
Cognate=1	0								
VST	.02	.01	.01	.04	6.06	.014	1.03	1.01	1.05
Corpus frequency	.51	.13	.25	.77	14.56	.000	1.66	1.28	2.15
FreqOccurrence	.23	.05	.12	.33	18.27	.000	1.25	1.13	1.39
Group=Captions * Imagery=0	.56	.34	-.11	1.24	2.67	.102	1.75	.89	3.44
Group= L1Subtitles * Imagery=0	.61	.34	-.06	1.29	3.16	.076	1.84	.94	3.62

Note: CI = Confidence interval; FreqOccurrence = frequency of occurrence; VST = score on Vocabulary Size Test

It should be noted though that the largest effect was found for the cognateness predictor. Cognates were almost 7 times more likely to be recognized than non-cognates (1/0.148 = 6.76). The other predictors, learners' prior vocabulary knowledge, frequency of occurrence, and corpus frequency, were all three positively correlated with word learning in the form recognition test.

Meaning recall test

The second regression analysis was run on 2392 observations (= words not known in the meaning pretest), of which 433 were correct responses and 1959 incorrect in the meaning posttest. The analysis showed that except for the interaction between Group and Imagery, all predictors contributed significantly to the model (see Table 6). Words with on-screen imagery were three times more likely to be learned than words without imagery (1/Exp(B) = 1/.331= 3.02). Second, the odds of a correct response in the meaning test were almost twice as high in the captions group compared to the no subtitles group (Exp(B)=1.97, p<.0001), which is lower than in the form recognition test. There was no difference between the L1 subtitles group and the no subtitles group (p=.46). Pairwise comparisons with Bonferroni adjustments showed that the captions group also performed better than the L1 subtitles group (p=.001).

In line with the results of the form recognition test, cognateness was the strongest predictor of word learning. The odds of a correct response were almost 7 times higher when the item was a cognate (1/Exp(B) = 1/.143 = 6.99). Learners' prior vocabulary size was positively correlated with word learning, as were frequency of occurrence and corpus frequency. It should be noted that the effect of corpus frequency was larger in the meaning recall test than in the form recognition test.

Table 6

Parameter estimates of GEE for meaning recall test

Parameter	B	SE	95% Wald CI		Wald Chi-Square	*p*	Exp(B)	95% Wald CI for Exp(B)	
(Intercept)	-4.40	.628	-5.63	-3.17	49.07	.000	.012	.00	.04

Group= Captions	.68	.145	.39	.96	21.73	.000	1.969	1.48	2.62
Group= L1Subtitles	.12	.161	-.20	.44	.56	.46	1.128	.82	1.55
Group= NoSubtitles	0								
Imagery=0	-1.10	.136	-1.37	-.84	65.64	.000	.331	.25	.43
Imagery=1	0								
Cognate=0	-1.95	.14	-2.23	-1.66	181.61	.000	.143	.108	.19
Cognate=1	0								
VST	.04	.01	.02	.05	27.52	.000	1.035	1.02	1.05
Corpus frequency	.66	.103	.46	.86	41.60	.000	1.940	1.59	2.37
FreqOccurrence	1.17	.0.28	.06	.17	17.23	.000	1.124	1.06	1.19

Note: CI = Confidence interval; FreqOccurrence = frequency of occurrence; VST = score on Vocabulary Size Test

Discussion

The present study extends previous research into vocabulary learning from audio-visual input by focusing on both imagery and three subtitles conditions (captions, L1 subtitles, no subtitles) while also taking into account learner- and word-related factors. The findings suggest that on-screen imagery is a facilitative factor for word learning. In addition, captions are more beneficial to foster word learning compared to L1 subtitles and no subtitles. Finally, other beneficial factors for learning are cognateness, frequency of occurrence, corpus frequency, and learners' prior vocabulary knowledge.

