Katholieke Universiteit Leuven

Faculteit Letteren

An empirical investigation into the acquisition and processing of L2 formulaic sequences in meaningful input

Proefschrift ingediend tot het behalen van de graad van Doctor

in de Taalkunde

door

Eva Puimège

Promotor: Prof. Dr. Elke Peters

Copromotor: Prof. Dr. Maribel Montero Perez

2022

Promotiecommissie:

Promotor:	Prof. Dr. Elke Peters
Copromotor:	Prof. Dr. Maribel Montero Perez
Leden:	Prof. Dr. Paul Pauwels
	Prof. Dr. Magali Paquot
	Prof. Dr. Geert Brône
	Prof. Dr. Ana Pellicer-Sánchez
Voorzitter:	Prof. Dr. Karen Lahousse

Acknowledgements / Dankwoord

So much to learn, so little time.

I wish to thank my supervisors Elke and Maribel for giving me the opportunity to spend the last four years learning, discovering, and researching. It has been great. Elke, thanks to your encouragement, feedback and advice I was able to bring this project to a (somewhat) satisfactory conclusion ☺. Maribel, although we weren't in the same location very often, I am truly grateful for your support. Your expertise and insights have been invaluable to this thesis. Thank you for the inspiration.

I am also grateful to all (former) members of the vocabulary group, who helped and motivated me throughout my PhD. Thank you so much, Paul, Ann-Sophie, Bert, Isabeau, Marion, Thao, Duy, Lieven, Elien, Eva, and Jolien. To my office mates, thanks for being there when I needed a break or positive energy (or candy). I promise I will take down the Christmas tree.

Merci aan alle collega's van Sint-Andries/Sint-Jacob. Een goeie werksfeer is niet te onderschatten. Jullie hebben allemaal bijgedragen aan dit doctoraat door het zo fijn en gezellig te maken op de campus. Inez, jij bent extra bedankt, voor al het eten en het uitstekende gezelschap de voorbije vier jaar.

Sebastian, formerly known as Dr. Geòrgia Pujadas Jorba, thanks for your friendship, it really spiced up my life. New (academic) adventures await.

Dank ook aan de prachtige mensen in mijn volleybalteam. Jullie gaven mij de energie die ik nodig had om dit af te werken. Love you.

Moeke, danku dat ge mijn huisgenoot wilde zijn tijdens de lockdown. Merci ook aan de rest van de familie voor de steun. Ik apprecieer jullie geduld en begrip tijdens de wat moeilijkere momenten van mijn doctoraat. De noveenkaars mag uit nu. Niemand heeft meer chance dan ik met zo'n schoon familie.

Mama, papa, Line en Heleen, ik zie jullie graag. En natuurlijk, ons Gloria. ♡

Finally, my eternal gratitude goes to Agnetha, Anni-Frid, Björn, and Benny. Thank you for the music.

Contents

Tables

Figures

1. Introduction

In recent years, second language acquisition (SLA) research has shown a considerable interest in the effects of informal and meaning-focused learning activities, in particular among English as a foreign language (EFL) learners (e.g., Sundqvist & Sylvén, 2016). A good example is the context of young learners in Flanders, who learn a large amount of English vocabulary purely from engaging in activities such as gaming and TV viewing (e.g., Puimège & Peters, 2019). Some Flemish children are able to produce whole phrases and sentences like "Watch out! Behind you is the enemy!" before having had a single English class (Bollansée, Puimège, & Peters, 2021). Findings like these suggest that learners are able to develop knowledge of a second language (L2) simply by engaging with language input incidentally, that is, while focusing on communication, not on language form (Schmitt, 2010).

The idea that informal engagement with language can foster L2 acquisition ties in well with usage-based approaches to language acquisition, which posit that a learner's linguistic knowledge is a cognitive reflection of their experience with language (Bybee, 1998; Divjak, 2019). In usage-based theory, language representation reflects the distributional properties of the language input, as evidenced, for example, by widely reported frequency effects in language use and processing (e.g., Ellis, 2002, 2006). These effects arise naturally in first language (L1) acquisition; learners implicitly use the statistical properties of their input to discover language structure (Divjak, 2019). Adult L2 learners may also acquire knowledge of words, multiword phrases, and grammatical structures through implicit learning from L2 exposure. A major strength of meaning-focused activities is that these can provide learners with authentic and *native-like* input, that is, L2 input which reflects the distributional properties of the L2 as it is known and used by native speakers (Northbrook & Conklin, 2019).

Books, television, the internet, and other sources of authentic input contain large samples of formulaic sequences, or recurrent, conventional phrases attached to communicative meanings (Siyanova-chanturia & Pellicer-Sánchez, 2019). Formulaic sequences make up a large proportion of L2 discourse, and cover a wide range of multiword phrases, including collocations (e.g., *strong tea*), binomials (e.g., *fish and chips*), and idioms (e.g., *a leopard can't change his shorts*). Native speakers have strong intuitive knowledge of these phrases, which they can retrieve quickly and automatically from memory. Formulaic knowledge is therefore believed to contribute to processing fluency and efficient communication in native speakers (Pawley & Syder, 1983). The benefits associated with knowledge of formulaic sequences also makes them highly relevant to SLA. Consequently, in recent years there has been an explosion of research into L2 learners' knowledge, use, and processing of formulaic sequences (e.g., Schmitt, 2004; Siyanova-chanturia & Pellicer-Sánchez, 2019; Wood, 2010).

Learning the formulaic patterns of a foreign language is fraught with difficulty, not least because of the sheer number of word sequences language users are assumed to rely on in their speech and writing (e.g., Ellis, Simpson-Vlach, & Maynard, 2008). Classroom time, though valuable, is too limited to teach all of the formulaic sequences a learner needs to know in order to be able to use the L2 with the same ease and efficiency as that of a native speaker (Vilkaitė, 2016). Therefore, incidental L2 activities have been promoted for use in instructed settings, where they may supplement deliberate learning activities. Through the wide availability of reading, listening, and viewing materials, learners have the opportunity to expand and strengthen their L2 formulaic knowledge relatively effortlessly outside the classroom.

Recent evidence suggests that knowledge of formulaic sequences is predicted by learners' engagement with the L2 in informal or out-of-school activities (González Fernández & Schmitt, 2015). Further, a number of studies have demonstrated that learners may pick up formulaic language from reading (e.g., Pellicer-Sánchez, 2017; Webb, Newton, & Chang, 2013), listening (Alotaibi, Pellicer-sánchez, & Conklin, 2021), and TV viewing (Majuddin, Siyanova-Chanturia, & Boers, 2021). By using a range of measures of formulaic knowledge, these studies have found that learners do not only recognize or recall the phrases they encounter in their L2 input (e.g., Pellicer-Sánchez, 2017), but may also develop fluent access to those phrases during L2 processing (Northbrook, Conklin, & Allen, 2021; Obermeier & Elgort, 2021; Toomer & Elgort, 2019).

Despite the benefits of communicative activities, purely meaning-focused learning in an L2 is typically less successful than in L1 acquisition (Ellis, 2015). Not all L2 input is converted into intake (Corder, 1967), and learning without consciously attending to form is considered a slow, incremental, error-prone process that relies on a large amount of input (e.g., Ellis, 1994, 2005). Because meaning-focused activities are typically associated with limited depth of processing and minimal attention to form (e.g., Laufer & Hulstijn, 2001), researchers have argued for methods of inducing or enhancing learners' conscious attention to form (e.g., Lightbown, Spada, & White, 1993; Schmidt, 1990; Sharwood Smith, 1993). A few studies have shown that highlighting formulaic sequences in written texts, for example through underlining or bolding, can enhance learners' attention for, and recall of those sequences (e.g., Choi, 2017). The positive effects of typographic enhancement have mainly been found in relation to declarative knowledge, but the effects on processing fluency or implicit knowledge of formulaic sequences are less clear (Northbrook et al., 2021; Toomer & Elgort, 2019).

At present, few studies have examined the effects of incidental learning activities on L2 knowledge of formulaic sequences, and existing research has focused primarily on written input. Before the start of my research project, no previous studies had investigated L2 learning of formulaic sequences from audiovisual input. Audiovisual input is a popular activity among EFL learners, and it differs from written text in various meaningful ways. For example, the cognitive demands of processing audiovisual input may affect the learning process (e.g., Gass, Winke, Isbell, & Ahn, 2019), and different kinds of formulaic language are associated with the genres and registers of written and audiovisual input (e.g., Lin, 2014). Following research on L2 vocabulary learning from viewing (e.g., Montero Perez, Van Den Noortgate, & Desmet, 2013), it seemed timely to investigate whether and to what extent formulaic sequences can be learned from L2 audiovisual input.

Further, before the start of my PhD, the effects of typographic enhancement on L2 learning of formulaic sequences had only been examined in written input (Sonbul & Schmitt, 2013; Szudarski & Carter, 2016). Although the use of enhanced audiovisual input (e.g., keyword captions, captions with highlighted keywords) had proven effective for other aspects of L2 knowledge (e.g., Lee & Révész, 2018; Montero Perez, Peters, & Desmet, 2015), the properties specific to formulaic sequences, such as their length and low salience in written text, warrant an investigation into the potential effects of enhanced captions on L2 processing and learning of formulaic sequences.

Finally, only a handful of studies have examined the effects of incidental learning activities on processing fluency or implicit knowledge of L2 formulaic sequences (Northbrook et al., 2021; Obermeier & Elgort, 2021; Toomer & Elgort, 2019). The use of online measures of processing such as eye-tracking may provide more insight into how learners engage with, and develop fluent access to formulaic sequences, during incidental learning.

1.1 Aims of the thesis

The general aim of the research reported in the present thesis was to investigate the effects of meaning-focused activities on L2 learners' knowledge of formulaic sequences, and to examine the role of attention in the learning process. The following broad research questions guided the research presented in this thesis:

1. Can learners acquire knowledge of formulaic sequences by watching L2 television, and, if so, which variables predict learning outcomes? (Study 1 & Study 2)
2. What is the effect of typographic enhancement on learners' attention to, and learning of, formulaic sequences when they watch captioned television? (Study 3)
3. What is the effect of typographic enhancement on learners' knowledge and processing fluency of formulaic sequences in written input? (Study 4)

Four empirical studies were conducted to examine the effects of incidental learning activities on learners' knowledge of formulaic sequences, with a focus on audiovisual and written input. In each study, data were collected from Flemish university students (L1 = Dutch) who learn English as a foreign language. Studies 1-2 examined the effect of viewing (non-captioned) audiovisual input on L2 knowledge of formulaic sequences. By using authentic viewing materials, the studies aimed to closely reflect the learning conditions that might occur in a natural learning context. Study 3 examined the effect of typographically enhanced captions on learners' processing and learning of formulaic sequences in audiovisual input. Finally, Study 4 used a more controlled design to examine the effects of incidental learning and typographic enhancement on declarative knowledge of formulaic sequences on the one hand, and processing fluency on the other. This study focused on written input to increase experimental control of variables such as phrase type and the number of exposures.

The four studies of this thesis used different instruments to measure learners' processing and learning of formulaic sequences, including pre- and posttests of declarative knowledge (Studies 1-4), questionnaires (Studies 1-2), eye-tracking (Studies 3-4), and interviews (Studies 3-4). The combination of online and offline measures allowed for a close examination of the learning process and learning outcomes of incidental activities.

Each of the four studies in this thesis makes a unique theoretical, pedagogical and methodological contribution to SLA research. Firstly, the thesis informs theories of SLA, by examining (a) the contributions of item -, learner -, and input variables in incidental learning of formulaic sequences from viewing (Studies 1-2), (b) the role of attention and attention enhancement in incidental acquisition of formulaic sequences (Studies 3-4), and (c) the effects of repeated exposure and attention enhancement in written input on learners' processing fluency of formulaic sequences (Study 4). Secondly, the studies inform L2 pedagogy, by measuring the effects of two different kinds of meaning-focused input of L2 learning of formulaic sequences (Studies 1-4), and by testing the effectiveness of typographic enhancement as a method of stimulating learning from meaning-focused input (Studies 3-4). Finally, the thesis makes a methodological contribution to SLA research, (a) by accounting for pretest effects using questionnaire and interview data (Studies 1-3), (b) by using eye-tracking to measure the cognitive processes underlying incidental learning of formulaic sequences (Studies 3-4), and (c) by using a counterbalanced, within-participants design to examine the effects of typographic enhancement (Studies 3-4).

1.2 Structure of the thesis

Chapter 2 provides the theoretical background of the four studies. The background section starts by defining formulaic sequences within phraseological and frequency-based approaches, and then describes the importance of formulaic language in the context of SLA. It then moves on to discuss how formulaic language is acquired, from the perspective of usage-based approaches to language acquisition, before reviewing psycholinguistic and corpus evidence for L2 learners' knowledge and use of formulaic sequences, and the role of salience and attention in L2 knowledge of formulaic language. Next, the chapter reviews research on incidental learning of formulaic sequences from meaning-focused input, first focusing on reading, then moving on to the role of input enhancement, and finally discussing studies on incidental learning from audiovisual input. The background section concludes with a brief overview of eye-tracking research and its applications in SLA, more specifically in L2 reading, incidental vocabulary learning, and multimodal processing.

Chapters 3-6 report the findings of the four studies of this thesis. Each study should be read as an independent report presenting the detailed description of the background literature, methodology, analysis, results, and findings.

Study 1 (Chapter 3), which was published in *The Language Learning Journal*, explored incidental learning of multiword expressions and single words through watching a single video. The study used a within-participants design to measure learning gains for words and phrases encountered in an L2 video, and compared learning gains for multiword expressions and single words through multiple regression.

Study 2 (Chapter 4), published in *Studies in Second Language Acquisition*, used a similar design, but focused on the effect of learners' prior vocabulary knowledge and item variables in incidental learning of multiword expressions from L2 television. This study improved on some of the methodological limitations of Study 1, for example by increasing the number of target items and reducing the number of vocabulary measures.

Study 3 (Chapter 5), published in *Second Language Research*, investigated the effects of typographic enhancement on L2 learning of formulaic sequences in captioned L2 video. This study also measured participants' eye movements to gain insight into the role of attention in the learning process. Mixed models were used to analyze the effect of typographic enhancement and other predictors (e.g., item variables) on learning gains.

Study 5 (Chapter 6), which is still under review, used a mixed design to investigate the durability of the effects of typographic enhancement in written input. In this study, participants took part in two reading sessions containing repeated exposures to L2 collocations, but typographic enhancement was only applied in the first reading session. By tracking participants' eye movements in both reading sessions, we could measure the effect of enhancement on learners' visual processing of collocations both in the initial, enhanced context, and in later, unenhanced contexts. In addition, the study used eye-tracking to measure the effect of reading on learners' processing fluency of the target collocations in decontextualized sentences. The effects of the learning treatment on learners' processing and knowledge of collocations were analyzed in mixed models.

Chapter 7 discusses the findings of the four studies in relation to the broader literature on incidental learning of formulaic sequences. This chapter also considers the methodological and pedagogical contributions of the thesis. Finally, the chapter specifies the limitations of the four studies and proposes a few potential directions for future research.

Table 1.1 Summary of the four studies

	Study 1	Study 2	Study 3	Study 4
Aim	- Learning single words and formulaic sequences from audiovisual input - Effects of prior vocabulary knowledge and item variables	- Learning formulaic sequences from audiovisual input - Effects of prior vocabulary knowledge and item variables	- Learning formulaic sequences from audiovisual input with captions - Effect of typographic enhancement	- Learning formulaic sequences from reading - Immediate and delayed effect of typographic enhancement
Participants	20 Flemish university students (L1 = Dutch)	77 Flemish university students (L1 = Dutch)	30 Flemish university students (L1 = Dutch)	61 Flemish university students (L1 = Dutch)
Type of study	- Within-participants - Pretest-posttest	- Within-participants - Pretest-posttest	- Within-participants - Pretest-posttest - Counterbalancing of typographic enhancement	- Mixed design - Counterbalancing of typographic enhancement
Type of input	- Audiovisual input	- Audiovisual input	- Audiovisual input + captions	- Written input
Dependent variable	- Learning of single words and formulaic sequences from pre- to posttest	- Learning of formulaic sequences from pre- to posttest	- Eye-tracking measures - Learning of formulaic sequences from pre- to posttest	- Eye-tracking measures during reading - Eye-tracking measures in sentence-reading posttest - Learning of collocations in Posttest
Independent variables	- Treatment (viewing) - Item variables	- Treatment (viewing) - Item variables	- Treatment (viewing) - Typographic enhancement - Reading times	- Treatment (reading) - Typographic enhancement - Exposure count

			- Item variables	
Statistical analysis	Generalized Estimating Equations	- Repeated measures MANOVA - Generalized Estimating Equations	Mixed effects models	Mixed effects models

1.3 Key terminology

Some of the key terms used in the thesis are explained below.

Formulaic sequence

A full definition for the term "formulaic sequence" is provided in Chapter 2. In the previous literature on formulaic language, a multitude of terms has been used to refer to formulaic sequences, including multiword expressions, idiomatic expressions, phrasal expressions, prefabs, chunks, and multiword phrases (Martinez & Schmitt, 2012; Wray, 2002). In this thesis, the terminology varies slightly between different studies. I will use the terms "formulaic sequence" (or FS) and "multiword unit" (or MWU) to refer to the all types of formulaic sequence examined in this thesis in general/collectively, and "collocation" to refer to the specific subtype of formulaic sequence examined in Study 4.

Collocation

Collocations are referred to throughout the thesis as a subcategory of formulaic language. Collocations are usually defined on the basis of statistical co-occurrence in language corpora, as words that co-occur more frequently in language than would be predicted by chance (Biber, Johansson, Leech, Conrad, & Finegan, 1999), although definitions based on the phraseological properties of collocations also exist (Granger & Paquot, 2008). In Study 1, the term "lexical collocation" is used to refer to combinations of two lexical words (e.g., adjective-noun combinations), whereas in the remaining studies, the term "collocation" is used to refer to combinations of two lexical words.

Semantic transparency

The semantic properties of formulaic sequences are well-researched, and various terms have been used throughout the literature to describe the nature of the relationship between formulaic sequences and their single word constituents. Semantic transparency, also termed "idiomaticity" by some authors, broadly refers to the degree to which the meaning of a formulaic sequence can be interpreted based on its single word constituents. Rather than a fixed property, transparency is highly subjective and changeable (Carrol, Littlemore, & Gillon, 2018). Throughout the thesis, "semantic transparency" is used interchangeably with "decomposability" and "compositionality".

Audiovisual input

Broadly speaking, audiovisual input refers to input that combines auditory and visual modalities or channels of communication (e.g., Peters, Heynen, & Puimège, 2016). In the present thesis it is used as a synonym of video, although in other contexts audiovisual input may also refer to other types of input (e.g., computer games).

Multimodal input

Multimodal input is any input that combines multiple modalities or sensory channels for communication (Montero Perez, 2022). The term may refer to any combination of verbal and nonverbal input (e.g., gesture and facial expressions in spoken interaction), spoken and written input (e.g., reading while listening), or text and images (e.g., an illustrated book). Audiovisual input can be considered a type of multimodal input, although various types of audiovisual input (e.g., with or without captions) may use a different combination of modalities.

Captions

Captions (or same-language subtitles, or L2 subtitles) are on-screen text in video input that appears in the same language as the audio (Vanderplank, 2010). Unlike subtitles (or standard subtitles, or L1 subtitles), which provide a translation of the auditory input, captions provide a (word-for-word) transcription of the auditory input stream.

1.4 Bibliographical information

Study 1 (Chapter 3). Learning L2 vocabulary from audiovisual input : an exploratory study into incidental learning of single words and formulaic sequences

Published in *The Language Learning Journal*

Reference:

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input : an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Study 2 (Chapter 4). Learning formulaic sequences through viewing L2 television and factors that affect learning

Published in *Studies in Second Language Acquisition*

Reference:

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 42*(3), 525–549. https://doi.org/10.1017/S027226311900055X

Study 3 (Chapter 5). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input : An eye-tracking study

Published in *Second Language Research* (online) on 11 October, 2021

Reference:

Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input : An eye-tracking study. *Second Language Research.* October 2021. https://doi.org/10.1177/02676583211049741

Study 4. (Chapter 6) The effects of typographic enhancement on L2 collocation processing and learning from reading: an eye-tracking study

Under review

2 Background literature

2.1 Formulaic sequences: definition, properties, and importance in second language acquisition

Language allows humans to express meaning in varying, novel and creative ways. In theory, adult speakers have the productive grammatical knowledge to generate an infinite number of unique utterances (Pinker, 1999). However, language users seldom take advantage of this creative potential (Bolinger, 1967). Instead, language in use is highly repetitive (Bybee, 1998), and language structure full of convention (Langacker, 1986). Phraseology is concerned with the study of conventional and frequent sequences that pervade language, or formulaic sequences. Phraseology has a long-standing tradition in descriptive linguistics and lexicography, and it is now an interdisciplinary field that borrows insights from, and contributes to sociolinguistics, psycholinguistics, neurolinguistics, corpus linguistics, and applied linguistics. Because phraseology has roots in different research traditions there have been many different takes on what constitutes formulaic language (Granger & Meunier, 2008; Wood, 2015). The examples below, taken from Wood (2015: 3) illustrate that formulaic language is not a unitary phenomenon.

> Good morning - Look up - On the other hand - Don’t let him take you for a ride - By and large - Haste makes waste - At top speed - Speed limit - Camera speed - Computer desk - In the case of - Up to date

These sequences vary widely in terms of semantic and pragmatic functions, formal and structural characteristics, and frequency and predictability of use. What they do have in common, is their conventionality to a language community, and their association with communicative meaning. These properties are captured in the broad, inclusive definition of Siyanova-chanturia and Pellicer-Sánchez (2019, p. 5):

> [Formulaic language], as conceived in this book, may comprise of strings of letters, words, sounds, or other elements, contiguous or non-contiguous, of any length, size, frequency, degree of compositionality, literality/figurativeness, abstractness and complexity, not necessarily assumed to be stored, retrieved or processed whole, but that necessarily enjoy a degree of conventionality or familiarity among (typical) speakers of a language community or group, and that hold a strong relationship in communicative meaning.

This definition differs from most others (Cowie, 1994; Granger & Paquot, 2008), because it includes single words that serve conventional functions, such as speech formulae (e.g., *yeah* or *bye*). However, most research that will be reviewed in this chapter has operationalized formulaic sequences as multiword expressions, i.e., sequences consisting of two or more words that are separated by a space in written text. I will adopt this restricted interpretation of formulaic sequences, because a focus on word sequences facilitates the examination of concepts that do not readily apply to single words, such as collocation strength (see below).

2.1.1 Phraseological and frequency-based criteria in the identification of formulaic sequences

As early as the 18th century, Russian scholars and lexicographers were interested in phrasal peculiarities in language, which they described in Russian dictionaries and phrasebooks devoted entirely to Russian idiomatic phrases (e.g., Barsov's "Collection of 4291 ancient Russian proverbs", 1770). In the 20th century, scholars began to lay the theoretical grounds for the description of phraseological patterns. These descriptions primarily focused on idiomatic or structurally deviant behavior of conventional expressions. For example, Vinogradov (1938) studied idiomatic uses of words, such as the Russian word for *drop*, which appears in phrases like *капля в море* (lit. *a drop in the sea*), meaning 'an insignificant amount' (Zykova, 2016: 132). The phraseological approach that emerged from the study of idiomatic language has defined and classified formulaic sequences on the basis of functional, semantic and structural criteria such as compositionality, figurativeness, and fixedness (e.g., Cowie, 2001; Howarth, 1998).

A central semantic criterion is compositionality, or "the degree to which the phrasal meaning, once known, can be analyzed in terms of the contributions of the idiom parts" (Nunberg, Sag, & Wasow, 1994: 498). Prototypical phraseological units, like idioms (e.g., *kick the bucket, bear in mind*, etc.) and phrasal verbs (e.g., *carry out, crop up*), tend to be non-compositional, i.e., they have meanings that cannot be predicted from the constituent words (Biber et al., 1999). Another criterion that has been employed to identify formulaic sequences is their restricted collocability or exchangeability (Erman & Warren, 2000; Howarth, 1998). Lexical or syntactic transformations leads to loss of idiomatic or conventional meaning. For example, in the idiom *kick the bucket*, the word *bucket* can be replaced by a synonym *pail*, but this renders an exclusively literal interpretation of the phrase. Less idiomatic types of phrase, such as binomials (e.g., *bride and groom, fish and chips*), and collocations (e.g., *small print, strong tea*) also have restricted collocability (Howarth, 1998). Alternative forms (e.g., *groom and bride, little print*) sound unnatural to native speakers.

Within the phraseological tradition, variables such as compositionality and fixedness are generally considered to be scalar in nature (Granger & Paquot, 2008). Formulaic sequences can be classified along continua of structural fixedness and semantic compositionality (e.g., Cowie, 1981; Gläser, 1998; Howarth, 1998). To illustrate, Howarth's phraseological continuum model (1998) classifies formulaic phrases according to their level of idiomaticity. His model combines the features of "restricted collocability, semantic specialization, and idiomaticity" (Howarth, 1998: 28) into four categories: free combinations (e.g., *blow a trumpet*), restricted collocations (e.g., *blow a fuse*), figurative idioms (e.g., *blow your own trumpet*), and pure idioms (e.g., *blow the gaff*). It is now widely agreed that there are no clear-cut boundaries between different phraseological categories, and researchers have increasingly relied on corpus-derived measures to study the formal and semantic properties of formulaic sequences in an empirical or bottom-up manner (e.g., Obermeier & Elgort, 2021; Tiv et al., 2019; Wulff, 2009).

Recently, the increasing availability of corpus tools and corpora has enabled the automatic extraction of formulaic sequences in large samples of discourse. An increasing number of studies have adopted a distributional or frequency-based approach to identify and analyze formulaic sequences. In this empirical approach, what counts as formulaic is determined on the basis of usage patterns discerned in language corpora. The distributional approach to phraseology was pioneered by Sinclair (1991) and the Cobuild project, a corpus-based learner dictionary of English. Following Firth's (1957) notion of meaning by collocation, the Cobuild approach to lexicography sought to describe words and their meanings based on contextual, syntactic, and collocational patterns in discourse. Their analyses showed that words enter in predictable collocation relationships in discourse (e.g., *of* is often immediately preceded by *kind, part,* and *sort*) (e.g., Sinclair & Renouf, 1988). Sinclair introduced the concept of significant collocation, or the "regular collocation between items such that they co-occur more often than their respective frequencies and the length of the text in which they appear would predict" (Jones & Sinclair, 1974: 19). The corpus-based study of formulaic sequences thus focuses on systematic co-occurrence patterns in language.

The nature of formulaic patterns retrieved using a corpus-based approach largely depends on the diagnostic criteria and corpora used (e.g., Gries, 2013). For example, there exist a wide variety of statistical measures which identify or rank collocations according to the strength of association or attraction between a node (i.e., the word under investigation) and its collocates (the words that co-occur with it within a given span of discourse). Commonly used collocation measures include t-score, mutual information (MI, Church & Hanks, 1990), logDice (Rychlý, 2008), and delta P (Gries, 2013). All of these measures identify significant collocations in some way or other, but there are considerable differences in the types of collocation they identify (Evert, 2008; Gablasova, Brezina, & McEnery, 2017). Measures such as t-score, which highly rank frequent word sequences (e.g., *this is*), have been criticized for not extracting those sequences that have psychological importance to language users (Gries, 2013). On the other hand, measures such as MI, which are biased towards more exclusive, low-frequency collocations (e.g., *bated breath*), tend to extract sequences that are not well represented in discourse (e.g., Gries, 2013; Öksüz, Brezina, & Rebuschat, 2020).

One strength of the distributional approach is that it has led to the discovery and analysis of previously unattested types of formulaic sequence, such as lexical bundles (e.g., Biber et al., 1999; Biber, Conrad, & Cortes, 2004), concgrams (Cheng, Greaves, Sinclair, & Warren, 2008), and collostructions (Stefanowitsch & Gries, 2003). For example, lexical bundles do not necessarily display any of the properties (e.g., non-compositionality, fixedness) associated with formulaic language in the traditional phraseological sense. Common examples in English are *I know I, Did you see that, Yes you do, he's such, It's really*. These phrases are often semantically transparent and structurally incomplete. Analyses of lexical bundles across text genres and registers have shown that lexical bundles appear with such systematicity in language output that they are unlikely to be "accidental" (Biber et al., 1999: 290).

The frequency-based approach has shifted the phraseological focus from sequences which could clearly be distinguished from general syntax on the basis of structural fixedness or semantic non-decomposability, to the inclusion of highly frequent sequences that are conventional in a more subtle way, and which provide support for the convergence of grammar and vocabulary (Granger & Paquot, 2008). This has led to a richer, more flexible understanding of what constitutes formulaic language, and how it relates to other aspects of language. For instance, Stefanowitsch and Gries' (2003) collostructional corpus analysis examines the patterning of words in grammatical constructions, as well as in formulaic sequences, or 'variable idioms' (Stefanowitsch & Gries, 2003: 211). In line with construction grammar and lexicogrammar, their approach highlights the strong ties between different levels of linguistic representation.

A challenge for frequency-based approaches is defining transparent and theoretically relevant criteria for the identification of formulaic sequences. As Wray (1999) pointed out, estimates of formulaic density, i.e., the proportion of discourse that is considered formulaic, vary widely depending on the word counting unit (e.g., types vs. tokens), the minimum number of occurrences (compare the estimates of Biber et al., 1999 to those of Altenberg, 1998), and the collocation span, among other variables (see also Gries, 2013). Further, several researchers have pointed out that frequency alone is not a sufficient criterion of degree of formulaicity (e.g., Siyanova-Chanturia & Pellicer-Sánchez, 2019). Longer sequences tend to occur with lower frequency than shorter sequences (e.g., Biber et al., 1999; Wray, 1999), and many highly idiomatic phrases that are widely perceived as conventional (e.g., *it's raining cats and dogs*, *curiosity killed the cat, make that the cat wise*) have low corpus frequency. Depending on the research aims, a hybrid approach combining the two perspectives may be preferable. Both low-frequency idiomatic and high-frequency phrases may be "psychologically real" to native speakers, in that they are both likely to be perceived and processed differently from novel language (e.g., Carrol & Conklin, 2020). In the context of SLA, many different kinds of formulaic phrase are likely to be of pedagogical relevance to L2 learners (e.g., Martinez & Schmitt, 2012).

Estimates of formulaic density have shown that large proportions of spoken and written discourse can be considered formulaic by phraseological and frequency-based criteria. For example, Biber et al. (1999) estimated that 30% of words occurring in spoken conversations, and 21% of words in academic prose, appeared in recurring lexical bundles. However, their estimates only included phrases consisting of at least three words. Other studies have yielded even higher estimates of formulaic density (e.g., Altenberg 1998; Erman & Warren, 2000). For example, Erman and Warren (2000) used the criterion of restricted exchangeability to identify "prefabs", or sequences they considered likely to be produced as fixed, conventional expressions instead of being generated freely at the time of production. They estimated that between 50 and 60% of a corpus of spoken English could be considered formulaic in this respect. Using a minimum occurrence frequency of 10 in a 500,000 word corpus, Altenberg (1998) estimated that 80% of spoken discourse is made up of three-word lexical bundles. The pervasiveness of formulaic sequences in language corpora provide support for the hypothesis that language acquisition relies to a large extent on memory for specific word combinations (Bybee, 1998; Pawley & Syder, 1983).

2.1.2 The importance of formulaic sequences for L2 learners

It has long been recognized that learning a second language involves more than studying its words and grammar rules. Nattinger (1980) observed that much of language processing is repetitive, in that language users routinely rely on context-appropriate "ready-made units", both in production and receptive processing (p. 341). Similarly, Pawley and Syder (1983) drew attention to the formulaic patterns of language in their discussion of native-like selection and native-like fluency. *Native-like selection* concerns the observation that native speakers have preferred ways of expressing meaning that cannot be predicted through abstract grammar rules, but which are acceptable purely because they are shared by the members of a speech community. They illustrate this with the conventional phrase *I want to marry you* which can be paraphrased in numerous ways (e.g., *I wish to be wedded to you*), none of which have the same communicative effect as the formulaic expression.

Examples like this show that formulaic sequences are not just preferred and natural forms of expression, but, through their close ties to communicative meaning, and their conventionality among native speakers, contribute to unambiguous, precise, and efficient communication (Nattinger & DeCarrico, 1992; Schmitt & Carter, 2004). Formulaic sequences allow learners to fulfill a wide range of discourse functions. In spoken interaction, such functions may include repetition, clarification, questioning, and focusing attention (Bygate, 1988). Formulaic sequences also serve to express concepts (e.g., *count on* someone) and routines (e.g., *How are you?*), organize discourse (e.g., *in conclusion*), and convey specialized or technical meanings (e.g., *gene therapy*) (Martinez & Schmitt, 2012). They also fulfill register- and genre-specific functions, for example in academic discourse and conversational speech (e.g., Biber et al., 1999). Knowing the formulaic sequences of the L2, and associated meanings and connotations, is therefore of considerable importance for successful comprehension and production.

Conventional forms of expression do not only govern natural-sounding and nativelike language, but are also central to fluent language processing. Pawley and Syder's discussion of *native-like fluency* refers to the hypothesis that access to formulaic sequences in memory requires less computational effort than creating utterances word for word. Because formulaic sequences are readily accessible in memory as units or as "automatically chained strings" (p. 205), they may be encoded and produced more rapidly and effortlessly than completely novel word sequences. Indirect evidence for this comes from the observation that formulaic sequences are more prevalent in discourse that occurs under high time pressure, such as sports commentaries (e.g., Kuiper & Haggo, 1984). This suggests that language users tend to rely more strongly on formulaic language when working memory resources are limited (Ellis, 2002a). The fluent retrieval of formulaic sequences may serve to free up cognitive resources for other aspects of speech, such as the planning of larger units of discourse (Pawley & Syder, 1983).

