CrossMark

Studies in Second Language Acquisition 42 (2020), 525–549
doi:10.1017/S027226311900055X

Research Article

LEARNING FORMULAIC SEQUENCES THROUGH VIEWING L2 TELEVISION AND FACTORS THAT AFFECT LEARNING

Eva Puimège*

KU Leuven

Elke Peters

KU Leuven

Abstract

The present study explores the incidental learning of formulaic sequences (FS) from audio-visual input and factors affecting the learning of FS. A pretest-posttest, within-participant design was adopted. English-as-a-foreign-language learners (L1 = Dutch; n = 42) watched a one-hour English-language documentary without subtitles. Learning gains were measured at the level of form recall and meaning recall. Significant learning gains were found in both the form recall and meaning recall tests, indicating that several knowledge aspects of FS can be learned incidentally from audio-visual input. The findings also revealed a positive relationship between learners' prior vocabulary knowledge and the learning gains. Further, several item-related factors affected the learning gains (e.g., collocate-node relationship, frequency). Finally, the findings showed the incremental nature of learning FS, as preknowledge of the meaning of the FS prior to the TV viewing was the strongest predictor of learning the form of FS.

INTRODUCTION

Formulaic language, long regarded as a peripheral phenomenon, is now considered an essential aspect of second/foreign (L2) vocabulary acquisition (Schmitt, 2010). SLA research has seen a large increase in studies examining the acquisition of formulaic sequences (FS) and collocations in particular (see Boers & Lindstromberg, 2012 for an overview). Studies have demonstrated that FS can be learned through explicit learning activities (e.g., Boers, Demecheleer, Coxhead, & Webb, 2014; Peters, 2014, 2016; Webb

*Correspondence concerning this article should be addressed to Eva Puimège, PhD Fellowship of the Research Foundation – Flanders, KU Leuven, Faculty of Arts, Campus Antwerp, Sint-Andriesstraat 2, B-2000 ANTWERP. E-mail: eva.puimege@kuleuven.be

Copyright © Cambridge University Press 2019

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

& Kagimoto, 2011). However, given that time spent on teaching vocabulary in the classroom is limited, the question arises whether FS can be learned incidentally from meaningful L2 input (reading, listening, TV viewing).

Most studies that have examined the incidental learning of FS have focused on reading (Pellicer-Sánchez, 2017; Szudarski, 2012; Szudarski & Carter, 2016) or reading-while-listening (Webb, Newton, & Chang, 2013). However, one type of input that has recently received increased attention in studies on incidental vocabulary acquisition is television. There is empirical evidence that TV viewing is an effective way of learning L2 vocabulary (e.g., Peters & Webb, 2018). Yet, research into the potential of TV viewing for FS is limited (Puimège & Peters, 2019a), even though television has been shown to be a rich source of formulaic language (Lin, 2014). At present, it remains unclear whether learners can pick up FS while watching a TV program. The present study aims to investigate incidental learning of FS through TV viewing, and which item-related and learner-related variables might mediate the learning of FS.

BACKGROUND

INCIDENTAL VOCABULARY ACQUISITION THROUGH TV VIEWING

There is growing evidence that exposure to authentic input plays an important role in L2 learning (e.g., Lindgren & Muñoz, 2013; Peters, 2018). Exposure to the L2 through reading, listening to music, video games, and TV viewing has been shown to benefit learners' L2 proficiency and vocabulary knowledge (e.g., Lindgren & Muñoz, 2013; Peters, 2018; Peters et al., 2019; Puimège & Peters, 2019b). One of the most common and influential types of L2 input is television. Lindgren and Muñoz (2013) found that TV viewing had a greater influence on reading and listening proficiency in young learners than other types of out-of-class exposure. Similarly, Peters (2018) found that watching television without subtitles was one of the most important predictors of vocabulary knowledge in secondary-school and university level learners.

Activities such as reading and TV viewing, in which the main focus is not on vocabulary acquisition, can be considered incidental learning activities (Schmitt, 2010). Following previous research (e.g., Hulstijn, 2003; Nation & Webb, 2011), the present study operationalizes incidental learning in terms of test announcement. A learning task is considered incidental when learners are not instructed to pay attention to vocabulary items or forewarned of a vocabulary test (for a thorough discussion, see Hulstijn, 2003). Research has demonstrated that single words can be acquired incidentally through TV viewing (Feng & Webb, 2019; Peters & Webb, 2018; Rodgers & Webb, 2019). Peters and Webb (2018) investigated incidental vocabulary learning from a single episode of a TV program. They performed two experiments in which participants watched a one-hour documentary. Learning gains were measured using a form recognition and meaning recall test in the first experiment, and a meaning recognition test in the second experiment. In both experiments, approximately four words were learned between the pre- and posttests.

In Rodgers and Webb's (2019) study, Japanese EFL learners watched 10 episodes of an English-language TV program, with a total viewing time of 7 hours. Fifty-six target words were tested before and after the experiment. The learning gains of the experimental group were significantly higher than those of the control group, who only

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

completed the tests. Moreover, a medium-sized correlation was found between frequency of occurrences of the target items and learning gains, showing the importance of frequency in incidental vocabulary learning.

Taken together, these studies show that TV viewing may result in incidental acquisition of single words. Further, it was found that factors that affect learning include exposure frequency, cognateness, and learners' prior vocabulary knowledge (Feng & Webb, 2019; Peters & Webb, 2018; Rodgers & Webb, 2019). However, not much is known about the potential of TV viewing for acquiring formulaic language.

LEARNING FORMULAIC SEQUENCES FROM INPUT

There are two approaches to the study of formulaic language: the phraseological approach (Granger & Paquot, 2008; Nesselhauf, 2005) and the distributional or frequency-based approach (e.g., Biber, 2009). For the purposes of this study, we adopt a definition of FS that combines both approaches. FS are defined as conventionalized and recurring word combinations (e.g., Van Lancker Sidtis, 2012), varying in length, frequency, semantic transparency, contiguity, and flexibility (Siyanova-Chanturia & Pellicer-Sánchez, 2019).

FS are widespread in language (e.g., Erman & Warren, 2000) and fulfill many communicative functions (Schmitt, 2010). Knowledge of FS has been shown to benefit language processing in both native speakers and L2 learners (e.g., Siyanova-Chanturia & Martinez, 2015; Underwood, Schmitt, & Galpin, 2004). The use of FS has also been found to be a key predictor of L2 writing and speaking proficiency (Boers et al., 2014; Crossley, Salsbury, & McNamara, 2015; Paquot, 2017). As such, knowledge of FS is considered to play a crucial role in attaining nativelike L2 proficiency. However, mastery of formulaic language has been shown to be a major challenge to L2 learners, particularly in terms of productive use. Compared to native speakers, learners make use of a smaller range of FS (e.g., Laufer & Waldman, 2011). They also tend to make many mistakes, often L1 induced, leading to their speech and writing being perceived as unnatural and nonnativelike (e.g., Laufer & Waldman, 2011; Nesselhauf, 2005).

A reason why formulaic language tends to be problematic for L2 learners is the low frequency of individual FS in language. Compared to their single-word components, many FS have a very low occurrence rate and are therefore less likely to be encountered in L2 input than single words (Bahns & Eldaw, 1993). The amount of exposure needed for incidental learning of FS is thus much higher than for the single words they are made up of (Webb et al., 2013). Because learners do not have access to the same amount of L2 input as native speakers, and the opportunities for encountering FS during reading and listening are limited (Cobb, 2007), some researchers have suggested that FS need to be taught explicitly (Boers & Lindstromberg, 2012). However, as Webb et al., (2013) argue, there is simply not enough time to teach all FS in the classroom. It thus seems warranted to also investigate the potential of exposure to L2 input and the learning gains that can be made in an incidental context.

Evidence for the potential of reading for learning FS incidentally is inconclusive. Szudarski (2012) investigated the acquisition of verb-noun collocations from reading.