Imagery in audio-visual input

Prior studies have concluded (Peters et al., 2016; Peters & Webb, 2018; Rodgers, 2018; Sydorenko, 2010) that imagery might be conducive to word learning from audio-visual input. The present study shows that words with on-screen imagery are almost three times more likely to be picked up incidentally than words without imagery. This findings holds for knowledge at the level of form recognition as well as of meaning recall. Our findings, thus, provide empirical evidence for Rodgers's (2018) claim that "the imagery present in documentary and narrative television co-occurs with words in the audio soundtrack in such a way that vocabulary learning may be supported" (p.202). As pointed out by Rodgers (2018), it is the presence of on-screen imagery in audio-visual input (TV viewing) that makes it potentially more beneficial for vocabulary learning compared to spoken input (listening). It is interesting to note that when answering the question which words the learners had learned from the video excerpt, the words that were listed most frequently were words with on-screen imagery. such as *cubs* or *den*. Further, learners also mentioned words that were not tested but that clearly had visual support, such as *a pack of wolves* or *caribou*. Some students even pointed out that they had never heard of caribou, so there was conceptual learning too.

The findings are consistent with the multimedia principle (Mayer, 2014). which states that learning will be better when visual and aural information are combined because learners make use of the auditory as well as the visual channel when processing information and building mental representations. Learners' access to the aural and pictorial form of words in audio-visual input helps them construct new knowledge, at least in the case of imagery-supported words. When the visual image occurs in close proximity (= within the timeframe of five seconds before or after the spoken occurrence), learners might be able to link the two sources of information and create a semantic match, which then results in an initial form-

meaning link of the word in the learners' mental lexicon. The on-screen imagery, thus, provides the learners with access to the meaning of words. This interpretation is supported by previous research which has shown that learners link the words they hear with the on-screen images (Sydorenko, 2010).

On-screen text in audio-visual input

The second research question about the effect of input modality on word learning can be answered affirmatively. On-screen text facilitates word learning at the level of form recognition and meaning recall, although it should be stressed that learning occurred in all three conditions. The findings of the current study suggest that captions are more beneficial for word learning than no on-screen text on the one hand and L1 subtitles on the other. Additionally, captions were slightly more beneficial for form recognition than for meaning recall.

The positive effects of captions compared to no on-screen text support the work of other studies (e.g., Jelani & Boers, 2018; Montero Perez et al., 2014; Sydorenko, 2010; Winke et al., 2010). Eye-tracking research has shown that foreign language learners' use of captions is high (Winke et al., 2013) and that captions are perceived as helpful for form-meaning mapping (Montero Perez et al., 2013; Winke et al., 2013). Similar to imagery, the combination of spoken and written language input gives language learners the opportunity to process word information through the auditory as well as the visual channel. As a result. captions support learners in segmenting the speech stream into words. which might make unfamiliar words more salient and consequently more noticeable. Further, no interaction between captions and imagery was found. meaning that the captions group did not learn more but also not fewer imagery-supported words. As shown in eye-tracking studies, the use of captions does not seem to come at the expense of imagery in the case of similar scripts, as

language learners seem quite capable of using the captions as well as the images in audio-visual input (Winke et al., 2013). Even though captions enhance vocabulary learning, it should be noted that TV viewing without captions also has advantages, especially for learning the aural form of words because previous research has shown that learners tend to rely more on the written than on the spoken form when captions are provided (Sydorenko, 2010).

The findings of the present study also suggest that captions are more helpful than L1 subtitles for word learning, which is in line with Bianchi and Ciabattoni (2008) and Frumuselu et al. (2015). In addition. as argued by Peters et al. (2016), captions might be particularly beneficial for learning the form of unfamiliar words, as was the case in the current study too. Captions have been proposed to be better suited for intermediate and advanced language learners (Danan, 2004; Vanderplank, 2016a). whereas L1 subtitles have generally been put forward for beginner learners (Danan, 2004). It should be acknowledged that the participants in the present study were intermediate learners of English (and some even advanced EFL learners) who were used to watching English-language TV. It could be that the effect of captions in the present study is to some extent related to the proficiency level of the participants. The lower scores of the L1 subtitles group might also be explained by learners' being more engaged in reading the L1 subtitles than in actively linking the auditory word form with its translation in the subtitles. When answering the question "Are there any new words that you learned from the video excerpt?", one participant in the L1 subtitles group pointed out that s/he did not learn any new words because s/he was focused more on the Dutch subtitles. Although this is only one reaction, it could point to more passive TV-viewing behavior when L1 subtitles are provided. L1 subtitles might to some extent distract learners' attention away from the spoken form of words which could result in shallower processing of the form-meaning connections of words.