In the context of L2 learning, formulaic sequences also play a role in facilitating speech fluency (e.g., Wood, 2006). Smooth access to formulaic sequences gives learners more time to focus on articulation and the monitoring of errors during L2 speaking tasks (e.g., Kormos, 2006; Skehan, 2009). Access to a set of memorized formulaic sequences may also help beginning learners meet their communicative demands in an optimally efficient and accurate manner (e.g., Bolander, 1989; Myles, Hooper, & Mitchell, 1998; Schmidt, 1983). Unlike native speakers, who can fully concentrate on message content, lower-level L2 learners have less automatic access to the lexical and grammatical representations, which, under the pressures of spoken interaction, could result in non-fluent, hesitant speech, or in production errors (Wray & Fitzpatrick, 2008). Memorizing formulaic sequences for use in particular social contexts may serve as a coping tactic to deal with the cognitive and linguistic demands of spoken interaction, which, in turn, could boost learners' confidence during L2 spoken interaction (Wray & Fitzpatrick, 2008).

Several studies have demonstrated that L2 proficiency and fluency are strongly associated with formulaic knowledge. Learners who make greater use of formulaic sequences in their L2 speech, tend to be perceived as more proficient and more fluent (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Stengers, Boers, Housen, & Eyckmans, 2010, 2011). In written production, L2 texts that contain more collocations, phrasal verbs, and other types of formula, receive higher ratings (Hsu, 2007; Ohlrogge, 2009). Further, Paquot (2019) found that phraseological complexity in L2 academic writing was better able to distinguish between different levels of the Common European Framework of References for Languages (CEFR; Council of Europe, 2001) than more traditional measures of grammatical and lexical complexity. Formulaic knowledge is also associated with reading comprehension (Kremmel, Brunfaut, & Alderson, 2017), and with perceived lexical proficiency in L2 speech and writing (Crossley, Salsbury, & Mcnamara, 2011). Taken together, these findings demonstrate that formulaic language is a central element of L2 proficiency.

2.2 How formulaic sequences are learned: frequency effects

Acquisition of formulaic sequences has often been explained within the context of usage-based models of linguistics (e.g., Bybee & Hopper, 2001; Goldberg, 2003; Tomasello, 2003). Usage-based theories assume that all language structure emerges from language users' experience. The individual learner acquires knowledge of grammar and vocabulary through the same general cognitive mechanisms that operate on the language input (e.g., Bybee & Hopper, 2001; Langacker, 1986). Formulaic sequences and collocation patterns likewise emerge from learners' cognitive engagement with their language input.

It is widely acknowledged that some degree of attention is required to form initial item representations in memory (e.g., Ellis, 1994; Schmidt, 1990). However, implicit processes also play an important role in language learning. In exemplar theory, learners are believed to retain a rich, detailed memory of each instance of language they are exposed to (e.g., Bybee, 1998). Every time a linguistic form is accessed in language use, it leaves a "memory trace", i.e., it updates relevant connections in memory (Divjak, 2019). The more often a unit of discourse is experienced, the more easily it is accessed later on, and the richer its phonetic, semantic, and collocational associations in memory (Bybee, 1998; Bybee & Hopper, 2001). As a result, learners more easily retrieve from memory those items that they have encountered many times. In line with general theories of learning, exemplar theory predicts that frequency of encounters is an important predictor of learning (Anderson, 1982; Ellis, 2002b; Hulstijn, 2001). Frequency effects are thought to arise automatically: language users do not consciously keep track of how many times they hear or produce specific words or contexts, but this statistical information is stored in memory such that it affects language representation.

Frequency also affects language processing: linguistic items that are encountered more often, are processed more fluently. The effects of frequency on word processing are widely attested in psychological research. Words that occur with greater frequency in language corpora, tend to be recognized and responded to faster in various kinds of processing task, e.g., in lexical decision tasks (e.g., Conklin, 2020; Harley, 2014). Corpus frequency is also one of the most robust predictors of learners' fixation times on words during reading (e.g., Inhoff & Rayner, 1986; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner & Duffy, 1986; Slattery, Pollatsek, & Rayner, 2007; White, 2008). Word frequency effects have been observed in both L1 and L2 learners (e.g., Diependaele, Lemhöfer, & Brysbaert, 2013), and vary depending on amount of language experience (see Conklin, 2020, for a discussion of L1 and L2 frequency effects).

Frequency effects have also been observed at the multiword level (e.g., Arnon & Priva, 2013; Arnon & Snider, 2010; Bannard & Matthews, 2008; Durrant & Doherty, 2010). In line with frequency effects found in word processing, language users also demonstrate faster processing of word sequences that appear more frequently in language input. Multiword frequency effects can be explained in relation to basic principles of automatization of skills, in which sequences of independent units merge together and become processed as chunks as a result of repetition or practice (e.g., Ellis, 1996, 2003). The process has also been likened to the functioning of web browsers, which store information about frequently accessed websites, so that new searches can be carried out without having to load the same large data files, resulting in faster and more efficient searches (Northbrook et al., 2021). Repeated exposure to word sequences thus leads to strengthening of associations between collocating words, so that these word sequences are retrieved from memory more fluently in receptive processing (e.g., Durrant & Doherty, 2010), and produced more easily in spoken discourse (e.g., Kuiper & Haggo, 1984; Pawley & Syder, 1983: 202). Another consequence of this process is that frequently collocating words prime each other in language processing (Hoey, 2005). The activation of a word such as *formulaic* in memory will trigger the activation of an associated word such as *sequence*, leading to faster recognition of *sequence* when it follows the word *formulaic* (e.g., Durrant & Doherty, 2010; Mckoon & Ratcliff, 1979).

The formulaic processing advantage does not only depend on the occurrence frequency of a collocation as a whole, but also on the occurrence patterns of its single word constituents. Words that are encountered in many different contexts and serve many different functions are, generally speaking, weak predictors of other words, which is why function words such as *the* have no entry in collocation dictionaries (e.g., https://www.freecollocation.com/). In the case of the collocation *formulaic sequence*, the priming effect may occur in both directions, that is, *sequence* may activate, and facilitate processing of *formulaic*. However, many collocations are asymmetrical in terms of their collocation strength. For example, the word *abject* is a much better predictor of the word *poverty*, than vice versa (Carrol & Conklin, 2019). This is illustrated in Figure 2.1, which shows the strongest collocates of both words based on the logDice measure of association strength (Rychlý, 2008), retrieved from the English Web 2015 corpus in Sketch Engine (Kilgarriff et al., 2014).

Language learners do not only store information about the co-occurrence strength of words, but also about the sequential patterns in which words occur. This is, for example, evidenced by the strong effect of transitional probability on word processing fluency, or the probability of encountering a word upon the presentation of a preceding word (McDonald & Shillcock, 2003). A clear example in formulaic processing is the sequential order effect in the processing advantage of binomials (*bride and groom*) over reversed binomials (*groom and bride*) (Carrol & Conklin, 2020; Siyanova-Chanturia, Conklin, Cafarra, Kaan, & Van Heuven, 2017; Siyanova-Chanturia, Conklin, & Van Heuven, 2011). To illustrate, in the English Web 2015 corpus, the number of occurrences of *salt and pepper* is substantially higher (17,717 occurrences, or 1.15 per million tokens) than that of *pepper and salt* (1,017 occurrences, or 0.07 per million tokens). Studies have shown that native speakers process frequent binomials faster than their reversed form in grammaticality judgement tasks (Siyanova-Chanturia et al., 2011), and eye-tracking experiments (e.g., Carrol & Conklin, 2020).

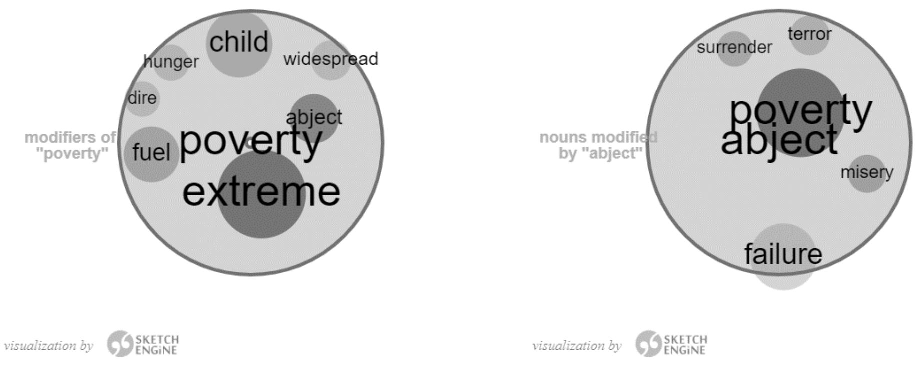

*Figure 2.1 Strongest collocates of poverty and abject in the English Web 2015 corpus, retrieved from Sketch Engine on 13 March 2022 (***http://www.sketchengine.eu***).*

Thus, knowledge of formulaic sequences may arise through an implicit, automatic learning process whereby language users store distributional patterns of the word sequences they encounter in their language input. Frequent or predictable word sequences become stored in memory either as an independent linguistic unit, with a separate, word-like representation in memory (e.g., Wray, 2002), or as a strong associative connection (e.g., Hoey, 2005). As a result, native speakers intuitively know which word sequences are common in discourse, without having to consciously memorize those sequences.

2.3 L2 knowledge of formulaic sequences: evidence from psycholinguistic experiments and learner corpora

2.3.1 L2 knowledge of formulaic sequences: psycholinguistic evidence

An increasing number of studies have examined whether L2 learners, too, develop intuitions for the formulaic patterns in the L2, and how variables such as L1 knowledge, proficiency level, and learner context, moderate learners' processing and use of formulaic sequences. Studies examining learner intuitions for L2 formulaic patterns have produced mixed findings (e.g., Carrol & Conklin, 2014; Gyllstad & Wolter, 2016; Northbrook & Conklin, 2019; Öksüz et al., 2020; Sonbul, 2015; Wolter & Gyllstad, 2011; Wolter & Yamashita, 2015, 2018; Yi, 2018). In a study comparing native and nonnative speakers' performance in an acceptability judgement task, Siyanova and Schmitt (2008) found that native speakers were faster and more accurate than nonnative speakers at judging the acceptability of adjective-noun collocations. In addition, only native speakers' reaction times differentiated between high- and medium-frequency collocations, showing that their processing fluency was more sensitive to small differences in frequency of occurrence. In a later study, Siyanova-Chanturia et al. (2011) used eye-tracking to examine learners' processing times for binomials (e.g., *church and state*) over their reversed, less conventional form (*state and church*). This study observed a formulaic processing advantage in native speakers and advanced learners, but not in intermediate-level learners, suggesting that a processing advantage reflected familiarity with the conventional word order from frequent exposure.

In a series of experiments, Ellis, Simpson-Vlach, and Maynard (2008) examined how native speakers and advanced learners of English process English academic phrases. They focused on various aspects of receptive and productive processing, and the relationship with corpus measures of frequency and collocation strength, as indexed by MI. MI measures the probability of observing the co-occurrence of two words compared to the probability of observing the independent occurrences of each word (Church & Hanks, 1990). This measure highlights exclusivity of collocation (Gablasova, 2017), and retrieves collocations that are semantically distinctive, e.g., the *citric acid cycle* (Ellis et al., 2008). Ellis et al. found that native speakers showed a stronger preference for high-MI phrases, whereas nonnative speakers were more responsive to the effect of phrase frequency. In contrast with Ellis et al.'s findings, recent studies have demonstrated that advanced L2 learners do show sensitivity to the same distributional variables as native speakers in their receptive collocation processing (e.g., Öksüz et al., 2020; Yi, 2018). Both Yi (2018) and Öksüz et al. (2020) observed that measures of association strength (MI in Yi's study, logDice in Öksüz et al.'s study) predicted native speakers' and nonnative speakers' response times in a grammaticality judgement task, which suggests that L2 learners may, in time, process formulaic sequences in a nativelike manner.

Recent evidence suggests that the distributional properties of L2 input strongly predict the processing preferences of L2 learners, in particular at beginning stages of learning. Northbrook and Conklin (2019) examined how Japanese learners of English as a foreign language processed lexical bundles extracted from their English textbooks. They found that response times in a grammaticality judgement task correlated with occurrence frequency of the lexical bundles in the learners' English textbooks. However, many of the sequences showing a processing advantage were not frequent in an L1 reference corpus (e.g., *how many CDs do you have*). The findings of this study indicate that beginning L2 learners develop a sensitivity for the word sequences they encounter in their textbooks, even if these sequences are not functionally relevant or frequent in the L2. The authors underscore the importance of providing learners with authentic, native-like input in textbook-based teaching, if the goal is to increase their knowledge of word sequences that are formulaic by native norms.

Finally, many studies on L2 processing of formulaic sequences have reported cross-language effects, or congruency effects (e.g., Carrol & Conklin, 2019; Carrol & Conklin, 2014; Carrol, Conklin, & Gyllstad, 2016; Pulido, 2021; Titone, Columbus, Whitford, Mercier, & Libben, 2015; Wolter & Gyllstad, 2011; Wolter & Yamashita, 2015; Yamashita & Jiang, 2010). The formulaic sequences a learner knows in their L1 may overlap to varying degrees with L2 phrases that have equivalent functions or meanings. Congruent formulaic sequences are sequences which have a literal, word-for-word translation equivalent in the L1 (e.g., *een deal sluiten* in Dutch, which can be translated word-for-word as *close a deal*) (Peters, 2016). Congruent sequences tend to be processed faster than phrases that share fewer similarities across languages (e.g., Titone et al., 2015; Wolter & Gyllstad, 2016). Congruency effects may interact with other factors of formulaic processing in complex ways, and appears to persist even at higher proficiency levels (see e.g., Wolter & Yamashita, 2018). This shows that not only variables associated with frequency and predictability, but also L1 experience may affect how learners process formulaic language in their L2.

2.3.2 L2 knowledge of formulaic sequences: corpus evidence

The studies reviewed above show that learners' processing of formulaic sequences responds to frequency of usage, and that this knowledge may arise implicitly and automatically from L2 exposure (e.g., Northbrook & Conklin, 2020). However, corpus evidence suggests that learners often lack the level of knowledge required for appropriate, nativelike, and fluent use of formulaic sequences in L2 speech and writing (Paquot & Granger, 2012). Studies comparing learner corpora to L1 reference corpora have found discrepancies in both the amount and the kind of formulaic language produced by native speakers and nonnative speakers.

A common source of L2 errors in use of formulaic language is L1 interference, akin to the congruency effects in formulaic processing described above (e.g., Laufer & Waldman, 2011; Nesselhauf, 2003). Laufer and Waldman (2011) examined EFL learners' (L1 = Hebrew) use of English verb-noun collocations in argumentative and descriptive essays. They found that a high proportion of the errors observed in the learner data resulted from direct, word-for-word translations of L1 collocations. L1 interference may also occur at the level of usage or pragmatic function. For instance, learners may wrongly transfer the functions of an L1 sequence to an L2 equivalent that has different contextual uses (Aijmer, 2009). Aside from L1-induced mistakes, learners may also overextend, blend or confuse collocations (e.g., Howarth, 1998; Nesselhauf, 2005), for example by extending the collocates of one word to a semantically similar word, e.g., **pay effort*, instead of *pay attention* (Howarth, 1998), or by overextending the literal meanings of formulaic sequences to figurative contexts, e.g., *weed out a crime* (Waibel, 2008).

Recent studies have investigated learners' use of formulaic sequences through measures of frequency and statistical co-occurrence (e.g., MI). Studies focusing on recurrent phrases such as lexical bundles (Chen & Baker, 2010; Groom, 2009) have found that learners tend to use more recurring sequences than native speakers, and that lower-proficiency learners show greater reliance on lexical bundles than higher-proficiency learners. With regard to co-occurrence strength, several studies have found that L2 learners underuse collocations that have high MI scores in a native speaker corpus (Durrant & Schmitt, 2009; Granger & Bestgen, 2014). Paquot (2019) suggested that this trend may reflect a low degree of sophistication in L2 production of formulaic sequences.

Researchers have also asked how L2 learners' knowledge of formulaic sequences may develop over time, by comparing texts produced by learners from different proficiency levels (e.g., Laufer & Waldman, 2011), or through longitudinal corpus research (e.g., Candarli, 2021; Garner & Crossley, 2018; Li & Schmitt, 2009; Siyanova-Chanturia, 2015; Siyanova-Chanturia & Spina, 2020; Spina, 2019). A common finding has been that higher-proficiency learners produce a higher rate of formulaic sequences (e.g., phrasal verbs, collocations), but that accuracy of use remains low (Paquot & Granger, 2012). However, findings are inconsistent and widely vary between samples and even individual learners. For example, Li and Schmitt (2010) observed greater use of adjective-noun collocations in some learners, and decreased use in others.

Findings have also varied depending on the types of phrase and discourse under investigation. For example, Spina (2019) investigated how phraseological errors in L2 writing developed over time in Chinese learners of Italian. Over a period of six months, learners produced fewer errors on noun-adjective phrases (e.g., *scuola elementare*), but more errors on adjective-noun phrases (e.g., *bel tempo*). In a larger-scale study, Siyanova-Chanturia and Spina (2019) compared the longitudinal development of Chinese learners' use of Italian noun-adjective phrases at different learner proficiency levels. They found that increased proficiency and greater L2 exposure did not necessarily yield more native-like use of formulaic sequences, in terms of frequency and association measures such as MI and delta P. On the contrary, with increased exposure, learners produced more weakly associated, low-frequency phrases. In contrast, Garner and Crossley (2018) observed a longitudinal increase in the rate of high-frequency lexical bundles in L2 speech. Siyanova-Chanturia and Spina (2019) suggested that the pressures associated with the online nature of speech might lead to greater use of high-frequency phrases in more proficient learners.

The insights afforded by analyses of L2 use and processing of formulaic sequences suggest that the acquisition of formulaic sequences is a complex and slow process. To achieve high-quality representations of formulaic sequences that can readily be retrieved during L2 processing, and used productively, learners may need many opportunities to consolidate their knowledge (Pellicer-Sánchez & Boers, 2019). However, L2 learners have considerably less experience with L2 input than native speakers, especially in classroom settings (e.g., Durrant & Schmitt, 2010; Ellis et al., 2008). This raises the question to what extent L2 learners may achieve the same levels of formulaic knowledge as native speakers purely from implicit learning through repeated exposure.

2.4 The role of attention and salience in L2 learning of formulaic sequences

Several theories of SLA have emphasized the role of attention in L2 acquisition (Ellis, 2005; Laufer & Hulstijn, 2001; Schmidt, 1990, 2001). Robinson (2003) distinguished three ways in which attention has been conceptualized within cognitive approaches to SLA. Firstly, attention can be seen as a selection mechanism that determines which information from the input is perceived and enters working memory ("attention as selection"). Secondly, attention can be studied in terms of the capacity of working memory resources, and how these are allocated to meet specific task demands ("attention as capacity"). Thirdly, the study of attention in SLA concerns the amount of effort associated with sustaining attention during a language task ("attention as effort") (Robinson, 2003: 632). SLA researchers may examine which elements of the language input become selected for processing in working memory (i.e., the transition from input to intake), and which information becomes stored in long-term memory.

Schmidt's Noticing Hypothesis (1990) predicts that noticing is a necessary and sufficient condition for the conversion of input to intake. In its strongest form, the noticing hypothesis holds that conscious attention to specific elements of the language input is necessary for those elements to become represented in memory. This implies that learners need to consciously process language features at all levels of linguistic representation: phonology, morphology, lexis, and grammar (Ellis, 2002b). However, in line with instance theory (Logan, 1988), once a linguistic stimulus has been encoded in memory, its representations can be strengthened automatically and implicitly, for example through frequent exposure as described above (Ellis, 1994, 2002b). In keeping with the Weak Interface hypothesis (Ellis, 2007), noticing may thus indirectly promote fluent, implicit, and automatic L2 knowledge.

A later, weaker version of the noticing hypothesis states that noticing, at the very least, facilitates L2 learning (Schmidt, 2001). The main advantage of conscious, effortful, and elaborate processing, is that it leaves a strong imprint on memory, and may thereby considerably speed up L2 development, in particular for those elements of language that are of low perceptual salience, such as function words and clitics (see below), or that may suffer from L1 interference (e.g., Ellis, 1994).

Salience refers to the tendency of certain elements in the input to pop out from their context (e.g., Cintrón-Valentín & Ellis, 2016; Ellis, 2006; Wulff & Ellis, 2018). The human visual and auditory systems are constantly bombarded with stimuli, and not all information that reaches our senses is also encoded and processed in working memory (Posner & Snyder, 1975). Elements of language input which are salient, i.e., which are preferably attended to over other elements of the input, have a better chance of being encoded, and of affecting representations in memory (e.g., Wulff & Ellis, 2018). For example, morphological tense markers tend to have low salience, especially when they appear in contexts that also contain a lexical cue with temporal reference (e.g., *yesterday he walked*) (Cintrón-Valentín & Ellis, 2016). Ellis (2018) and Cintrón-Valentín and Ellis (2016) make a distinction between aspects of bottom-up (stimulus- or input-specific) salience vs. top-down (learner-internal or externally induced) salience.

Psychophysical salience refers to the tendency of certain signals to naturally arouse the attention of the human processing system. In speech comprehension, variables of perceptual salience include "acoustical marking", "phonetic substance, stress level, usual serial position, and so on" (e.g., Brown, 1973: 463, cited in Ellis, 2018: 25). In the spoken input stream, linguistic units of low perceptual salience include short words, words that are phonetically reduced, words that are not stressed in speech, and words that appear in larger prosodic units not separated by pauses (including formulaic sequences, see Lin, 2012). Linguistic units such as function words (e.g., *by, for,* etc.) and clitics (e.g., /s/ in 'he's') are thought to be particularly difficult to identify during listening, which contributes to their difficulty in L2 acquisition (Ellis, 2006; Goldschneider & DeKeyser, 2001). Further non-adjacent dependencies in grammar (e.g., third-person marking *she writes*) and in collocations (e.g., *design a number of different experiments*) may be less salient due to their being broken up in the input stream (Vilkaitė, 2017).

Salience can also be modality-specific. In spoken input, formulaic sequences often coincide with prosodic units, that is, chunks of speech that are prosodically marked through e.g., pauses (Lin, 2012, 2019). Lin (2012) argued that this prosodic marking might enhance the salience of formulaic sequences as linguistic units. On the other hand, there is evidence that frequent word sequences are phonetically reduced in speech (e.g., *I don't know* → *[I] dunno*), both in elicitation tasks and in spontaneous speech (Arnon & Priva, 2013; Bannard & Matthews, 2008). The tendency of high-frequency words and word sequences to be less audible and clearly articulated, might lower their perceptibility, both as whole chunks, and in terms of their internal phonetic and lexical structure (Boers, Lindstromberg, & Eyckmans, 2014). Short sequences, such as verb-particle combinations (e.g., *give up*) may be particularly difficult to notice in spoken input, due to their limited "phonetic substance" (Brown, 1973: 463; see also Garnier & Schmitt, 2016).

Written input, on the other hand, presents the full, unreduced form of formulaic sequences, which allows for lexical and structural decoding. However, formulaic sequences are also broken up by spaces in written text, which may negatively influence their perceptual salience as chunks (e.g., Boers et al., 2014; Wray, 2002). It has therefore been argued that literacy and exposure to written input might detract from adult language learners' ability to identify and retain knowledge of formulaic sequences (Kurvers & Uri, 2006). The effects of visual salience may be particularly relevant in the context of adult L2 learning, and L2 classroom settings, which tend to make heavier use of written input compared to child L1 acquisition and L2 immersion contexts (Ellis, 2003).

Top-down salience, or salience that is associated with "memory-dependent, expectation-driven processing", is also highly relevant to L2 acquisition (Cintrón-Valentín & Ellis, 2016: 2). This form of salience is related to the informativity or psychological importance of a cue given a language user's expectations (Ellis, 2018). In associative learning theory, what is perceived as important depends to some extent on a learner's prior knowledge (Rescorla & Wagner, 1972). When learners process language input for meaning, they predict upcoming information on the basis of their current knowledge (Christiansen & Chater, 2016). Linguistic items that cause confusion or comprehension difficulties, such as novel word forms, are likely to induce noticing (Godfroid, Boers, & Housen, 2013). On the other hand, cues that are contextually redundant, i.e., do not need to be attended to for global text comprehension, are unlikely to receive attention (Schmidt, 2001; VanPatten, 1996).

Many formulaic sequences are semantically transparent and consist of familiar words. As soon as the single word components of a formulaic sequence can smoothly carry the comprehension process, there is no communicative need to pay special attention to collocational restrictions (Boers et al., 2014). This means that formulaic sequences which have low compositionality, such as idioms (e.g., *once in a blue moon*), may be more easily noticed than communicatively redundant word sequences, such as lexical bundles or collocations. The same may be true for collocations that contain a delexical verb, i.e., a verb that is low in semantic information value due to its high frequency in language. In collocations such as *take a photo*, or *give birth*, all semantic meaning is carried by the noun, which might render the verb less noticeable or memorable (Altenberg & Granger, 2001; Szudarski, 2012). This semantic asymmetry aligns to some extent with unidirectional association strength, as described in the previous section. Like semantic and collocational properties of formulaic sequences (e.g., Granger & Paquot, 2008), elements of salience and frequency may overlap.

As Boers and Lindstromberg (2009) point out, the semantic transparency of certain types of formulaic sequence, notably collocations, may not result in comprehension difficulties, but is likely to cause problems in L2 production. When learners do not consciously attend to L2-specific collocation patterns, they might unknowingly produce errors in their L2 speech and writing, for example as a result of L1 transfer (Laufer & Girsai, 2008; Laufer & Waldman, 2011) or "intra-lexical" interference, i.e., the inappropriate substitution of a single-word constituent with a semantically similar word (Boers, 2020: 150). The low salience of certain types of formulaic sequence may thus help explain why learners do not achieve knowledge of formulaic sequences on par with that of native speakers, at least at the level of L2 production.

2.5 Learning formulaic sequences from L2 input

Aside from input-related variables such as frequency and salience, learners' chances of acquiring knowledge of the formulaic sequences in the L2 are also affected by learners' cognitive processes, including attention, noticing and depth of processing (e.g., Craik & Lockhart, 1972; Schmidt, 1990). There is growing evidence that the quality of L2 processing is to a large extent determined by the goals and demands of a learning activity (e.g., Laufer & Hulstijn, 2001; Schmidt, 1990). A common distinction made in this respect is that between deliberate, or intentional learning, and incidental learning (Webb, 2020).

Deliberate learning includes activities which are explicitly aimed at improving knowledge of linguistic features, for example through decontextualized memorization or rehearsal of word lists. These activities involve a deliberate effort on the part of the learner to acquire or consolidate knowledge of formulaic sequences. Various studies have examined the effectiveness of deliberate learning activities to promote learners' uptake of formulaic sequences, with the aim of informing and improving L2 teaching methods and materials (e.g., Boers, Demecheleer, He, & Deconinck, 2017; Boers et al., 2014; Boers & Lindstromberg, 2009; Boers & Webb, 2018; Laufer & Girsai, 2008; Webb & Kagimoto, 2009; Wray & Fitzpatrick, 2008). It is beyond the scope of the current chapter to provide an extensive overview of research on deliberate learning activities, but see Boers, 2020, or Pellicer-Sánchez and Boers, 2019, for recent reviews.

Incidental learning happens as a by-product of meaning-focused or communicative activities not explicitly aimed at improving knowledge of linguistic features (Schmitt, 2010). Learners may meet new vocabulary or grammatical constructions while reading a novel (e.g., Horst, Cobb, & Meara, 1998), watching television (e.g., Peters & Webb, 2018), or attending a university lecture (Vidal, 2003, 2011). During these activities, the primary focus is on communication, but knowledge of words and other linguistic features may be picked up in the process. The learning conditions of incidental activities are such that they encourage attention to communicative meaning, although learners may still make deliberate attempts to commit linguistic features to memory (Pellicer-Sánchez & Boers, 2019). Many vocabulary studies have operationalized incidental learning through the absence of an explicit instruction to learn, or of an announced posttest (Hulstijn, 2001; Nation & Webb, 2011).

2.5.1 Incidental vocabulary learning

The bulk of research into incidental L2 vocabulary learning has focused on single word vocabulary (e.g., Brown, Waring, & Donkaewbua, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Waring & Takaki, 2003; Webb, 2007; Zahar, Cobb, & Spada, 2001). Following predictions of learning theory (e.g., Anderson, 1982), frequency is one of the most widely researched variables in studies examining incidental vocabulary learning. The number of times a word occurs in the context of L2 input (e.g., books, TV series, course materials, etc.) strongly predicts learners' memory for the word's form, meaning, and use (e.g., Uchihara, Webb, & Yanagisawa, 2019). Incidental learning from repeated exposure is considered a slow, incremental process (Webb, 2020). Significant gains in vocabulary knowledge are generally only found from around 6-10 exposures (e.g., Pellicer-Sánchez & Schmitt, 2010), although much depends on the kinds of test used to measure vocabulary growth (see e.g., Elgort & Warren, 2014).

Other variables that have been found to predict incidental vocabulary learning include learners' amount of engagement with the new words, and quality of contextual information. For example, the presence of more contextual cues has been found to enhance learning of word meaning (Webb, 2008). On the other hand, studies have suggested that learners do not remember words well when these are provided in rich contexts (Mondria & Wit-de Boer, 1991). Hulstijn and Laufer (2001) similarly found that giving learners direct access to the meanings of target words, for example through glosses presenting the translation in the margin of a text, did not necessarily promote retention of those words' meanings. These contrasting findings can be explained in relation to communicative needs and learner engagement. A new word that appears in an uninformative context may open up a conceptual gap, or a communicative need to learn the word's meaning (Zahar et al., 2001). However, an opaque context does not provide the semantic information required to address that knowledge gap. Informative contexts, on the other hand, contain semantic cues, but are less likely to arouse learners' curiosity. Zahar et al. (2001) suggested that a combination of informative and uninformative contexts may provide the best conditions for picking up new vocabulary from reading.

Incidental vocabulary learning not only depends on the information provided in the input, but also on learners' quality of processing or engagement with new vocabulary (Laufer & Hulstijn, 2001). Engagement has most commonly been operationalized by the inclusion of additional tasks, such as asking questions, note-taking, dictionary use (e.g., Vidal, 2003; Laufer & Girsai, 2008; Laufer & Rozovski-Roitblat, 2015). When learners take notes, or negotiate the meaning of a new word, they show higher vocabulary gains than when they simply read a text for meaning (e.g., Knight, 1994; Luppescu & Day, 1993). These findings are in line with Laufer and Hulstijn's (2001) Involvement Load Hypothesis, which predicts that the effectiveness of a communicative activity depends on the quality of processing induced by the learning activity. Thus, in incidental learning, new vocabulary has a better chance of being committed to memory when it is actively engaged with, for example because of its communicative relevance or importance in understanding the overall content of a reading text (D. Pulido, 2007).

Finally, learner variables such as working memory capacity and proficiency level have also been found to affect incidental word learning (Bisson, 2021). Learners of higher proficiency levels acquire new words more easily from meaning-focused activities, and tend to rely less on repeated exposure than lower-proficiency learners (Uchihara et al., 2019; Zahar et al., 2001). Several studies have observed a positive correlation between incidental vocabulary gains and vocabulary size, also termed the Matthew effect (Stanovich, 1986). The widely observed phenomenon in which "the rich get richer" has various explanations. Firstly, prior vocabulary knowledge is associated with text comprehension (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Noreillie, Kestemont, Heylen, Desmet, & Peters, 2018; Schmitt, Jiang, & Grabe, 2011). Good comprehension of the L2 input is likely to facilitate incidental learning processes such as inferring the meaning of new words from context (e.g., Webb & Chang, 2015). Further, more proficient learners may process L2 input more efficiently and automatically, which allows them to focus their attention on vocabulary learning or the development of other linguistic features (LaBerge & Samuels, 1974). Finally, more proficient learners are likely to have more extensive semantic networks, which might facilitate L2 form-meaning mapping (Bisson, 2021; Keuleers, Stevens, Mandera, & Brysbaert, 2015).

Vocabulary gains from communicative activities are typically small, incremental, and largely dependent on input- , task-, and learner-related factors. Studies comparing the effectiveness of incidental conditions to activities encouraging deliberate learning (e.g., pre-teaching, dictionary use, flash cards, matching exercises, etc.), have typically found stronger and more immediate effects in intentional conditions (e.g., Paribakht & Wesche, 1997). However, the strength of communicative learning activities is that they may foster the development of aspects of vocabulary knowledge beyond the levels and aspects of knowledge tested (Webb, 2020). Although many studies have focused on the acquisition of entirely novel words and their associated meanings, activities such as reading may also consolidate knowledge of familiar or partially known vocabulary items, for instance by strengthening semantic and collocational networks of those items, or by increasing fluency of lexical access (Horst et al., 1998). Longitudinal and cross-sectional studies have shown that EFL learners who more frequently engage with English-language media perform better on tests of English proficiency and vocabulary knowledge (Verspoor, de Bot, & van Rein, 2011). Engagement with the L2 in informal settings may even promote L2 development before formal instruction has begun (Bollansée et al., 2020; De Wilde, Brysbaert, & Eyckmans, 2020; Puimège & Peters, 2019).

2.5.2 Learning formulaic sequences incidentally from reading

Recently, there has been an increased interest in how learners develop knowledge of formulaic sequences from meaning-focused activities. Most previous research has examined incidental learning of collocations, although recent studies have also looked at other types of formulaic sequence, including binomials (Alotaibi et al., 2021), lexical bundles (Northbrook et al., 2021), and idioms (Obermeier & Elgort, 2021). Further, studies examining incidental learning of formulaic sequences have primarily focused on learning from reading (e.g., Obermeier & Elgort, 2021; Pellicer-Sánchez, 2017; Szudarski & Carter, 2016), although more recent studies have also explored the effects of other types of activity, including reading-while-listening (e.g., Webb et al., 2013), and TV viewing (Majuddin et al., 2021). A number of recent studies have compared the effects of different input modes on learning of formulaic sequences (e.g., Alotaibi et al., 2021; Vu & Peters, 2020, 2021, 2022).