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

Two experimental groups read a text containing target collocations. The first group only read the text, while the second group also completed form-focused activities. On average between 7% and 14% of collocations were learned in the reading-only group. No significant difference in learning gains was found between the reading-only group and a control group (which only completed the vocabulary tests), indicating that reading only did not lead to considerable improvement in collocational knowledge. However, the items targeted included verb-noun collocations containing delexicalized verbs such as *take office* and *do damage*, which have been shown to be particularly difficult for L2 learners (e.g., Peters, 2016).

In a highly controlled study, Pellicer-Sánchez (2017) focused on the learning of adjective - pseudo-noun collocations for five word knowledge aspects when EFL learners read a story. Learning gains were found for all aspects of knowledge, but were higher for receptive knowledge (3.20 out of 6 words on average) than for productive knowledge (.90 out of 6 words) of collocations. In three case studies, Macis (2018) investigated the learning of "duplex collocations," or collocations that have both a figurative and a literal meaning such as *drop the ball* and *hit the roof*, through reading a modified novel. In individual interviews, three participants were asked to recall the meaning(s) of the target collocations before and after they read the novel. Between 20 and 44 percent of target collocations ($n = 76$) were learned between the pretest and the delayed posttest administered 3 weeks after the participants had read the novel.

One study has investigated the learning of FS through reading-while-listening. In Webb, Newton, and Chang's study (2013), EFL learners read and listened to four modified versions of a graded reader, each including a different number of occurrences of 18 verb-noun collocations. The results showed that learners were able to pick up receptive and productive knowledge of the form and meaning of the collocations. The lowest posttest scores were found in a productive form test for items encountered once (M = .60 out of 18 items), and the highest scores were found in a receptive form test for items encountered 15 times (M = 15.58 out of 18 items).

There is reason to believe that collocations and other types of FS can also be acquired from watching L2 television. Corpus research has shown that formulaic language is ubiquitous in spoken language (e.g., Erman & Warren, 2000), and that the distribution of formulaic language in television is similar to that in everyday speech (Lin, 2014). Television might therefore be expected to give adequate exposure to L2 FS for incidental learning. Only one study (as far as we know) has addressed the incidental learning of FS from watching nonsubtitled television. In an exploratory study, Puimège and Peters (2019a) investigated the potential of TV viewing for learning single words and FS. Twenty intermediate to high-proficiency EFL learners watched a 20-minute episode of an English-language TV program. Learning gains were found for both single words and FS. In the form recall test, 2 out of 15 single words and 3.95 out of 20 FS were learned on average. These gains were slightly higher than in the meaning recall test (M = 1.05 for single words, M = 2.4 for FS), but the latter was not further analyzed because of a potential test effect. Given the limited sample size and the limited number of FS ($n = 20$) tested, more research into the effect of TV viewing on learning FS is warranted.

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

FACTORS AFFECTING THE LEARNING BURDEN OF FORMULAIC SEQUENCES

Frequency and Association Strength

An important consideration in incidental vocabulary acquisition is the frequency of lexical items. It is widely acknowledged that high-frequency words tend to be learned before lower-frequency words (e.g., Ellis, 2002; Nation, 2001). Learners' sensitivity to frequency patterns in language arises through statistical learning mechanisms in which knowledge is built through repeated encounters in varied contexts (Ellis, 2006). However, the relationship between knowledge and frequency is not as straightforward for FS (Durrant, 2014). Apart from the frequency of a FS as a whole, the frequency of its single-word components is likely to determine its learnability. Further, contingency and association strength, i.e., the frequency of co-occurrence of the single-word components in relation to their frequency of occurrence in other word combinations, might also affect learners' knowledge of FS. One consistent finding in previous research is that FS that are frequent in use and contain high-frequency components tend to be known better than low-frequency FS containing low-frequency components (e.g., Durrant & Schmitt, 2010; González Fernández & Schmitt, 2015; Nguyen & Webb, 2016). For example, González Fernández and Schmitt (2015) tested advanced learners' knowledge of collocations, and examined the relationship between knowledge and the association strength of the collocations, based on corpus-derived measures. It was found that collocations with high t-scores (e.g., *good example*), which tend to be high-frequency collocations, were more likely to be known than collocations with high MI scores (e.g., *densely populated*), which tend to have lower frequency (Evert, 2008).[1] A similar finding was attested by Durrant and Schmitt (2010), who focused on learners' productive knowledge of collocations. They found that learners make extensive use of high t-score collocations, but produce fewer high MI collocations than native speakers.

Puimège and Peters (2019a) looked into the effect of association strength on the incidental learning of FS through TV viewing and found that there was a negative relationship between the MI scores of the target items and learning. This finding tentatively suggests that items with lower MI scores might be more easily learned incidentally probably due to the low frequency of their components. However, more research into the effect of frequency and association strength is needed.

Learners' frequency of encounters with FS has also been investigated more directly in intervention studies on incidental learning of FS. Three studies found a significant effect for frequency of occurrence of target collocations in the input (Macis, 2018; Szudarski & Carter, 2016; Webb et al., 2013), but Pellicer-Sánchez (2017) did not. The reason for this may be that frequency was operationalized differently in these studies, and that different types of FS (e.g., adjective-noun collocations in Pellicer-Sánchez [2017] and verb-noun in Webb et al. [2013]) were tested.

Collocate-Node Relationship

Another variable found to affect the learning burden of FS is their collocate-node relationship, or the combination of word classes or parts of speech of the constituent parts (Peters, 2016; Puimège & Peters, 2019a). In a study on deliberate learning of collocations (Peters, 2016), verb-particle combinations (e.g., *head back*) were by far the most difficult

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

to acquire incidentally through TV viewing, followed by verb-noun combinations. In spite of the scarcity of studies into the factors that might affect the learning of FS, it seems that word combinations containing a verb might be more difficult to acquire, possibly because many verb-noun or noun-verb collocations contain delexicalized verbs, such as *make*, *do*, and *take*, which tend to have low salience, as they often do not carry the meaning of a FS (e.g., *take a chance*, *go crazy*, *make acquaintance*) (Boers et al., 2014). This could explain why they might not be noticed when learners encounter them during reading (Boers et al., 2014). Second, FS containing verbs have greater morphological variation than, for example, adjective-noun combinations (Peters, 2016). Finally, it should be noted that phrasal verbs (e.g., *give up*, *take off*) may be even more difficult because they can be highly polysemous, idiomatic, and semantically opaque (Garnier & Schmitt, 2016).

Semantic Transparency

FS also vary in their level of fixedness and semantic transparency or compositionality, i.e., the extent to which the meaning of a FS as a whole can be derived from the meaning of its constituents (Lewis, 1993). While some types of FS, namely collocations and lexical bundles tend to be transparent in meaning (e.g., *shift the blame*, *what do you think*), idioms and proverbs can be semantically opaque (e.g., *beat around the bush*) or ambiguous (e.g., *kick the bucket*).

Research examining the effect of semantic transparency on L2 learners' knowledge of FS has yielded conflicting findings. Macis and Schmitt (2017) investigated learners' knowledge of polysemous collocations that have both a figurative and a literal meaning. They found that intermediate EFL learners lacked knowledge of the figurative meanings of the target collocations, with only 33% of figurative meanings being known on average. However, semantic transparency, operationalized through ratings, was not a significant predictor of knowledge. Garnier and Schmitt (2016) measured learners' knowledge of different meaning senses of frequent English phrasal verbs. No relationship was found between semantic transparency and knowledge of phrasal verbs. The effect of transparency has not been examined in studies on incidental learning of FS.