Other beneficial factors for word-learning

Even though the main focus of this study was not on the effect of cognateness, frequency of occurrence, corpus frequency, and learners' prior vocabulary knowledge, the findings indicated that all these factors predicted word learning from audio-visual input.

Cognateness was the most powerful predictor of word learning. The findings support previous research that has shown that cognates are more likely to be picked up incidentally than non-cognates (Peters & Webb, 2018; Vidal, 2003, 2011). It has been shown before that cognateness might be more important in spoken than in written input because they are salient in the spoken input (Vidal, 2011). Given the challenges of decoding and segmenting the speech stream, learners might rely more on words that are similar to words in their L1 when exposed to aural input (Vidal, 2011). It should be noted that d'Ydewalle and Van de Poel (1999) also showed that linguistic similarity might affect vocabulary learning from audio-visual input because Dutch-speaking children learned more Danish than French words when watching TV due to greater similarities between Danish and Dutch vocabulary.

Frequency of occurrence was also positively correlated with word learning, which is in line with previous research (Peters et al., 2016; Peters & Webb, 2018). Documentaries in particular seem to be characterized by repetition of the same words because "the same referent is likely to be talked about for a prolonged period" (Rodgers, 2018, p.205). The combination of repeated encounters and on-screen imagery appears to make some words highly salient in the input and good candidates for learning. However, it should be noted that the repeated viewing of the clip will have enhanced the effect of frequency of occurrence. A word's corpus frequency was also found to predict word learning, as was also shown in Puimège and Peters (2019). Corpus frequency can be considered a proxy for word familiarity (Kuperman & Van Dyke, 2013) and thus for previous encounters with words. Words occurring more often in the language will be encountered more frequently. The learners in the

present study could already have had partial knowledge of some of the more frequent items, which could not be tapped into with the pretest. It, thus, seems reasonable to assume that the correlation found for corpus frequency might to some extent reflect familiarity with the target words.

Finally, in accordance with previous results (Montero Perez et al., 2014; Peters et al., 2016; Peters & Webb, 2018). learners' prior vocabulary knowledge was positively correlated with word learning at the level of form recall as well as meaning recall. The more words learners know, the higher their lexical coverage of the audio-visual input will be, making it easier to segment the spoken input. Consequently, the higher the lexical coverage, the more words a learner will pick up.

Pedagogic implications

It is well established that exposure to FL input is beneficial for language learning. The question is how audio-visual input differs from other types of input, such as written and spoken input, and how these differences might affect vocabulary learning. Audio-visual input has three advantages compared to written and spoken input: (1) it is motivating, (2) it contains visual support, and (3) it recycles low-frequency words. First, research into the effects of out-of-school exposure on vocabulary learning (e.g., Peters, 2018; Peters et al, , 2019) has shown that learners of different age groups engage more frequently in English-language TV viewing than in reading English books. As pointed out before, learners are motivated to learn a language through watching TV and movies (Colwell and Ipince Braschi, 2006. as cited in Webb & Rodgers, 2009b; Sockett & Kusyk, 2015). In the questionnaire, learners mentioned that they enjoyed watching the documentary and that they found the video interesting, engaging, informative, and beautiful. Secondly, audio-visual input presents information in two or more forms, i.e. spoken text combined with visual images. As the

present study has shown, on-screen imagery is beneficial for vocabulary learning, as it allows learners to process information through the auditory and visual channel. Third, in addition to providing learners with authentic, spoken English-language input (Webb, 2015), audio-visual input has the advantage that low-frequency words are repeated more often in a relatively small amount of TV viewing time compared to reading (Webb & Rodgers, 2009a). This is especially the case in related TV programs, such as episodes from the same TV show (Rodgers. & Webb, 2011), and documentaries (Rodgers, 2018). It is the combination of these three features that makes audio-visual input a particularly valuable source for language learning with great potential for vocabulary.

Finally. captioned audio-visual input has the advantage that it also provides language learners with the written form of words, which might help learners decode the rapid speech of audio-visual input (Vanderplank, 2016a). The present study adds to the growing body of evidence that captions are indeed useful for learning vocabulary. However. as pointed out by Sydorenko (2010), "different types of video input seem to provide different benefits" (p.64). When the goal is listening or learning the aural form, it might be better to use audio-visual input without captions. Teachers should support EFL learners in making use of captions outside of the classroom and in becoming independent language learners. As put forward by Webb (2015) and as shown in the research into out-of-school exposure (Peters, 2018; Sockett & Kusyk, 2015), the greatest potential of TV viewing probably lies outside of the classroom. Given the increased availability of audio-visual input (TV shows, movies, documentaries, …) via the internet or streaming services, most EFL learners should have relatively easy access to large amounts of authentic, spoken input.