Several studies have investigated how knowledge of L2 formulaic sequences develops from reading. In a between-participants study, Laufer and Girsai (2008) examined the effectiveness of reading verb-noun collocations (e.g., fulfill an ambition) in English texts compared to two conditions that combined reading with exercises focusing on the form and meaning of the target collocations (e.g., translation, fill-in-the-gaps). One of the conditions also included contrastive analysis of the collocations, aimed at raising the learners' awareness of differences between the L1 (Hebrew) and L2 collocations. One day and one week after the treatment, participants completed translation tests measuring receptive (L2 to L1) and productive (L2 to L1) knowledge of the target collocations. The highest gains were found in the condition combining exercises and contrastive analysis, showing that learners may benefit from an awareness-raising intervention to improve knowledge of non-congruent collocations. Almost no learning was found in the reading-only group, which suggests that the benefits of reading may be minimal when learners' focus is on comprehension.

In a similar design, Szudarski (2012) compared the effects of a purely meaning-focused reading condition (reading + comprehension questions), and a condition combining reading with decontextualized exercises, on learning of non-congruent verb-noun collocations containing a delexical verb (e.g., *make a mistake, make money*). Advanced EFL learners (L1 = Polish) read stories containing a total of six occurrences of the target collocations over a period of three weeks. Two weeks after reading, participants were tested on their knowledge of the target collocations in a form recall test (L1 to L2 translation), a cued recall test (providing the correct verb collocate for the noun), and a form recognition test (selecting the correct verb from four options). In the meaning-focused reading condition, learners only improved their collocation knowledge at the levels of cued recall and form recognition, showing that reading with a focus on communicative meaning did not promote productive recall of the target collocations. The combination of reading and collocation exercises led to gains at all levels of collocation knowledge, and significantly higher gains than in the reading-only condition. Szudarski suggested that more than six exposures may be needed to find strong, durable gains in collocation knowledge in purely incidental reading conditions.

Webb, Newton, and Chang (2013) investigated whether reading while listening to collocations in L2 texts could result in durable knowledge of those collocations. They also examined the effect of repetition on collocation learning. In this study, Taiwanese EFL learners read short stories from a graded reader while they listened to an audio recording of the stories. Participants were assigned to one of five groups: a control group that only took part in the testing phase of the experiment, and four experimental groups, each of which read a different version of the text containing a different number of occurrences (1, 5, 10, or 15) of 18 verb-noun collocations (e.g., *meet demand, shed light, lose touch*). To establish a baseline of collocation knowledge, learners were given a pretest measuring their ability to recognize the verb of each collocation from four options. After reading, participants completed posttests measuring various aspects of productive and receptive collocation knowledge.

Compared to studies examining incidental word learning from reading (Waring & Takaki, 2003), Webb et al. found exceptionally strong effects of their learning treatment on collocation knowledge. Various design features might explain the optimistic results found in their study. Firstly, the collocations consisted of frequent single word constituents. Prior word knowledge may have affected learners' ability to decode and interpret the contextual meaning of the novel collocations. Further, participants demonstrated knowledge of some of the target collocations in the pretest of form recognition. Because the pretest only measured one aspect of collocation knowledge (form recognition), posttest scores at other levels of knowledge could not be directly compared to this baseline. The large effect sizes reported by Webb et al. may thus, to some extent, reflect prior knowledge (Webb et al., 2013). The authors further argued that the provision of auditory support may have enhanced learning gains, compared to reading-only conditions reported in previous studies (the beneficial effects of bimodal and multimodal input are described in the next section). Finally, the use of multiple posttests may have led to a test order effect whereby information presented in the first test (e.g., the node word of each collocation) was remembered and put to use in later tests.

Aside from high learning gains overall, Webb et al. also reported a strong repetition effect. Posttest scores were considerably higher when participants read texts in which the target collocations appeared more often. Whereas a single occurrence did not result in significant gains, 15 exposures led to receptive collocation knowledge of around 76%, and form-meaning knowledge of 73% and 83% at productive and receptive levels, respectively. However, even in the 15-exposures condition, productive recall scores remained low, at only 18.55%, indicating that developing productive collocation knowledge may require a greater amount of L2 exposure.

Unlike Webb et al., Pellicer-Sánchez (2017) did not find a strong repetition effect on learners' uptake of L2 collocations from reading. In her study, learners read a short story containing 4 or 8 occurrences of collocations consisting of an adjective and a pseudoword (e.g., *wooden glabe, small nuse, magic salp*). The study thus examined the effect of reading a novel noun repeatedly with the same adjective collocate, and compared the differential effects of occurrence rate (4 vs. 8 repetitions). One week after reading, participants were tested on their knowledge of the newly learned pseudowords. The posttests included measures of form recognition (selecting the correct pseudoword from four options), meaning recall (describing the meaning of the pseudoword in an interview), meaning recognition (selecting the correct synonym from four options), collocation recall (providing the adjective that appeared with the target word in an interview), and collocation recognition (selecting the correct collocate from four options).

Significant gains in pseudoword knowledge were found. Participants scored lower on tests of recall, that is, tests measuring learners' ability to retrieve a meaning or collocate from memory without access to a set of options to choose from, than on tests of recognition (i.e., multiple-choice) (e.g., González-Fernández & Schmitt, 2020). Participants recognized around 50% and recalled around 11% of collocates, showing that reading can improve learners' collocation knowledge. The study found no significant difference between test scores representing knowledge of meaning and collocation knowledge. Further, test scores of meaning and collocation knowledge correlated significantly, with a medium-strong effect size (Plonsky & Oswald, 2014). These findings suggests that collocation knowledge and semantic knowledge may develop in parallel, at least at the initial stages of word learning. It is worth noting that Pellicer-Sánchez' study focused on learning of novel words through meaning inferencing. Deriving the meaning of a new word during reading typically involves processing its collocates and wider context, which might explain why gains in collocation knowledge were comparable to gains in semantic knowledge.

Vilkaitė (2017) was interested in the effect of adjacency on learners' odds of picking up collocational knowledge during reading. In a between-participants design, she compared the effects of reading academic texts containing 15 verb-noun collocations in adjacent form (*design experiments*), or in non-adjacent form (*design a number of different experiments*), on collocation knowledge in tests of form recall (= productive knowledge) and form recognition (= receptive knowledge). The target collocations each appeared four times in the reading texts. Learners were tested on their knowledge of the collocations one week before the treatment, immediately after reading, and after a two-week delay. The results showed negligible gains at the form recall level, and modest gains at the level of form recognition, with 20%, or two collocations learned on average. The effects of the treatment were considerably smaller than those reported by Webb et al. (2013) and by Pellicer-Sánchez (2017), the latter of whom included a comparable number of occurrences, but a different type of collocation (i.e., word-pseudoword collocations). Further, form recognition gains were very similar in the adjacent and non-adjacent conditions, indicating that learners developed an intuition for the co-occurrence of words regardless of whether these appeared in continuous or discontinuous form. Vilkaitė also reported a positive effect of vocabulary size on form recognition scores, suggesting that learners with a greater prior vocabulary knowledge learn new collocations more easily during reading.

Macis (2018) examined the effect of reading on L2 acquisition of "duplex collocations", or collocations that have both a literal and a figurative sense (e.g., *kick the bucket, fat cat, take a hike*). Three ESL learners with different L1s read an entire English novel, which contained repeated occurrences of 38 adjacent target collocations. Macis interviewed participants about their semantic knowledge of the collocations before and after reading. The participants learned one or both meanings of more than half of the target collocations. However, the study did not observe a clear correlation between occurrence rate and learning gain, some of the most frequent phrases (20-25 occurrences) were not learned by any of the participants. Despite its small scale, this study demonstrates that figurative expressions can be learned from repeated exposure during reading.

Overall, it seems that reading in the L2 can foster learners' intuitive, receptive knowledge of formulaic sequences, in particular when those sequences appear many times in the input (Webb et al., 2013), or when they contain a pseudoword constituent (Pellicer-

Sánchez, 2017). Sequences that are more challenging seem to be those with low salience, or with firmly established lexical associations already in place, such as collocations containing delexical verbs (Szudarski, 2012). On the other hand, adjacency does not appear to make a difference at initial stages of learning (Vilkaitė, 2017). The short reading interventions examined in most of the studies reviewed above do not typically yield the strong representations needed for productive recall. However, this does not mean that reading cannot, in the long term, lead to productive collocation knowledge, as longitudinal studies have demonstrated (see below, Vu & Peters, 2021, 2022).

2.5.3 Input enhancement

While purely incidental activities can result in the uptake of formulaic sequences, the success rate of such activities is moderated by item variables (e.g., compositionality), learning conditions (e.g., modality) and learner variables (e.g., proficiency level). As outlined above, a strict focus on meaning comprehension may hinder the perception and retention of low-salience formulae, such as transparent collocations, or collocations consisting of high-frequency constituents (e.g., Szudarski & Carter, 2012). Further, there is a high risk of fossilization of formulaic errors (e.g., by L1 interference) in purely informal learning conditions, as learners are not made aware of their mistakes through explicit instruction or corrective feedback that can “retune” their selective attention (e.g., Ellis, 2005: 118). In those cases where incidental learning activities might be inefficient or unsuccessful, researchers have argued for *focus on form* (e.g., Long, 1991; Sharwood Smith, 1993).

Focus on form refers to methods of increasing learners’ engagement with linguistic forms in communicative activities such as reading or viewing. Focus on form can be achieved (a) through teacher instructions or pre-learning activities that proactively alter learners’ processing strategies, (b) through post-learning activities following a communicative activity (e.g., Laufer & Girsai, 2008; Szudarski, 2012), or (c) through the manipulation of elements of the language input that draw attention to relevant linguistic features (Robinson, 2003). The third kind of focus on form has also been termed *input enhancement* (Sharwood-Smith, 1993).

One type of input enhancement that appeared in the previous section is input flooding, which involves increasing the density of repeated occurrences of formulaic sequences within an L2 text. Increasing the occurrence rate of formulaic sequences has been found effective in some (Durrant & Schmitt, 2010; Webb et al., 2013), but not all studies on incidental learning through reading (Pellicer-Sánchez, 2017). Input flooding has also received several criticisms. Firstly, seeding a text with multiple occurrences of a formulaic sequence in semantically fitting contexts can be challenging and labor-intensive. If not done properly, input flooding may damage the integrity or readability of the text, detracting from the pleasurable nature of incidental activities (Cobb, 2019). Cobb (2019) further argued that input flooding misrepresents the distribution characteristics of formulaic sequences, which may be detrimental to developing accurate, native-like knowledge of formulaic patterns in the L2.

An arguably less obtrusive form of input enhancement is typographic enhancement (TE), which involves the use of typographic techniques such as bolding, underlining, or coloring, to make L2 features more visually salient in written text. Typographic enhancement has a long tradition in L2 grammar acquisition (see e.g., Leow & Martin, 2017). Although eye-tracking research has demonstrated that TE attracts learners' attention to enhanced text elements (Indrarathne & Kormos, 2017; Simard & Foucambert, 2013; Winke, 2013), the effects on grammar learning are often minimal (e.g., Winke, 2013). This suggests that TE may not trigger the kind of elaborate processing and abstraction over exemplars that is required for grammar acquisition (Leow & Martin, 2017; Winke, 2013). Leow and Martin (2017) have therefore argued that L2 grammar learning may only benefit from the use of TE if it is combined with some form of explicit instruction or additional attention-enhancing method. Other researchers (e.g., Cintrón-valentín et al., 2019) have suggested that the beneficial effects of TE may vary depending on the level of difficulty or complexity of the targeted structure. Specifically, more difficult or abstract grammar knowledge may benefit less strongly from salience-raising than more superficial or simpler structures.

A number of studies have found positive effects of TE on L2 processing and learning of formulaic sequences from reading (Choi, 2017; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016; Toomer & Elgort, 2019; Vu & Peters, 2020, 2021, 2022). Peters (2012) examined the effect of TE on L2 learning of German formulaic phrases (e.g., *Bier zapfen, tief bewegt*) , and single words from reading. She found that TE guided learners' attention to the target collocations, resulting in significant gains at the level of form recall. Further, the effect was stronger for multiword phrases than for single words, suggesting that TE may be particularly useful for drawing attention to formulaic sequences.

Similar results were reported by Choi (2017), who combined eye-tracking and offline tests to examine the effects of TE on learners' attention to, and learning of collocations during reading. Like Peters (2012), he found a positive effect of TE on collocation knowledge at the level of form recall. The eye-tracking data suggested that TE increased learners' noticing of new collocations, as indicated by longer fixation times for collocations that were not known in the pretest. These findings show that TE can lead to significant improvements in collocation knowledge even after a single exposure. However, both studies only used an immediate posttest, and therefore did not reveal whether TE may also promote retention of L2 collocations in the longer term.

Szudarski and Carter (2016) compared the effects of input flooding and input enhancement on L2 learners' uptake of verb-noun and adjective-noun collocations from reading. In three reading sessions, over a three-week period, EFL learners read stories containing low-frequency collocations (e.g., *take a shortcut*), each occurring six or twelve times in total. Collocation knowledge was measured in productive and receptive collocation knowledge after a two-week delay. In line with Szudarski's (2012) findings, the results showed no significant gains in collocation knowledge when collocations were not enhanced in the input, indicating that repeated exposure alone was not enough to result in durable collocation knowledge. On the other hand, input enhancement (underlining) in combination with input flooding resulted in significant gains at the levels of form recall and form recognition, but not at the level of meaning recall. These results suggest that enhancing the visual salience of collocations may promote attention to form, but may not result in elaborate semantic processing.

Vu and Peters (2021, 2022) investigated the longitudinal effects of different input modes (reading-only, reading with TE, reading-while-listening, reading-while-listening with TE) on L2 knowledge of verb-noun collocations (e.g., *make an attempt*, *obey an order*). Participants read three graded readers containing new L2 collocations. In all reading conditions, significant gains were found in delayed posttest of form recall. Further, conditions that involved TE led to the highest gains in both studies. Taken together, the results of these studies suggest that TE may promote productive collocation knowledge, which is an encouraging finding, considering the limited gains typically found at this level in purely incidental conditions (e.g., Webb et al., 2013).

All of the aforementioned studies used measures of declarative, or explicit knowledge (e.g., form recall, form recognition, meaning recognition, etc.). Recently, researchers have advocated the adoption of implicit or tacit measures of L2 vocabulary knowledge, to complement more traditional tests (e.g., Elgort & Warren, 2014). Although explicit tests can give an idea of learners' intuitions for collocation patterns, as well as their ability to retrieve the correct collocation from memory, more tacit measures are useful for investigating the effect of a learning activity on processing fluency, or automaticity of use. This aspect of knowledge is perhaps particularly relevant for formulaic sequences, which, as illustrated in the previous sections, are associated with both receptive and productive processing advantages (e.g., Carrol & Conklin, 2020).

A few studies have examined whether the beneficial effects of TE generalize to implicit collocation knowledge. Sonbul and Schmitt (2013) compared the effects of different kinds of learning activity (reading without enhancement, reading with textual enhancement, reading with glosses, deliberate study) on L1 and L2 speakers' implicit and explicit knowledge of low-frequency medical collocations (e.g., *cloud baby*). Implicit knowledge was measured by means of a primed lexical decision task. Faster response times for target words primed by their collocate (e.g., *cloud* → *baby*), compared to a non-associated prime (e.g., *steam* → *baby*), were interpreted as evidence for a structural association between the two collocating words. The results showed that TE was associated with greater recall and recognition of collocations in immediate and delayed posttests. However, no evidence of collocation priming was found in any of the experimental conditions.

In a partial replication of Sonbul and Schmitt (2013), Toomer and Elgort (2019) increased the amount of input (multiple texts instead of a single text), the number of exposures to target collocations (nine instead of three), and the duration of the experiment (two days instead of one), to optimize the conditions for implicit collocation learning. Their study confirmed the positive effect of TE on collocation recall and recognition. However, evidence of collocation priming was only found in the unenhanced condition. The authors suggested that enhancing collocations in each of the nine exposures may have obstructed the creation of an "accurate mental model" of the contexts in which the collocations appeared, thereby obstructing the expansion of lexical and semantic representations of words in the input, including the implicit associations between those words (Toomer & Elgort, 2019: 426). Possibly, the combination of repeated exposure and TE disrupted the reading process, and hindered the development of implicit collocation knowledge. Because in learning materials such as university textbooks and graded readers, key terms and important vocabulary are often only enhanced once (e.g., Field, 2013), it could be interesting to examine how this form of TE affects L2 learners' processing and retention of collocations.

Northbrook, Allen and Conklin (2021) examined the effects of repetition and TE on learners' response times to lexical bundles (e.g., *set off home, tired and hungry*) in a phrase acceptability judgement task. Beginning EFL learners (L1 = Japanese) read English stories that were interspersed with iterations of a phrasal judgement task in which participants were instructed to decide as quickly as possible whether phrases were acceptable in English or not. By presenting this task in between stories, processing fluency could be measured after each new contextual exposure.

Northbrook et al. found that response times for lexical bundles encountered in the stories were significantly shorter after the first exposure, and kept decreasing with every new exposure. TE was associated with faster response times, but only in the first three exposures, suggesting that the effects of TE on recognition of lexical bundles were quickly overridden by the effect of repeated exposure. In a delayed phrasal judgement task one week after the treatment, the repetition effect was somewhat diminished but still observable. However, the difference between response times for word sequences repeated in the stories, and sequences encountered only in the phrasal judgement task disappeared in the delayed test. This could mean that long-term retention benefited mainly from the judgement task itself, not from the additional exposures in the stories. Because the phrasal judgement task presented the target phrases in isolation, without any distracting contextual information, this task may have had a stronger effect on learning than contextualized exposure during reading (see also Durrant & Schmitt, 2010).

Nevertheless, the results of Northbrook et al.'s experiment showed that repetition had a significant effect on processing fluency of word sequences after only a few exposures. Unlike Sonbul and Schmitt (2013) and Toomer and Elgort (2019), this shows that implicit knowledge of formulaic sequences can arise quickly. An important difference with the two previous studies is the proficiency level of the participant sample. Given the asymptotic nature of frequency effects on processing (Conklin, 2020), the effect of repeated exposure on processing fluency might be more evident in beginning learners. It is possible that the phrasal judgement task used in Northbrook et al.'s experiment was more sensitive to small increments in processing fluency. Alternatively, it is also possible that a phrasal judgement task is less suitable for measuring implicit knowledge, because it involves a conscious decision on the formulaic status or conventionality of a phrase (Öksüz et al., 2020).

With regard to implicit knowledge, the first few exposures in written input can already result in measurable increases in processing fluency in beginning learners (Northbrook & Conklin, 2019), although implicit associations needed to find a collocation priming effect appear to take more exposure and/or time to develop in intermediate L2 learners (Sonbul & Schmitt, 2013; Toomer & Elgort, 2019).

2.6 Learning formulaic sequences from audiovisual input

The studies above all focused on incidental learning from reading. Other activities have also been found to promote incidental vocabulary acquisition, including listening (e.g., Vidal, 2003; Van Zeeland & Schmitt, 2013), and TV viewing (e.g., Montero Perez, Van Den Noortgate, & Desmet, 2013; Peters, 2019; Peters & Webb, 2018; Rodgers, 2018). The media-specific properties of different activities may affect how learners perceive and process the L2 input, which L2 features are likely to be learned, and at what rate those features are learned.

One promising medium for incidental L2 acquisition is audiovisual input. Watching television, film, and online videos are very popular L2 activities, especially among EFL learners (De Wilde et al., 2020; Kuppens, 2010; Lindgren & Muñoz, 2013; Peters, 2018; Peters, Noreillie, Heylen, Bulté, & Desmet, 2019; Sundqvist & Sylvén, 2016). With technological innovation, audiovisual input is becoming increasingly common and accessible to L2 learners, for example via the internet. Television is generally considered a comprehensible source of L2 input for learners of various proficiency levels, because it contains a large proportion of high-frequency words (Rodgers & Webb, 2011; Webb & Rodgers, 2009). As a result, learners do not need as extensive a vocabulary size in order to comprehend the content of L2 videos, compared to reading (Durbahn, Rodgers, & Peters, 2020).

Extensive out-of-school engagement with L2 vocabulary in audiovisual input may compensate for the limited classroom time that can be devoted to vocabulary teaching (Webb & Nation, 2017). Television offers opportunity for frequent exposure to L2 words. Firstly, TV genres and series contain repeated uses of topic- and genre-specific vocabulary (Webb, 2015). Further, learners may choose to repeatedly view videos that are of interest to them, offering verbatim repetition that could promote the uptake of new vocabulary (Majuddin et al., 2020; Winke, Gass, & Sydorenko, 2010). Television is also a rich source of formulaic language (Bednarek, 2012; Lin, 2014), which occurs with greater density in spoken discourse than in written texts (e.g., Biber et al., 1999). Viewing online videos, for example, can offer access to casual, everyday expressions that are less likely to be covered in more formal learning contexts (Lin & Siyanova-Chanturia, 2014).

A number of intervention studies similar to those reviewed in the previous section have shown that learners can pick up new vocabulary through viewing without subtitles (e.g., Peters & Webb, 2018). Further, vocabulary can be learned through extensive viewing of full

TV series (Muñoz, Pujadas, & Pattemore, 2021; Pujadas & Muñoz, 2019). Factors that predict vocabulary gains from audiovisual input include learners' L2 proficiency (Suárez & Gesa, 2019), prior L2 vocabulary knowledge (e.g., Peters & Webb, 2018), working memory capacity (e.g., Suárez, Gilabert, & Moskvina, 2021), the presence of imagery (Peters, 2019), TV genre (Suárez et al., 2021), and the presence and types of on-screen text (e.g., Montero Perez, 2022; Peters, Heynen, & Puimège, 2016).

2.6.1 Theories of multimedia learning

The benefits of audiovisual input for L2 acquisition can be understood in relation to cognitive theories of multimodal processing. The multimedia principle holds that, generally speaking, "people learn more deeply from words and pictures than from words alone" (Mayer, 2014: 43). This principle underlies the Cognitive Theory of Multimedia Learning, which makes three assumptions. The first assumption is that processing of visual and auditory input (Baddeley, 1992), or verbal and nonverbal input (Paivio, 1986) preferentially happens in two different information processing channels in working memory. However, information processed in one channel can also become represented in the other channel (see Mayer, 2014: 48 for examples). According to Paivio's (1986) Dual-Coding Theory, when text and images are simultaneously presented, the information from these two input sources can be integrated as cross-channel representations in working memory. These cross-modal associations lead to greater long-term retention.

Multimedia theory also assumes that working memory capacity is limited. In line with theories of attention (e.g., Robinson, 1995), and skill automatization theory (e.g., Miller, 1956), humans can only hold a limited amount of information in working memory. Therefore, when the working memory demands of a task are too high, different elements of the input may compete for attention. This could lead to a cognitive overload, or to split attention effects (e.g., Ayres & Sweller, 2014), whereby learners have to ignore certain features of the input, e.g., low-salience or task-irrelevant information. For example, in communicative tasks, learners may prioritize attention to meaning over form, which could negatively affect processing and retention of linguistic features (Cintrón-Valentín & Ellis, 2016; Skehan, 1998; VanPatten, 1996).

Thirdly, multimedia theory assumes that learners actively process multimodal input. To make sense of information coming simultaneously from different input modalities, learners have to use their attentional control to build a mental model or knowledge structure in line with their expectations and task-related aims. In language learning, active processing involves selecting relevant stimuli from the input (e.g., phonological and visual cues), organizing this information (e.g., linking a new lexical form to an image), and integrating it into existing knowledge structures (e.g., existing semantic and world knowledge) (Mayer, 2014: 51).

Based on these three assumptions, multimedia learning theory proposes that learning materials should ideally induce active processing of information from different input modes without overloading working memory (Mayer, 2014). This may be achieved by presenting auditory information at an appropriate speed to allow for the organization and integration of relevant input features, or by avoiding the presentation of extraneous information that could distract from learning. The cognitive theory of multimedia learning also predicts that individual differences in prior knowledge and working memory capacity may determine the effectiveness of multimodal learning (Mayer, 2014).

2.6.2 The role of captions in incidental L2 learning

Although audiovisual input by itself can provide a good source of L2 input, it does not guarantee successful comprehension and acquisition (e.g., Baltova, 1999). In line with multimedia learning theory, videos can be efficiently enhanced while taking into account learners' working memory limitations and background knowledge. One way of enhancing audiovisual input is through on-screen text, more specifically L1 subtitles or captions. Captions, or same-language subtitles, have been defined as "redundant text that matches spoken audio signals and appears in the same language as the target audio" (Vandergrift, 2007, p. 79). Captions originally served as visual support for the deaf and hearing-impaired, but are increasingly used in L2 educational settings to support language teaching and learning (Montero Perez et al., 2013; Price, 1983; Vanderplank, 2010). In recent years, captions have also commonly appeared in online videos, for example on social media and in video streaming services.

Captions may help learners cope with the working memory demands of L2 speech comprehension, which is considerably more difficult than L1 speech comprehension (Gass, Winke, Isbell, & Ahn, 2019). The transient nature of spoken input does not allow for much reanalysis or additional processing, and, most of the time, listeners have little control over the presentation speed of the aural input stream. This poses a challenge for L2 learners, who generally have less automatized processing skills (de Groot & van Hell, 2005). Learners may therefore struggle with decoding and interpreting spoken L2 input (Goh, 2000). The simultaneous presentation of spoken and written text may support decoding, for instance by showing word boundaries, thereby facilitating speech segmentation (Mitterer & McQueen, 2009; Wisniewska & Mora, 2018). Further, experimental research has shown that processing the written and spoken form of words simultaneously may increase decoding speed as well as phonological decoding accuracy, thereby contributing to processing efficiency and elaborate semantic processing (Bird & Williams, 2002). This may also enhance learners' memory for novel words (Bird & Williams, 2002).

By increasing processing efficiency, captions may also help learners connect and integrate information from the aural input stream and the imagery on screen (Gass et al., 2019). Indeed, most studies comparing the differential effects of viewing in captions and no-captions conditions have found that the inclusion of captions leads to improved L2 comprehension and greater L2 vocabulary gains (e.g., Montero Perez et al., 2013). However, in line with predictions of the cognitive theory of multimedia learning, the benefits of captions strongly depend on learner variables such as L1 background (Winke et al., 2010; Winke, Sydorenko, & Gass, 2013), proficiency level (Leveridge & Yang, 2013; Pujola, 2002) and working memory capacity (Gass et al., 2019). For example, Gass et al. (2019) found that learners with greater working memory capacity relied less on captions for comprehension. Winke et al. (2010) observed differences in how learners from different L1 backgrounds used captions to learn L2 vocabulary, suggesting that orthographic similarity between the L1 and the L2 may affect incidental vocabulary learning from captioned video.

Input difficulty may also moderate the beneficial effects of captions. For example, when the presentation speed of captioned video is too high, L2 learners make less use of captions for comprehension (Mayer, Lee, & Peebles, 2014). This suggests that under higher cognitive pressure, learners may ignore one of the input modalities and focus their attention on a single modality. Similarly, Winke et al. (2013) observed that processing difficulties associated with unfamiliar topics or unknown vocabulary led to decreased use of captions in favor of the imagery in a video, or vice versa. Findings such as these show that learners may not always successfully integrate information from different input modalities, and suggest that learning success from captioned audiovisual input may depend on properties of the input as well as learner variables.

Ideally, to optimize opportunities for L2 learning, captioned audiovisual input should be carefully selected so that it is closely matched with learners' processing needs. This means taking into account variables that may be difficult to control, such as complexity and pace of the input, topic familiarity, and L1-L2 script differences. Alternatively, teachers may rely on captions that are designed specifically for the purpose of promoting L2 acquisition. A few studies have explored the use of enhanced captions that are aimed at stimulating learners' listening skills (Mirzaei, Meshgi, Akita, & Kawahara, 2017), and their learning of L2 words (Cintrón-Valentín et al., 2019; Cintrón-Valentín & García-Amaya, 2021; Montero Perez, Peters, Clarebout, & Desmet, 2014; Montero Perez et al., 2015; Montero Perez, Peters, & Desmet, 2018), formulaic sequences (Majuddin et al., 2021), and abstract grammatical constructions (Cintrón-Valentín et al., 2019; Lee & Révész, 2018, 2020).

2.6.3 Enhanced captions

Montero Perez et al. (2014) compared the effects of different caption conditions on incidental learning of L2 words during viewing. Dutch-speaking learners of French watched three French videos twice, in one of four viewing conditions: no captions, full or regular captions, keyword captions (captions that only contain the target vocabulary), and full captions with highlighted keywords. Target vocabulary included both words and multiword expressions (e.g., *frôlé le naufrage*). Using a pretest-treatment-posttest design, learning was measures for four aspects of knowledge (form recognition, meaning recognition, meaning recall, and clip association, or linking a word to the video in which it appeared). In this study, all captioning groups outperformed the no-captions group in form recognition and clip association, but only the keyword captions and captions with highlighted keywords led to greater meaning recognition compared to the no-captions group. Meaning recall, on the other hand, was not affected by caption condition. These results suggest that enhancing the salience of new vocabulary in captions may promote elaborate processing of these words so as to increase learning of word meaning, at least at the level of meaning recognition. Meaning recall may have been too challenging a level of word knowledge, considering the relatively short learning intervention, and the minimal exposure to the target items and their contextual meanings.

Using similar vocabulary measures, Montero Perez, Peters, and Desmet (2015) compared the effects of keyword captions and full captions in incidental (no vocabulary test announcement) and intentional (vocabulary test announcement) learning conditions. In contrast with Montero Perez et al.'s (2014) findings, caption condition affected learning at the level of form recognition, with significantly higher gains found in the keyword captions conditions. No significant differences were found in any of the other vocabulary tests, which might be because learners were only allowed to watch the videos a single time. In Montero Perez et al.'s (2014) experiment, viewing the video twice may have given learners in all caption conditions the same opportunity for learning new word forms, whereas in Montero Perez et al.'s study (2015), the salience-raising effect of keyword captions benefited learning of form from a single viewing. No differences between caption conditions were found at the level of meaning recall, in line with Montero Perez et al.'s (2014) results.

Cintrón-Valentín, García-Amaya and Ellis (2019), and Cintrón-Valentín and García-Amaya (2021) compared the effects of three viewing conditions: no captions, full captions with highlighted vocabulary, and full captions with highlighted grammar constructions on learners' L2 vocabulary and grammar knowledge. In both studies, intermediate-level learners of Spanish watched short video clips developed by the researchers to contain repeated uses of Spanish words, and three Spanish constructions (*gustar*-type verbs, the preterite and imperfect, and the uses of *ser* and *estar*). In immediate posttests, scores on receptive and productive word knowledge were found to be significantly higher when word forms were highlighted in the captions. Unlike Montero Perez et al.'s (2014, 2015) findings, significant gains were also found at a relatively high level of word knowledge (form recall). One explanation for this difference may be the higher number of occurrences (four) in Cintrón-Valentín et al.'s (2019) experiment. However, Cintrón-Valentín and García-Amaya (2021) found that most of the learned vocabulary items were not retained in a delayed posttest.

Interestingly, the two studies also found a significant effect of caption condition on the learning of *gustar*-type verbs, but not of the other grammatical structures. The authors concluded that item-specific salience and complexity may have moderated the effects of the enhanced captions. More difficult and less salient L2 features may benefit less from enhanced and repeated exposure during viewing. Cintrón-Valentín and García-Amaya (2021) further suggested that structures which are challenging for learners due to their contextual use (e.g., the conditional), the limited support provided by typographic enhancement may not fully direct learners' attention to information relevant to the target feature. Finally, we could also interpret this finding in relation to multimedia learning theory, which predicts that multimodal processing is less beneficial when the input is highly complex and cognitively demanding (e.g., Gass et al., 2019; Mayer, 2014).

Lee and Révész (2019) studied the effect of enhanced captions, compared to unenhanced captions, on Korean EFL learners' processing and learning of pronominal anaphoric reference. This is a relatively complex grammatical feature to learn incidentally, as it requires the integration of two separate sources of information: the referential antecedent (e.g., *Maurice*) and its matching pronoun (e.g., *he*). Lee and Révész found a small positive effect of caption condition on learners' accuracy in a grammaticality judgement task, which measured learners' ability to distinguish correct and incorrect uses of the newly learned grammatical structure. The findings of this study suggest that learning of more complex grammar may, in some cases, also benefit from the use of enhanced captions. Lee and Révész pointed out that the target structure was already partially known by participants at pretest, which may have contributed to the learning effect.

In a more recent study, Lee and Révész (2021) found that enhanced captions promoted learning of the use of the present perfect versus the simple past. In this study also, prior knowledge and the relatively high salience of the target structures may have contributed to the positive effect of enhancement. learning gains. Interestingly, Lee and Révész (2021) found that typographic enhancement promoted both declarative and procedural knowledge of the target construction, which suggests that more elaborate processing of the target construction also enhanced automatic use during task performance. Lee and Révész suggested that the typographic enhancement may have a stronger effect on L2 grammar learning in multimodal input than in written texts. In multimodal input, learners may ignore information in the captions that is redundant to comprehension of the oral input. Enhancement might cause learners to pay more attention to target structures presented in the captions, leading to a stronger salience-raising effect compared to the use of typographic enhancement in written text.