Congruency

Another variable that might affect the learning burden of FS is L1 congruency. Congruency is "the presence or absence of a literal L1 translation equivalent" (Peters, 2016, p. 114). For example, for Dutch-speaking learners, the English collocation *go for a walk* is incongruent, as the Dutch equivalent is translated as *make a walk*. An example of a congruent collocation for Dutch-speaking learners of English is *close a deal*. These are clear examples, but it should be noted that congruency is not always that straightforward, for example, when one of the components in the FS is not used in its prototypical sense (e.g., "pull" in *pull a muscle*) (Nesselhauf, 2005; Peters, 2016). Nevertheless, congruency has been shown to affect the deliberate learning of L2 collocations. Peters (2016) found that congruent collocations were recalled more easily than incongruent ones, and some of the mistakes in recalling target collocations were L1 induced, showing the importance of L1 interference in L2 production. Further, learners have also been shown

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

to make more errors for incongruent than for congruent collocations in acceptability judgment tasks (Yamashita & Jiang, 2010), as well as in production of FS in writing (e.g., Laufer & Waldman, 2011; Nesselhauf, 2005). Finally, research has also shown that congruent FS have a processing advantage over incongruent ones (Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018). It seems reasonable to assume that congruent FS will be easier to learn incidentally, given their resemblance to L1 word combinations. Yet, there is no empirical evidence for this hypothesis.

THE PRESENT STUDY

Compared to the number of studies on single words, research into the incidental acquisition of FS is still scarce. Further, even though studies on the potential of TV viewing for language learning are gaining traction, little is known about the learning of FS from audio-visual input (Puimège & Peters, 2019a). Given that TV is an important source of input for language learning, more studies are needed that explore the potential of TV for different aspects of vocabulary learning. To the best of our knowledge, no study has explored the learning of FS from one full-length TV episode. The present study aims to investigate the incidental acquisition of FS through watching one full-length nonsubtitled TV episode by testing a large number of FS varying in corpus frequency, degree of semantic transparency, congruency, and collocate-node relationship. Research into the incidental learning of FS from audio-visual input is important if we want to understand the item-related and learner-related mechanisms underlying this learning process. The following research questions were addressed:

1. What is the effect of watching a one-hour TV program on the learning of FS?
2. What is the relationship between learner- and item-related factors and vocabulary learning gains?

A pretest-posttest within-participants design was adopted to answer our research questions, as within-participants designs have been argued to have more power than between-participants designs. Further, fewer participants are needed in within-participants studies and no random assignment of participants to conditions is needed (Charness, Gneezy, & Kuhn, 2012). Because we did not include a control group (no exposure to audio-visual input), we used distractor items not occurring in the input in our tests to control for a test effect. Participants watched a one-hour documentary during an English class and were tested on their knowledge of the form and meaning of 56 FS appearing in the input and of 9 distractor items not occurring in the input.

METHODOLOGY

PARTICIPANTS

Seventy-seven Flemish EFL learners (L1 = Dutch) in their first year at university participated in the present study. Data from 35 participants were removed because they did not attend all four sessions of the experiment or because they did not have Dutch as their L1, bringing the total number of participants to 42 (32 female; 10 male;

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

M_{age} = 18.72 (range = 11)). The participants could be considered upper-intermediate learners of English (B1-B2 according to the European Framework of Reference). Their results on the Vocabulary Levels Test (VLT; Schmitt et al., 2001)[2] showed that they were familiar with the 5,000 most frequent word families in English (see "Results" section for more details). Given the present study's focus on incidental vocabulary learning from audio-visual input, it should be noted that EFL learners in Flanders, the Dutch-speaking part of Belgium, are used to watching English-language TV programs and movies with and without subtitles (Peters, 2018; Peters et al., 2019). As was shown in a questionnaire about the participants' viewing habits, the learners in the present study regularly watched English-language television with and without subtitles (see also "Results").

INPUT

An authentic full-length documentary that was relevant to the language students was selected as the audio-visual input, namely the episode "Uses and Abuses" of Stephen Fry's *Planet Word*. The documentary focuses on the role of swearing and euphemisms in language. Like Peters and Webb (2018), we used a full-length, one-hour TV program because such viewing conditions better reflect authentic viewing behavior. Additionally, the use of a one-hour TV program allowed us to select a sufficiently large set of target FS to explore the role of a range of item-related variables, which would not have been possible with a short video clip. The input contained 8,124 tokens and 1,085 types, of which 91% belonged to the 2,000 most frequent word families in the British National Corpus (BNC) and Corpus for Contemporary American English (COCA), and 93.76% belonged to the 3,000 most frequent word families (determined by Compleat Lexical Tutor; Cobb, n.d.). These percentages were considered acceptable for comprehension, given the learners' vocabulary knowledge as measured by the VLT.

TARGET ITEMS

Fifty-six FS occurring in the documentary were selected as target items. These are listed in Appendix 1 in the Supporting Information. The first selection of target items was done quasirandomly. We made a rough selection of word combinations that we considered to be formulaic. Following previous studies (e.g., Macis & Schmitt, 2017), we then checked whether their MI score in the OpenSubtitles corpus (Tiedemann, 2012) was higher than 3 by using SketchEngine (Kilgarriff et al., 2014). All FS contained lexical words, so that knowledge of the form-meaning link could easily be tested. This meant that lexical bundles, proverbs, complex prepositions, and other types of structural or communicative FS (Granger & Paquot, 2008) were not included.

Unlike previous studies (e.g., Pellicer-Sánchez, 2017; Webb et al., 2013), which typically focused on a small selection of FS, a large number of target items were selected to capture as much incidental learning as possible. By including both low-frequency (e.g., *jazz up*, *grist to the mill*) and high-frequency items (e.g., *cell phone*, *depend on*), we could detect learning of FS that may have been partially known. The final selection of target items included 12 compounds (e.g., *ice cube*, *cell phone*), 6 phrasal verbs (e.g., *tell off*, *jazz up*), 7 idioms (e.g., *grist to the mill*, *spend a penny*), 23 lexical collocations (e.g., *common denominator*, *keep a diary*), 6 grammatical collocations (e.g., *drive off*,

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

chip in), 1 simile (*laugh like hell*), and 1 binomial (*fight or flight response*) (Granger & Paquot, 2008). Except for eight items, all FS occurred only once in the documentary.

In addition to the 56 target FS, nine distractor items were added to control for a test effect. Distractor items were FS taken from the OpenSubtitles2011 corpus based on the same criteria as the target items. They are shown in Table 2, Appendix 1 in the Supporting Information. We piloted 70 FS (60 target items and 10 distractors) with a group of second-year students at the same university as the participants to test the difficulty of the target items and to avoid a ceiling effect. Items that were considered too easy (= known by all participants in the pilot) were not included in the final study, resulting in 56 target items and 9 distractors.

Because different types of FS were examined, we analyzed the effects of a number of item-related factors that have been shown to affect learning: collocate-node relationship, congruency, transparency, corpus frequency, and MI. Table 1 in Appendix 1 (Supporting Information) lists the values for all target items.

Collocate-Node Relationship

Collocate-node relationship was determined in terms of the parts-of-speech of the single-word components. Six types of collocate-node relationship were identified: noun-noun (e.g., *pain reliever*), verb-noun (e.g., *cause offence*), adjective-noun (e.g., *collateral damage*), verb-particle (e.g., *spark off*), verb-preposition (e.g., *tap into*), and "other" (*beyond the pale*, *in the end*, *well versed in*). For target items consisting of three or more single-word components, two main components were identified. For example, the main components of the simile *laugh like hell* were *laugh* (verb) and *hell* (noun).

Congruency

To analyze the influence of the participants' L1, target FS were labeled as congruent or noncongruent by four raters, resulting in a dichotomous variable for congruency/L1 influence. Raters were native speakers of Dutch with a master's degree in English. Items for which there was disagreement received the label chosen by three of the raters. Interrater reliability, calculated using both Krippendorff's alpha and Fleiss's kappa, was rather low: .62 and .65, respectively.

Transparency

Following Gyllstad and Wolter (2016), semantic transparency of the target items was determined through Howarth's Phraseological Continuum Model (1998). We categorized the 56 FS according to their level of transparency by means of the Continuum Model: free combinations (e.g., *sexual depravity*), restricted collocations (e.g., *mince words*), figurative idioms (e.g., *white lie*), and pure idioms (e.g., *beyond the pale*).