Limitations and future research

Even though this study adds to our understanding of the factors that might boost vocabulary learning from audio-visual input, the study is not without its limitations. First, the study did not control for a testing effect because we did not have a group who were not exposed to the input and only took the tests. Because of practical constraints, only an immediate posttest was administered. Consequently, the present study cannot give evidence for long-term retention of the target words. It should be noted that this study only used a short excerpt from one documentary. Future research should investigate the effects of L1 subtitles and captions as well as the effects of imagery longitudinally. Although both the written and spoken form were provided in the tests, learners in the captions group might have relied mostly on the written form. Previous research (Jelani & Boers, 2018; Sydorenko, 2010) has shown that the benefits of captions might be due to written prompts. Future research should investigate to what extent captions promote the learning of the written or the spoken word form. Even though care was taken to control for guessing in the form recognition test by including non-words, future research might use a test format that is less sensitive to guessing. Given the differences in on-screen imagery between different TV genres (Rodgers, 2018). further research should be undertaken to compare the effect of imagery in documentaries with narrative genres for instance. Additionally, Rodgers (2018) found that 7% of the nouns in the documentary analysed did not occur simultaneously with the image, but within a timeframe of five seconds. It would be worthwhile to investigate whether the timeframe (concurrent or asynchronous presentation [2 or 5 seconds before/after] of the image with the aural form, as in Rodgers, 2018) mediates the effect of imagery on word learning. Even though teachers will often show a video excerpt twice, it should be noted that the immediate repeated viewing in the present study will have enhanced learning. Finally, the participants in this study were used to watching (subtitled) English-language TV programs. More research is needed with

different participant profiles (e.g., young learners, learners not used to watching subtitled TV) and other foreign languages in order to generalize the findings of the present study.

Conclusion

Among other factors the present study focused in particular on the effect of on-screen imagery and on-screen text on word learning from audio-visual input. The study is the first to show that on-screen imagery is conducive to word learning at the level of form recognition as well as meaning recall. Even though learning occurred in all three viewing modalities (captions, L1 subtitles, no subtitles), the findings further support previous research that has suggested that captions have the potential to boost vocabulary learning from audio-visual input. In addition to imagery and captions, word learning was also enhanced by word-related factors, such as cognateness, frequency of occurrence, and corpus frequency, as well as learners' prior vocabulary knowledge.

Endnotes

1. Cognateness was determined by having five raters determine whether the target word was a cognate or not. If items were considered cognate by the majority of raters (3 or more out of five), they were labeled as a cognate item. There was 90% agreement between the five raters.
2. Non-words were taken from the LexTutor website (Cobb, s.d.): *almanical. to cantileen. to combustulate. dogmatile. to galpin. to humberoid. nickling. to rudge.*
3. It was verified that the Spearman correlations between the predictors were all lower than .70.

Acknowledgements

I am grateful to the three reviewers whose feedback has greatly improved this article. I would also like to thank Brecht Rubens for his help in the data collection and Marion Durbahn Quinteros for her help in analyzing the relationship between on-screen imagery and audio.

Bio

Elke Peters is Associate Professor at KU Leuven, Belgium. Her research interests focus on incidental and deliberate learning of single words and formulaic sequences in a foreign language. She is interested in how different types of input can contribute to vocabulary learning.

References

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 206-226). Cambridge: Cambridge University Press.

Baddeley, A. (2007). *Working memory. thought. and action*. Oxford: Oxford University Press.

Bianchi, F., & Ciabattoni, T. (2008). Captions and subtitles in EFL learning: an investigative study in a comprehensive computer environment. *From Didactas to Ecolingua: An Ongoing Research Project on Translation and Corpus Linguistics*, 69–90.

Bisson, M.-J., Van Heuven, W. J. B., Conklin. K., & Tunney, R. J. (2012). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*, 1–20. https://doi.org/10.1017/S0142716412000434

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word

frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46(3),* 904–911. doi: 10.3758/s13428–013–0403–5.