2.6.4 Learning formulaic sequences from captioned video

Only one previous study has explored the use of enhanced captions to promote L2 learning of formulaic sequences (Majuddin et al., 2021). Majuddin, Siyanova-Chanturia and Boers (2021) examined the effects of repeated viewing and caption condition (no captions, full captions, full captions with highlighted words) on incidental learning of formulaic sequences from watching L2 video. In this study, Malaysian EFL learners watched a full episode of an English-language TV program. In the enhanced captions condition, twenty formulaic sequences (e.g., *slippery slope, on the same page*, etc.) were underlined in the captions. Learning of the target items was tested at the level of form recall (gap fill test) in a pretest and immediate posttest. In a delayed posttest phase, participants were additionally tested on aspects of form-meaning knowledge. Significantly higher gains emerged in the captioned conditions compared to the no-captions group. However, enhancement did not significantly affect learning, whereas repeated viewing did. Majuddin et al. suggested that the learners may not have had enough processing time to make efficient use of the enhanced text during viewing. However, this contradicts previous findings that enhanced captions may support L2 learning (e.g., Lee & Révész, 2019; Montero

Perez et al., 2015). Importantly, Majuddin et al. measured productive knowledge of target items, which is more difficult than the levels of knowledge tested in most of the preceding studies (e.g., Laufer & Goldstein, 2004; González-Fernández & Schmitt, 2020). An alternative explanation offered by Majuddin et al. is that the word sequences may have been too long (up to five words) for enhancement to have affected their processing in the brief time that the captions were on screen. However, because the study did not use an online measure of processing (unlike e.g., Lee and Révész, 2019), these explanations remain speculative.

2.7 Measuring L2 processing: eye-tracking

Although studies discussed above have suggested that variables related to the salience of formulaic sequences may affect how they are processed and learned, there is very little empirical research into the role of attention in L2 learning of formulaic sequences. Within the broader context of SLA, many studies have examined the relationship between degree of attention, or conscious awareness, and learning, for example by means of concurrent and retrospective verbal reports (e.g., think-aloud protocols, questionnaires, interviews, stimulated recall) (see e.g., Robinson, 2003).

An increasingly popular tool to investigate the cognitive processes involved in L2 learning is eye-tracking. The advantage of eye-tracking over other concurrent measures of attention, such as think-aloud protocols, is its relatively low degree of reactivity (Godfroid & Spino, 2015). Although eye-tracking experiments happen in a lab setting that differs in many ways from natural learning contexts (Spinner, Gass, & Behney, 2013; Tinker, 1936), it does not require learners to perform a secondary task that may alter learners' processing of L2 target structures. Eye-tracking can therefore provide a more faithful reflection of learners' thought processes as they would occur in a more natural setting (Godfroid & Spino, 2015). Further, as a concurrent measure of attention, eye-tracking can provide a more complete, less memory-dependent account of learners' cognitive processes compared to retrospective measures such as interviews and questionnaires (Godfroid & Schmidtke, 2013). Another important difference between eye-tracking and verbal reports, is that eye-tracking can examine attention as a continuous variable, unlike the absolute, binary operationalization of awareness in verbal reports (Godfroid et al., 2013). In what follows, I explain how eye-tracking has been used in SLA research, and discuss the underlying assumptions and applications of this method in some detail.

2.7.1 Eye-tracking research in SLA

The increasing popularity of eye-tracking is evident from the growing number of methodological guides and reviews devoted to eye-tracking in SLA (Carrol & Conklin, 2014a; Conklin, Pellicer-Sánchez, & Carrol, 2018; Godfroid, 2020; Godfroid & Hui, 2020; Pellicer-Sánchez & Conklin, 2019; Roberts & Siyanova-Chanturia, 2013). A multitude of research topics in SLA have benefited from the use of eye-tracking. To give a few examples, eye-tracking has been employed in studies examining construct validity in L2 assessment (McCray & Brunfaut, 2018), task complexity in L2 writing processes (Révész, Michel, & Lee, 2019), and the role of attention in L2 acquisition (e.g., Lee & Révész, 2018; Montero Perez et al., 2015).

Eye-tracking research makes the general assumption of an eye-mind link, i.e., that eye movements provide a window into L2 learners' cognitive processes (Just & Carpenter, 1980). This assumption captures two more specific beliefs. Firstly, what the eye gaze fixates on, is what is being considered at that moment (i.e., the locus of the eye fixation is the focus of attention). Secondly, the duration of an eye fixation reflects the amount of cognitive effort involved in processing the fixated region (i.e., processing time reflects processing effort). Using eye-tracking, researchers can infer what part of the L2 input a language learners is attending, and how much cognitive effort is being used in the process (Pellicer-Sánchez & Conklin, 2019).

Visual scenes, images, and texts that are more difficult, complex, or dense, attract more and longer *fixations* (i.e., the point at which the eye stops moving to take in visual information), and shorter *saccades* (i.e., eye movements between fixations during which no visual information is taken in (Castelhano & Rayner, 2008). In reading, increased processing effort is also associated with less linear reading, as reflected, for example, by a greater number of regressions, or saccades that go back to a previous part of the text. Task demands also affect eye movement measures during L2 processing. For example, reading a text for gist requires less effort, and thus, fewer fixations and regressions, than reading for comprehension (Pellicer-Sánchez & Conklin, 2019).

2.7.2 Eye movements in reading

At the word level, a distinction can be made between the processes of *word identification* and *context integration* (Clifton, Staub, & Rayner, 2007), which roughly correspond with early and late eye movement measures (Pellicer-Sánchez & Conklin, 2019). Early measures (e.g., first fixation duration, gaze duration or first pass time) are thought to capture automatic, low-level processes involved in word identification and lexical access, whereas late measures (e.g., total reading time, second pass time, fixation count, regression path duration, regressions in/out) reflect conscious, controlled, or strategic processes involved in context integration (Clifton et al., 2007).

Variables that have been found to predict word identification include, among others, corpus frequency (e.g., Rayner & Duffy, 1986), recency of occurrence (e.g., Rayner, Raney, & Pollatsek, 1995), word familiarity (e.g., Williams & Morris, 2004), and age of acquisition (e.g., Juhasz & Rayner, 2006). Late measures have been associated with syntactic or contextual processing difficulties, e.g., syntactic or semantic complexity or anomaly (e.g., Frazier & Rayner, 1982), or garden-path effects (e.g., Ferreira & Clifton, 1986)).

A few studies have examined how native speakers process novel words during reading. Chaffin (1997) examined L1 reading of novel and familiar words in sentence contexts. He observed that novel and less familiar words received longer fixations than familiar words during the first pass (first fixation duration, gaze duration), suggesting that lexical access is affected by familiarity. Further, Chaffin (1997) and Chaffin, Morris and Seely (2001) found that, when target words appeared in semantically supportive contexts, reading times and number of regressions-in were significantly higher for novel words, indicating that novel words induce processing efforts associated with context integration.

A comparatively small number of studies have used eye-tracking to examine the L2 reading process. Compared to reading in the L1, L2 reading is associated with more fixations, longer reading times at the sentence level, and less word skipping (Conklin, Alotaibi, & Vilkaitė, 2020; Cop, Drieghe, & Duyck, 2015). Notable theoretical explanations for the slower nature of L2 processing are the Weaker Links Hypothesis (Gollan, Montoya, Cera, & Sandoval, 2008), and the Bilingual Interactive Activation Plus Model, BIA+ (Dijkstra & Van Heuven, 2002). These models are both based on the assumption that more limited usage or exposure to the L2 compared to the L1 results in weaker lexical representations, and, thus, less efficient lexical access (Conklin, 2020).

Overall, based on current evidence, the same word-related variables appear to affect processing times in both L1 and L2 readers (e.g., Conklin et al., 2020; Godfroid et al., 2018). Like native speakers, L2 learners tend to read more frequent and more familiar words faster than novel and low-frequency words. Further, several L2 studies have found a significant correlation between familiarity and reading times on words presented in continuous texts (Elgort, Brysbaert, Stevens, & Van Assche, 2018; Godfroid et al., 2013, 2018; Pellicer-Sánchez, 2016). Using a counterbalanced design, Godfroid et al. (2013) showed that learners spent significantly longer fixating novel words (nonce words) compared to familiar control words that appeared in the same contexts in short L2 reading passages. Similar findings were reported by Pellicer-Sánchez (2016), Godfroid et al. (2018), and Elgort et al. (2018).

2.7.3 Eye movements in incidental vocabulary learning

Several studies have examined the relationship between learners' eye movements during reading and incidental L2 vocabulary learning (e.g., Elgort et al., 2018; Godfroid et al., 2013; Mohamed, 2017; Pellicer-Sánchez, 2016). Godfroid, Boers and Housen (2013) examined whether noticing, operationalized as the *amount or duration of overt attention for* novel words during reading led to increased memory for those words. They found a significant correlation between learners' total fixation times and scores in a posttest of form recognition. This study provided strong evidence for the positive effect of attention on word learning from context. Similar results were found by Pellicer-Sánchez (2016), who reported exceptionally high learning gains of form recognition (85%), meaning recognition (78%), and meaning recall (61%) after eight repeated exposures to novel words during reading.

Additionally, Pellicer-Sánchez observed that, in the course of reading, repeated encounters with novel words in high-constraining contexts led to a significant decrease in reading times on those words. Both novel and familiar words were read faster with repeated exposure, but the repetition effect showed up earlier and was evident in more eye-tracking measures (gaze duration, fixation count, total reading time) for novel words than for familiar words (fixation count, total reading time). By the eighth, final encounter, novel words were read at similar speed as familiar words, which Pellicer-Sánchez interpreted as evidence for increased familiarity with the novel word forms and their contextual meanings. Godfroid et al. (2018) reported a repetition effect on L1 and L2 readers' fixation times for unfamiliar Dari words appearing in five chapters of an English novel. In both groups of learners, repeated exposure led to a curvilinear decrease in reading times. In this study, both total reading times and fixation counts correlated with learners' scores on posttests of form recognition, meaning recognition, and meaning recall.

Elgort et al. (2018) measured the effect of repeated exposure during reading on learners' processing and knowledge of novel words. In this study, L2 learners encountered novel and familiar words 10 times in meaningful contexts within a continuous L2 text. In order to gain more insight into the effect of contextual exposure on various aspects of word knowledge, the study additionally used a sentence-reading posttest to measure learners' processing of the target words in semantically neutral contexts. Similar to Pellicer-Sánchez (2016), this study observed a steep decrease in reading times on novel words. After 5-10 exposures, novel words were read at similar speed as familiar words. The effect was mainly evident in early measures (first-fixation duration, gaze duration, regressions-in). However, in late measures (go-past time, total reading time, fixation count, number of regressions), a significant difference between the novel words and the familiar words remained until the final exposure. Further, when the novel words were encountered in neutral sentence contexts, only first fixation durations of the novel words were comparable to those of the familiar words. The authors suggested that the strong decreases in reading times primarily reflected familiarity with orthographic form, whereas fluent access to meaning remained difficult after 10 contextual exposures. Elgort et al. (2018) found no strong correlation between reading times in the sentence-reading posttest and scores on a posttest of meaning recall, suggesting that accessing meaning during reading may involve different cognitive processes than retrieving meaning from memory in a test of explicit recall.

The studies reviewed here illustrate the potential of eye-tracking to measure the effects of processing effort or degree of attention during reading on declarative L2 vocabulary knowledge (e.g., Godfroid et al., 2013, 2018). Moreover, eye-tracking measures provide a window into the effect of repeated exposure on the fluency with which learners access and integrate new words into context during reading (e.g., Elgort et al., 2018; Pellicer-Sánchez, 2016). This form of implicit knowledge has also been of profound interest to researchers studying L2 knowledge of formulaic sequences, which, as discussed in a previous section, is characterized by a processing advantage over novel word sequences (e.g., Carrol & Conklin, 2020).

Only a few studies to date have used eye-tracking to investigate how formulaic sequences are processed in continuous texts (Choi, 2017; Conklin & Carrol, 2021). Choi (2017) examined the effect of typographic enhancement on learners' attention to, and learning of L2 adjective-noun collocations encountered during reading. Participants in the enhanced group read texts containing collocations in bold typeface. These learners showed significantly longer fixation times (total reading time) for novel collocations than participants who read the same texts without enhancement. The enhanced group also performed significantly better on a posttest of collocation knowledge, indicating that the positive effect of attention on incidental vocabulary learning extends to learning of collocation.

Conklin and Carrol (2020) investigated how quickly native speakers develop a processing advantage for novel word co-occurrence patterns during reading. The premise of the study was the well-established processing advantage found for conventional binomials (*time and money*) over their reversed order (*money and time*) (e.g., Siyanova-Chanturia et al., 2011). This advantage is believed to reflect the effect of occurrence frequency in language input on learners' representation of binomials, i.e., a multiword frequency effect. Conklin and Carrol aimed to find out how many exposures in written input it takes for learners to preferentially process novel binomials over their reversed form during reading.

To that end, participants' eye movements were tracked while they read a story containing conventional binomials (e.g., *time and money*), and novel word sequences of the same structure (e.g., *wires and pipes*). The novel binomials appeared between one and five times in the reading texts, whereas their reversed form (e.g., *pipes and wires*) appeared only once at the end of the story. To avoid that any pre-experimental preferences for word order would affect processing fluency, the two forms of each novel binomial were counterbalanced across two presentation lists. For example, participants assigned to list 1 read the form *pipes and wires* throughout the story and *wires and pipes* as the noncongruent alternative form at the end of the story, whereas participants assigned to list 2 read *wires and pipes* repeatedly, and *pipes and wires* as the matched noncongruent form.

In line with previous findings in sentence-reading experiments (e.g., Carrol & Conklin, 2020), the study found that conventional binomials like *time and money* were read significantly faster than their reversed form, showing that the processing advantage of formulaic sequences occurs when native speakers have to access and integrate binomials in larger discourse contexts. Another interesting finding was that a single exposure to a new binomial led to increased processing fluency of its reversed form. That is, reading the item *pipes and wires* once in the story led to significantly faster reading times for *wires and pipes*. This shows that reading the words *wires* and *pipes* together led to facilitated recognition of this word pattern in a later context. Alternatively, this effect may also indicate a word-level repetition priming effect, whereby *wires* and *pipes* were read faster because these words were still active in memory due to their recent occurrence (e.g., Bybee, 1998). Additionally, Conklin and Carrol found a significant processing advantage of novel binomials over their reversed form after only three (first pass time) to four (total reading time) exposures. The eye movement data collected in this study thus demonstrate that native speakers quickly develop a preference for sequential patterns in the course of reading. Carrol and Conklin's eye movement data showed how native speakers develop knowledge of word sequences automatically and implicitly during reading. Whether L2 learners develop a processing advantage for co-occurrence patterns as quickly, remains unclear (but see Northbrook et al., 2021).

2.7.4 Eye movements in multimodal L2 processing

Eye-tracking has been used to provide insight into how language users distribute their attention between multiple sources or channels of input, including verbal vs. nonverbal input (Brône & Oben, 2018; McDonough, Trofimovich, Lu, & Abashidze, 2020), written vs. spoken input (e.g., Conklin et al., 2020), or text vs. images (Bisson, van Heuven, Conklin, & Tunney, 2014). Of direct relevance to the current thesis, and to Chapter 5 in particular, are studies examining L2 learners' visual processing of video with captions (Bisson, van Heuven, Conklin, & Tunney, 2014; Gass et al., 2019; Lee & Révész, 2018; Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022; Winke, Godfroid, & Gass, 2013; Wisniewska & Mora, 2018). Because of the complexity of attending to multiple input modes, traditional measures of attention (e.g., interviews) have limited value in this context. Learners may not always be able to describe accurately and in detail how they process different input modalities (Winke et al., 2013). Eye-tracking can provide a more complete and detailed account of how learners deal with the processing demands of captioned audiovisual input.

Eye movement patterns during processing of captioned audiovisual input are very different from standard reading. Firstly, caption reading may be influenced by the timed nature of audiovisual input. Unlike traditional texts, captions disappear after a few seconds. Eye-tracking indices such as fixation times and fixation counts may therefore be influenced by the amount of time the caption remains on screen, as well as the number of words in a caption. To account for these effects, studies investigating L2 caption reading behavior have analyzed eye movement measured normalized for caption duration (e.g., Bisson et al., 2014; Winke et al., 2013). Further, because captions overlay moving images, some studies examining L2 caption reading have also included a control group who watched the same video without captions, to make sure that fixation times reflect attention to the captions or subtitles, and not attention to other visual stimuli (e.g., Bisson et al., 2014). The timed nature of captions also means that learners do not have much time for rereading, and that the number of regressions may be considerably lower compared to traditional reading (Godfroid, 2020).

The multimodal nature of captioned video also affects eye movement behavior in other ways. For example, eye fixations may reflect processing of the visual text as well as the aural input stream. A recent study investigating eye movements during bimodal processing (i.e., reading while listening) found that reading times and word skipping were affected by the simultaneous presentation of visual and auditory text, both in native speakers and in nonnative speakers (Conklin et al., 2020). Further, both native speakers and nonnative speakers tended to read slightly ahead of the audio. Similarly, in a study focusing on L2 caption reading, Wizniewska and Mora (2018) found that learners tended to process words in the captions prior to the presentation of the corresponding aural word form. The fact that learners can process words visually before they hear them aligns well with previous claims that captions may enhance learners' processing efficiency (Bird & Williams, 2002).

Learners also have to divide their attention between the captions and imagery. Several studies have investigated how L2 learners divide their visual attention during viewing with captions. For example, Bisson et al. (2014) compared L2 learners' processing of different subtitling conditions (standard subtitles, reversed subtitles, captions). They found significant differences between the subtitling conditions in terms of the amount of time learners' spent fixating the captions and imagery regions. For example, learners who watched a video with Dutch audio spent significantly more time fixating the captions region. The presence of salient images or shot changes may also affect caption reading behavior (Kruger, Szarkowska, & Krejtz, 2015).

As discussed in the previous section, the usefulness of captions for L2 acquisition may vary depending on learner variables (e.g., Gass et al., 2019) and variables of the input (e.g., Montero Perez et al., 2015). Likewise, how learners process captions depends on factors such as prior vocabulary knowledge, working memory capacity, L1 background, and topic familiarity (Gass et al., 2019; Wang & Pellicer-Sánchez, 2022; Winke et al., 2013).

In L2 vocabulary learning, the use of eye-tracking has provided insight into how learners may cope with the timed, multimodal nature of subtitled or captioned audiovisual input, and how this affects L2 vocabulary learning (e.g., Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022). A few studies have found that learners' reading times on novel words vary depending on the type of captions/subtitles used (Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022). These studies have also shown that reading times on novel words presented in the captions predict vocabulary gains. However, the relationship between reading times and learning gains may depend on the type of captioning, as well as the aspect of word knowledge tested (Montero Perez et al., 2015, 2018; Wang & Pellicer-Sánchez, 2022).

The relationship between caption processing and vocabulary gains may sometimes be difficult to interpret. For example, Wang and Pellicer-Sánchez (2022) found that learners spent significantly longer fixating L1 translations than L2 words in bilingual captions, i.e., L1 and L2 captions presented simultaneously. However, reading times on L2 word forms, not their L1 translation, predicted learning of word meaning. This contradictory finding shows that eye-tracking measures cannot always pinpoint the specific processes that underlie incidental vocabulary learning, in particular in the complex processing conditions of multimodal input (Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022).

2.7.5 Limitations of eye-tracking in SLA research

Although eye-tracking clearly has merits, it also comes with a number of limitations and challenges. Firstly, in spite of the growing user-friendliness of eye-tracking tools (Rayner, 1998), the reliability of eye-tracking data relies strongly on the specifications of the eye-tracking equipment, and the technical know-how of the researcher using it (Conklin et al., 2018; Godfroid & Hui, 2020). Further, the correct interpretation of eye movement data stands or falls with experimental design, which ideally takes on a confirmatory, hypothesis-testing approach to the research topic under investigation (Godfroid & Hui, 2020). A basic understanding of the theoretical assumptions underlying eye movement control in processes like reading and visual search is also recommendable. The appropriate use of eye-tracking thus requires some training on the part of both the research team and the individual researcher. Also worth noting is the fact that eye-tracking is restricted to the measurement of visual attention, which means that researchers have relied on inventive but restricted paradigms for measuring auditory L2 processing (e.g., the visual world paradigm, see Godfroid, 2020). Finally, because of the limited insight eye-tracking can provide into various aspects of cognitive processing, researchers have called for the triangulation of eye-tracking with declarative measures of processing, such as stimulated recall (e.g., Jung & Révész, 2018).

3 Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences

Previous research has shown that learners should ideally know the most frequent 8,000–9,000 word families for reading (Nation, 2006) and 3,000 for TV viewing (Webb & Rodgers, 2009). Because only so much time can be spent teaching vocabulary in the classroom, researchers have encouraged incidental vocabulary learning through exposure to L2 input as a means to expand learners' vocabulary knowledge (e.g., Pellicer-Sánchez, 2017; Peters, 2018; Peters & Webb, 2018; Peters et al., 2019). Incidental vocabulary learning has primarily been researched in the context of reading (e.g., Horst et al., 1998; Pellicer-Sánchez & Schmitt, 2010; Saragi, Nation, & Meister, 1978). However, the increasing popularity of television and movies has incited scholars to investigate the potential of audiovisual input as a source of L2 vocabulary (Webb & Rodgers, 2009). Although some recent studies (Montero Perez et al., 2014; Peters & Webb, 2018) have shown that words can be learned incidentally through watching TV, the research into vocabulary learning from audiovisual input is still scarce.

Most studies into incidental vocabulary acquisition have focused on single words (Nation, 2001). However, little is known about the incidental learning of formulaic sequences (FS). A few studies have shown that FS can be learned incidentally through reading a single text (Choi, 2017; Macis, 2018; Pellicer-Sánchez, 2017) and through reading-while-listening (Webb et al., 2013). Yet it remains unclear what the potential of TV viewing is for learning FS. This is surprising, given that a corpus study showed that the distribution of FS on television is similar to that in everyday speech (Lin, 2014). The study suggests that television might be a good source of L2 FS, particularly for those FS occurring in spoken discourse. However, there is as yet no conclusive evidence that learners can pick up FS from audiovisual input.

This article reports on an exploratory study that investigated the incidental learning of single words and FS through TV viewing. The first aim was to explore whether FS, like single words, can be picked up incidentally. Secondly, the study aimed to explore some factors that might mediate the learning process of single words and FS.

3.1 Background

3.1.1 Incidental learning from audiovisual input

Because learners need to acquire large vocabularies for L2 reading, listening and viewing, it has been argued that L2 vocabulary acquisition should to some extent occur outside the classroom (Peters, 2018; Peters et al., 2019). Although most studies on incidental vocabulary learning have been carried out in the context of reading (Nation, 2001), there are a number of reasons to consider the role of television as a source of L2 vocabulary. First, viewing is a very popular activity among the majority of L2 learners. A 2017 survey of the European

Commission (2018) shows that television is the preferred medium among EU citizens, with 84% watching television almost every day. Watching English-language TV is also a popular activity among EFL learners (Peters, 2018; Peters et al., 2019). Peters and colleagues found that EFL learners spend considerably more time watching television than reading in English. Moreover, due to its entertainment value, L2 television may be more effective in lowering learners' anxiety than written input, given that learners find TV viewing more accessible than written texts (Larsen-Freeman, 1983; Neuman & Koskinen, 1992). A third reason why watching television might be beneficial for vocabulary learning is that vocabulary demands for TV viewing (3,000 word families) are lower than for reading (Webb & Rodgers, 2009).

One of the earliest studies on incidental vocabulary learning from audiovisual input (Neuman & Koskinen, 1992) examined whether children could pick up English vocabulary by watching short educational videos with and without captions, and through reading-while-listening. Higher learning gains were found for both viewing groups than for the reading-while-listening group. A recent study by Peters and Webb (2018) examined incidental learning from watching an entire one-hour episode of a TV program. They found that incidental vocabulary learning through watching TV is possible at the level of meaning recall and meaning recognition. Further, it was shown that word-related factors (cognateness, frequency of occurrence) as well as learners' prior vocabulary knowledge were positively correlated with learning.

3.1.2 Learning formulaic sequences

The term 'formulaic sequence' encompasses a broad range of word combinations and expressions, such as collocations, idioms, phrasal verbs, proverbs and lexical bundles. Two major approaches have been used to define and categorize formulaic sequences: the phraseological approach and the distributional or frequency-based approach (Granger & Paquot 2008). The phraseological approach identifies formulaic sequences based on linguistic criteria, such as syntactic fixedness and semantic opacity. The frequency-based approach, on the other hand, is based on corpus-derived measures such as the frequency and association strength of word combinations. Because previous research has found that both distributional (González Fernández & Schmitt, 2015) and phraseological (Peters, 2016) properties may affect the learning burden of FS, we combine both approaches in our methodology and loosely define FS as conventionalized, recurring word combinations.

FS constitute up to 60% of spoken and written discourse (Erman & Warren, 2000), fulfil many pragmatic functions and make up virtually all conventionalized speech acts (Schmitt, 2010). Yet, despite the ubiquity and prominence of FS in language, it has been shown that L2 learners struggle with the appropriate use of FS. Corpus studies have revealed that even advanced learners tend to overuse or misuse FS, and produce many errors, often caused by L1 interference (Laufer & Waldman 2011). Furthermore, L2 learners' knowledge of FS has been shown to lag behind their knowledge of single words (Bahns & Eldaw, 1993; Granger, 1998; Nesselhauf, 2003, 2005).

One reason why L2 learners struggle with FS concerns their lack of exposure. It has been proposed that the acquisition of FS essentially relies on extensive exposure (González Fernández & Schmitt, 2015). However, although natural language is highly formulaic, individual formulaic sequences do not occur as frequently as single words (e.g., Moon, 1998). One way that learners could improve their knowledge of FS is through incidental learning activities. However, only a few studies have examined incidental learning of FS from exposure to L2 input (Choi, 2017; Frumuselu, De Maeyer, Donche, & Gutiérrez Colon Plana, 2015; Macis, 2018; Pellicer-Sánchez, 2017; Peters, 2012; Webb et al., 2013). It has been shown that FS can be learned incidentally from reading texts.

To our knowledge, only one study has explored incidental learning of FS through viewing (subtitled TV). Frumuselu et al. (2015) conducted a longitudinal study in which L2 learners with various L1s and levels of proficiency watched 13 episodes of an English TV program with L1 subtitles (captions) or L2 subtitles. The findings showed that it was possible to learn FS from subtitled TV. However, it remains unclear whether FS can be picked up incidentally from watching TV without subtitles or captions. Further research into the incidental learning of FS from audiovisual input thus seems warranted.

3.1.3 Incidental learning: the role of prior knowledge and item-related factors

Incidental learning is a slow and incremental process, which relies heavily on a learner's level of proficiency and prior vocabulary knowledge (Hulstijn, 2013). Moreover, since individual words and FS each have their own learning burden (Schmitt, 2004), the effectiveness of incidental learning might additionally depend on item-related factors. Some of the factors that have been shown to affect the learning of single words are:

- prior vocabulary knowledge (Montero Perez et al., 2014; Peters & Webb, 2018; Peters et al., 2016): learners with a larger vocabulary size tend to learn more words incidentally.
- corpus frequency (Vidal, 2003): frequent words tend to be learned more easily than infrequent words.
- concreteness (Pichette, De Serres, & Lafontaine, 2012): concreteness is the degree to which a word refers to a perceptible entity (Brysbaert, Warriner, & Kuperman, 2014). Concreteness has been shown to facilitate processing and learning of words.
- word length (Campoy, 2008): short words tend to be learned more easily than long words.
- part of speech (Campoy, 2008): nouns tend to be more concrete, imageable and meaningful than verbs (Crossley, Subtirelu, & Salsbury, 2013), and may also be acquired more easily.

With regard to the learning of FS, phraseological as well as distributional factors have been shown to mediate learning. Some of these are:

- collocate-node relationship (Peters, 2016): adjective–noun combinations tend to be learned more easily than other types of FS. Verb-particle combinations such as phrasal verbs are challenging to L2 learners, because they lack semantic transparency and often have multiple meanings (Garnier & Schmitt, 2016).
- association strength (Nguyen & Webb, 2017): association strength is the strength of co-occurrence of two or more words, based on corpus-derived measures such as mutual information (MI) and t-score. Mutual information measures the strength of attraction between two words by comparing their observed frequency of co-occurrence to the expected frequency of co-occurrence. An MI score above 0 corresponds to higher co-occurrence than is expected by chance, i.e., attraction between the words, whereas a score below 0 corresponds to lower co-occurrence than is expected by chance, or repulsion (Evert, 2008). MI has been shown to assign inflated scores to word combinations that contain low-frequency components (Evert, 2008). As a result, FS with high MI scores tend to be more difficult to acquire because they often contain low-frequency words (Nguyen & Webb, 2017).

The factors mentioned above have mainly been studied in the context of intentional vocabulary learning (e.g., Peters, 2016) or incidental learning from written texts (Pellicer-Sánchez, 2017). Not much is known about the effects of item-related factors in incidental vocabulary learning from audiovisual input.

3.2 Rationale and research questions

To date, very little is known about the incidental learning of formulaic sequences and the potential of TV viewing for learning FS. There is some evidence now that single words can be learned from watching TV. However, it remains unclear whether FS can be learned to the same extent. As a result, more research into the learning of FS is needed, if we want obtain a fuller picture of the learning process of FS and the factors affecting this process. Given the abundance of FS in everyday spoken language and audiovisual input (Lin, 2014), television could be an important source of formulaic language for L2 learners. The aim of this exploratory study is to investigate whether, like single words, FS can be learned from watching a television program and which item-related factors might affect the learning process. The following research questions were addressed in this study:

1. Can single words and formulaic sequences be learned from watching L2 television?
2. Which item-related and learner-related factors affect the incidental learning of single words from watching L2 television?
3. Which item-related and learner-related factors affect the incidental learning of formulaic sequences from watching L2 television?

To answer the research questions, an exploratory experiment was conducted adopting a pre-test post-test within-subject design. Twenty participants watched an English TV program without captions or subtitles. English words and formulaic sequences occurring in the program were tested before and after the treatment. There was no control group, but distractor items were added to the pre- and post-tests in order to control for a test effect.

3.3 Method

3.3.1 Participants

Twenty first-year business students at a Flemish university (L1 = Dutch) aged between 19 and 21 (average age: 19.5) took part in the experiment. All participants had received at least five years of instruction in English, as the starting age for English in Flanders is 13–14 years old. Their level of proficiency could be considered B1 to B2 of the Common European Framework of Reference. However, in order to take into consideration individual differences in proficiency, the participants' vocabulary size was measured by means of Nation and Beglar's (2007) Vocabulary Size Test (see Instruments section). Their vocabulary size ranged between 8,200 and 15,000 word families (average 11,450). Given the study's focus on audio-visual input, it should be mentioned that learners in the present study were used to watching English language TV, as was shown in a questionnaire tapping into learners' TV viewing habits.

3.3.2 Materials

3.3.2.1 Audiovisual input

The audiovisual input selected for this study was a 30-minute excerpt of Dragon's Den, a British reality TV program in which budding entrepreneurs pitch their business ideas to a panel of successful business people. The program is abundant with business English which makes it an interesting source of vocabulary for business students. By entering the script of the excerpt into the Vocabprofile section of the Compleat lexical tutor (Cobb, n.d.), it was found that 87.32% of the words in the input belong to the 1000 most frequent words in the British National Corpus/Corpus for Contemporary American English (BNC/COCA), and 92.85% belonged to the 2,000 most frequent words.

3.3.2.2 Target items

A diverse set of 15 single words and 20 FS was selected from the input (see Table 3.1 for the single words and Table 3.2 for the formulaic sequences). Unlike in some reading studies (Pellicer-Sánchez, 2017; Webb et al., 2013), we did not manipulate the input or target items for reasons of ecological validity. However, because a number of factors have been found to affect the learning of single words and FS, some of these were taken into account in the analyses. The following factors were taken into account in the analyses for the learning of single words:

- corpus frequency
- item length (= number of syllables)
- part of speech
- concreteness

Raw corpus frequencies were obtained from the Spoken and Fiction subcorpora of the COCA (Davies, 2008). Concreteness was obtained from Brysbaert, Warriner and Kuperman's (2014) norms. The norms are based on ratings from 4,000 participants obtained through internet crowdsourcing. As opposed to other concreteness norms (e.g., Paivio, Yuille, & Madigan, 1968), the ratings are available for a large number of lemmas (approx. 40,000), and are based on all modes of perception. The norms can be found on Marc Brysbaert's website (http://crr.ugent.be/archives/1330). Because only one word was a Dutch cognate (*empathize*), this factor was not taken into account. Frequency of occurrence in the input was also not taken into account since only three items occurred more than once.