Statistical Measures

Two statistical measures of FS were considered: corpus frequency and MI. For all target items, raw frequency values and MI scores were obtained from the OpenSubtitles2011

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

corpus using SketchEngine. The frequencies were logarithmically transformed. The least frequent item (*well versed in*) occurred 54 times in the OpenSubtitles corpus, the most frequent item (*find out*) occurred 833,102 times.

Prior Knowledge of Target Items at the Level of Meaning Recall

Learners' prior knowledge of the meaning of the target item was included as an item-related variable in the analysis of the form recall test results. Incidental vocabulary learning is an incremental process in which partial knowledge is gained through single exposures, each encounter potentially strengthening knowledge gained from previous exposures (e.g., Hulstijn, 2013). Given that form recall and meaning recall represent two levels of strength of form-meaning knowledge (see "Vocabulary Tests" section), the learning threshold for form recall should be lower if there is already knowledge of meaning recall. In other words, if participants demonstrated partial knowledge of a target item in the meaning recall test before the experiment, this could affect their learning gains in the form recall test. To account for a potential effect of partial knowledge on learning, we included as a variable whether the meaning of a target item was known in the pretest. This variable was dichotomous, reflecting the binary scoring on the meaning recall test, that is, the item was either known or unknown at the level of meaning recall before the experiment.

VOCABULARY TESTS

Two vocabulary tests were designed to measure knowledge of the form-meaning link at the meaning recall and form recall levels of strength. Form recall represents a more difficult level of form-meaning knowledge than meaning recall as it measures learners' production of the L2 form, whereas meaning recall requires receptive knowledge of form and meaning (e.g., Laufer & Goldstein, 2004). These levels of strength have been said to be implicational: knowledge at the level of form recall implies or presupposes knowledge at the meaning recall level. The test formats, both paper-and-pencil, were adapted from Puimège and Peters's (2019a) exploratory study. In the form recall test, participants were asked to provide the written form of the target items based on a Dutch equivalent or meaning description. The first letter and number of letters of each component was given because for some FS there were correct synonyms (e.g., *cell phone/mobile phone*, *pain relievers/pain killers*), and the purpose of the test was to elicit the FS appearing in the input. An example item from the test follows. Reliability of the test was measured using Cronbach's alpha. Both for the pretest ($\alpha = .85$) and posttest ($\alpha = .87$), reliability was high.

- de boodschap of het nieuws verspreiden = s_ _ _ _ _ t _ _ w _ _ _ _

In the meaning recall test, participants were asked to provide the meaning of the target items. Because the input was aural, only the spoken form of the items was given. The target items were presented twice by a native speaker of English, and 4 seconds were left between the first and second presentation of each item. The duration of the aural presentation of the items was 20 minutes and 43 seconds. To allow participants sufficient

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

time to formulate their answers, and to go back to a previous item on the test if necessary, they were allowed to write down the FS. Reliability was high for both the pretest (α = .85) and posttest (α = .87).

QUESTIONNAIRES

We gave the participants two short questionnaires: one about learners' TV viewing habits and the content of the documentary, and one about the experimental treatment. The first questionnaire had two purposes. First, we wanted to briefly test learners' minimal comprehension of the input. Participants were asked what they had learned from the documentary in terms of content and grammar/vocabulary. They were also asked to rate if the input was interesting on a five-point Likert scale, relevant to the course, easy to understand, and of appropriate length. The second questionnaire, administered after the posttests, was designed to explore how participants had experienced the experimental treatment and the pre- and posttests. The questionnaire was used to check whether learners had been aware of the purpose of the study during the treatment, whether they recognized some of the target items while watching the documentary, and if there may have been a test effect. Both questionnaires were in English and can be found in Appendix 2 in the Supporting Information.

PROCEDURE

The procedure consisted of four sessions (see Table 1). Three weeks before the learning session, participants completed the VLT (Schmitt et al., 2001). The pretests were administered in a second session, 1 week after the VLT and 2 weeks before the learning session. Both the form recall test and the meaning recall test were administered in group. The form recall test was completed first. To avoid a test effect from the form recall test to the meaning recall test, participants had to do grammar exercises for 1 hour, after the form recall test and before taking the meaning recall test. The order of the items was different in the two tests, also to control for a test effect. The total duration of the pretest session was approximately 2 hours, grammar exercises included.

Two weeks after the pretests, participants watched the documentary in group, without captions or subtitles. Immediately after watching the documentary, participants completed the first questionnaire about their TV viewing habits and the content of the documentary. Because a test effect from immediate to delayed posttest has been reported in vocabulary studies (Peters & Webb, 2018; Webb et al., 2013), we decided to administer the posttests in only one session, 5 days after the learning session. As in the pretest session, grammar exercises were provided between the form recall test and the meaning recall test to avoid

TABLE 1. Procedure

Session 1: week 1	VLT test
Session 2: week 2	Pretests
Session 3: week 4	Learning session + questionnaire 1
Session 4: week 5	Posttests + questionnaire 2

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

a test effect. After completing the posttests, the second short questionnaire was administered.

SCORING AND ANALYSES

All data were scored dichotomously by two raters. Because only the aural forms of the FS were presented during the learning session, spelling mistakes were tolerated. Responses that were judged to be phonetically correct (e.g., *ginea pig*, *ginney pig*, *pain releavers*, *free speach*, *faul language*) received a score of 1, incorrect responses received a score of 0. Interrater reliability was 1.00 for the form recall pretest, 0.99 for the form recall posttest, 0.97 for the meaning recall pretest, and 0.99 for the meaning recall posttest.

To answer the first research question, a repeated measures multivariate analysis of variance (MANOVA) was conducted with three within-subject variables: Time (pre- or posttest), Test (form or meaning test), and Item (target or distractor). Skewness and kurtosis values showed that all but one of the data subsets were normally distributed (form recall pretest [distractors]: $W = .93$, $p = .01$, skewness = .771 ($SE = .365$), kurtosis = .718 ($SE = .717$)). Because of the light skewness of the data, a parametric test was still considered reliable for analyzing our results. Although we did not assess multivariate normality directly, the fact that nearly all data were normally distributed suggest that this assumption was probably met (Grice & Iwasaki, 2007).

Repeated measures logistic regressions (Generalized Estimating Equation or GEE in SPSS, version 25) were conducted to analyze the relationship between the dependent variable, FS learned, and the predictors, learners' prior vocabulary knowledge and the five item-related factors. The advantage of GEE is that models can include variables at both the learner level and the item level. One GEE model was run per test. The following parameters were entered in each model: prior vocabulary knowledge, collocate-node relationship, congruency, transparency, corpus frequency, and MI score. Prior knowledge of target items in the meaning recall test was also included as a parameter for the analysis of the form recall test. The categorical variable collocate-node relationship was dummy coded using "other" as reference level (Field, 2009, pp. 253–260). A stepwise procedure was followed to remove nonsignificant parameters from the model. The model was rerun until only significant predictors were left.

Although nine target items occurred more than once in the input, frequency of occurrence was not entered as a variable in the analyses because we did not have enough FS per frequency level (four items occurred twice, four items three times, one four times), so that including a parameter for occurrence frequency might give unreliable results in the GEE analyses.

RESULTS

VOCABULARY LEVELS TEST

The descriptive statistics of the VLT are shown in Table 2. The high average scores on the 2K, 3K, and 5K levels suggest that the participants were familiar with these frequency bands. A level was considered mastered when a score of 26 out of 30 was reached (Schmitt et al., 2001).