Cobb, T. (n.d.). Vocabprofile [Computer program]. Retrieved from http://www.lextutor.ca

Cobb, T. *Plausible non-words*. Accessed 2 December 2016 at https://www.lextutor.ca/freq/lists_download/pnwords.html.

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

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. https://doi.org/10.1023/A:1023202130625

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,* 1–11. https://doi.org/10.1016/j.linged.2015.10.001

Jelani. N. A. M., & Boers. F. (2018). Examining incidental vocabulary acquisition from captioned video. Does test modality matter? *ITL - International Journal of Applied Linguistics, 169*(1), 169–190. https://doi.org/10.1075/itl.00011.jel

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. https://doi.org/10.1007/BF02299476

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology. Human Perception and Performance, 39*(3), 802–823. doi: 10.1037/a0030859

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757–786. https://doi.org/10.1002/tesq.194

Mayer, R. (2014). Introduction to multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-24). Cambridge: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. doi: 10.1037/0022-0663.90.2.312

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. http://dx.doi.org/10125/44357

Montero Perez. M., Peters. E., & Desmet. P. (2013b). Is less more? Effectiveness and perceived usefulness of keyword and full captioned video for L2 listening comprehension. *ReCALL, 26*(1), 21–43. https://doi.org/10.1017/S0958344013000256

Montero Perez. M., Peters. E., & Desmet. P. (2018). Vocabulary learning through viewing video: the effect of two enhancement techniques. *Computer Assisted Language Learning, 31*(1–2), 1–26. https://doi.org/10.1080/09588221.2017.1375960

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

Nation. P., & Beglar. D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13.

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), 94–106.

Pellicer-Sánchez. A., & Schmitt. N. (2010). Incidental vocabulary acquisition from an authentic novel: do things fall apart? *Reading in a Foreign Language, 22*(1), 31–55.

Peters. E. (2018). The effect of out-of-class exposure to English language media on learners ' vocabulary knowledge. *ITL - International Journal of Applied Linguistics, 169*(1), 142–168. https://doi.org/https://doi.org/10.1075/itl.00010.pet

Peters. E.. Heynen. E.. & Puimège. E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System, 63*, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Peters, E., Noreillie, A., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, 69, 747–782. https://doi.org/10.1111/lang.12351

Peters. E.. & Webb. S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Puimège, E., & Peters. E. (2019). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *Language Learning Journal*. https://doi.org/10.1080/09571736.2019.1638630

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: a study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, 47, 479–496.. https://doi.org/10.1080/09571736.2019.1616806

Rodgers, M. P. H. (2018). The images in television programs and the potential for learning unknown words The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics*, *169*(1), 191–211. https://doi.org/10.1075/itl.00012.rod

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television programs. *TESOL Quarterly*, *45*(4), 689–717. https://doi.org/10.5054/tq.2011.268062

Rodgers, M. P. H., & Webb, S. (2017). The Effects of Captions on EFL Learners ' Comprehension of English-Language Television Programs. *Calico Journal*, *34*(1), 20–38. https://doi.org/10.1558/cj.29522

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning from viewing television. *ITL - International Journal of Applied Linguistics.* Published online 25 June 2019. https://doi.org/10.1075/itl.18034.rod

Sockett, G., & Kusyk, M. (2015). Online informal learning of English: frequency effects in the uptake of chunks of language from participation in web-based activities. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 153-177). Berlin/Boston: Mouton De Gruyter.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73. http://dx.doi.org/10125/44214

Vanderplank, R. (2016a). "Effects of" and "effects with" captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, *49*(2), 235–250. https://doi.org/10.1017/S0261444813000207

Vanderplank, R. (2016b). *Captioned media in foreign language learning and teaching*. London: Palgrave MacMillan. https://doi.org/10.1057/978-1-137-50045-8

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

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56–89. https://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. https://doi.org/10.1111/j.1467-9922.2010.00593.x

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: Routledge.

Webb, S., & Chang, A. C. S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*(4), 651–675. https://doi.org/10.1017/S0272263114000606

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

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

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. http://dx.doi.org/10125/44214

Winke. P., Gass. S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal*, *97*(1), 254–275. https://doi.org/10.1111/j.1540-4781.2013.01432.x