Table 3.1 Target items – single words

	Corpus frequency	Frequency of occurrence	Length (syllables)	Part of speech	Concreteness
quirky	1963	2	2	Adjective	2.12
engaging	2568	1	3	Adjective	1.72
fledgling	1785	1	2	Adjective	2.76
articulate	1388	1	4	Adjective	2.31
scalable	216	1	3	Adjective	2.5
to entice	2207	1	2	Verb	1.12
to empathize	800	1	3	Verb	2.23
to forecast	3786	2	2	Verb	2.78
interrogate	1957	1	4	Verb	2.48
quandary	829	1	3	Noun	1.75
longevity	2626	1	4	Noun	1.9
ardor	426	1	2	Noun	2.59
clarification	1714	1	5	Noun	1.93
retail	12200	8	2	Noun	2.92
venture	12314	1	2	Noun	2.6

Table 3.2 Target items – formulaic sequences

	Occurrence frequency	Corpus frequency	Length	Collocate-node relationship	MI	Type
corporate event	63	1	4	Adjective-noun	3.37	lexical coll.
advance booking	19	1	4	Adjective-noun	8.77	lexical coll.
subliminal message	78	1	6	Adjective-noun	10.53	lexical coll.
raw material	2351	1	5	Adjective-noun	9.91	lexical coll.
vast number	498	2	3	Adjective-noun	5.08	lexical coll.
master × art	366	1	5	Verb-noun	5.53	lexical coll.
capture × imagination	395	1	10	Verb-noun	6.73	lexical coll.
hold × own	1422	1	4	Verb-noun	2.1	idiom
catch × eye	2495	1	3	Verb-noun	3.89	idiom
take × punt	7	2	4	Verb-noun	0.41	lexical coll.
run × course	599	1	7	Verb-noun	0.7	lexical coll.
tap into	2879	1	3	Verb-particle	5.51	gramm. coll.
come about	2830	1	3	Verb-particle	0.54	phrasal verb
head back	3687	1	2	Verb-particle	5.59	phrasal verb
shy away	1472	1	3	Verb-particle	10.63	phrasal verb
turn over	5724	3	3	Verb-particle	3.87	gramm. coll.
drill down	203	1	2	Verb-particle	4.76	phrasal verb
down the line	1888	1	2	Prep.-noun	2.95	idiom
up and running	1287	1	3	Prep.-adjective	1.33	binomial
economies of scale	497	1	6	Noun-noun	5.09	compound

For the learning of FS, the following factors were analyzed:

- corpus frequency
- item length (= number of syllables)
- phraseological type
- collocate-node relationship
- mutual information (MI)

The phraseological type of each FS was determined based on Granger and Paquot's (2008) classification. Five types were distinguished: lexical collocations (e.g., *subliminal message*), idioms (e.g., *catch × eye*), binomials (*up and running*), grammatical collocations (e.g., *turn over*) and phrasal verbs (e.g., *head back*). In terms of collocate-node relationship, the following types were included: verb-noun, verb-particle, adjective–noun, noun-noun, preposition-noun, and preposition-adjective. Mutual information (MI) scores ranged between 0.41 and 10.63 (average 4.75). Mutual information tends to favor low-frequency and specialized word combinations (Gablasova et al., 2017). Because we wanted to include different types of FS, including those containing high-frequency words, we did not adopt the commonly held threshold of 3 (Schmitt 2010), but instead included FS with lower MI scores if these could be considered conventional based on their corpus frequency.

All of the selected target items were incongruent, i.e., did not have a literal word-for-word translation in Dutch. We could not find word-for-word translations for any of the items in online bilingual dictionaries or translation websites, although one item (*subliminal message*) has a single-word literal translation. Congruency was therefore not included as a variable possibly affecting the learning of FS. Other factors that were not taken into account include adjacency of the component words in the input and grammatical variability. Because only three FS occurred more than once in the input, occurrence frequency was also not included in the analyses. Corpus frequencies of the component words were not included as a separate factor because this would likely cause multi-collinearity with MI and possibly with corpus frequency of the FS as a whole.

3.3.3 Instruments

3.3.3.1 Vocabulary size test

Because learners' prior vocabulary knowledge has been shown to be an important predictor of incidental vocabulary learning from meaning-focused input (e.g., Peters & Webb, 2018), a vocabulary size test was administered to the participants, viz. Nation and Beglar's (2007) Vocabulary Size Test. The test is a frequency-based multiple choice test, containing 100 items that measure receptive knowledge of the form-meaning link. The score multiplied by 200 gives an estimate of the participant's total vocabulary size (Nation & Beglar, 2007).

3.3.3.2 Vocabulary tests

Because vocabulary learning might take place at various levels of knowledge, three vocabulary tests were used to measure knowledge of form and meaning at different levels of strength (form recognition, meaning recall, form recall). To control for a test effect, five low-frequency single words that did not occur in the input were added to the tests as distractors (*lunacy, to anneal, to bray, bereft, insipid*).

The first test was a form recognition test which measured the ability to recognize the meaning of a spoken or written form, through a multiple-choice cloze format. Participants were asked to complete an English sentence by selecting a missing word or formulaic sequence out of four options. The distractor options in this test were either non-words (for the single words) or incorrect formulaic sequences, i.e., word combinations that could not be found in the COCA. The distractor options for formulaic sequences were composed by replacing one single-word component with a word assigned to the same part of speech and related to the original word in meaning, in order to avoid deduction of the correct combination based on semantic association. The correct FS could only be recognized based on the form of the collocation as a whole. Below are two examples of sentences included in the form recognition test:

(1) The adverts……the customer into buying things they don't really want.

□ barlone □ delead □ *entice* □ insuade

(2) He tried to give the……that he was a man of the people.

□ astute message □ *subliminal message* □ circumspect message □ contemplative message

The second test measured knowledge at the level of form recall, i.e., the ability to supply the form of a given meaning. As shown below, participants were asked to write the form of the target words and formulaic sequences based on a Dutch translation or short description. The first letter of each component word was given to avoid elicitation of other potential word combinations.

schaalvoordelen = e _ _ _ _ _ _ _ _ of s _ _ _ _

Given that the target items were presented orally in the input, spelling mistakes were ignored as long as the response was comprehensible (e.g., *corperate event, sublimenal message, interoggate*), following Sonbul and Schmitt (2013).

The third test measured meaning recall, i.e., the ability to supply the meaning of a given form. Participants were asked to translate the English words and formulaic sequences into Dutch or describe their meaning. Both the spoken and written form of the target items were given:

economies of scale = …

3.3.3.3 Questionnaire

In addition to the vocabulary tests, the participants completed a short questionnaire about their comprehension of the gist of the content, their general comprehension, TV viewing habits (e.g., how often they watch English television without subtitles), and familiarity with the television program. The questionnaire can be found in Appendix 1.

3.3.4 Procedure

The data were collected in two sessions. Two weeks before the treatment the vocabulary size test and pre-tests were administered. To minimize a potential test effect, the form recall test was administered first, followed by the form recognition test, the vocabulary size test, and finally the meaning recognition test. The vocabulary size test was completed within 35 minutes, the pre-tests took 15 minutes each. Two weeks after the pre-tests, participants watched the excerpt from Dragon's Den, after which they completed the questionnaire and the immediate post-tests. The post-tests were administered in the same order as the pre-tests. To avoid a test order effect, the items comprising the post-tests appeared in a different order from the items of the pre-tests. Five target items were known by 18 of the participants on the form recall pre-test, and were therefore no longer included in the posttests; *scalable, interrogate, venture, raw materials, economies of scale*. After the post-tests, the participants were debriefed about the aims of the study.

3.4 Results

3.4.1 Questionnaire

The answers to the questionnaire indicated that the participants had no trouble understanding the gist of the content and that they had prior experience of viewing English television. All participants indicated that they watched English film or television with Dutch subtitles on a daily or a weekly basis. They also watched English video without subtitles, albeit less often (once a week/month). All participants were familiar with the television program Dragon's Den and had watched at least one episode before the treatment. However, none of the participants had watched the episode used in the treatment.

3.4.2 Vocabulary tests

The vocabulary tests were scored dichotomously; a correct answer received a score of 1, an incorrect answer received a score of 0. The descriptive results are reported in Tables 3.3 and 3.4. Learning gains were found for all levels of vocabulary knowledge.

Although scores on the form recall test were lower than on the meaning recall and meaning recognition tests, absolute learning gains (i.e., the difference between the pre- and post-test scores) were the highest for this area of knowledge. Learning gains were higher for formulaic sequences than for single words, but some learning gains were also made for the distractors, possibly due to a test effect. The 95% confidence intervals indicate that differences between the pre- and post-tests were nevertheless larger for the target items than for the distractors. Closer examination of the individual results of the form recognition test revealed that many scores per item were higher on the pre-test than on the post-test, which may be due to guessing. Guessing effects are typical of multiple-choice tests and often lead to overestimation of scores (Gyllstad, Vilkaitė, & Schmitt, 2015). Therefore, the results of the form recognition test were not further analyzed.

3.4.3 Research question 1: Can single words and formulaic sequences be learned from watching L2 television?

Logistic regression analyses with repeated measures (Generalized Estimating Equations (GEE) in SPSS) were performed on the scores of each test to find out if there were significantly more learning gains for target items than for distractors and if the treatment (watching TV) had an effect. The dependent variable in the models (one model for the meaning recall test, and one for the form recall test) was the score on the post-test given that the score on the pre-test equaled 0. Whether or not a test item was a distractor was entered as a dichotomous factor in the analyses. The participants' vocabulary size was entered as covariate.

On the form recall test, 'type of item (= target or distractor)' was a significant predictor of learning ($p < .001$, $B = 1.429$, $\text{Exp}(B) = 4.174$), as was vocabulary size ($p < .001$, $B = 0.069$, $\text{Exp}(B) = 1.071$). This indicates that target items were learned significantly better than distractors. Learning gains could therefore be ascribed to the treatment, and not to the vocabulary tests.

On the meaning recall test, the factor 'type of item' did not significantly predict learning (p = .24), while vocabulary size did (p = .036, B = 0.035, Exp(B) = 1.036). This indicates that both target items and distractors were learned to the same extent on the meaning recall test. For that reason, the effects of item-related and learner-related factors were only analyzed for the form recall test results.

Table 3.3 Learning gains – target items and distractors

	Pre-test *M* (*SD*)	95% CI	Post-test *M* (*SD*)	95% CI	Abs. gains *M* (*SD*)	95% CI
Target items (max. = 35)						
Form recognition	20.45 (4.97)	18.12–22.78	23.7 (3.7)	21.97–25.43	3.45 (2.78)	2.15–4.75
Form recall	3.35 (2.97)	1.96–4.74	9.3 (5.67)	6.65–11.95	5.95 (3.73)	4.20–7.70
Meaning recall	12.15 (6.16)	9.27–15.03	15.6 (5.71)	12.93–18.27	3.6 (2.44)	2.46–4.74
Distractors (max. = 5)						
Form recognition	1.65 (1.09)	1.14–2.16	2.25 (1.29)	1.65–2.85	0.6 (1.57)	-0.13–1.33
Form recall	0.05 (0.22)	0.39–1.31	0.4 (0.5)	0.93–2.17	0.35 (0.49)	0.36–1.04
Meaning recall	0.85 (0.99)	-0.05–0.15	1.55 (1.32)	0.16–0.64	0.7 (0.73)	0.12–0.58

Table 3.4 Learning gains – single words and formulaic sequences

	Pre-test *M* (*SD*)	95% CI	Post-test *M* (*SD*)	95% CI	Abs. gains *M* (*SD*)	95% CI
Single words (max. = 15)						
Form recognition	7.80 (2.24)	6.75–8.85	8.85 (1.84)	7.99–9.71	1.05 (2.35)	-0.05–2.15
Form recall	1.25 (1.16)	0.71–1.79	3.25 (2.34)	2.16–4.34	2 (1.59)	1.26–2.74
Meaning recall	4.20 (2.26)	3.14–5.26	5.52 (2.29)	4.18–6.32	1.05 (1.54)	0.33–1.77
FS (max. = 20)						
Form recognition	12.65 (3.42)	11.05–14.25	14.85 (2.23)	13.81–15.89	2.20 (2.26)	1.14–3.26
Form recall	2.1 (2.13)	1.11–3.09	6.05 (3.43)	4.45–7.65	3.95 (2.48)	2.79–5.11
Meaning recall	7.95 (4.25)	5.96–9.94	10.35 (3.91)	8.52–12.18	2.4 (2.06)	1.43–3.37

3.4.4 Research question 2: Which item-related and learner-related factors affected learning of single words?

Logistic regression analyses with repeated measures were performed on the form recall test results in order to investigate the relationship between vocabulary learning and item-related factors. Two models were designed: one for single words, one for formulaic sequences. We checked for the assumptions of logistic regressions (multi-collinearity, linearity of independent variables and log odds, ratio of the number of observations and the number of variables).

For the single words, the following variables were entered into the regression model: corpus frequency (logarithmically transformed), length in syllables, part of speech, concreteness. Through a backward stepwise selection, the non-significant predictors were removed one by one, resulting in the final models including only predictors with a p-value lower than .05. The results are presented in Table 3.5. Four parameters were significantly related to learning: corpus frequency, item length, concreteness, and learners' vocabulary size. Part of speech did not significantly predict learning. Concreteness was the most important predictor of learning (see Exp(*B*) in Table 3.5); the odds of a correct response on the form recall post-test were 5.03 times higher if the concreteness level increased by 1 unit. Vocabulary size was also positively correlated with learning. The odds of a correct response were 1.12 times higher if a participant's score on the vocabulary size test increased by one. This means that the odds of a correct response were 3.19 times higher if a participant's score on the vocabulary size test increased by ten (Exp(10*B)), i.e., if their vocabulary size increased by 2,000 word families. Large differences in vocabulary size could thus account for differences in learning gains. The effects of corpus frequency and length were also positively correlated with learning.

3.4.5 Research question 3: Which item-related and learner-related factors affected learning of formulaic sequences?

For the learning of formulaic sequences, the following covariates and factors were included in the logistic regression: corpus frequency, length in syllables (as appearing in the input), collocate-node relationship, mutual information, phraseological type. Again through a backward stepwise selection, the non-significant predictors were removed one by one, resulting in the final models including only predictors with a p-value lower than .05. Because phraseological type and collocate-node relationship were highly correlated (>.70), only collocate-node relationship was included in the analyses given its higher correlation with the posttest scores.

Table 3.5 Generalized estimating equations (GEE) for single words

	B	*p*	Exp (*B*)
Vocabulary size	0.12	.001	1.12
Length	0.56	.002	1.76
Concreteness	1.62	<.001	5.03
Corpus frequency	1.39	<.001	4.00
Part of speech		ns	

All factors were significantly related to learning in the form recall test (see Table 3.6). The most important predictor of learning was the item's collocate-node relationship. The odds of a correct response were 30.64 times higher if the FS was an adjective–noun combination. The effect size was also very large for verb-noun combinations (12.05), indicating that verb-particle combinations (phrasal verbs and grammatical collocations) had much smaller learning gains than the other two types. For mutual information, a negative correlation was found. The stronger the association between the component words, the smaller the learning gains. Similarly, a negative relationship was found between item length and learning. Specifically, the odds of a correct response were 1.25 (1/Exp(*B*)) times higher if the FS was one syllable shorter. As was the case for the single words, learners' vocabulary size was positively correlated with learning FS. The odds of a correct response were 2.18 times higher if the participant's score on the vocabulary size test increased by ten, i.e., if their vocabulary size increased by 2,000 word families.

3.5 Discussion

3.5.1 Research question 1: Can single words and formulaic sequences be learned incidentally from L2 television?

Our findings tentatively suggest that both single words and formulaic sequences can be learned incidentally through TV viewing in terms of form recall. On the form recall test, learning gains were significantly higher for target items than for distractors. We can therefore assume that the items were learned from the treatment. Even though previous research has shown that incidental vocabulary learning is a slow and incremental process (Nation, 2001), and that one encounter with a word or FS is unlikely to result in a strong form-meaning link (Hulstijn, 2013), our findings suggest that different word knowledge aspects can be learned even from a single exposure. Previous research into the effects of TV viewing has mainly focused on learning the meaning of target items. Our study adds that learning might also occur at the level of form recall, a knowledge aspect that is generally considered to be more difficult than meaning recall and recognition (Nation, 2001).

Table 3.6 Generalized estimating equations (GEE) for formulaic sequences

	Form recall		
	B	*p*	Exp. (*B*)
Vocabulary size	0.08	<.001	1.08
Collocate-node = adjective-noun	3.42	<.001	30.64
Collocate-node = verb-noun	2.49	<.001	12.05
Length	−0.23	.04	0.80
MI	−0.16	.001	0.85
Corpus frequency	0.55	.016	1.74

On the meaning recall test, however, no significant difference was found between the learning gains for target items and distractors. The lack of this significant difference might be explained by a test effect. Because the form recall test, in which the target items' definitions were provided, was administered before the meaning recall test, it is possible that participants associated some of the target items' forms on the meaning recall test with definitions they remembered from the form recall test. A second explanation might be the low number of participants, which resulted in considerable variation in learning gains. Finally, the low number of target items and distractor items in particular might also have played a role. Consequently, no conclusions can be drawn for the effect of TV viewing at the level of meaning recall.

3.5.2 Research question 2: Which item-related and learner-related factors affect the incidental learning of single words from watching L2 television?

In answer to the second research question, the results of the logistic regression analysis show that item-related variables as well as the participants' vocabulary size played a role in the learning of single words. Our study adds to the growing body of evidence that learners' vocabulary size is positively correlated with learning. The more words learners know, the more likely they are to pick up new words incidentally from viewing. This is in line with previous findings that prior vocabulary knowledge predicts incidental learning (Montero Perez et al. 2014; Peters et al., 2016; Peters & Webb, 2018).

Secondly, a positive correlation was also found for concreteness. More concrete target words (e.g., *empathize, quirky*) were learned better than more abstract words (e.g., *engaging, to entice*), which supports findings of previous studies that concrete words are easier to learn incidentally (Pichette et al., 2012).

Thirdly, corpus frequency was positively correlated with learning. In psycholinguistic research, corpus frequency is considered a good proxy for word familiarity (Kuperman & Dyke, 2013). Because words that occur frequently in a language tend to be encountered more often, they are also more likely to be acquired. It is possible that the participants in our study already had partial knowledge of the most frequent items (e.g., *quirky, engaging*) before the experiment, which they strengthened at the level of form recall. The correlation found for corpus frequency might thus to some extent reflect familiarity with the target items.

Finally, there was a small positive correlation between learning and word length. Longer target words may have been more salient in the aural input and therefore more likely to be learned. However, some of the longer items such as *articulate* and *quandary* were not learned well, indicating that the role of length was less clear than that of concreteness and vocabulary size.

3.5.3 Research question 3: Which item-related and learner-related factors affect the incidental learning of formulaic sequences from watching L2 television?

All item-related factors vocabulary size played a role in the incidental learning of FS. First, a positive correlation was found between the participants' vocabulary size and learning, which suggests that learners' prior vocabulary knowledge may facilitate the incidental learning of FS from viewing. The facilitating role of prior vocabulary knowledge on the acquisition of FS has also been found in one previous study (Peters, 2016).

Secondly, collocate-node relationship was an important predictor of learning. In line with previous findings (Peters, 2016), the highest learning gains were found for adjective–noun combinations, whereas the lowest learning gains were found for verb-particle combinations. As has been mentioned in previous studies (Garnier & Schmitt, 2016), phrasal verbs often have opaque or idiomatic meanings (e.g., *drill down, come about*). The verb-particle combinations in our study may therefore have been semantically less accessible than other types of FS. Moreover, the results of the form recall test show that the participants were sometimes able to provide the particle, but not the verb. Verbs such as *head* in *head back* and *shy* in *shy away* do not often occur in other word combinations, and may have been unfamiliar to the participants.

Thirdly, a negative correlation was found between length and the learning of FS. This is surprising, given that the longest FS were adjective–noun combinations and some of the shortest FS were verb-particle combinations. The effect of item length may be due to a few outliers that were not learned well, such as *subliminal message*, and to the limited number of target items in general, making it difficult to draw any strong conclusions.

Another factor that negatively predicted learning was MI. In their study on intentional learning of collocations Nguyen and Webb (2017) found a negative correlation between mutual information (MI) and learning, which they ascribed to the low-frequency bias of MI, because FS that consist of low-frequency words tend to have higher MI scores. In our study too, FS with low-frequency components had high MI scores and were not learned well, possibly because of participants' lack of knowledge of the component words (e.g., *shy away, subliminal message*). Finally, a small correlation was found for corpus frequency. It has been suggested that corpus frequency does not predict L2 knowledge of FS to the same extent as it predicts single word knowledge (González Fernández & Schmitt, 2015), which might explain the small correlation found in our study. However, given the small sample size and the possible effect of prior knowledge, no strong conclusions should be drawn.

3.5.4 Limitations and suggestions for future research

The study reported in this article is based on a small-scale experiment and has a number of limitations. Due to the small sample of participants, the generalizability of our findings is limited. It should also be noted that the participants in our experiment are all intermediate to high-proficiency EFL learners who are used to watching English television in their spare time. Therefore, the findings of this study cannot be generalized to other EFL contexts. Another limitation concerns the limited number of target items. A larger sample of items might decrease the chance of a test effect. Furthermore, the findings of our study might underestimate the amount of learning from the input because the results of the form recognition and meaning recall test were not interpreted. The results of the form recognition test were unreliable due to a guessing effect, which has also been found in other studies for this test format (Peters & Webb, 2018). The results of the meaning recall test may have been distorted due to a test effect. Nevertheless, it is likely that some vocabulary was learned at these two levels of knowledge. Future studies should include a larger sample of participants and target items, in order to be better able to generalize the findings. As previous research has shown, formulaic language is a vast and complex phenomenon that cannot be represented by a sample of 15 FS. Finally, given that formulaic language is not all the same, more fine-grained analyses are needed to further investigate the role of influencing factors such as collocate-node relationship and strength of association.

3.6 Conclusions

The aim of this study was to explore the incidental learning of single words and formulaic sequences through TV viewing. The results of our experiment show that words and formulaic sequences can be learned from exposure to L2 television, even at the level of form recall. Learning gains were affected by prior vocabulary knowledge and item-related factors such as concreteness and collocate-node relationship. Given that formulaic language is ubiquitous in spoken discourse and on television, watching L2 television might be an effective way for learners to expand their knowledge of formulaic sequences outside the classroom. More research is needed to explore the potential of L2 television for incidental learning of FS, and to improve our understanding of the factors affecting the learning process of single words and formulaic language.

Note

1. The Kolmogorov-Smirnov test for normality showed that only the scores on the form recall post-test were normally distributed ($p < .146$). All other data deviated from normality (form recall pretest: $p < .001$, meaning recall pretest: $p = .014$, meaning recall posttest: $p = .041$).

4 Learning formulaic sequences through viewing L2 television and factors that affect learning

Formulaic language, long regarded as a peripheral phenomenon, is now considered an essential aspect of 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 (Boers, Demecheleer, Coxhead, & Webb, 2014; Peters, 2014, 2016; Webb & 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, 2019b), 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.

4.1 Background

4.1.1 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, 2020; Peters & Webb, 2018; Rodgers & Webb, 2020). 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 re- call 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 (2020) 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 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, 2020; Peters & Webb, 2018; Rodgers & Webb, 2020). However, not much is known about the potential of TV viewing for acquiring formulaic language.

4.1.2 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 (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, Demecheleer, 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. 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 (0.90 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* = 0.60 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 non-subtitled 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.

4.1.3 Factors affecting the learning burden of formulaic sequences

4.1.3.1 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 (Durrant & Schmitt, 2010; González Fernández & Schmitt, 2015; Nguyen & Webb, 2017). 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). 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 (2019b) 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.

4.1.3.2 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, 2019b). In a study on deliberate learning of collocations (Peters, 2016), verb-particle combinations (e.g., *head back*) were by far the most difficult 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).

4.1.3.3 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.

4.1.3.4 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 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.

4.2 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, 2019b). 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 non-subtitled 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.

4.3 Methodology

4.3.1 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; M_{age} 5 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, Schmitt, & Clapham, 2001)[1] 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").

4.3.2 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.

4.3.3 Target items

Fifty-six FS occurring in the documentary were selected as target items. These are listed in Appendix 2. The first selection of target items was done quasi-randomly. 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, 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 can be found in Appendix 2. 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 (5 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. The values for all item variables can be found in Appendix 2.

4.3.3.1 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).

4.3.3.2 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.

4.3.3.3 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*).

4.3.3.4 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 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.

4.3.3.5 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.

4.3.4 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' (2019b) 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 (α = 5.85) and posttest (α = 5.87), 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 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 = 5.85) and posttest (α = 5.87).

4.3.5 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.

4.3.6 Procedure

The procedure consisted of four sessions (see Table 4.1). 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 a test effect. After completing the posttests, the second short questionnaire was administered.

Table 4.1 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

4.3.7 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 suggests 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.

4.4 Results

4.4.1 Vocabulary levels test

The descriptive statistics of the VLT are shown in Table 4.2. 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).

Table 4.2 Mean scores and standard deviations (in parentheses) 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)

4.4.2 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 4.3 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*.

Table 4.3 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

4.4.3 RQ1. What is the effect of watching a one-hour TV program on the learning of FS?

4.4.3.1 Form Recall

As can be seen in Table 4.4 and Figure 4.1, 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.

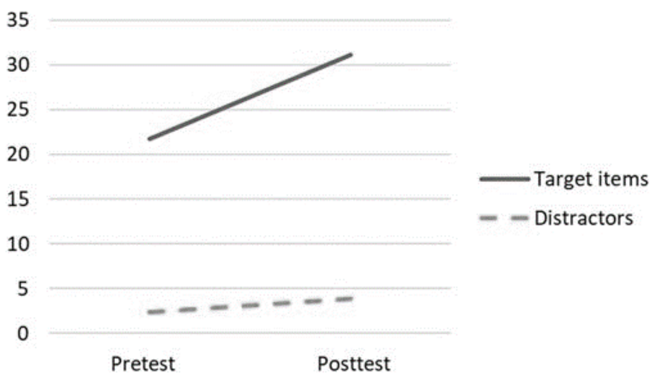

Figure 4.1 Learning gains from form recall pretest to form recall posttest

4.4.3.2 Meaning Recall

At the level of meaning recall, learning gains were made between the pre- and posttests (see Table 4.4 and Figure 4.2). 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 4.2).

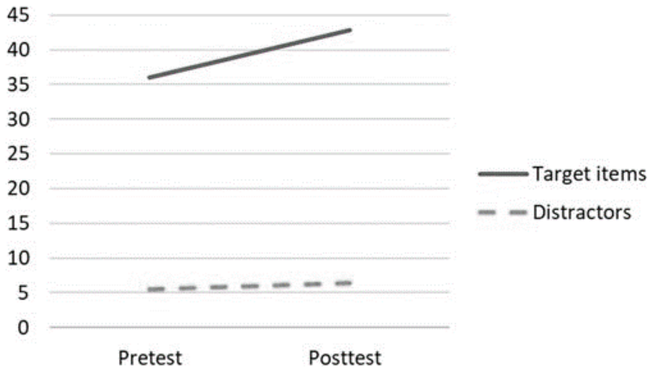

Figure 4.2 Learning gains from meaning recall pretest to meaning recall posttest

The MANOVA with repeated measures (see Table 4.5) 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 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.

Table 4.4 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
Distractors (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

Table 4.5 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	.3	.98
Time*Item	326.5	< .001	.90	1.00
Test*Item	371.76	< .001	.89	1.00
Time*Test*Item	9.9	.003	.2	.87

4.4.4 RQ2. What is the relationship between learner- and item-related factors and vocabulary learning gains?

4.4.4.1 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 4.6. 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, 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.

Table 4.6 GEE: Form recall

Parameter	Wald Chi-Square	*df*	*p*	*B*	Exp(*B*)	95% CI Lower	95% CI Upper
Intercept	86.50	1	< .001	-4.9	0.01	0.00	0.02
Prior vocabulary knowledge (=VLT)	32.17	1	< .001	0.02	1.02	1.01	1.03
Raw 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*).

4.4.4.2 Meaning Recall

A GEE analysis was conducted for 843 observations (5 unknown items in the pretest). The results are shown in Table 4.7. There were fewer observations than in the form recall 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 other collocate-node relationships were taken as the reference category confirmed this result.

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 co-efficient for "figurative idiom." Additional GEE models confirmed that free combinations were learned best, followed by restricted collocations, pure idioms, and figurative idioms.

Table 4.7 GEE: Meaning recall

Parameter	Wald Chi-Square	*df*	*p*	*B*	Exp(*B*)	95% CI Lower	95% CI Upper
Intercept	26.62	1	< .001	-4.97	0.01	0.00	0.04
VLT score	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
Raw 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 Comb.	4.96	1	.026	0.53	1.7	1.07	2.70
Transparency = Restr. Coll.	0.07	1	.796	0.08	1.08	0.6	1.96
Transparency = Fig. 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*).

4.5 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.

4.5.1 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 television. In line with the findings of Puimège and Peters' study (2019b), 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, 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.

4.5.2 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 (2019b), 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 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, 2019b). 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 et al., 2014; Peters et al., 2016; Peters & Webb, 2018). Similar to the findings of Puimège and Peters's (2019b) 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*).[2] This seems to lend evidence to Grant & Bauer (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 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 5.3) 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.

4.5.3 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 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.

4.6 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.

Notes

1. 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.
2. 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."

5 Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study

In order to acquire a large L2 vocabulary, learners need a great amount of exposure to L2 input. However, not all input is converted to intake (Corder, 1967), and attentional mechanisms determine which elements of the input are processed and encoded in memory (e.g., Robinson, 1995; Schmidt, 2001). The mediating role of attention in vocabulary learning from meaningful input has recently been demonstrated in a number of experiments using behavioral measures of attention such as eye-tracking and verbal reports (e.g., Godfroid et al., 2018; Pellicer-Sánchez, 2016). Their findings support the notion that attention is an important, if not essential factor in vocabulary learning from L2 input.

The central role of attention also provides an explanation for the typically small gains found in studies on incidental learning. Although studies have demonstrated that learners can develop lexical knowledge as a by-product of reading (e.g., Horst et al., 1998), listening (e.g., Van Zeeland & Schmitt, 2013), or TV viewing (e.g., Peters & Webb, 2018), incidental learning has been described as a slow, error-prone process in which knowledge develops in small increments (e.g., Webb, 2020). This may be because incidental learning activities do not require learners to notice or elaborately process new vocabulary (e.g., Laufer and Hulstijn, 2001). Attentional mechanisms may also partly explain the low acquisition rate of multiword units, many of which are discontinuous, semantically transparent, and therefore less salient than unknown words (Boers et al., 2017).

A few studies have explored ways of promoting learners' attention to and learning of multiword units through input enhancement (e.g., Choi, 2017). Input enhancement involves instructional methods that render linguistic elements more salient (Sharwood Smith, 1993). Words and phrases can be made more salient through top-down methods of enhancement (e.g., pre-teaching items appearing in the input) or by increasing bottom-up salience (e.g., underlining, boldfacing). Previous research has already demonstrated a relationship between bottom-up enhancement, henceforth textual enhancement (TE), and learning of multiword units from written texts (e.g., Boers et al., 2017; Choi, 2017; Peters, 2012).

Only one study to date has examined the effect of TE on learning multiword units from audiovisual input with captions (= L2 on-screen text) (Majuddin et al., 2021). Since watching L2 television may be an effective way of acquiring knowledge of multiword units (Lin & Siyanova, 2014; Puimège & Peters, 2019b, 2020), it is worth investigating whether TE might also support learning of multiword units from audiovisual input. The use of TE in captioned television has proved effective in single-word acquisition (e.g., Montero Perez et al., 2014, 2015, 2018), and may also be a useful method of promoting learning of multiword units (Majuddin et al., 2021). However, to the best of our knowledge, no study has investigated how TE affects learners' processing of multiword units while watching captioned television.

The present study explores the effect of textually enhanced captions on processing and learning of multiword units while watching L2 audiovisual input. The first aim is to examine the effect of TE on form recall of multiword units. Second, using eye-tracking, we examine how TE (underlining) affects online processing of multiword units, specifically whether TE leads to an increased amount of visual attention. Finally, we aim to investigate the relationship between attention and learning, by linking online processing of multiword units to learning gains in the form recall posttest.

5.1 Background

5.1.1 Incidental learning of multiword units

To achieve fluency in their L2, learners need to acquire a large number of formulaic sequences, that is, frequently recurring words, phrases and word combinations assumed to be familiar and conventional to native speakers (Siyanova-Chanturia & Pellicer-Sánchez, 2019). In SLA research, the term formulaic sequence is commonly used to cover a wide range of constructions such as lexical bundles, collocations, and idioms. The current study focuses on multiword units (MWUs), or formulaic sequences consisting of multiple words. MWUs fulfill many communicative functions and are very widespread in language. However, despite their ubiquity, many L2 learners have difficulty using MWUs (e.g., Laufer & Waldman, 2011). This has been explained in terms of learners' limited exposure to L2 input compared to native speakers. Each individual MWU occurs less frequently than its single-word components, and especially low-frequency MWUs are unlikely to be encountered repeatedly in a short time span (Boers & Lindstromberg, 2009). Learners may therefore lack the necessary amount of L2 exposure to learn the associative links between words (e.g., Durrant & Schmitt, 2010).

Another reason why learners might struggle with MWUs is related to the role of attention and salience in learning from meaningful input. There is some evidence that MWUs can be learned incidentally through meaning-focused activities such as reading (e.g., Pellicer-Sánchez, 2017; Szudarski, 2012; Szudarski & Carter, 2016), reading-while- listening (Webb, Newton & Chang, 2013), and TV viewing (Majuddin et al., 2021; Puimège & Peters, 2019b; 2020). In most studies on incidental learning (for a notable exception, see Szudarski, 2012), relatively brief L2 exposure led to significant improvements in formulaic knowledge. However, learning gains were generally small and depended strongly on factors such as frequency of encounters (e.g., Webb et al., 2013) and semantic transparency (e.g., Puimège & Peters, 2019). Because MWUs vary widely in terms of their semantic and formal properties, not all MWUs will receive the same amount of attention when encountered in L2 input. For instance, MWUs can be highly schematic (allowing for internal variation, e.g., verb-noun collocations) or discontinuous (e.g., *provide information* vs. *provide* some *information*) (Vilkaitė & Schmitt, 2019).

5.1.2 Increasing the salience of multiword units: Textual enhancement

Recent studies (Boers et al., 2017; Choi, 2017; Peters, 2012; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016; Toomer & Elgort, 2019) have used TE to promote learners' noticing of unknown MWUs in L2 written input. These studies all found a positive relationship between TE and incidental learning, but usually only when knowledge of form was measured. To give one example, Szudarski and Carter (2016) found superior learning outcomes for enhanced collocations in a form recognition and a form recall test, but not in a meaning recognition test. The positive effect of TE on form learning suggests that drawing learners' attention to the target forms, and the lexical makeup of the MWUs, may result in a more durable memory trace (Boers et al., 2017). However, increased attention to form does not necessarily result in retention of semantic knowledge, which requires a greater level of analysis or more elaborate processing (e.g., Leow & Martin, 2017).