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

PARTICIPANTS' VIEWING HABITS

The first questionnaire tapped into learners' TV viewing habits and comprehension of the documentary. The results show that participants were used to watching English-language television on a regular basis, although there was some variation between participants. Table 3 shows how many hours per week the participants reported to watch TV. With regard to comprehension of the input, learners' responses show that they had at least basic comprehension of the content. Specifically, the questionnaire items "what have you learned in terms of content?" and "what have you learned in terms of language?" elicited references to central topics in the documentary, for example, "I have learned that taboo words are very powerful and that slang evolves quickly," "Pain-relieving effects of swearing (experiment with the ice cubes)," and "How our brain functions when producing bad language." Participants also named expressions and words they had learned from the documentary, for example, "spend a penny," "to the end of the rainbow," "I'm sorry for offending you," versus "I'm sorry if any offense was caused."

RQ1. WHAT IS THE EFFECT OF WATCHING A ONE-HOUR TV PROGRAM ON THE LEARNING OF FS?

Form Recall

As can be seen in Table 4 and Figure 1, learning gains were made between the pre- and posttest. The mean total score for the target items increased by 9.4 points between the pre- and posttests, indicating that on average 9.4 FS were learned at the level of form recall. The minimum number of items learned was 3, and the maximum number was 16. The mean score for the distractor items also increased, but not to the same extent.

Meaning Recall

At the level of meaning recall, learning gains were made between the pre- and posttests (see Table 4 and Figure 2). On average, 6.88 FS were learned, with a minimum learning gain of zero and a maximum of 18 items. As in the form recall test, there was some learning of distractor items, but again to a lesser extent (see Figure 2).

The MANOVA with repeated measures (see Table 5) revealed significant main effects for Time ($F(1,41) = 433.54$, $p < .0001$, $\eta_p^2 = .914$), Test ($F(1,41) = 638.2$, $p < .0001$, $\eta_p^2 = .94$), and Item ($F(1,41) = 1598.61$, $p < .0001$, $\eta_p^2 = .975$). A significant interaction was found between Time and Item ($F(1,41) = 326.5$, $p < .0001$, $\eta_p^2 = .901$), with higher learning gains for target items than for distractor items. This shows that learning was primarily due to the treatment, rather than to the tests. There was also a significant

TABLE 2. Mean scores and standard deviations (in brackets) for all participants

	2K (Max = 30)	3K (Max = 30)	5K (Max = 30)	10K (Max = 30)	Total (Max=120)
All participants (N = 42)	29.43 (.88)	28.40 (2.51)	26.69 (3.61)	18.38 (6.39)	102.86 (11.39)

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

TABLE 3. Number of hours spent per week watching English-language television

	0h	2h	4h	8h	16h	32h	>32	Total
TV without subtitles	1	11	9	7	9	4	1	42
TV Dutch subtitles	11	14	9	6	2	0	0	42
TV English subtitles	14	16	6	3	3	0	0	42
TV other subtitles	37	3	2	0	0	0	0	42

interaction between Time and Test ($F(1,41) = 17.19$, $p < .0001$, $\eta_p^2 = .295$), indicating that learning gains were significantly higher in the form recall test than in the meaning recall test. Significant interactions were also found for Test and Item ($F(1,41) = 371.76$, $p < .0001$, $\eta_p^2 = .888$), and Time, Test, and Item ($F(1,41) =$, $p = .003$, $\eta_p^2 = .195$). The three-way interaction indicates that the difference in learning gains between target items and distractors was larger in the form recall test than in the meaning recall test.

RQ2. WHAT IS THE RELATIONSHIP BETWEEN LEARNER- AND ITEM-RELATED FACTORS AND VOCABULARY LEARNING GAINS?

Form Recall

To answer the second research question, a GEE with repeated measures was conducted with the data of the pre- and posttests. The results are summarized in Table 6. The analysis was done for 1,442 cases. Only cases for which there was no knowledge of the target item in the pretest were analyzed. The following parameters contributed significantly to the model: prior vocabulary knowledge, corpus frequency, knowledge of meaning in the pretest, and collocate-node relationship. Congruency, transparency, and MI score did not significantly contribute to the model.

Prior knowledge of the meaning of the FS was the most important predictor of learning the form of FS. The odds of a correct response in the form recall test were almost four times higher when the item was known in the meaning pretest. Second, the odds of a correct response increased by nearly 54% when the logarithmically transformed corpus frequency increased by one unit (Exp(B) = 1.537). For prior vocabulary knowledge,

TABLE 4. Descriptive statistics for the form recall and meaning recall tests

Item type	Form pretest		Form posttest		Meaning pretest		Meaning posttest	
	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
Distractor items (Max. = 9)	2.40 (1.59)	1.91, 2.90	3.83 (1.89)	3.25, 4.42	5.52 (1.72)	4.99, 6.05	6.36 (1.60)	5.87, 6.85
Target items (Max.= 56)	21.67 (6.14)	19.76, 23.57	31.07 (6.59)	29.02, 33.13	35.95 (6.48)	33.96, 37.95	42.83 (5.72)	41.07, 44.59
All items (Max.= 65)	24.07 (7.34)	21.78, 26.36	34.9 (8.08)	32.39, 37.42	41.45 (7.89)	38.99, 43.91	49.17 (7.03)	46.98, 51.36

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

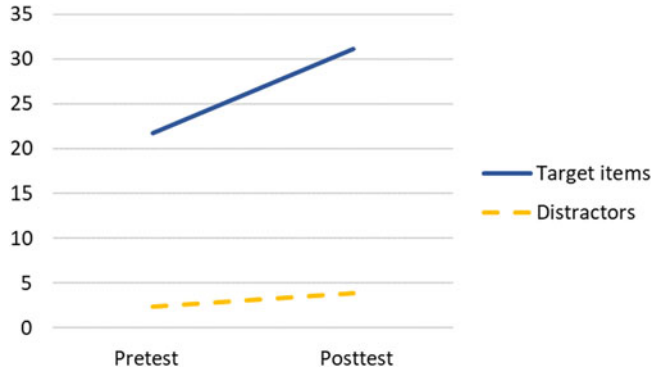


FIGURE 1. Learning gains from form recall pretest to form recall posttest.

when score on the VLT increased by 1, the odds of a correct response increased by 2%. Thus, if the score increased by 10, the odds would increase by 21% ($\exp^{10*B} = 1.21$).

For collocate-node relationship, when "other" was taken as the reference category to which other categories of collocate-node relationship were compared, a positive relationship was found between learning and collocate-node = "adjective-noun," collocate-node = "noun-noun," collocate-node = "verb-noun," and collocate-node = "verb-particle." A negative relationship was found between learning and collocate-node "verb-preposition." The odds ratios indicate that, compared to "other," adjective-noun combinations were learned the best, followed by noun-noun, verb-noun, verb-particle, and verb-prepositions combinations. Additional analyses in which the other types were taken as reference categories confirm this result. The results of the additional models can be found in Appendix 3 (Supporting Information).

Meaning Recall

A GEE analysis was conducted for 843 observations (= unknown items in the pretest). The results are shown in Table 7. There were fewer observations than in the form recall

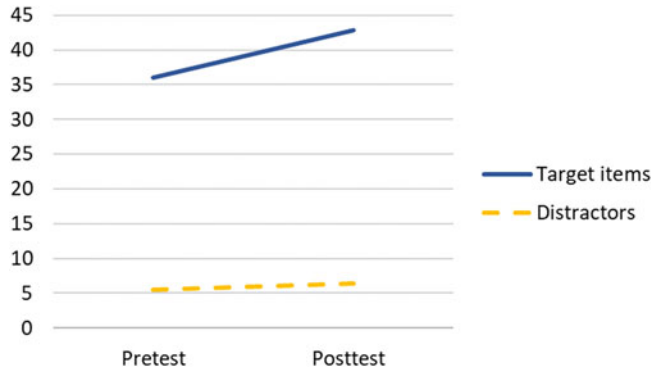


FIGURE 2. Learning gains from meaning recall pretest to meaning recall posttest.