One study by Choi (2017) used eye-tracking to measure learners' online processing of unknown MWUs during reading in a textually enhanced (boldfaced) and an unenhanced condition. Advanced learners of English (L1 = Korean) read a text containing 14 semantically transparent phrases. Learners in the enhanced group performed significantly better on a posttest of form recall than learners in the baseline text group, who read the same text without enhancement one week after the experiment. Choi also found an interaction between pretest knowledge and TE in predicting learners' visual processing of the collocations. Items that were not known in the form recall pretest received more fixations and were fixated longer in both conditions, but more so in the TE group. This finding suggests that TE helped learners notice MWUs that were unknown. However, results on a separate cued recall test indicated a trade-off between attention to the enhanced and unenhanced content in the reading passage, with lower retention of unenhanced information in the TE group.

The results of previous research indicate that TE can be used to promote learners' attention to unknown MWUs. All of the aforementioned studies employed TE in written texts, but little is known about the use of TE to increase the salience of MWUs in multimodal input, e.g., television with captions.

5.1.3 Vocabulary learning from audiovisual input

There has been a recent surge in research investigating L2 vocabulary acquisition from audiovisual input (see, e.g., Peters & Webb, 2018). There is growing evidence that vocabulary learning can occur from watching L2 television (e.g., Feng & Webb, 2020; Peters & Webb, 2018). Further, captions, or subtitles in the language of the audio track, have been found to promote vocabulary acquisition from TV viewing (e.g., Peters, 2019; Peters et al., 2016; Pujadas & Muñoz, 2019).

A number of studies have demonstrated that enhanced captions can increase the salience of unknown vocabulary in audiovisual input (Cintrón-Valentín et al., 2019; Montero Perez et al., 2014, 2015, 2018). For example, Montero Perez et al. (2015) used vocabulary tests and eye-tracking to measure the relationship between learners' attention to, and learning of unknown vocabulary for two types of audiovisual input (full captions, keyword captions). The keyword captions group outperformed the full captions group in a form recognition test, but not in a form-meaning test. The authors only found a significant positive relation- ship between eye movement indices (second pass time, total reading time) and form recognition in the full captions group. They concluded that their eye-tracking measures may only have captured low-level attention, but not elaborate processing of meaning.

We should mention that some of the assumptions underlying eye-tracking research in reading may not be directly applicable to captioned audiovisual input, due to its multimodal, dynamic nature (for eye-tracking research on learners' caption use, see, for example, Bisson et al., 2014; Gass et al., 2019; Muñoz, 2017; Winke et al., 2013). First, the relationship between reading times and L2 learning may be very different when processing is constrained by the pace of the video. Further, due to the overlap between written and spoken input, processing may happen in various input modes simultaneously. This means that eye movement data in captioned audiovisual input have to be interpreted with caution and may not always be comparable to findings in reading research.

To the best of our knowledge, only one study has investigated the effect of TE on learning MWUs from captioned television. Majuddin et al. (2021) examined the effect of two types of captions on learners' recall of MWUs. Participants were divided into six groups, based on the number of repetitions (viewing the same video once or twice) and captioning type (normal captions, enhanced captions, and no captions). In the enhanced condition, MWUs were bolded and underlined. A positive relationship was found between captioning and form recall in a cued gap fill test, with higher gains made from pre- to posttest in the enhanced and unenhanced captions groups compared to the no- captions group. However, enhancing the items did not further increase learning gains. The authors argued that the real-time nature of video (as opposed to written text), as well as the length of the MWUs (some of their target items contained five words), might explain why TE did not have a noticeable effect on learning.

5.2 Rationale and research questions

Previous research has shown that TE can be used to draw learners' attention to unknown vocabulary in captioned television (e.g., Montero Perez et al., 2015). However, still little is known about how TE affects learners' processing of MWUs during TV viewing. Current evidence suggests that TE may not support learning of MWUs from captioned audiovisual input, possibly due to the brief, fleeting presentation of the items (Majuddin et al., 2021). A closer examination of learners' attention to unknown MWUs with and without TE could provide more insight into how learners interact with enhanced items in captioned video. This might in turn improve our understanding of how TE can be used in different input modalities to support learning.

The present study examines in what way TE affects learners' processing and learning of MWUs when watching captioned television. First, we use eye-tracking to examine how learners' visual processing of MWUs in captioned television is affected by TE. A second aim is to study the relationship between TE, visual attention, and learning gains. We aim to determine whether TE and attention are positively associated with learners' performance on a form recall posttest. The current study addresses the following research questions:

1. Does TE affect learners' visual attention to MWUs in the captions of audiovisual input?
2. Do TE and visual attention affect form recall of MWUs encountered in captioned audiovisual input?

5.3 Method

5.3.1 Participants

Thirty Flemish students (L1 = Dutch, M_{age} = 22, 7 males, 23 females) were recruited for the experiment. Participants were enrolled in an applied language studies program (with the exception of four students from other programs). Sixteen participants were majoring in English. The lexical profile of the documentary (see Materials) suggests that knowledge of the 3,000 most frequent English words was necessary to reach 95% coverage of the documentary, which is assumed to be sufficient for adequate com- prehension in television (Durbahn et al., 2020; Webb & Rodgers, 2009). We excluded data from four participants who did not reach a score of 27/30 on the 3K level (3,000 level) of the Vocabulary Levels Test (VLT; Schmitt et al., 2001), leaving a sample of 26 participants. Participants' results per frequency band in the VLT are presented in Table 5.1. The students either received course credit or monetary reimbursement for participating.

*Table 5.1 Mean vocabulary levels test scores and standard deviations per frequency band (*n *= 26)*

Frequency band	*M*	*SD*
2K	29.5	0.76
3K	28.92	0.80
Academic	28.96	1.15
5K	27.38	1.81
10K	17.42	5.89
Total	**132.19**	**8.14**

5.3.2 Materials

Video and captions. As audiovisual input we used the first 30 minutes of an episode of Fry's Planet Word (BBC), a documentary series about language. The episode 'Uses and abuses', about swear words, slang and taboos, was chosen for its relevance to language students, to ensure that participants would pay attention to the content. Running the script of the documentary through the Compleat Lexical Tutor (Cobb, n.d.) showed that 95% of the running words in the text belonged to the 3,000 most frequent English word families. Captions were taken from the original DVD track using Aegisub, and were adapted to a word-for-word transcription of the spoken input. Of the 549 captions appearing in the video, 35 (6%) were presented across two lines. Spacing between the two lines was 10 mm. Average caption duration was 3,285 ms (*SD* = 1,088 ms). Each caption contained 9 words on average (*SD* = 3).

Target items. Multiword units were taken from Puimège & Peters' (2020) study, who used the same viewing material. From their sample, we selected 22 items from a wide range of MWUs, including idioms (e.g., *beyond the pale*), collocations (e.g., *supernatural powers*), compounds (e.g., *guinea pig*), and phrasal verbs (e.g., *tap into*). All items have an MI score higher than 3 in the Contemporary Corpus of American English (COCA), which is a commonly used cut-off of collocation strength (e.g., Durrant & Schmitt, 2010). MI reflects how strongly two words attract by comparing their co-occurrence rate in a corpus to their co-occurrence rate by chance (Schneider, 2018). Semantic decomposability was rated on a seven-point scale by 33 learners from the same study program as the participants. Raters were given the following question in English: 'How easy is it to guess the meaning of these phrases based only on their single-word components, on a 7-point scale?: 1 = impossible, 7 = very easy'. They were instructed to only rate items for which they knew the meanings of both single word components. This resulted in missing values for some items (minimum number of ratings was 18 for *racial epithet*). Because of the similarity in participant profile in the current study and in Puimège and Peters' (2020) study, we discarded items that were known in the pretest by 80% of their sample of participants. To reach a sample of 28 items, we added low-frequency MWUs not appearing in Puimège and Peters' (2020) sample. Two target items (*took umbrage, abusive language*) appeared in two-line captions. All of the MWUs appeared only once in the input. The full list of items can be found in Table 5.2. Information on item variables (corpus frequency, mutual information, and semantic decomposability) can be found in Appendix 3.

Experimental conditions. To represent the two experimental conditions, a counterbalanced within-participants design was adopted so that all participants were exposed to both conditions (enhanced and unenhanced) and all items appeared in both conditions in the full data set (Godfroid, 2020; Nicklin & Vitta, 2021). Participants were allocated to one of two versions of the same video. Items that were enhanced in version 1 were unenhanced in version 2, and vice versa. By adopting a counterbalanced design, we could ensure that there was no confound between our treatment and knowledge of constituent words or linguistic complexity of the items (see also Godfroid, 2020 on the value of within-participant designs in eye-tracking research). Instead of dispersing or alternating the conditions chronologically, for instance by underlining every other item in the captions, we split each version of the video into an unenhanced and an enhanced part (see Table 5.2). Participants assigned to version 1 saw the first 14 items underlined, and participants assigned to version 2 saw the last 14 items underlined. This approach was chosen to avoid an attentional trade-off effect. In a study on L2 reading (Choi, 2017), TE was not just associated with increased attention to enhanced items, but it also led to decreased attention to unenhanced information appearing near enhanced items. Splitting the video in two parts could ensure that a greater amount of attention in the TE condition would reflect enhanced attention compared to attention in an unenhanced captioned video.

5.3.3 Form recall test

Learning of MWUs was measured by means of a form recall test. We used only one measure of knowledge to avoid a learning effect from one test to another, as was found in previous studies (e.g., Puimège & Peters, 2020). The form recall format was chosen to minimize an effect of the pretest on learners' attention to target items during the experiment. Participants had to provide the form of the English MWUs based on a Dutch translation. The first letter of each single word component was given to avoid elicitation of other plausible word combinations. Participants were asked to give the written and spoken form of each item. Spoken responses were recorded. Participants were also asked to provide single word components where they could, even if they did not know the full MWUs. To make sure that learning could be ascribed to the treatment, 30 MWUs from Puimège and Peters' (2020) study which did not appear in the part of the documentary used in the current study were included as distractor items. Learning of these items would signal potential test effects (e.g., guessing based on the first letter of the single word components) or learning outside the treatment (e.g., looking up MWUs at home). Two versions of the test were made in which the items appeared in a different, semi-random order. Both versions included all target items. Participants who completed version 1 in the pretest, received version 2 in the posttest, and vice versa. The test was not timed and took on average 18 minutes to complete. The full tests are available in Appendix 3.

Examples from the form recall test:

lichaamssappen (bloed, zweet, enz.) = b _ _ _ _ _ f _ _ _ _ _ (*bodily fluids*)

een spier verrekken = p _ _ _ a m _ _ _ _ _(*pull a muscle*)

5.3.4 Procedure

One week before watching the documentary, participants filled out an informed consent form and completed the form recall pretest and the VLT. They were not forewarned of a vocabulary posttest, nor that any of the MWUs in the pretest would appear in the video. The eye-tracking experiment was conducted using an Eyelink Portable Duo, Version 1.0.2 (SR Research). The video was presented on a 1280 × 1024 monitor with a refresh rate of 60 Hz. Display dimensions were 190 × 330 mm. Captions appeared in Arial (proportional), with a character size of 38pt, corresponding to approximately 5 mm on the monitor and 0.39° of visual angle.[1] Participants were seated in front of the monitor at a viewing distance of 720 mm. A desk-mounted chin rest was used to stabilize head position. Participants' dominant eye was tracked. Sampling rate was 500–2,000 Hz.[2] After setup, a nine-point calibration and validation procedure was performed. The maximum calibration error was 0.8° of visual angle. To keep track of accuracy during the experiment, the 30-minute documentary was split at scene changes into seven short (3–6 minute) video clips, each representing a separate trial in the experiment. The target items were distributed between the shorter videos as shown in Table 5.2.

Table 5.2 Distribution of target items across the input in the two counterbalanced conditions

Video	Target items	Condition version 1	Condition version 2
Video 1	*highest echelons, foul language, bodily fluids, common denominator, supernatural powers, sexual depravity, mutual pleasure, heck of a lot, pass over (into)*	enhanced	unenhanced
Video 2	*tell off*		
Video 3	*tap into, fair description, end up, evolutionary advantage*		
BREAK			
Video 4	*abusive language, guinea pig, spark your interest, unleash a torrent of, subliminal effect*		
Video 5	*take into account, pain relievers*	unenhanced	enhanced
Video 6	*jab line, turning point, win the right*		
Video 7	*beyond the pale, sheer coincidence, racial epithet, take umbrage*		

At the start of the experiment, participants watched a practice video with captions, to let them adjust to the experimental procedure and to correct their behavior (e.g., head movements). After each trial, a drift check was performed and the calibration procedure was repeated when necessary. A 5-minute break and recalibration was inserted for each participant after the first 4 videos (approximately 15 minutes). This is also where the captions changed from enhanced to unenhanced or vice versa.

Timed interest areas were created for the 28 MWUs as a whole, and their single word components. Each interest area remained on screen for the duration of the caption in which it appeared (*M* = 3,523 ms, *SD* = 849 ms). Margins of approximately 6 mm were added at the top and bottom of each interest area. Interest areas for target items had an average size of 60 × 25 mm. Five eye-tracking measures were used to examine online processing of MWUs: first pass reading time, rereading (binary measure), rereading time, single-word skipping (binary measure), and total reading time. First pass reading time is a durational measure which sums all fixations during the first visit, before the eye gaze leaves the interest area. It captures early stages of processing, and may be sensitive to low-level visual, orthographic and frequency-related factors (Conklin et al., 2018; Godfroid, 2020). Textual enhancement could therefore be expected to affect first pass reading time, although previous studies investigating the effect of TE on grammar learning did not find such an effect (Lee & Révész, 2018, 2020). Rereading often results from processing difficulties related to comprehension or contextual integration (Conklin et al., 2018; Godfroid, 2020), and might also be affected by caption duration. We analyzed rereading both as a binary event and as a durational measure. Our analysis of binary rereading captures the odds of rereading an item, rather than the amount of time spent on rereading. Rereading time sums all fixation durations in an interest area after the first pass. We included a binary variable for single-word skipping, to examine whether participants fixated both single-word components of each MWU. This measure was included because TE might cause learners to distribute their attention more evenly across both words in a MWU. Final-word skipping has been interpreted as an indication of more fluent reading (see for example Carrol & Conklin, 2019), but in the context of the current study, single-word skipping could also reflect the amount of attention to the lexical composition of a MWU. Total reading time was included as a late measure encompassing both first pass reading time and rereading time. This measure was included because it has produced strong associations with learning gains in previous studies (e.g., Godfroid et al., 2018).

After watching the documentary, participants completed a short questionnaire about the content of the video and about their general viewing habits. The questionnaire can be found in Appendix 3. The questionnaire was immediately followed by the form recall post-test. After the posttest, participants were interviewed about their explicit recall of target items, based on their responses on the form recall test. This was done to gain more insight into participants' conscious noticing of the target items, and to check if learners had guessed in the form recall test, or had learned items outside the experiment. The interviews were not recorded, but the interviewer took notes which were used to help interpret the quantitative results of eye-tracking measures and learning gains. Finally, participants completed another short questionnaire about their awareness of the purpose of the experiment and a potential effect of the pretest on their conscious viewing behavior.

5.3.5 Data preprocessing and cleaning

Eye movements were parsed according to the default cognitive configuration of Eyelink. Following Godfroid and Hui (2020), the output of the algorithm was visually inspected in the DataViewer Temporal Graph Trial View. The initial pool of eye-tracking data contained 728 data points at phrase level (28 MWUs × 26 participants). Trials where track loss occurred were removed (10%). Data were cleaned using the default settings in the four-stage cleaning procedure of Eyelink DataViewer. Items with a fixation time of 0 ms for the full phrase (not at word level) (n = 60) were excluded from the analyses because they would lead to skewed reading times, even after log transformation.[3] The analysis of rereading time also excluded zeroes (n = 400) to reduce skew.

5.3.6 Scoring and analysis

5.3.6.1 Eye-tracking measures as outcome variable

In all analyses, the binary factor TE was the main independent variable of interest. Form recall pretest score (binary), phrase frequency per million, mutual information (MI), length, caption duration, frequency of the least frequent single word component, and semantic decomposability were entered as control variables. Any continuous variables that were not normally distributed, were log transformed with base 2. Continuous control variables were also centered around the grand mean.

For the analysis of first pass reading time, rereading time, and total reading time, linear mixed-effects models were fit using the lmer() function in the package lme4 (Version 1.1-21) in R (Version 3.6.1). Because distributions of these measures were positively skewed, the data were log transformed. In each of the mixed effects models, the same procedure was followed. First, a null model was constructed containing only random intercepts for item and participant. Fixed effects and an interaction term for TE and pretest score variables were then added to the model. Any non-significant variables that did not improve the model fit were removed one by one. Model fit was estimated through log-likelihood ratio tests and comparison of AIC values. The final model was the most parsimonious model (i.e., with the fewest covariates) with the lowest AIC value. Restricted Maximum Likelihood was used for model fitting. Because our regression models for different eye-tracking measures could be said to test the same hypothesis (Godfroid & Hui, 2020; von der Malsburg & Angele, 2018), we applied a Bonferroni adjustment, $\alpha = 0.01$. After adding the fixed effects, we added a random slope for TE at participant-level, and a correlation between the random slope and the random intercept. If the random slope (+ correlation) did not improve the model fit, it was removed. Finally, a sensitivity analysis was performed to check the influence of outliers (Godfroid, 2020).

Each lmer() model was rerun without outliers (studentized residuals with an absolute value higher than 2.5). For the binary outcome variables (skipping, rereading), generalized linear mixed models were fit using the glmer() function in the lme4 package.

5.3.6.2 Form recall as outcome variable

In the form recall test, items that were pronounced or spelled correctly, received a score of 1, incorrect items received a score of 0. Half of the tests were scored by a second rater. Interrater agreement was 98% for both the pre- and posttests. For the other 2%, the first rater revisited the test responses to make sure the criteria described above were applied correctly. Remaining disagreements were solved through discussion. To analyze learning gain at the item level, the glmer() function in the lme4 package in R was used. First, models were fit to analyze the effect of the treatment on form recall, by comparing learning gains for target items and distractor items. Main effects for time (pretest vs. posttest) and item type (target item vs. distractor) were entered into the first model, as well as an interaction term between these two variables. We also analyzed learning of the target items and distractors in two separate models. The main analysis included only the binary posttest scores for target items. Items that were known in the pretest, and distractor items, were excluded from the analyses, leaving 484 data points. In this analysis, TE and the eye-tracking measures were the main independent variables of interest. Because previous studies did not include any eye-tracking measures, we first fit a model with only TE as independent variable. Then, total reading time was added to the model. Control variables were phrase frequency, MI score, length, caption duration, and decomposability. Participants' score on the VLT could not be added as a fixed effect, because its inclusion led to inflated odds ratios and convergence problems, possibly due to the low number of unique values for this predictor (n = 16).

5.4 Results

5.4.1 Eye-tracking measures as outcome variable

To find out if TE affected reading of MWUs encountered in the input (see research question 1), we first analyzed the eye-tracking measures. The descriptive results are summarized in Table 5.3.

The Bonferroni-adjusted results of the mixed effects models (see Tables 5.4 and 5.5) indicate that TE was associated with significantly longer first pass and total reading times, as well as less single-word skipping. Enhancement also led to higher odds of rereading, but did not significantly predict rereading time. Pretest knowledge was a significant predictor of rereading time, total reading time, and single-word skipping. Items that were not known in the pretest tended to receive longer reading times. Item length and decomposability predicted first pass reading time and total reading time. Longer and less decomposable items received longer reading times, particularly during the first pass. Caption duration predicted rereading time, binary rereading, and total reading time. Items that were unknown in the pretest and stayed on screen longer were more likely to be reread and had longer reading times, particularly after the first pass. Mutual information, frequency of the full MWU and of the least frequent component predicted single word skipping, with greater odds of skipping for higher-frequency and less strongly associated MWUs. The effect of VLT score was not significant for any of the eye-tracking measures.

Table 5.3 Means and standard deviations (in parentheses) for eye-tracking measures, per condition

	Unknown in pretest		Known in pretest		All	
	unenhanced	enhanced	unenhanced	enhanced	unenhanced	enhanced
FPR	471.6 (332.3)	662.5 (466.5)	371.0 (217.9)	494.1 (384.4)	439.5 (303.8)	604.1 (446.4)
RRT	631.9 (401.1)	587.4 (401.5)	385.4 (225.4)	521.8 (390.7)	562.4 (376.3)	568.4 (398.1)
TRT	725.9 (468)	968.8 (500.3)	501.2 (278.7)	702.8 (476.2)	654.2 (429.5)	876.4 (507.3)
Skipping	.17 (.1)	.08 (.05)	.10 (.05)	.09 (.06)	.27 (.10)	.17 (.07)
RRR	.14 (.06)	.17 (.08)	.05 (.05)	.07 (.05)	.19 (.07)	.24 (.09)

Note. FPR = First pass reading time, RRT = rereading time, TRT = total reading time, Skipping = rate of single-word skipping, RRR = rereading rate. Rates for skipping and rereading were calculated by dividing the number of items that were reread, or in which one word was skipped, by the total number of fixated items (Conklin et al., 2018).

Table 5.4 Best fitting models for the continuous eye-tracking measures

	First pass time (n = 587)				*Rereading time* (n = 254)				*Total reading time* (n = 590)			
Fixed effects	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Level 1												
Intercept	8.5**	0.07	116.72	< 0.001	8.93**	0.09	98.74	< 0.001	9.08**	0.09	102.89	< 0.001
TE	0.35*	0.1	3.59	0.002					0.54**	0.09	5.82	< 0.001
Pretest					-0.35*	0.13	-2.64	0.01	-0.32**	0.08	-3.74	< 0.001
MI	0.16	0.07	2.39	0.025					0.13	0.07	2.05	0.053
Length	0.66**	0.11	5.97	< 0.001					0.52**	0.11	4.81	< 0.001
Decomp.	-0.44**	0.11	-3.98	< 0.001					-0.39*	0.12	-3.63	0.002
Duration					0.59*	0.19	3.09	0.006	0.46*	0.14	3.24	0.004
Level 2												
VLT score	-1.79	0.7	-2.54	0.018	-0.51	0.89	-0.57	0.571				
Random effects	*Variance*	*SD*			*Variance*	*SD*			*Variance*	*SD*		
(1 | item)	0.04	0.04			0.03	0.18			0.03	0.18		
(1 | subject)	0.05	0.05			0.06	0.24			0.1	0.31		
TE	0.14	0.37							0.11	0.33		
Residual	0.56	0.75			0.72	0.85			0.56	0.75		
Marginal /Cond. R^2	0.18 / 0.36				0.08 / 0.18				0.22 / 0.39			
AIC	1437.7				678.18				1449.05			

Note. ** $p < .001$, * $p < .01$. Level 1 = item level, level 2 = participant level. The reference level for textual enhancement was “no enhancement”, the reference level for pretest score was “no pretest knowledge”.

Table 5.5 Best fitting models for the binary eye-tracking measures

	Single word skipping (n = *595)*				*Rereading* (n = *595)*			
Fixed effects	*B*	*SE*	*z*	*p*	*B*	*SE*	*z*	*p*
Intercept	-0.04	0.21	-0.21	0.835				
TE = 1	-0.96**	0.2	-4.84	< 0.001	0.63**	0.18	3.49	< 0.001
Pretest score = 1	0.66*	0.27	2.48	0.01				
Corpus frequency	0.16*	0.06	2.59	0.009				
SW frequency	0.15*	0.05	2.82	0.005				
MI	-0.84**	0.24	-3.54	< 0.001				
Duration					1.20*	0.39	3.13	0.002
Random effects	*Variance*	*SD*			*Variance*	*SD*		
(1 l item)	0.56	0.75			0.25	0.5		
(1 l subject)					0.15	0.39		
Adjusted ICC / conditional ICC	0.15 / 0.11				0.11 / 0.10			
AIC	689.85				778.85			

Notes. ** $p < .001$, * $p < .01$. SW frequency = frequency of the least frequent single word component.

5.4.2 Form recall as outcome variable

In the second part of the analyses, we examined participants' scores on the form recall test, to see if (a) learning had occurred from pre- to posttest, (b) learning could be ascribed to the treatment, and (c) a relationship could be found between learning gains, TE, and amount of attention (research question 2). Scores on the form recall tests are summarized in Table 5.6.

The results of the first mixed effects model show that there was a significant main effect of time on the binary outcome variable form recall score ($B = 1.19$, $SE = 0.16$, $z = 7.5$, $p < .001$), indicating that items were more likely to be known in the posttest than in the pretest. The interaction between time (reference level = pretest) and item type (reference level = target item) was also significant ($B = -0.68$, $SE = 0.21$, $z = -3.22$, $p = 0.001$). Predicted probability of form recall knowledge was similar for target and distractor items in the pretest, but significantly higher for target items in the posttest. The main effect of item type was not significant ($p = .82$). In two additional models, the effect of time was analyzed separately for target items and distractors. The results confirm that the effect was stronger for target items ($B = 1.2$, $SE = 0.16$, $z = 7.48$, $p < 0.001$) than for distractors ($B = 0.55$, $SE = 0.15$, $z = 3.75$, $p < 0.001$). The results of the three models are reported in full in the Appendix 3.

We took the significant interaction between time and item type as evidence for learning from watching the captioned video, and went on to analyze the relationship between TE and learning by comparing learning gains between the enhanced and unenhanced target items. Results of the second mixed effects model show that TE significantly predicted learning from pre- to posttest, with greater odds of learning in the enhanced condition than in the unenhanced condition ($B = 0.66$, $SE = 0.26$, $z = 2.51$, $p = .01$). When total reading time was added to the model, TE was no longer significant. In the final model, summarized in Table 5.7, total reading time significantly predicted learning. A doubling in total reading time (the variable was log transformed with base 2) was associated with 63% (Exp(*B*) = 1.63) higher odds of learning an item. Other variables that significantly predicted learning were item length, caption duration, and decomposability.

Table 5.6 Mean scores, standard deviations (in parentheses), and gains on the form recall tests

	target items			distractors
	unenhanced (max. = 14)	*enhanced* (max. = 14)	*all* (max. = 28)	(max. = 30)
pretest score	4.58 (1.90)	4.81 (2.26)	9.39 (3.01)	10.77 (4.17)
posttest score	6.42 (2.28)	6.81 (2.73)	13.23 (3.35)	12.69 (4.43)
absolute gain	1.85 (1.57)	2 (1.33)	3.85 (2.05)	1.92 (1.52)
normalized gain	.20 (.16)	.24 (.20)	.21 (.11)	.10 (.09)

Note. Normalized gains were calculated using the following formula: (post - pre)/(total number of test items - pre) (Horst et al., 1998).

Table 5.7 Best fitting model for form recall (n *= 397)*

Fixed effects	*B*	*OR*	*SE*	*z*	*p*
Intercept	-1.40***	0.25	0.24	-5.73	< 0.001
Total reading time	0.49**	1.63	0.17	2.89	0.004
Duration	-1.24*	0.29	0.61	-2.02	0.04
Length	-1.61**	1.2	0.51	-3.14	0.002
Decomposability	2.25***	9.49	0.54	4.15	< 0.001
Random effects			*Variance*	*SD*	
(1 | item)			0.64	0.8	
(1 | subject)			0.22	0.47	
Adjusted ICC / conditional ICC	0.21 / 0.17				
AIC	392.26				

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. The reference level for textual enhancement was "no enhancement", the reference level for pretest score was "no pretest knowledge".

5.4.3 Questionnaires and interview data

A questionnaire concerning learners' TV viewing habits and input comprehension was used to check for obvious comprehension issues with regard to the content of the episode. This helped us identify participants who had trouble understanding the input (see Section IV.1). The questionnaire also showed that all participants were used to watching subtitled or captioned television.

A second questionnaire was used to check if the pretest had affected learners' conscious attention to target items during the experiment. Participants were also asked if they had looked up any of the target items after the pretest. The results showed that the majority of participants (n = 16) had been aware while watching the documentary that some items had appeared in the pretest. This was true for both enhanced and unenhanced items. Further, almost half of the participants (n = 12) indicated that this had made them pay more conscious attention to MWUs.

Finally, notes taken by the researcher during the interviews were examined to gain more insight into learners' noticing of items that were learned from pretest to posttest. A few trends emerged from these data. First, all participants reported that they remembered seeing items underlined in the input. However, most of the time, participants could not remember whether a specific item had been underlined. Often participants could not explain why they had been able to recall an item in the posttest. In some cases, participants reported that they had guessed the correct response in the form recall test based on the first letters of the single word components. For example, one participant had guessed the distractor *white lie* in both the pretest and posttest. Some participants reported that they already knew items before the experiment, but had not been able to recall them in the pretest.

5.5 Discussion

5.5.1 Research question 1: Does TE affect learners' attention to MWUs in captioned audiovisual input?

To address the first research question, we used five eye-tracking indices measuring learners' processing of 28 MWUs encountered with or without TE in the captions of a 30-minute documentary. Eye-tracking measures in the two conditions were compared using mixed effects models. The results show that participants spent significantly more time fixating on target items that were underlined in the captions. Enhanced items were also more likely to be reread, although rereading time was not significantly correlated with TE. The odds of fixating both single word components were also higher for enhanced items.

The positive relationship between TE and first pass reading time suggests that underlined items caught learners' attention, leading to increased initial processing. This interpretation is supported by the results of the questionnaires and interviews, in which participants reported noticing underlined items in the input. It is important to note that studies investigating learners' attention to grammatical structures in captioned input did not find a significant effect of TE on first pass reading time (Lee & Révész, 2018, 2020). A possible explanation is that analyzing a grammatical structure in parallel with sentence comprehension may be more cognitively demanding than processing a MWU during the first pass. The effect of TE may therefore occur later for grammatical structures (Alsadoon & Heift, 2015; Lee & Révész, 2018, 2020; Winke, 2013). It would be interesting to further investigate how learning burden interacts with TE in different stages of processing.

TE also significantly predicted rereading. Learners were more likely to revisit enhanced items, suggesting that the salience-raising effect of TE led to increased reanalysis. Despite its effect on rereading, no significant relationship was found between TE and rereading time. This may be due to the limited amount of time learners had to read the captions. Visual inspection of the eye movement data confirmed that, in many cases, learners could only briefly revisit the target items, limiting the opportunity for late processing. Caption duration was a significant predictor of both rereading measures, which indicates that rereading time depended first and foremost on how long an item remained on screen.

Total reading time was also positively associated with TE. This is perhaps not surprising, given that the measure incorporates first pass reading time and rereading time, and correlated with both of these measures (total reading time and first pass reading time: r = .61, $p < .001$; total reading time and rereading time: r = .73, $p < .001$). Total reading time was also associated with all variables that predicted first pass reading time and rereading time.

Finally, single-word skipping was affected by TE as well as by frequency-related variables (item frequency, mutual information, and single word frequency). In line with findings in unimodal reading (e.g., Rayner, 1998), highly frequent component words (e.g., *into, off*, . . .) tended to be skipped more often. The effect of mutual information suggests that a weak association between the words in a MWU was also likely to result in skipping. TE was associated with increased odds of reading both words in the MWU, which suggests that TE led to increased visual attention to the written form of the entire phrase.

Taken together, the results for the duration measures indicate that learners spent more time fixating underlined target items, but due to time limits, they mainly processed them during the first pass. It is important to mention that our analysis only focused on visual attention, although learners could also process the target items in spoken form. We need to be careful in drawing strong conclusions about early and late processes in caption reading, as the assumptions about different stages of processing may not hold when auditory input is presented simultaneously (see for example Conklin et al., 2020). For example, we did not analyze to what extent learners integrated information from both input modalities, and how this may have affected word identification and meaning integration. It is possible that the auditory support changed the degree or quality of processing of the target items. Caption duration was an important predictor of reading times, but exactly which elements of word processing could occur in the limited presentation time of the captions remains open for investigation. Finally, in line with previous findings (Choi, 2017), our results show that pretest knowledge of the MWUs affected reading times. Items unknown in the pretest tended to be fixated longer, which suggests that the novelty of certain phrases may have rendered them more salient. Novel words have been shown to attract attention in reading (e.g., Godfroid et al., 2013), and our results suggest that novelty may affect multiword processing as well. However, many of the MWUs contained low-frequency words (e.g., *epithet, depravity*), which may have contributed to this novelty effect. Another explanation for the effect of pretest knowledge could be that learners paid special attention to previously (partially) unknown items because they remembered them from the form recall pretest. In the form recall test, we found that knowledge improved for some of the distractor items, indicating learning from the pre- test itself. In addition, participants reported that the pretest caused them to pay closer attention to certain MWUs, sometimes in anticipation of an (unannounced) posttest. These findings indicate that pretesting can affect learners' engagement with linguistic items during the learning treatment, and may, to some extent, even undermine the construct validity of incidental learning, which is often defined in terms of a primary focus on meaning (Swanborn & de Glopper, 1999). In the current study, the pretest (in addition to TE) may have enhanced learners' attention to MWUs in the input, even with a one-week interval between the pretest and the treatment.

5.5.2 Research question 2: Do TE and visual attention affect form recall of MWUs encountered in captioned audiovisual input?

To examine the effects of TE and visual attention on learners' recall of unknown MWUs encountered in the input, we used mixed effects logistic regression. The results show that TE only contributed to learning odds when reading times were not taken into account. Once total reading time was entered into the model, the effect of TE was no longer significant. This suggests that the amount of (visual) attention was more important than the experimental manipulation, or that the effect of the learning intervention depended on how learners engaged with the items in the input. As discussed above, TE can clearly affect visual processing, and therefore has the potential to promote vocabulary learning. However, remembering the form and meaning of MWUs from a single exposure is likely to rely on the degree and quality of processing, which cannot be controlled directly by means of TE (see also Leow & Martin, 2017). Other variables (such as the inclusion of a pretest in the current study) may contribute to depth of processing as well.

Nevertheless, the significant relationship between TE and learning gains before total reading time was entered into the regression models confirms that TE can support incidental learning of MWUs under certain conditions (e.g., Choi, 2017). In Majuddin et al.'s (2021) study, higher average scores were found for enhanced captions compared to unenhanced captions, but the difference did not reach significance when other variables (number of viewings, Vocabulary Size Test score) were taken into account. The researchers offered a number of explanations why the effect of TE might not be as outspoken in audiovisual input, such as the limited amount of time for caption reading, the length of the MWUs in their experiment (which included items of 5 words), and the distribution of attentional resources between the captions and imagery. These explanations are consistent with the findings of the current study, and they reveal the complex and multifaceted nature of audiovisual input. In line with studies that emphasize the importance of cognitive load in multimodal processing (e.g., Gass et al., 2019), the effectiveness of enhanced captions for learning MWUs is likely to depend on variables related to the input and the learners. It appears that TE might benefit learning provided that learners can fluently read the L2 captions, and distribute their attention efficiently.

Overall, our findings suggest that TE can promote learners' attention to unknown MWUs encountered in captioned video, but that engagement with the input more strongly affects learning than TE by itself. Further, because semantic decomposability played an important role in predicting learning gains, we cannot draw strong conclusions about form-meaning mapping. Our results show that participants could remember the form of MWUs that were semantically transparent, such as evolutionary advantage. However, MWUs with low decomposability, which in our sample were generally idiomatic items or items containing low-frequency single words (e.g., *beyond the pale*), may require more contextual support and possibly also longer exposure time to allow for more elaborate semantic processing. The small gains found for non-decomposable items suggest that TE alone was insufficient to overcome the high learning burden of semantically less accessible MWUs from a single exposure in captioned video. One unexpected finding was a negative correlation between caption duration and learning odds. This finding may seem counterintuitive, especially as longer reading times led to higher gains. We cannot offer a clear explanation based on our data, but it is possible that auditory and visual information were integrated differently in shorter versus longer captions. In multimodal input, L2 learning may not just rely on the amount of (visual) attention to relevant information, but also on the way in which information from different input modalities is combined in memory. Currently, there is hardly any research that has investigated the effects of auditory processing on reading patterns and learning gains in multimodal input with moving imagery (for an exception, see Wisniewska & Mora, 2018). This could be an interesting avenue for future research.

5.5.3 Limitations

Our study has several limitations. Although the goal of the experiment was to measure meaning-focused or incidental learning, and how TE affects this process, the results of our questionnaire show that some participants noticed items from the pretest while they were watching the documentary. We suggest that future studies consider other methods to control for prior item knowledge (see, for example, Sonbul & Schmitt, 2013).

Another limitation is our reliance on a small sample of participants and a small, varied sample of MWUs. Further, although we focused on learners' visual processing of MWUs, the target items were also presented in spoken form, which may have affected processing. In addition, although the input chosen in the current study did not contain any explicit visual cues to the meaning of the target items, transitions between the imagery region and the caption region may have affected processing of the target items (e.g., Bisson et al., 2014).

5.6 Conclusions

Despite these limitations, our study provides further empirical support for the beneficial effect of TE in captioned audiovisual input (Cintrón-Valentín et al., 2019; Lee & Révész, 2018, 2020; Montero Perez et al., 2014, 2015, 2018), and extends the findings of previous studies by examining how TE can affect learners' visual processing of MWUs in captioned audiovisual input. It seems that the attention-raising effect of TE has the potential to increase the likelihood that MWUs are picked up. However, the effectiveness of TE in captioned television may depend on factors related to item difficulty and processing load. Further research is needed to examine how different input modalities are integrated during L2 processing of captioned audiovisual input, and how this might affect the acquisition of MWUs.

Notes

1. Because captions were presented in a proportional font instead of a monospaced font, spatial dimensions of characters had to be estimated (see Godfroid, 2020: 175–76).
2. Lowering the sampling rate made it easier to track the eyes of some participants. This means that the sampling error was not the same for all participants, increasing the amount of individual variability in the data. However, simulations show that differences in fixation durations at different sampling rates tend to be negligible (Andersson et al., 2010). Further, because conditions were compared within participants, we do not expect that different sampling rates led to any systematic differences that might confound the effect of the treatment.
3. Of the 28 MWUs, 22 were not fixated by at least one participant. We could not discern any patterns in full phrase skipping (at the item nor the participant level), and the number of cases (n = 60) was too small for a statistical analysis.

6 The effects of typographic enhancement on L2 collocation processing and learning from reading: an eye-tracking study[1]

Collocations have been of considerable interest to applied linguistics and SLA researchers, due to their pervasiveness in language. In corpus research, collocations are broadly defined as word sequences that appear more frequently in language than the occurrence frequency of their constituent words would predict (Biber et al., 1999). Typical examples in English are *heavy rain* and *strong wind*. As conventional forms of expression, collocations may be used to express concepts in an efficient manner, or serve as phrasal terminology in technical, scientific, or academic discourse (e.g., Schmitt & Carter, 2004).

Collocations have also provided a window into implicit and statistical learning mechanisms in language processing and acquisition (e.g., Hoey, 2005). In usage-based theories of language acquisition, collocation learning is considered a slow, cumulative, and largely unconscious process in which learners build representations of word sequences on an exposure-by-exposure basis, resulting in intuitive knowledge of word co-occurrence patterns (Bybee, 2002; Ellis, 2003; Hoey, 2005). Experimental studies have demonstrated that native speakers process and recognize collocations faster than less frequent or predictable word sequences (e.g., Carrol & Conklin, 2020; Öksüz et al., 2020). This processing advantage is thought to contribute to fluent and efficient language use in interaction (Ellis, 2003).

Despite the advantages of collocation knowledge, L2 learners (also advanced L2 learners) tend to produce many collocations that are deviant or non-nativelike (e.g., **make a photo*) (e.g., Laufer & Waldman, 2010). Although the evidence is far from conclusive, it is generally thought that L2 collocation knowledge develops slowly, and does not always come to the surface at the level of productive use (e.g., Durrant & Schmitt, 2010; Ellis, 2008). A better understanding of why this might be the case requires a closer look at the variables that may affect the learnability of collocations.

6.1 Background

6.1.1 Learning collocations from meaning-focused input

Input frequency is thought to play an important role in L2 collocation learning (e.g., Durrant & Schmitt, 2010). Recent studies have demonstrated that the occurrence frequency of word sequences in L2 input predicts how fast learners recognize or process those word sequences (Conklin & Carrol, 2020; Northbrook & Conklin, 2019; Northbrook et al., 2021). There is also increasing evidence that advanced L2 learners are sensitive to the distributional properties of L2 collocations in measures of receptive knowledge and processing (e.g., Ellis et al., 2008; Öksüz et al., 2020). However, learning materials used in classroom-based contexts do not always provide a faithful reflection of the collocation patterns that are conventional or common to native speakers (Northbrook et al., 2019). Further, learners tend to have less L2 exposure than native speakers. Therefore, the low occurrence rate of collocations might hinder the implicit development of collocation knowledge through repeated exposure (Durrant & Schmitt, 2010). Forming the nuanced representations of word associations that lead to strong, productive collocation knowledge is therefore likely to be a slow process that requires a substantial amount of linguistic evidence (Ellis et al., 2008).

One way of supporting the uptake of collocations from authentic L2 input is through meaning-focused activities, such as reading, reading-while-listening, and viewing. Several studies have shown that learners can pick up collocation knowledge from these activities (e.g., Pellicer-Sánchez, 2017; Webb et al. 2013), although purely incidental exposure does not always result in durable gains (e.g., Szudarski & Carter 2016). One explanation for this is the low salience of collocations in L2 input. Many theories of SLA hold that linguistic representations respond not just to exposure frequency, but also to degree of attention and elaborate processing (e.g., Ellis, 2005; Laufer & Hulstijn, 2001; Schmidt, 2001). L2 features which are salient, i.e., which attract conscious attention, are likely to become processed elaborately and retained in memory, as for instance evidenced in eye-tracking studies (e.g., Godfroid et al., 2013). Collocations, which are often semantically transparent (e.g., *make a mistake*) or semi-transparent (e.g., *catch a disease*) do not tend to be perceived as functional or meaningful, especially when they consist of highly frequent or familiar words (Boers et al., 2014). Collocation knowledge may therefore take longer to develop in meaning-focused activities compared to words and more salient formulaic sequences (e.g., idioms), in particular with a view to achieving the strong, permanent representations that are required for productive use.

6.1.2 Promoting L2 collocation learning: typographic enhancement

Recent studies have explored the use of typographic enhancement (TE) to increase learners' noticing and uptake of L2 collocations in meaning-focused activities (e.g., Szudarski & Carter, 2016; Majuddin et al., 2021; Puimège, Montero Perez, & Peters, 2021; Vu & Peters, 2020; 2022). Typographic enhancement involves the use of typographic techniques such as bolding, underlining, or coloring, to make elements of a written text more visually salient. There is some evidence to suggest that TE may stimulate learners' attention for, and learning of collocations encountered during reading.

Peters (2012) examined the effect of TE on learners' uptake of German formulaic phrases during reading. The results showed that TE guided learners' attention to the target collocations, resulting in improved form recall immediately after reading. Similar results were found by Choi (2017), who combined eye-tracking and offline tests to examine the effects of TE on learners' attention to, and learning of collocations during reading. He found that TE helped learners notice novel collocations in the reading passage, and that this positively affected immediate recall of the target collocations.

A few studies have found that TE may also promote long-term retention of L2 collocation knowledge. In a study by Szudarski and Carter (2016), combining TE with repetition in reading texts had a positive effect on recall and recognition after a two-week delay, whereas repeated exposure by itself did not result in durable collocation knowledge. In two longitudinal studies, Vu and Peters (2020, 2022) investigated the effects of different input modes (reading-only, reading with TE, reading-while-listening, reading-while-listening with TE) on L2 collocation learning over a nine-week period. In both studies, enhancement led to the highest gains. These studies suggest that TE may promote the development of productive collocation knowledge, when repeated exposure by itself does not have a durable effect on learning. Szudarski and Carter (2016) interpret this finding in relation to the Involvement Load Hypothesis (Laufer & Hulstijn, 2001), which underscores the importance of degree of elaboration in vocabulary learning.

The studies reviewed above found that attention enhancement may promote collocation knowledge at the level of form recall. However, the beneficial effects of TE may not generalize to collocation knowledge captured by implicit measures, or measures that focus on receptive processing speed. Sonbul and Schmitt (2013) examined the effect of TE on L1 and L2 speakers' implicit and explicit knowledge of medical collocations (e.g., *cloud baby*). Implicit collocation knowledge was measured in a primed lexical decision task comparing response times for words primed by their collocate (e.g., *cloud* → *baby*), compared to a non-associated prime (e.g., *steam* → *baby*). Although TE promoted recall and recognition of collocations, the study found no evidence of collocation priming.

In a partial replication of Sonbul and Schmitt (2013), Toomer and Elgort (2019) increased the amount of input and the duration of the experiment to optimize conditions for implicit collocation learning. Collocations were encountered nine times either in enhanced (bold) or unenhanced form. Evidence of collocation priming was only found in the unenhanced condition. The authors suggested that the repeated use of TE may have obstructed the creation of an 'accurate mental model' of the contexts in which the collocations appeared, thereby preventing the expansion of lexical and semantic representations of words (Toomer & Elgort, 2019: 426).

In a recent study, Northbrook, Allen and Conklin (2021) examined the effects of repetition and TE (underlining) on learners' response times to lexical bundles (e.g., *set off home*, *tired and hungry*) in a phrase acceptability judgement task. Participants read stories interspersed with iterations of the phrasal judgement task, so that processing fluency could be measured after each exposure. The results showed that repetition had a significant effect on response times after only a few exposures, in contrast to the findings of Toomer and Elgort (2019). Further, the effect of TE was quickly overridden by the effect of repetition, suggesting that processing fluency of word sequences may benefit from repeated exposure regardless of the degree conscious attention.

Conklin and Carrol (2020) used eye-tracking to examine the effect of a reading task on native speakers' online processing of binomials. The study observed a processing advantage for newly learned binomials over their reversed form (e.g., *wires and pipes* vs. *pipes and wires*) after only four to five contextualized exposures. The study showed that native speakers rapidly develop a sensitivity to the word co-occurrences they encounter during reading. However, it is unclear whether a similar advantage can arise as rapidly for L2 learners, and for different types of formulaic sequence. Thus far, no previous studies have used eye-tracking to measure the effect of reading on the development of L2 collocation knowledge.

Most studies that have found strong, positive effects of TE on collocation knowledge used a design that combined TE with repetition, by providing multiple enhanced exposures in written texts. However, it is unclear whether the effect of TE is equally strong and durable when it is not combined with repetition in this way. Because of the low occurrence rate of collocations in naturalistic discourse, it may not be feasible for teachers to visually enhance every occurrence of a new collocation in learners' L2 input. Further, in learning materials such as university textbooks and graded readers, key terms and important vocabulary are often only enhanced once (Toomer & Elgort, 2019). If the aim of TE is to support or complement collocation learning from meaning-focused input, then we need to ask whether its effects are durable in combination with unenhanced contextual encounters.

6.2 Research questions

The present study examines to what extent TE affects learners' processing of collocations in a first enhanced contextual exposure, compared to later, unenhanced exposures. It also examines the effects of repetition with and without a first enhanced exposure on collocation knowledge. The study addresses following research questions:

1. What are the immediate and delayed effects of TE on learners' reading times on novel L2 collocations when only the first occurrence is enhanced?
   a. What is the immediate effect of TE on learners' reading times for novel L2 collocations?
   b. What is the effect of TE on learners' reading times for novel L2 collocations in later, unenhanced exposures?
2. What are the effects of repeated exposure and TE on learners' knowledge of novel L2 collocations when only the first occurrence is enhanced?
   a. Does (enhanced) repeated exposure to novel L2 collocations during reading lead to a processing advantage for target collocations over matched control collocations?
   b. Does (enhanced) repeated exposure affect learners' recall and recognition of novel L2 collocations?

6.3 Methodology

6.3.1 Participants

Sixty-one Dutch-speaking students from various study programs at Flemish universities took part in the experiment. Students participated for course credit or a monetary compensation.

Their prior vocabulary knowledge was estimated using the Vocabulary Levels Test (VLT; Schmitt et al., 2001). Participant-related information is summarized in Table 6.1. Data from four participants were excluded due to absence in one of the experimental sessions, or track loss in the eye movement data.

Table 6.1 Summary of participant information

	Age	Years of English	VLT 2K	VLT 3K	VLT 5K	VLT 10K
Experimental group (*n* = 41)						
Mean	21.88	6.63	28.82	27.52	24.35	14.75
SD	2.37	1.78	2.21	3.72	5.78	7.80
Range	18-30	4-11	20-30	14-30	7-30	4-29
Control group (*n* = 16)						
Mean	26.56	7.5	29.43	28.14	25.29	16.86
SD	9.89	1.60	1.05	3.05	4.53	7.21
Range	19-56	5-10	26-30	19-30	13-30	0-29

Note. Years of English is the total number of years of formal English instruction participants had received in secondary school and higher education; VLT 2K, 3K, 5K, and 10K are participants' scores on each 1,000 word level of the VLT.

6.3.2 Reading materials and collocations

Participants read 10 English texts containing eight occurrences of 24 collocations. The first text (approx. 2,000 words) consisted of edited passages of a popular science book (Godfrey-Smith, 2016). The remaining nine texts (each approx. 400 words) contained edited passages of TED-ed scripts similar in topic and lexical profile to the first reading text. The experimental texts were edited so that (a) at least 95% of word tokens belonged to the 3,000 most frequent word families in English (Cobb, n.d.), and (b) each of the target items appeared exactly eight times in total: twice in Text 1, and six times in Texts 2-9. The edited texts were checked for naturalness and grammatical correctness by three native speakers of English.

Target items were 24 modifier-noun collocations with technical or context-specific meanings. Most collocations were found in the original texts, although a few (e.g., *sensitive cells, reflecting skin*) were adapted from their original form (e.g., *light-detecting cells, reflecting cells*), so that the same single word component did not appear in more than one target item. Most target collocations consisted of words with high occurrence frequency in the COCA (*M* = 92 per million, *SD* = 77). Three lower-frequency words were either cognates (*reflex*) or transparent compounds (*workspace, pinhole*). Prior knowledge of the target collocations was checked in post-experiment interviews. One target item (*action potential*) was excluded, because nine participants reported having prior knowledge of this collocation. The reading materials and target collocations can be found in Appendix 5.

6.3.3 Treatment conditions

The study adopted a mixed design combining between-participants and within-participants comparisons (see Table 6.2). Participants were assigned to one of the conditions according to their order of participation (participant 1 was assigned to condition 1, participant 2 to condition 2, etc.). In order to examine the effect of the treatment on collocation knowledge, exposure to the target collocations was a between-participants variable. Participants in the experimental group (*n* = 41) read English texts containing the target collocations, while participants in the comparison group (*n* = 16) read a version of the same texts not containing target collocations. In this version, we deleted the modifiers of the collocations (e.g., 'reflecting skin' → 'skin'), or through paraphrasing where needed (e.g., 'sensitive cells' → 'cells that capture light').

Textual enhancement was a counterbalanced within-participants variable. Participants in the experimental group were assigned to one of two versions of the first of ten texts, each containing 12 enhanced (bold font) and 12 unenhanced collocations. Collocations enhanced in version 1, were unenhanced in version 2, and vice versa, so that all 24 target items appeared in both conditions in the full data set. Enhanced and unenhanced items were alternated in the texts. Note that in the enhanced condition, collocations only appeared in bold in their first occurrence in the reading texts. The remaining 7 occurrences were unenhanced for all items. This allowed us to investigate the effect of TE on later, unenhanced exposures (RQ1).

Table 6.2 Examples of the first item occurrence for each level of the between-participants variable (Group) and the counterbalanced within-participants variable (TE)

	Experimental group		Comparison group
Target item	Version 1 (*n* = 21)	Version 2 (*n* = 20)	Version 3 (*n* = 16)
central brain	The **central brain** only contains about 10% of the neurons	The central brain only contains about 10% of the neurons	The brain only contains about 10% of the neurons
capturing surface	These cells are located in the retina, the capturing surface of the eye.	These cells are located in the retina, the **capturing surface** of the eye.	These cells are located in the retina, the surface found at the back of the eye.
shared control	This unique form of **shared control** gives it incredible flexibility	This unique form of shared control gives it incredible flexibility	This unique form of control gives it incredible flexibility

6.3.4 Instruments

The data collection instruments consisted of a sentence-reading task, a form recall test and form recognition test. The three posttests can be found in Appendix 5. To establish whether repeated exposure to the target collocations had an effect on reading speed (RQ2a), participants completed a sentence-reading task immediately after the experimental treatment. The task contained three types of target item: (1) target collocations from the reading texts, (2) matched control items, and (3) distractor items (see Table 6.3). Matched control items contained the same modifier as their matched target item (e.g., *human*) but a different noun (e.g., *weight*). The nouns of target-control pairs were matched for number of characters and occurrence frequency in the COCA. However, the nouns of target collocations appeared in the input with considerably higher frequency (M = 17.33, SD = 14.25) than the nouns of control collocations (M = 2.67, SD = 3.55). This was the case in both the experimental group and the comparison group.

Table 6.3 Examples of items in the sentence-reading task

Sentence	Item type
1. A great number of dead zones have appeared, causing some fish to leave the area.	target item
2. A great number of dead belts have appeared, causing some fish to leave the area.	control item
3. By means of a dry bite the snakes can send a message to warn their victim.	distractor
4. Claws are ideal for piercing and hooking, but their points make grabbing difficult.	filler sentence

Note. Underlined items were not underlined the sentence-reading task.

Distractor items were added to control for a learning effect from the sentence-reading posttest to the delayed posttests of form recall and form recognition. Some of these were taken from Sonbul and Schmitt's (2013) list of medical collocations (e.g., *cloud baby, split hand*), while others were found in TED-ed videos and other texts about scientific topics (e.g., *marine snow, dry bite*). We created two presentation lists, so that participants did not encounter the same sentence contexts twice. Each list contained 12 target items, 12 control items, 12 distractor items, and 24 filler sentences not containing a modifier-noun collocation. Participants were assigned to one of the lists so that the final data set contained a comparable number of observations in every experimental condition.

Participants' eye movements were recorded while they read the 60 sentences one by one. Reading times on target and control items were compared in order to establish whether a processing advantage arose for target collocations as a result of collocational knowledge acquired during reading. Significant differences in reading times between target and control items were only interpreted as evidence for a processing advantage associated with collocation knowledge if the effect was significantly larger in the experimental group than in the comparison group.

The second posttest measured collocation knowledge at the level of form recall. The test had a sentence cloze format. Target collocations appeared in informative sentences describing their meaning, and participants were instructed to provide the correct modifier (e.g., *dead*) based on the associated noun (e.g., *zone*) and the sentence context. The test contained all target collocations, and all distractors from the sentence-reading posttest. Two example items are presented in Table 6.4.

Table 6.4 Example items of the form recall test

1. All animals, even insects, show a(n) _ **response** in harmful or painful situations: this is a direct reaction of the nervous system.	? response (*reflex*)
2. A(n) _ **baby** is a young child that spreads infectious diseases very efficiently.	? baby (*cloud*)

Note. Correct responses are in parentheses.

The form recognition test contained the same sentences as the form recall test, but participants had to select the correct modifier from four options. Participants were allowed to guess or select an option randomly when they did not know the correct response. We coded different response types after the data collection, based on participants' responses in a retrospective interview (see Interviews). Two examples are presented in Table 6.5.

Table 6.5 Example items of the form recognition test

1. All animals, even insects, show a(n) _ **response** in harmful or painful situations: this is a direct reaction of the nervous system.			
1. trigger	2. impulse	3. instinct	4. *reflex*
2. A _ **baby** is a young child that spreads infectious diseases very efficiently.			
1. steam	2. mist	3. *cloud*	4. fog

Note. Correct responses are italicized.

After completing the posttests, participants were interviewed. First, they were asked to explain their motivation for each response in the form recognition posttest. This helped ensure that test scores truly reflected collocational knowledge remembered from the reading texts, and not, for example, guessing based on semantic knowledge of the single word components. The responses were coded based on audio recordings of the interviews, and the results were used in the quantitative analysis of the form recognition posttest results (see Data cleaning and scoring).

Participants were also interviewed about their perception of the experimental treatment. They reported whether they remembered encountering enhanced collocations in the first text, whether they had tried to memorize those collocations during reading, and whether they had noticed the recurrence of collocations in the experimental texts. This part of the interview was not analyzed quantitatively, but the results will be briefly discussed in relation to the eye-tracking data.

6.3.5 Procedure

Data were collected individually in three sessions, over a period of three weeks. In Session 1, participants read the first of ten experimental texts while their eye movements were recorded with the Eyelink Portable Duo of SR Research. Participants were seated in front of a 1,280 x 1,024 resolution monitor, their head resting on a table-mounted chin rest. The text was presented in Courier, 16-pt font, at a viewing distance of 73 cm, so that each character corresponded to 0.3° of visual angle. Eye movement data were recorded for the dominant eye only, at a sampling rate of 2,000 Hz. The text was presented across 25 screens, each followed by a drift check, to allow for recalibration of the eye-tracker when necessary. A short break and recalibration were inserted after 12 screens. The 24 target collocations appeared twice in the first reading text. None of the target collocations immediately preceded or followed a line break or punctuation. Participants were instructed to read the text attentively. They were also informed that they would complete questions about the content afterwards. After reading the text, they completed three open-ended comprehension questions, followed by the 2K, 3K, 5K and 10K levels of the VLT.

After a one-week interval, participants read the remaining nine texts containing six unenhanced occurrences of each target collocation, in random order. Each text was presented across five screens, followed by a true-or-false comprehension question. Two short breaks were inserted after the third and sixth text. Afterwards, participants completed the sentence-reading posttest, in which they read 60 sentences, one by one. Following previous studies (e.g., Carrol & Conklin, 2020), one-third of the sentences were followed by a true-or-false question (at random intervals) to encourage attentive reading. Sentences were presented in random order, each on a separate screen, followed by a drift check. After 30 sentences, participants received a short break, and the eye-tracker was recalibrated.

In the third session, one week later, participants completed the form recall and form recognition tests. They were also interviewed about their responses in the form recognition posttest, and about their perception of the experimental treatment. Finally, participants were debriefed about the aims of the experiment.

6.3.6 Data cleaning and scoring

A binary scoring was applied to responses in the posttests. In the form recall test, misspelled or approximate responses (e.g., ‘reflective’ instead of ‘reflecting’) were coded as incorrect (0). Because most collocations consisted of highly frequent single word constituents, the number of misspelled items was very low, and a more lenient scoring system would have yielded very similar results.

Participants' interview responses were coded by one of the researchers, resulting in a four-level nominal variable. For each form recognition test item, responses were coded as 'guess' when participants reported selecting an item based on a pure guess, or based on semantic knowledge or information in the sentence contexts. Responses were coded as 'certain' when participants reported remembering the correct collocation from the reading texts with certainty, as 'uncertain' when they vaguely remembered the collocation from the reading texts, but with less certainty. Finally, items were coded as 'prior knowledge' when participants reported having learned a target or distractor item before reading the experimental texts. Results for one target item (*action potential*) and two distractor items (*mass number, shell shock*) were excluded from analysis, because at least 15% of participants reported having prior knowledge of these collocations.

Eye-tracking data were visually inspected in the Eyelink Data Viewer. Vertical drift was corrected manually, and lines of text or trials where considerable drift occurred were deleted. Additionally, data from two participants were discarded due to track loss. This resulted in the removal of 6% of interest areas. Next, the default four-step cleaning procedure of Data Viewer was used to merge or remove overly short or long fixations. The final data set only included eye-tracking data for the target collocations and component words.

6.3.7 Analysis

All data were analyzed in mixed effects models, using the lmer and glmer functions of the lme4 package in R (version 1.1-21, Bates, Mächler, Bolker, & Walker, 2015). Each model included random intercepts for target items and participants, but no random slopes (adding random slopes for Item condition/TE caused convergence errors in lme4). Model assumptions (linearity, homogeneity of variance, normal distribution of residuals) were checked, and outliers were identified based on the studentized residuals of the mixed models. Residuals with a value above 2.5 were removed.

We first analyzed the effect of TE on visual processing of target collocations during reading (RQ1). Log-transformed total reading time (sum of fixations for the entire collocation) was the continuous outcome variable. The model included fixed effects for TE (enhanced, unenhanced), Exposure count (eight-level factor), and their interaction. Item length was also included as log-transformed, grand mean centered covariate. To interpret the interaction between TE and Exposure count, pairwise comparisons were conducted using the emmeans function in the emmeans package (version 1.4, Lenth, 2018).

Next, we analyzed participants' fixation times in the immediate sentence-reading posttest (RQ2a). Outcome variables were first pass time (sum of fixation durations during the first visit of an interest area), and total reading time. First pass time, being an early measure, might give an indication of the extent of a processing advantage at the early stages of processing (e.g., lexical activation), whereas total reading time may provide more insight into the effects on meaning integration in the sentence contexts (e.g., Carrol & Conklin, 2015). Additionally, we analyzed total reading time of the two component words separately. A first model included fixed effects for Group (experimental, comparison), Item condition (target, control) and their interaction. As there was no enhanced condition in the comparison group, the effect of TE was analyzed in a second model including only the results of target items in the experimental group. Here, the main independent variable of interest was TE (enhanced, unenhanced). Log-transformed, mean-centered item length was included as a covariate in both models.

Finally, we analyzed the effect of the treatment on binary scores in the form recognition and form recall posttests (RQ2b), using the glmer function. Because only 7.7 percent of responses (= 193 observations) in the form recall test were coded as correct, we only report descriptive results for this test. In the form recognition test, the majority of correct items (56%) were guesses, as reported by the participants in the interviews (see Results). All guessed responses were recoded to 'incorrect'. The first model contained fixed effects for Group (experimental, control), Item condition (target item, distractor item), and their interaction. Pairwise comparisons were conducted to break down the interaction between Group and Item condition. A second model focused on the results of the experimental group, and contained a fixed effect for TE (enhanced, unenhanced).

6.4 Results

6.4.1 Descriptive results

Descriptive results of eye-tracking measures are presented in Tables 6.6 and 6.7. Descriptive results of the form recall and form recognition tests are presented in Table 6.8.

Table 6.6 Means and standard deviations (in parentheses) for total reading time during the experimental treatment

	Session 1		Session 2					
	Exposure 1	*Exposure 2*	*Exposure 3*	*Exposure 4*	*Exposure 5*	*Exposure 6*	*Exposure 7*	*Exposure 8*
Enhanced	1243 (1096)	866 (734)	886 (620)	808 (588)	763 (521)	771 (477)	782 (508)	713 (531)
Unenhanced	943 (675)	877 (680)	927 (691)	842 (547)	787 (593)	786 (619)	779 (573)	755 (519)

Table 6.7 Means and standard deviations (in parentheses) for the sentence-reading posttest

	TRT	*FPT*	*TRT word 1*	*TRT word 2*
Experimental group				
Control item	1220.95 (657.47)	427.16 (338.19)	635.90 (379.06)	616.27 (364.32)
Target item	989.62 (536.54)	369.88 (290.99)	543.87 (338.83)	492.35 (300.63)
Enhanced	984.04 (532.52)	356.14 (285.63)	550.10 (335.90)	480.00 (305.88)
Unenhanced	995.31 (541.73)	383.92 (296.35)	537.61 (342.38)	505.22 (295.23)
Comparison group				
Control item	1337.31 (826.68)	424.96 (319.35)	693.43 (477.49)	662.32 (462.37)
Target item	1141.72 (580.95)	411.84 (343.00)	633.92 (368.75)	529.53 (323.78)

Note. TRT is total reading time, FPT is first pass time, and RRT is rereading time.

Table 6.8 Means, standard deviations, and percentages for form recall and form recognition posttests

	Recall						Recognition					
	Experimental			Comparison			Experimental			Comparison		
	M	*SD*	%	*M*	*SD*	%	*M*	*SD*	%	*M*	*SD*	%
Total	4.00	2.36	9	1.93	21.68	4	10.17	4.84	23	6.67	3.50	15
Distractors	0.80	1.01	4	0.71	6.71	3	1.46	1.42	7	2.07	1.94	9
Target items	3.29	1.81	14	1.13	14.98	5	8.71	4.05	38	4.6	2.26	20
Enhanced	1.61	1.18	14		7.78		4.80	0.51	42			
Unenhanced	1.68	1.06	15		7.20		3.90	0.51	34			

Note. Experimental = Experimental group (n = 41), Comparison = Comparison group (n = 16). In the form recognition test, guesses were recoded to 0 (incorrect).

6.4.2 Research question 1

Figure 6.1 presents the results of the mixed model with log-transformed total reading time during the experimental treatment as dependent variable. The full output can be found in Appendix 5. With Exposure 1 as the reference category of Exposure count, there were significant main effects of Exposure count and TE, and a significant interaction between these variables. Pairwise comparisons indicate that, in both conditions, reading times decreased significantly from the first to the final exposure (B = 0.21, p < .001), and TE significantly predicted reading times in the first and fourth exposures. In the first exposure of Session 1, enhanced items were associated with significantly longer reading times (B = 0.22, p < .001), with an estimated difference of 25% (exp(B) - 1). In Session 2, predicted reading times were shorter for enhanced collocations in Exposure 4 (B = -0.07, p = .032). Here, the estimated difference between enhanced and unenhanced items was 7%.

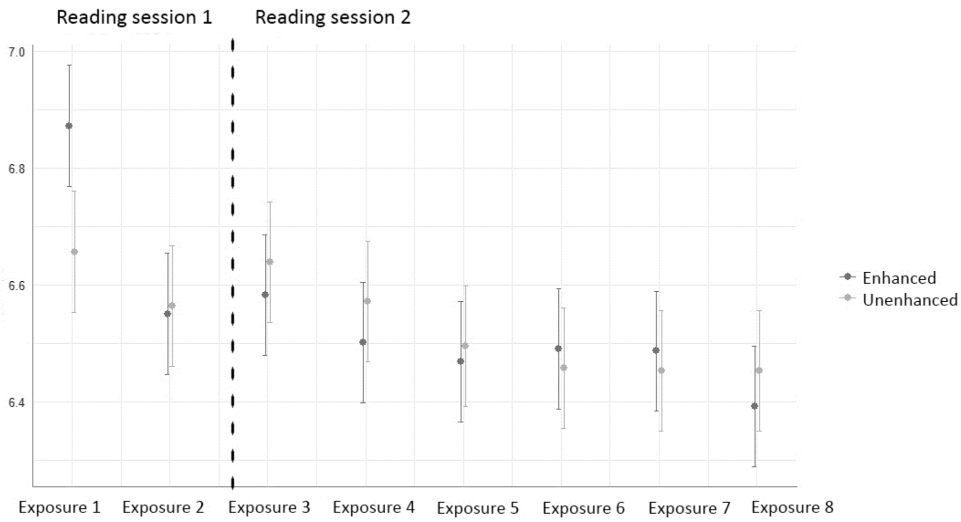

Figure 6.1 Predicted values of log-transformed total reading time by Exposure count and TE

6.4.3 Research question 2

The results of the mixed models for the sentence-reading posttest are presented in Table 6.9 and Figure 6.2. There was a significant main effect of Item condition in total reading times on the full collocation and the final word, but not in first pass times and total reading times on the first word. The main effect of Group, and the interaction between Item condition and Group, were nonsignificant in all models. Pairwise comparisons indicated that significant differences between target and control items existed in both groups of participants. Estimated differences were larger in the experimental group, in particular for total reading time of the full collocation (16% in the comparison group (p = .013), 24% in the experimental group (p < .001)), and total reading time of the first word (9% in the comparison group (p = .181), 17% in the experimental group (p = .001). Differences in total reading time on the final word were more similar for the two groups (26% in the comparison group (p < .001), 28% in the experimental group (p < .001)). Finally, there were no significant differences between enhanced and unenhanced items for any of the reading time measures (the output of models focusing on the experimental group can be found in Appendix 5.