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

TABLE 5. Results of the repeated measures MANOVA

	F(1, 41)	p	η_p^2	Observed power (α = .05)
Time (pretest/posttest)	433.54	< .001	.91	1.00
Test (meaning recall/form recall)	638.2	< .001	.94	1.00
Item (target/distractor)	1598.61	< .001	.98	1.00
Time*Test	17.19	< .001	.30	.98
Time*Item	326.5	< .001	.90	1.00
Test*Item	371.76	< .001	.89	1.00
Time*Test*Item	9.9	.003	.20	.87

test because more items were already known in the meaning recall pretest. Positive correlations with learning were found for frequency, transparency, MI score, and prior vocabulary knowledge. Congruency did not significantly affect learning.

As was the case in the form test, corpus frequency positively affected learning, but its effect was slightly larger. The odds ratio was 1.68, indicating that the odds of a correct response increased by 68% when frequency increased by one unit. Similarly, there was a positive relationship between prior vocabulary knowledge and learning at the level of meaning recall. The effect of prior vocabulary knowledge was also larger than in the form recall test. The odds of a correct response were 3% higher when one extra word was known in the VLT (or 31% when 10 more words were known).

For collocate-node relationship, when "other" was taken as reference category, there were positive coefficients for the categories "adjective-noun" and "noun-noun." For the other types, negative coefficients were found, indicating that these were not learned well. Noun-noun combinations were learned best, followed by adjective-noun, other, verb-noun, verb-preposition, and verb-particle combinations. Additional analyses in which

TABLE 6. GEE: Form recall

						95% CI	
Parameter	Wald Chi-Square	*df*	Sign	B	Exp(B)	Lower	Upper
Intercept	86.50	1	< .001	−4.90	0.01	0.00	0.02
Prior vocabulary knowledge (=VLT)	32.17	1	< .001	0.02	1.02	1.01	1.03
Corpus frequency	36.23	1	< .001	0.43	1.54	1.34	1.77
Meaning known in pretest	123.14	1	< .001	1.35	3.85	3.04	4.89
Meaning not known in pretest							
Collocate-node = Adj.-noun	3.29	1	.070	0.57	1.76	0.96	3.25
Collocate-node = Verb-noun	0.49	1	.485	0.24	1.27	0.65	2.47
Collocate-node = Noun-noun	1.88	1	.171	0.47	1.6	0.82	3.12
Collocate-node = Verb-preposition	2.49	1	.118	−0.66	0.52	0.23	1.18
Collocate-node = Verb-particle	0.70	1	.403	0.34	1.40	0.64	3.08
Collocate-node = Other							

Note: B = coefficient, Exp(B) = odds ratio, 95% CI = 95% confidence interval for Exp(B).

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

TABLE 7. GEE: Meaning recall

Parameter	Wald Chi-Square	*df*	Sign	B	Exp(B)	95% CI Lower	95% CI Upper
Intercept	26.62	1	< .001	−4.97	0.01	0.00	0.04
Prior vocabulary knowledge (=VLT)	12.26	1	< .001	0.03	1.03	1.01	1.05
MI score	11.84	1	.001	0.09	1.1	1.04	1.15
Corpus frequency	39.66	1	< .001	0.52	1.68	1.43	2
Collocate-node = Adj.-noun	2.67	1	.102	0.49	1.64	0.91	2.96
Collocate-node = Verb-noun	1.43	1	.232	−0.36	0.7	0.39	1.26
Collocate-node = Noun-noun	4.70	1	.030	0.78	2.18	1.08	4.42
Collocate-node = Verb-prep.	1.65	1	.199	−0.47	0.63	0.31	1.28
Collocate-node = Verb-particle	2.92	1	.087	−0.73	0.48	0.21	1.11
Collocate-node = other							
Transparency = Free combination	4.96	1	.026	0.53	1.7	1.07	2.70
Transparency = Restricted Collocation	0.07	1	.796	0.08	1.08	0.6	1.96
Transparency = Figurative idiom	15.37	1	< .001	−1	0.37	0.22	0.61
Transparency = pure idiom							

Note: B = coefficient, Exp(B) = odds ratio, 95% CI = 95% confidence interval for Exp(B).

other collocate-node relationships were taken as the reference category confirmed this result (see Appendix 3, Supporting Information).

Unlike in the form test, MI was a significant predictor of learning the meaning of the FS. An odds ratio of 1.04 for MI score indicates that the odds of learning an item in the meaning recall test increased by 4% when the MI score of the item increased by one unit. For transparency, when "pure idiom" was the reference category, there were positive coefficients for "free combination" and "restricted collocation," and a negative coefficient for "figurative idiom." Additional GEE models (see Appendix 3, Supporting Information) confirmed that free combinations were learned best, followed by restricted collocations, pure idioms, and figurative idioms.

DISCUSSION

The aim of the present study was to investigate whether FS could be learned from watching a full-length TV program and whether learner-related and item-related factors would contribute to the learning. The study expands on previous research by using a full-length TV episode and by measuring a large number of FS. Given the small learning gains typically found in incidental learning, this allowed us to more accurately research incidental learning gains (Peters & Webb, 2018). Moreover, by including a large variety of FS we could examine the effects of item-related factors on incidental learning.

THE EFFECT OF WATCHING A ONE-HOUR TV PROGRAM ON THE LEARNING OF FS

In answer to the first research question, the results of the pre- and posttests show that knowledge of both the form and meaning of FS can be gained from watching L2

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

television. In line with the findings of Puimège and Peters's study (2019a), significant learning gains were found for two knowledge aspects, form recall and meaning recall, showing that even knowledge at the level of form recall can be acquired from TV viewing. Approximately 10 items on average were learned at the level of form recall, and nearly 7 items were learned at the level of meaning recall.

Previous studies have shown that television is a rich source of vocabulary for L2 learners (e.g., Lin, 2014; Webb & Rodgers, 2009), and that TV viewing can result in considerable learning of single-word knowledge (e.g., Peters & Webb, 2018). The present study shows that watching L2 television also has great potential for the incidental learning of formulaic language. Considering the amount of knowledge that can be gained from a one-hour documentary, watching L2 television extensively might be an effective way for learners to enhance their knowledge of FS in the long term, especially for learners who otherwise have little exposure to aural L2 input, as argued by Webb and Chang (2015).

The learning gains found in this study are larger than those found in most previous studies on incidental learning of collocations (e.g., Szudarski, 2012). Moreover, it is likely that even more learning occurred than accounted for in the vocabulary tests. In the first questionnaire, participants were asked what they had learned from the input in terms of grammar or vocabulary. Apart from items occurring in the vocabulary tests, participants also listed words and expressions that were not tested (e.g., *to go to our Rose Cottage* or *our rainbow's end*).

One reason why such high learning gains were found compared to previous research might be that the participants in the present study were upper-intermediate learners studying English at university. As the questionnaires showed, they were motivated participants for whom the content of the video as well as the linguistic elements appearing in it were relevant. The participants may therefore have been more oriented toward learning new words and expressions from the input than learners in previous studies. Furthermore, the participants indicated that they found the content of the documentary interesting. They may thus have been more engaged with the learning materials than learners in previous studies. It should also be noted that the FS tested in this and the exploratory study varied in semantic transparency and collocate-node relationship, whereas previous studies (e.g., Pellicer-Sánchez, 2017; Webb et al., 2013) only tested lexical collocations. Moreover, in contrast to previous studies, we included a large number of target items, allowing us to measure a large proportion of formulaic language that could be potentially learned from the input. Finally, the type of input used might also partly explain differences in learning gains between this and previous studies. Television is not only a highly engaging form of L2 input (e.g., Rodgers & Webb, 2019), it also offers contextual clues to the meaning of unknown words through imagery (Peters, 2019; Rodgers, 2018). Further, there is more correspondence between text and imagery in documentaries than in other genres, as was shown by Rodgers (2018). This might also have contributed to the learning gains found in the present study.

As in Pellicer-Sánchez's study (2017), we used delayed posttests instead of immediate posttests to measure learning gains. Our results show that items learned were retained for at least 5 days after watching the documentary. This is an encouraging finding, especially considering that individual FS tend to occur infrequently in language input and are unlikely to be encountered multiple times in a short time span. As Webb et al. (2013,

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

p. 113) pointed out, encounter rates with FS tend to be infrequent, and "it may be that any knowledge gained through one encounter is forgotten before the next encounter." The learning gains found in the present study indicate that learners do retain knowledge of FS for some amount of time. Moreover, our findings suggest that learners already partially knew some of the FS before the experiment and increased their knowledge of these items (see the next section).

Rather unexpectedly, we found higher learning gains in the form recall test than in the meaning recall test. Although this finding seems counterintuitive at first given that form recall represents a higher level of knowledge, it can be explained by learners' knowledge of some of the items' meaning in the meaning recall pretest. Items that were already known at the level of meaning recall would be easier to learn at the level of form recall. Another possible explanation is that in spite of our efforts to flush learners' memory after the form test there might have been a test effect between the form recall test and the meaning recall test in the pretest session, as some participants duplicated some of the definitions appearing in the form recall test in the meaning recall test, which suggests that they tried to link definitions from the first test to items of the second test. This was also confirmed in the questionnaire data because some participants indicated that the form recall test had helped them when taking the meaning recall test. However, it is not clear if completing the form recall test resulted in learning, or if the test merely primed "the knowledge area" of the items, making these items easier in the form recall test (Nation & Webb, 2011, p. 284). If knowledge of meaning was indeed gained during the pretest session, this could explain why learning gains between the pre- and posttest were smaller for meaning recall than for form recall.

THE RELATIONSHIP BETWEEN LEARNER- AND ITEM-RELATED FACTORS AND VOCABULARY LEARNING GAINS

The results of the GEE analyses show that the learner-related variable, learners' prior vocabulary knowledge (VLT score), and a number of item-related factors had an effect on the learning of FS from TV viewing.

The most important finding is probably that this study shows the incremental nature of vocabulary learning as a result of repeated encounters with lexical items in meaningful input. In the form test, the largest effect was found for learners' knowledge of the meaning of the FS prior to the treatment. Items for which the meaning was already known in the pretest were 3.5 times more likely to be learned in the form recall test. This seems to indicate that the incidental acquisition of FS is an incremental process, in which learners gain partial knowledge of FS through repeated encounters in meaningful input. Our findings suggest that, despite the fact that individual FS are not frequently encountered in L2 input, learners can build up knowledge of FS over time, given that they have access to the meaning of the FS and notice them.

Factors that were found to affect learning both at the form and meaning recall level were corpus frequency, collocate-node relationship, and learners' VLT score. Corpus frequency of the target FS influenced learning, with more frequent FS being learned more easily. This finding is in line with Peters (2019) and Puimège and Peters (2019a), who suggested that corpus frequency could be considered a proxy of learners' previous encounters with target items. They argued that learners might have partial knowledge of

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

vocabulary items that they have encountered previously. It is likely that items with higher corpus frequency will have been encountered more frequently and will therefore be more familiar to learners. The relationship between corpus frequency and learning gains could also be due to the fact that items with the lowest frequency of occurrence in the OpenSubtitles corpus tended to have low-frequency single-word components (e.g., *sexual depravity*, *grist to the mill*). These were also the items for which the smallest learning gains were found. Words like *depravity* and *grist* are so infrequent that they were likely to be unknown by most of the participants. The presence of unknown components could have reduced the semantic transparency of FS, regardless of their level of idiomaticity or figurativeness. Furthermore, when a FS contained an unknown component, learners may have focused solely on deriving the meaning of the unknown word, instead of paying attention to the FS as a whole (Webb et al., 2013).

A relationship was also found between learning and collocate-node relationship in the form recall test and in the meaning recall test. At both levels of knowledge, noun-noun and adjective-noun combinations were learned better than other types. In line with previous findings, combinations containing verbs were the most difficult to acquire (e.g., Peters, 2016; Puimège & Peters, 2019a). Particularly verb-preposition combinations were not learned well. This may be because they contain grammatical words, which tend to be less salient than lexical words (Cintrón-Valentín & Ellis, 2016). Furthermore, combinations that contain verbs have higher morphological variation than other types of FS (Peters, 2016), which might add to their learning difficulty.

Learners' VLT score also predicted learning in both tests. Our study provides further evidence for "the rich get richer hypothesis" (Horst & Meara, 1999), in which learners who have a larger vocabulary size tend to have greater vocabulary gains from exposure to L2 audio-visual input (Montero Perez, Peters, Clarebout, & Desmet, 2014; Peters, Heynen, & Puimège, 2016; Peters & Webb, 2018). Similar to the findings of Puimège and Peters's (2019a) study, FS were more likely to be learned when participants had greater higher scores on the VLT.

Unlike in the form test, MI and transparency predicted learning in the meaning test. Items with higher MI scores were learned better than items with low MI scores. This finding contradicts earlier research showing that FS with high MI scores tend to be infrequent and therefore less likely to be known by L2 learners (e.g., González Fernández & Schmitt, 2015). A closer examination of the target items showed that in our study items with low MI scores were mostly verb-particle (e.g., *tell off*) and verb-preposition (*draw from*) combinations. This makes sense given the low-frequency bias of MI (Evert, 2008), and the fact that particles and prepositions tend to have high corpus frequency. The perceived effect of MI score might thus be due to the collocate-node relationship of items with low MI-scores, rather than the association strength or overall frequency of the items.

Transparency was also found to affect learning in the meaning recall test. Fully or highly transparent items had the highest learning gains, followed by semitransparent restricted collocations. Pure and figurative idioms were more difficult. It is possible that learners could not derive the meaning of the most idiomatic FS, except for those that were explicitly explained in the input (e.g., *pushing up daisies*, *spend a penny*).[3] This seems to lend evidence to Grant and Bauer's (2004) claim that opaque FS are the most difficult to acquire because their meaning cannot be derived from their components. Even if opaque FS are more salient, they still need to be understood to be learned. It is likely that the

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

learning burden of FS in incidental contexts is determined by a compromise between salience and transparency of meaning.

No significant effect was found for congruency, which seems surprising given that incongruent FS have been found to cause problems in L2 learners' writing (e.g., Nesselhauf, 2005). However, the previous studies focused on processing or deliberate learning of FS and did not consider other factors that might affect learning. Our findings suggest that other factors might be more important in the incidental learning of FS. Furthermore, even though congruency of the target items was determined by multiple raters, interrater reliability was low, indicating that congruency was a matter of subjective judgment, as was suggested by Peters (2016). In addition, as was the case with semantic transparency, the presence of unknown single-word components might have rendered some of the FS more opaque and more difficult to learn, even if they had a literal L1 translation.

Finally, it is worth mentioning that although there was some variation in participants' English TV viewing habits with regard to the presence of subtitles or captions, our questionnaire data (see Table 3) indicate that all participants were used to hearing English on TV, and regularly watched English-language television with or without subtitles. We therefore did not consider this variable in our GEE models.

LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

We adopted a within-participants design because such a design needs fewer participants and has greater power than between-participants designs (Charness et al., 2012). Nevertheless, the number of distractors was limited, which could be considered a limitation of the present study. Moreover, the participants in this study were all upper-intermediate learners of English who were used to watching English-language television. More research is needed to generalize the findings to other learning contexts and other TV genres.

Another limitation might be the use of authentic input. Because we did not want to compromise the ecological validity by manipulating the audio track of the documentary, we did not control for the number of occurrences of target items in the input. It might be interesting to investigate the effect of frequency of encounters on the learning of FS through TV viewing. Other factors that were not controlled for were length of the FS, morphological and syntactic flexibility, and adjacency. However, we decided to examine only those factors that have been found to relate to the learning burden of FS, and that have been considered important in previous studies on the incidental acquisition of FS.

Finally, a noteworthy methodological issue in the present study pertains to the use of multiple vocabulary tests to measure learning gains. Testing two levels of word knowledge allowed us to tap into partial knowledge and the incremental nature of vocabulary from meaning knowledge to form knowledge. However, when multiple tests are administered, a test effect cannot be completely avoided. Even with an hour in between the two tests, it seems that learners were still alerted to certain items in the meaning recall test, as they managed to link some words to the definitions provided in the form recall test. Because some learning may have occurred during test completion, the results of the present study might overestimate the learning gains made through the treatment. The test effect found in our study shows that a great deal of care should be

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

taken when testing the same target items for multiple aspects or levels of knowledge, as the risk of priming exists "even when a difficult test … comes before an easier test" (Nation & Webb, 2011, p. 284). A possible way to avoid test effects in the future is by including a larger proportion of distractors in the pre- and posttests.

CONCLUSION

The aim of the present study was to examine the effect of TV viewing on the incidental acquisition of L2 FS. The results indicate that a considerable number of FS can be learned incidentally from watching a one-hour documentary, at the levels of form recall and meaning recall. This study adds to the growing body of evidence that TV viewing can be a highly effective method for acquiring L2 vocabulary. It shows that TV viewing can aid learners' acquisition of L2 vocabulary beyond the single word, particularly for items that are already partially known, and for highly frequent FS, which are likely to have been encountered before. It was also found that semantically transparent items, which are more accessible in terms of form-meaning mapping, were learned better. Finally, in line with previous findings, learners with larger vocabularies were more likely to acquire knowledge of FS from the input.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S027226311900055X

NOTES

[1]MI is an effect size measure of association. It measures how strongly two words are attracted to each other by comparing their observed and expected frequency of co-occurrence (Evert, 2008). T-score is a measure of significance, which measures the probability of two words occurring together, or the likelihood of two words being associated, regardless of effect size (Evert, 2008).

[2]Only the four word-frequency levels (2,000, 3,000, 5,000, and 10,000) of the VLT were tested. We did not administer the test section related to academic vocabulary.

[3]The meanings of some of the target items were explained explicitly in the input. For example, it was explained how someone had interpreted *spend a penny* literally because they did not know the figurative meaning. The meaning of *pushing up daisies* was also mentioned as a euphemism for "dead."

REFERENCES

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*, 101–114.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*, 275–311. doi: 10.1075/ijcl.14.3.08bib

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research*, *18*, 54–74. doi: 10.1177/1362168813505389

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83–109. doi: 10.1017/S0267190512000050

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*, 1–8. doi: 10.1016/j.jebo.2011.08.009

Cintrón-Valentín, M. C., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in Psychology*, *7*, 1284. doi: 10.3389/fpsyg.2016.01284

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

Cobb, T. (n.d.). Compleat Lexical Tutor (Version 8.3) [Online resource]. Retrieved from http://www.lextutor.ca/.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*, 38–63.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*, 570–590. doi: 10.1093/applin/amt056

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics*, *19*, 443–477. doi: 10.1075/ijcl.19.4.01dur

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*, 163–188. doi: 10.1177/0267658309349431

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N.C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24. doi: 10.1093/applin/ami038

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text – Interdisciplinary Journal for the Study of Discourse*, *20*, 29–62. doi: 10.1515/text.1.2000.20.1.29

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö, (Eds.), Corpus linguistics. An international handbook (pp. 1212–1248). Berlin: Mouton de Gruyter.

Feng, Y., & Webb, S. (2019). Learning vocabulary through reading, listening, and viewing. Which mode of input is most effective? *Studies in Second Language Acquisition*. Advance online publication. doi: 10.1017/S0272263119000494

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: SAGE.

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29–44. doi: 10.1016/j.system.2016.04.004

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, *166*, 94–126. doi: 10.1075/itl.166.1.03fer

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), Phraseology: An interdisciplinary perspective (pp. 27–49). Amsterdam, The Netherlands, and Philadelphia, PA: Benjamins.

Grant, L., & Bauer, L. (2004). Criteria for redefining idioms: Are we barking up the wrong tree? *Applied Linguistics*, *25*, 38–61. doi: 10.1093/applin/25.1.38

Grice, J. W., & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. *Applied Multivariate Research*, *12*, 199–226. doi: 10.22329/amr.v12i3.660

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*, 296–323. doi: 10.1111/lang.12143

Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review*, *56*, 308–328. doi: 10.3138/cmlr.56.2.308

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell. doi: 10.1002/9780470756492.ch12

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), The encyclopedia of applied linguistics (pp. 2632–2640). Chichester, UK: Wiley-Blackwell. doi: 10.1002/9781405198431.wbeal0530

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*, 7–36. doi: 10.1007/s40607-014-0009-9

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436. doi: 10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*, 647–672. doi: 10.1111/j.1467-9922.2010.00621.x

Lewis, M. (1993). The lexical approach: The state of ELT and a way forward. Hove, UK: Language Teaching Publications.

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

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*, 105–129. https://doi.org/10.1080/14790718.2012.679275

Macis, M. (2018). Incidental learning of duplex collocations from reading: Three case studies. *Reading in a Foreign Language*, *30*, 48–75.

Macis, M., & Schmitt, N. (2017). Not just "small potatoes": Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, *21*, 321–340. doi: 10.1177/1362168816645957

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*, 118–141. doi: 10125/44357

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge, UK: Cambridge University Press.

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

Nesselhauf, N. (2005). Collocations in a learner corpus. Amsterdam, The Netherlands: John Benjamins.

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*, 94–106. doi: 10.2307/747835

Nguyen, T. M. H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*, 298–320. doi: 10.1177/1362168816639619

Paquot, M. (2017). L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL learner writing. *Second Language Research*, *33*, 13–32. doi: 10.1177/0267658315620265

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*, 381–402. doi: 10.1177/1362168815618428

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*, 75–94. doi: 10.1177/1362168813505384

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*, 113–138. doi: 10.1177/1362168814568131

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*, 142–168. doi: 10.1075/itl.00010.pet

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*. Advance online publication. doi: 10.1002/tesq.531

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, *63*, 134–148. doi: 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. doi: 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*, 551–557. doi: 10.1017/S0272263117000407

Puimège, E., & Peters, E. (2019a). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*, 424–438. doi: 10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2019b). Learners' English vocabulary knowledge prior to formal instruction: The effect of learner-related and word-related factors. *Language Learning*, *69*, 943–977. doi: 10.1111/lang.12364

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*, 191–211. doi: 10.1075/itl.00012.rod

Rodgers, M. P. H., & Webb, S. (2019). Incidental vocabulary learning from viewing television. *ITL – International Journal of Applied Linguistics.* Advance online publication. doi: 10.1075/itl.18034.rod

Schmitt, N. (2010). Researching vocabulary: A vocabulary research manual. London, UK: Palgrave Macmillan.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*, 55–88. doi: 10.1177/026553220101800103

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *26*, 549–569. http://doi.org/10.1093/applin/amt054

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). Understanding formulaic language: A second language acquisition perspective. London, UK: Routledge.

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press

Szudarski, P. (2012). Effects of meaning- and formed-focused instruction on the acquisition of verb–noun collocations in L2 English. *Journal of Second Language Teaching and Research*, *1*, 3–37.

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*, 245–265. doi: 10.1111/ijal.12092

Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In: Proceedings of the 8th international conference on language resources and evaluation. Istanbul: LREC.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), Formulaic sequences (pp. 155–172). Amsterdam, The Netherlands: John Benjamins.

Van Lancker Sidtis, D. (2012). Formulaic language and language disorders. *The Annual Review of Applied Linguistics*, *32*, 62–80. doi: 10.1017/S0267190512000104

Webb, S., & Chang, A. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*, 667–686. doi: 10.1177/1362168814559800

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*, 259–276. doi: 10.1093/applin/amq051

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

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

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*, 451–482. doi: 10.1017/S0272263113000107

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, *40*, 395–416. doi: 10.1017/S0272263117000237

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*, 647–668. doi: 10.2307/27896758

https://doi.org/10.1017/S027226311900055X Published online by Cambridge University Press