Table 6.9 Results of best fitting models for reading times in the sentence-reading posttest

	TRT (*n* = 1,286)			FPT (*n* = 1,276)			TRT word 1 (*n* = 1,241)			TRT word 2 (*n* = 1,222)		
	Intercept	*T*	*p*	Intercept	*T*	*p*	Intercept	*T*	*p*	Intercept	*T*	*p*
Fixed effects												
(Intercept)	7.06	92.19	< .001	5.76	52.75	< .001	6.38	86.91	< .001	6.33	86.80	< .001
Item condition	-0.15	-2.51	.013	-0.13	-1.51	.133	-0.09	-1.34	.181	-0.24	-3.89	< .001
Group	-0.09	-1.01	.316	-0.04	-0.32	.747	-0.09	-1.12	.266	-0.05	-0.67	.503
Item length	0.68	5.65	< .001	-0.04	0.29	.777	0.70	8.21	< .001	0.57	7.88	< .001
Item condition*Group	-0.06	-1.17	.241	0.02	0.22	.828	-0.08	-1.18	.240	-0.02	-0.28	.783
Random effects	*Variance*			*Variance*			*Variance*			*Variance*		
(1|Item)	0.07			0.13			0.05			0.06		
(1|Participant)	0.01			0.01			0.01			0.01		
Residual	0.20			0.64			0.26			0.23		
AIC	1775.17			3190.26			1997.39			1840.32		

Note. The reference level of Group is 'comparison group', and the reference level of Item condition is 'control item'. TRT is total reading time, FPT is first pass time.

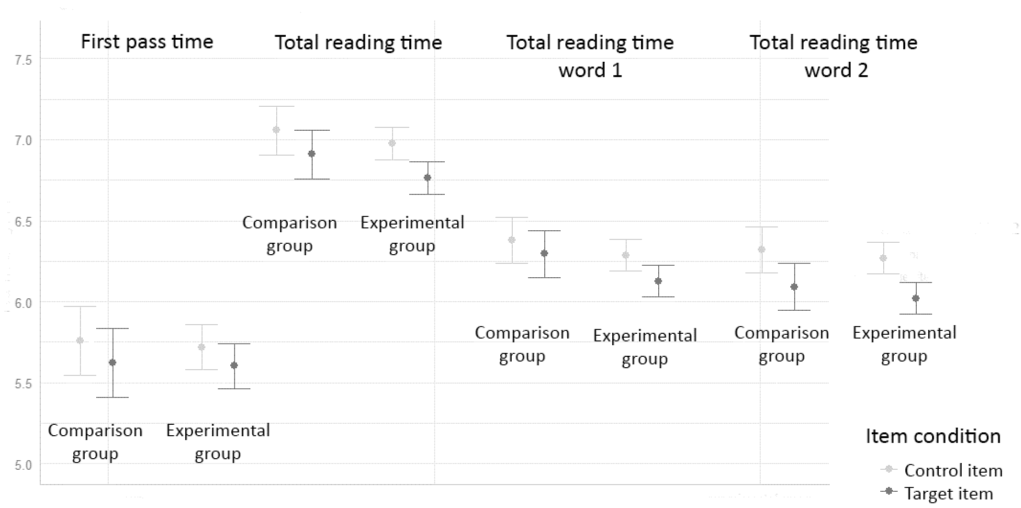

Figure 6.2 Predicted probabilities of log-transformed reading times in the sentence-reading posttest by Group and Item condition

The results of the mixed models for scores in the form recognition test are summarized in Table 6.10. The main effect of Item condition, and the interaction between Item condition and Group were significant. Pairwise comparisons showed that the estimated odds of a correct response in the form recognition posttest were 2.34 times higher for target items than for distractors in the comparison group (p = .049), and 12.88 times higher in the experimental group (p < .001). In a separate model focusing on the results of the experimental group, the effect of TE yielded a small, significant difference between enhanced and unenhanced items (B = -0.52, p = .002).

Table 6.10 Results of best fitting models for form recognition posttest scores

Model 1: all participants (n = 2,520)					Model 2: experimental group (n = 943)				
	Intercept	SE	*Z*	*p*		Intercept	SE	*Z*	*p*
Fixed effects					**Fixed effects**				
Intercept	-2.80	0.39	-7.20	< .001	Intercept	-0.53	0.39	-1.35	.178
Item condition	0.85	0.43	1.97	.049	TE	-0.52	0.17	-3.03	.002
Group	-0.48	0.34	-1.38	.166					
Item condition*Group	1.70	0.29	5.78	< .001					
Random effects	*Variance*	*SD*			**Random effects**	*Variance*	*SD*		
Item	1.36	1.17			Item	2.52	1.59		
Participant	0.67	0.82			Participant	1.13	1.06		
AIC	1918.72				AIC	972.56			

Note. Reference levels in Model 1 are 'distractor' (Item condition) and 'comparison group' (Group). The reference level in Model 2 is 'enhanced' (TE).

6.5 Discussion

6.5.1 Research question 1

The first aim of the study was to examine how TE (bolding) may affect collocation processing during reading, both in a first, enhanced exposure, and in later, unenhanced exposures. To that end, learners' reading times were measured for 12 enhanced and 12 unenhanced collocations, each of which occurred 8 times in the reading materials, across two reading sessions. The findings show that, in the first exposure, TE led to significantly longer total reading times on target collocations. This confirms previous findings that TE promotes overt attention to collocations during reading (e.g., Choi, 2017). In the second exposure within the same reading text, there were no clear differences in reading times between the enhanced and unenhanced conditions, which suggests that the effect of TE on visual processing was local and short-lived. It is important to note that the first two exposures occurred in close approximation within the same reading text. It may not have been necessary for learners to deeply engage with the collocations in the second exposure, as the information gathered from the first encounter was still fresh in the mind.

In session 2, no clear differences in reading times emerged between previously enhanced and unenhanced collocations. It seems that TE temporarily affected learners' attention for target collocations (see also Choi, 2017), but did not have a strong delayed effect on processing times in new contexts. Further, in both conditions, reading times decreased from the first until the fifth exposure. This is in line with previous findings that repetition has a facilitative effect on online processing of words and phrases (Conklin & Carrol, 2020;

Northbrook et al., 2021; Pellicer-Sánchez, 2016). Decreases in reading times have been explained in relation to theories of automatic processing or skill acquisition, which assume that processing speeds up due to storage of exemplars or instances in memory (e.g., Logan 1988). In L2 collocation learning, every new encoding of a word sequence has an effect on its representation in memory, and on its subsequent retrieval (Hoey, 2005). Northbrook et al. (2021) demonstrated that such an effect may arise instantaneously when L2 learners encounter new lexical bundles during reading. However, the current experiment did not compare target collocations to matched controls during the reading task. Therefore, the repetition effect evident in the reading times during the second session cannot confidently be interpreted as evidence of learners' increasing familiarity with the target collocations. or collocation priming, as opposed to, for example, a repetition priming effect of the constituent words (e.g., Ledoux, Camblin, Swaab, & Gordon, 2006).

The post-experiment interviews reveal a few trends in participants' general perception of the reading task. Firstly, participants consistently reported that they had read the experimental texts in function of the upcoming comprehension questions. The bolded collocations were generally interpreted as key words meant to draw attention to relevant concepts in the text, similar to how enhancement has been used in university textbooks (e.g., Toomer & Elgort, 2019). Although participants reported that TE made them pay closer attention to the meanings associated with the collocations, they did not deliberately attempt to memorize their form. Participants also reported that they did not notice the recurrence of target collocations in the experimental texts, even though they remembered that the texts dealt with similar topics, and sometimes overlapped in content. This confirms that participants did not consciously link collocations appearing in the second reading session to those encountered the week before, at least, as far as participants could remember one week after the final reading session.

Taken together, the results of the reading task suggest that visually enhancing a single occurrence of a collocation does not seem to have a durable effect on the processing of collocations in later contexts. Enhancing various instances of L2 collocations across different texts, as was done in previous studies (e.g., Szudarski & Carter, 2016; Toomer & Elgort, 2019), may more effectively promote noticing of collocations as recurring or important word sequences. However, it is also possible that features specific to the current experiment, such as the announcement of comprehension questions, the use of bolding (as opposed to underlining, highlighting in color, etc.), or the scientific topics of the texts, may have negatively affected conscious attention to the form of the target collocations during reading. As has been found in L2 grammar acquisition (e.g., Cintrón-Valentín et al., 2019), TE does not guarantee elaborate processing of low-salience linguistic features.

6.5.2 Research question 2

Another aim was to find out whether repeated exposure and TE affected learners' reading speed, recognition, and recall of target collocations. The results of the sentence-reading posttest revealed a processing advantage of target collocations over matched control collocations, in particular in total reading times. This indicates that reading the experimental texts facilitated the integration of the target collocations into new sentence contexts. Importantly, a processing advantage was found in both the experimental group and the comparison group. As learners in the comparison group did not read any of the target collocations in their version of the experimental texts, their reading times could not reflect memory of the target collocations. A more plausible explanation is that learners were primed for the constituent words of the target collocations in the experimental texts. The effect was more evident in late measures, which suggests that participants familiarized themselves with specialized uses of words, and integrated these with greater ease in the sentences of the posttest (Ledoux et al., 2006).

Our failure to observe a clear processing advantage resulting from collocation knowledge concurs with the findings of Sonbul and Schmitt (2013) and Toomer and Elgort (2019), who reported inconsistent evidence of implicit collocation knowledge in a primed lexical decision task. The results of the present study appear to support Toomer and Elgort's conclusion that the development of strong and robust associations between words may require a considerable amount of L2 input. In contrast, two recent studies have demonstrated that a small number of contextualized exposures can measurably affect processing times of multiword phrases such as lexical bundles (Northbrook et al., 2021), and binomials (Conklin & Carrol, 2020). Importantly, these studies carefully controlled for extraneous variables such as input frequency of the single word constituents of target phrases. Although we do not exclude the possibility that L2 learners' sensitivity to co-occurrence frequency may develop faster for some types of formulaic sequence than for others, it is clear that more controlled designs are necessary to further examine the effects of repeated exposure during reading on L2 collocation processing.

Results of the delayed tests showed that participants were unable to recall most of the target collocations after one week. However, they could recognize the correct modifier of one-third of the collocations, on average. The mixed effects analysis showed that scores in the form recognition test were considerably higher in the experimental group, indicating that learners recognized the target collocations from encountering them in the reading texts. The analysis also showed that TE did not have a strong effect on form recognition scores, which suggests that form recognition was not affected by the visual salience of the target items in the first text. In line with previous findings (e.g., Pellicer-Sánchez, 2017), the results of the recognition test indicate that learners retain memory of the collocations they encounter repeatedly during reading, even if they make no deliberate attempts to memorize these collocations. On the other hand, the negligible gains found at the level of form recall seem to support Szudarski and

Carter's (2016) conclusion that elaborate processing may considerably raise the odds of finding a long-term effect of an incidental learning activity such as reading on productive collocation knowledge. A much greater amount of L2 input may be necessary to find durable productive collocation knowledge in purely incidental conditions.

6.5.3 Limitations and suggestions for future research

Our study has a number of important methodological limitations. Firstly, we did not include control collocations in the reading passages of the experimental treatment. Decreases in reading times within session 2 may have been caused by repetition priming of the single word components, or increasing topic familiarity. Therefore, we could not analyze if, or at what rate, frequency of the target collocations affected reading times in the experimental texts. Secondly, the treatment effect found in the sentence-reading posttest seems to reflect repetition priming of the single word constituents, rather than, or in addition to, a sensitivity to specific word co-occurrences. Although we found small between-group differences, we could not clearly separate the effects of co-occurrence frequency and single word frequency in this test.

These issues may be addressed by controlling for lexical and contextual factors. Similar to Conklin and Carrol's (2020) design, this might be achieved through a within-participants design that uses two or more counterbalanced texts in which a target collocation for one participant serves as control for another participant. In the current study, we prioritized including collocations that referred to real scientific concepts, and which therefore had some pedagogical relevance for the participants. However, we concede that a more controlled method might have been more informative in terms of finding evidence of collocation knowledge.

We must also point out that the complex, scientific content of the experimental texts may have affected learning gains. The collocations in our experiment introduced new scientific concepts, similar to the medical collocations in Sonbul and Schmitt's (2013) and Toomer and Elgort's (2019) studies. This may have introduced an additional cognitive burden that inhibited the encoding of collocational information. More research is needed to see how well current findings hold up when collocation learning is measured in less challenging or cognitively demanding tasks.

6.6 Conclusion

This study examined the effects of TE on online processing of collocations during reading, and on L2 collocation knowledge. The eye-tracking results indicate that the initial attention-enhancing effect of TE did not carry over to later, unenhanced exposures. Results of post-experiment interviews suggested that learners' primary focus was on meaning comprehension, and that TE did not induce conscious attention to the form of the target collocations. In an immediate posttest, we observed a treatment effect on reading times for target collocations in new sentence contexts. However, the test did not provide strong evidence of a processing advantage for collocations. One week after the treatment, participants could recognize the correct form of target collocations, but they could not recall them. We conclude that TE does not guarantee elaborate processing of collocations during reading, and that the development of L2 collocation knowledge does not necessarily benefit from TE when only a single occurrence is enhanced.

Note

1. A pilot experiment was conducted in preparation of Study 4. This pilot, which led to considerable changes to the design, materials, and instruments of the study, is reported in detail in Appendix 4.

7 Conclusions

In this chapter I discuss the general findings of the four empirical studies, each of which have yielded results that may contribute to our understanding of how formulaic sequences are learned from meaningful input. Below I discuss how the findings of this thesis contribute to theories on incidental learning of formulaic sequences. I will also discuss a few pedagogical implications, as well as the limitations of each study and possible directions for future research.

7.1 Summary of the four studies

In this thesis, I set out to investigate how L2 formulaic sequences are processed during, and learned from meaning-focused activities, specifically viewing and reading. To that end, four empirical studies were conducted combining offline measures (vocabulary tests, interviews, questionnaires), and eye-tracking to gain insight into the learning process and learning outcomes. The four studies are summarized in Table 7.1.

Table 7.1 Summary of the four studies

	Study 1	Study 2	Study 3	Study 4
Aim	- Learning single words and formulaic sequences from audiovisual input - Effects of prior vocabulary knowledge and item variables	- Learning formulaic sequences from audiovisual input - Effects of prior vocabulary knowledge and item variables	- Learning formulaic sequences from audiovisual input with captions - Effect of typographic enhancement	- Learning formulaic sequences from written input - Immediate and delayed effect of typographic enhancement
Participants	20 Flemish university students (L1 = Dutch)	77 Flemish university students (L1 = Dutch)	30 Flemish university students (L1 = Dutch)	61 Flemish university students (L1 = Dutch)
Type of study	- Within-participants - Pretest-posttest	- Within-participants - Pretest-posttest	- Within-participants - Pretest-posttest - Counterbalancing	- Mixed design - Counterbalancing of typographic enhancement

			of typographic enhancement	
Type of input	- Audiovisual input	- Audiovisual input	- Audiovisual input + captions	- Written input
Dependent variable	- Learning of single words and formulaic sequences from pre- to posttest	- Learning of formulaic sequences from pre- to posttest	- Eye-tracking measures - Learning of formulaic sequences from pre- to posttest	- Eye-tracking measures during reading - Eye-tracking measures in sentence-reading posttest - Learning of collocations in posttests
Independent variables	- Treatment (viewing) - Item variables	- Treatment (viewing) - Item variables	- Treatment (viewing) - Typographic enhancement - Reading times - Item variables	- Treatment (reading) - Typographic enhancement - Exposure count
Statistical analysis	Generalized Estimating Equations	- Repeated measures MANOVA - Generalized Estimating Equations	Mixed effects models	Mixed effects models

Study 1 (Chapter 3) explored whether formulaic sequences can be learned from watching television without subtitles or captions, and whether learning gains are comparable to those of single word vocabulary. Twenty EFL learners watched a 30-minute authentic English-language video containing 35 low-frequency words and formulaic sequences. Participants were tested on their knowledge of the target items before and after viewing. The

study adopted a within-participants design, in which learning from pre- to posttest was compared between target items from the video and distractor items occurring only in the tests. The study also explored the contribution of a number of item variables that have been found to predict formulaic knowledge (e.g., corpus frequency, part of speech, concreteness), and which might influence incidental learning from viewing. The results showed that learners' ability to productively recall the form of words and formulaic sequences improved significantly (6 out of 35 of the tested items were learned on average, or 17%). However, no significant gains were found in other tests (meaning recall, form recognition). No large differences emerged between learning gains of words and word sequences. We refrained from drawing strong conclusions about the effect of viewing on different knowledge aspects, because we suspected that test scores at the level of meaning recall partially reflected learning from one test to the next. Further, even though there was a two-week delay between the pre- and posttests, we still found (small) gains for distractor items, suggesting that part of the gains found in this study may have been due to learning from the tests.

Study 2 (Chapter 4) study focused on the learning of formulaic sequences from audiovisual input. The study was similar in design to Study 1, but improved on some of the limitations of the first study, (a) by expanding the sample of target items and distractors in order to increase statistical power, (b) by reducing the number of pre- and posttests in order to lower the chance of pretest effects, and (c) by using delayed instead of immediate posttests in order to measure durable gains. In line with the findings of Study 1, learners significantly improved their knowledge of the formulaic sequences tested in the pre- and posttests, with 10 items (18%) learned in the form recall test, and 7 items (13%) learned in the meaning recall test, on average. Learners also demonstrated knowledge of phrases that were not tested but which were freely recalled immediately after viewing. Importantly, the gains found in this study resulted from a single exposure during TV viewing without captions, and knowledge was retained for at least five days after viewing. In terms of the factors predicting learning gains, the statistical analyses showed that learning correlated with learners' vocabulary size, partial item knowledge (i.e., pretest score), corpus measures (frequency and MI), and semantic transparency.

Study 3 (Chapter 5) examined the effects of enhanced and unenhanced captions on incidental learning of formulaic sequences from audiovisual input. Participants watched a video containing captions in which some of the formulaic sequences were underlined. A form recall test was administered one week before and immediately after the treatment to measure learning. In addition, eye-tracking was used to examine learners' online processing of formulaic sequences in the two counterbalanced captioning conditions. We analyzed five eye movement indices (first pass time, rereading time, total reading time binary rereading, and binary skipping) in relation to the captioning conditions and learning gains. As was found in the previous studies, learners gained knowledge of formulaic sequences from watching the video, with 3.8 sequences (or 14%) learned on average. The eye-tracking results showed that learners spent longer reading underlined sequences, but, possibly due to the limited time the captions remained on screen, the effect was only evident in first pass times and rereading odds, but not in rereading times. Pretest knowledge also predicted reading times, either pointing to increased processing difficulty/effort, or to a salience-raising effect of the pretest. Finally, the results indicated that processing times more strongly predicted learning than enhancement.

Study 4 (Chapter 6) aimed to measure the effects of input enhancement in a more controlled way by focusing on (a) written input, (b) a single type of formulaic sequence (modifier-noun collocations), (c) multiple occurrences of target items in realistic contexts, and (d) low-frequency items that could be assumed to be unknown prior to the study (e.g., *capturing surface, pinhole effect*). This study asked whether the beneficial effect of typographic enhancement found in studies combining enhancement with input flooding (e.g., Szudarski & Carter, 2016) would hold when items were only enhanced in their first occurrence. Eye-tracking was used to examine the immediate and delayed effects of enhancement on learners' processing times during reading. In addition to form recall and form recognition posttests, a sentence-reading posttest similar to that used by Elgort et al. (2018) was included to measure the effect of the reading experiment on learners' receptive processing fluency of the target collocations. The eye movement data collected during reading indicated that the effect of enhancement on learners' reading times did not transfer to later exposures after a one-week delay. The sentence-reading posttest did not yield evidence of a processing advantage for collocations, although it did show a clear treatment effect at the level of single word processing, possibly reflecting a recency effect. Considerable learning gains were found at the level of form recognition (9 collocations or 38% learned on average), but not at the level of form recall (3 collocations or 14% learned on average). The interview data suggest that learners' strict focus on communicative meaning may have prevented the development of collocation knowledge.

7.2 Theoretical contribution of the four studies

7.2.1 Can formulaic sequences be learned simply by watching videos?

In the second chapter, I discussed empirical research showing that L1 and L2 learners alike are sensitive to sequential patterns in their language input, which may lead to an intuition for the frequency and conventionality of formulaic phrases (e.g., Northbrook & Conklin, 2021). However, formulaic knowledge has not been found at all levels of proficiency (e.g., Siyanova-Chanturia et al., 2011), and is not always evident at the level of productive use, as shown by studies examining formulaic patterns in L2 speech and writing (e.g., Paquot & Granger, 2012). Theories that emphasize the role of salience (e.g., Ellis, 2018) and attention/engagement in SLA (e.g., Laufer & Hulstijn, 2001) predict that the effects of incidental learning activities may vary strongly depending on how learners engage with the L2 input as well as on item and learner variables.

The first three studies of this thesis provided evidence that learners can acquire knowledge of formulaic sequences by viewing L2 television (see also Majuddin et al., 2021). The relevance of this type of input lies in its rich formulaic content (e.g., Lin, 2014), and its popularity and accessibility as a leisure activity among EFL learners (Peters et al., 2019; Peters, 2018). The learning gains of the three studies showed that, when audiovisual input is highly engaging, learners can learn formulaic sequences simply by watching television. The potential of audiovisual input for L2 acquisition had previously been established for learning single words (e.g., Peters & Webb, 2018), and the results of the three studies show that its beneficial effects extend to formulaic language. Formulaic sequences are not, by definition, too low-profile or structurally complex to be learned under the multimodal conditions of audiovisual input. However, item -, learner- , and input properties clearly played a role. Table 7.2 summarizes which variables predicted learning in the form recall test in each study, and the direction of the effect. The form recall test is the only measure of knowledge for which significant learning gains were found in all three studies.

Table 7.2 Effects of item variables in the form recall tests of Studies 1-3

	Study 1	Study 2	Study 3
Item variables			
Corpus frequency	+	+	n.s.
Mutual information	-	n.s.	n.s.
Collocate-node relationship (= verb-particle)	-	n.s.	
Semantic transparency		n.s.	+
Length	-		-
Phrase type (e.g., idiom, phrasal verb)	n.s.		
Congruency		n.s.	
Learner variables			
Pretest knowledge (meaning recall)		+	
Prior vocabulary knowledge	+	+	
Total reading time			+
Input variables			
Caption duration			-
Typographic enhancement			n.s.

Note. "+" indicates a positive estimate, "-" indicates a negative estimate, "n.s." indicates a

Looking at the contributions of predictor variables across studies, one clear trend emerges, namely the role of learners' pre-existing knowledge. In Study 2, the effect of pretest knowledge suggests that when some semantic knowledge of formulaic sequences is already in place, learners are more likely to develop knowledge at the level of form recall. Form recall of a vocabulary item is generally only learned once lower-level semantic knowledge has been mastered (Gyllstad, 2007). The finding that meaning recall predicted learning is in line with claims that incidental learning activities may serve to strengthen existing L2 representations (e.g., Horst et al., 1998). Thus, like L2 word knowledge (e.g., Webb, 2020), formulaic knowledge may develop incrementally from incidental exposure, reaching the level of productive recall once lower-level representations are in place.

Studies examining incidental collocation learning from reading have likewise reported a potential role for prior (partial) knowledge (e.g., Webb et al., 2013). Learning gains might thus be considerably higher for partially known than for entirely novel word sequences, although what novelty entails in the context of formulaic language is debatable. For example, collocation learning may be operationalized as learning a new association between two familiar words (e.g., Durrant & Schmitt, 2010), or as the combination of a familiar word and a novel (pseudo)word (Pellicer-Sánchez, 2017). At the semantic level, new word sequences may be matched with familiar concepts or communicative functions (e.g., Pellicer-Sánchez, 2017), or with entirely novel (e.g., technical) meanings (e.g., Sonbul & Schmitt, 2013; see also Study 4). In all of these situations, I would argue that some degree of prior knowledge may affect the learning process.

In line with findings for incidental collocation learning from reading (Vilkaitė, 2017; Vu & Peters, 2021), learners' prior vocabulary knowledge (VLT or VST score) significantly predicted learning gains (in studies 1 and 2). Lexical knowledge may affect incidental learning in various ways, as discussed in Chapter 2. Firstly, a larger vocabulary size may be facilitative of processing efficiency and automaticity, which may be of particular importance under the processing demands of spoken and audiovisual input (e.g., Goh, 2000). Further, learners with larger L2 vocabularies tend to have more extensive semantic networks, which may facilitate access to word meanings relevant to learning novel vocabulary (Bisson, 2021). With respect to L2 formulaic sequences, more lexically proficient learners may be familiar with the meanings of less frequent constituent words (e.g., *subliminal* in *subliminal effect*), or they may know more different word senses, which might facilitate learning of the figurative uses of frequent words (e.g., *pull* in *pull a muscle*).

Thirdly, corpus frequency of formulaic sequences positively predicted learning odds, in line with single word frequency effects on L2 processing and learning (Spätgens & Schoonen, 2020). This frequency effect suggests that learning was at least partly determined by how many times learners had encountered target items or their single word constituents prior to the learning intervention (see also Northbrook et al., 2021). Although corpus frequency is only a crude measure of prior knowledge, and cannot perfectly reflect the individual learner's language experience (e.g., Durrant & Schmitt, 2010), its positive effect shows that formulaic sequences which commonly appear in language input tend to be more familiar, which makes them more easily recognized and recalled after a single incidental exposure. Conversely, formulaic sequences that have never been encountered before, or that contain unfamiliar single word components, are unlikely to be recalled after a single exposure in audiovisual input. Considering the incremental nature of incidental learning, it is likely that knowledge of lower-frequency items will develop gradually, each new contextual exposure contributing to its representation (e.g., Hoey, 2005).

However, corpus frequency was not the only item variable predicting learning gains. As discussed in the second chapter, formulaic sequences vary along a number of phraseological dimensions (see e.g., Granger & Paquot, 2008; Siyanova-Chanturia & Van Lancker Sidtis, 2019). The selection of a wide variety of formulaic sequences allowed us to measure learning gains from a good number of phrases without having to manipulate the content of the authentic viewing input. However, the large variability in learning odds between items illustrates that formulaic sequences cannot be treated as a homogenous group.

Aside from corpus frequency, collocate-node relationship and semantic transparency significantly predicted learning odds. In studies 2 and 3, learners more easily recalled items (or their meanings) that were more compositional. Non-compositional phrases, in particular idioms, have been labeled as word-like in their form-meaning associations because they communicate "unitary semantic concepts" (Wulff, 2010: 225). Similar to novel words (Godfroid et al., 2013), idioms may attract learners' attention through their opaque form-meaning associations, but they are unlikely to be picked up when their meaning cannot be inferred from context or derived from their single word constituents. Because we measured learning from a single contextual exposure, semantic decoding was a crucial factor in successful recall. If learners are given more opportunity for semantic elaboration, for example through repeated contextualized exposure, the high salience of idioms might be beneficial for long-term retention. However, in a single encounter, semantic non-compositionality is clearly an inhibiting factor in incidental acquisition of word sequences from viewing.

In summary, it seems that what can be learned from viewing L2 audiovisual input largely depends on what learners already know about the lexical and semantic properties of the formulaic sequences they encounter. In line with associative learning theory, what learners already known will determine how they perceive the L2 input, and which elements of the input are noticed and learned (e.g., Cintrón-Valentín & Ellis, 2016). Learning formulaic sequences during viewing is no exception, although the temporal and multimodal conditions of audiovisual input may additionally impact how learners process formulaic sequences. This was evident in Study 3, where we observed that attention enhancement only had a minor effect on learning gains.

7.2.2 What is the role of attention in learning formulaic sequences from viewing?

By analyzing learners' eye movements during caption reading, Study 3 provided more insight into the role of attention in the incidental learning process. The results showed that typographic enhancement increased learners' visual attention for target items, but only had a minor effect on learning gains, compared to item variables (e.g., semantic compositionality) and amount of attention.

Our findings partly concur with those of Majuddin et al. (2021), who found that the effect of enhanced captions on form recall of formulaic sequences was minimal both in an immediate and a delayed posttest. Together, the findings of the two studies suggest that typographic enhancement has a less outspoken effect on L2 learning of formulaic sequences in audiovisual input than in written text (e.g., Choi, 2017; Szudarski & Carter, 2016). Majuddin et al. suggested that, compared to reading, the "real-time nature of viewing entails that learners have less time to fixate anything, including typographically enhanced items". In line with this suggestion, our eye-tracking data showed that learners had little time to elaborately process the formulaic sequences during caption reading. Learners only reread around 23% of captions, and the main predictor of rereading times was caption duration, indicating that the opportunity for reanalyzing difficult formulaic sequences was limited. Further, if we compare the average reading times on enhanced (876 milliseconds) and unenhanced (654 milliseconds) items in Study 3, to the average reading times (1,243 milliseconds for enhanced items, 943 milliseconds for unenhanced items) for the first exposure in Study 4, we see that learners with a similar profile took more time to process both enhanced and unenhanced word sequences in written texts. Of course, phrase properties and learner variables (e.g., proficiency level) may partly explain differences in processing times of Study 3 and 4. Still, it seems that enhancement alone was not enough to promote retention of novel formulaic sequences under the processing demands of captioned audiovisual input.

The analysis of learning odds in Study 3 also showed that semantic compositionality was a strong predictor of learning. Less compositional sequences were less likely to be recalled after viewing, regardless of whether they were visually enhanced. This concurs with Montero Perez et al.'s (2014, 2015) finding that typographically enhanced captions may not promote the development of semantic knowledge of novel words. The transient nature of spoken input implies that learners must rapidly process relevant information in working memory before it is overwritten by new information (Christiansen & Chater, 2016). The opportunity to engage with semantically opaque items such as novel words and idioms during viewing is therefore limited. Even if learners notice a difficult phrase while watching captioned video, they may not attempt to find its meaning, unless they can pause the video and look up the meaning online or in a dictionary (Montero Perez et al., 2018).

The fact that semantic transparency was a stronger predictor of learning gains than typographic enhancement is also in line with Cintrón-Valentín et al.'s (2019) suggestion that the effectiveness of enhancement in captioned video may depend on the complexity of the linguistic features under investigation. Another noteworthy finding is that reading times were significantly predicted by pretest scores. Items that were unknown in the pretest received longer first pass times in both the enhanced and unenhanced condition, indicating that unknown sequences were more difficult to process (see also Choi, 2017). Further, in post-experiment interviews, some learners reported that they recognized items from the pretest in the captions. It seems that learners were able to use the enhanced captions efficiently to study unknown word sequences more closely. At the same time, this finding also suggests that attention enhancement by itself only had a minimal effect on learners' processing of formulaic sequences. Other factors, such as degree of familiarity and semantic compositionality more strongly affected how elaborately learners processed target sequences.

In spite of the negligible effect of typographic enhancement on learning odds, we did find a significant, positive relationship between learning odds and total reading times on target items, which suggests that learners' amount of engagement with formulaic sequences determined the likelihood of recalling those sequences after viewing. This finding is in line with theories of SLA that emphasize the importance of engagement and elaborate processing during incidental learning (e.g., Laufer & Hulstijn, 2001; Schmidt, 1990). It also corresponds with Choi's (2017) finding of a positive relationship between total reading time and collocation learning during reading, as well as with the positive effects of visual attention found for incidental learning of L2 vocabulary (e.g., Godfroid et al., 2013, 2018) and grammar (e.g., Lee & Révész, 2019). It appears that increased engagement with, or rehearsal of word combinations may promote retention of the lexical components of formulaic sequences, resulting in improved form recall immediately after viewing.

Because both Majuddin et al.'s (2021) study and our own study focused on a range of different phrase types, the non-significant effect of enhancement might partially be explained by the high level of difficulty of more idiomatic phrases. It seems plausible that typographic enhancement of formulaic sequences might benefit learning of transparent sequences (e.g., collocations) that remain on screen long enough to induce additional processing or rehearsal in working memory. An important caveat in the study of fixation times during L2 processing is the ambiguous relationship between processing duration and learning. In one of the first studies to examine the relationship between visual attention and incidental vocabulary learning, Godfroid et al. (2013) operationalized amount of attention as total fixation time, based on the assumed importance of elaborate processing, or rehearsal, in the formation of long-term memory traces (Robinson, 1995; 2003). Godfroid et al. (2013) predicted that a positive correlation would arise between total reading times on novel word forms and long-term retention of those word forms. At the same time, fixation times also reflect processing difficulty: more familiar and predictable elements of a text receive shorter reading times (e.g., Conklin et al., 2018). Accordingly, Godfroid et al. (2013) also found significantly higher reading times for novel words compared to familiar words.

Because increased processing difficulty and amount of engagement may both induce longer reading times, the relationship between eye-tracking measures and learning of L2 features cannot always be interpreted in a straightforward manner. One the one hand, processing effort or elaboration is likely to increase learning odds of novel words and structures. On the other hand, elements of a text that are processed more easily (e.g., familiar word sequences) may also be memorized and/or recalled more easily. This makes the interpretation of eye-tracking data particularly difficult when target items vary in terms of their degree of familiarity or novelty, as was the case in Study 3. Further, by no means does elaborate processing guarantee retention of target forms or structures (e.g., Winke, 2013). When learners experience greater difficulty in processing a phrase or grammatical structure, they may not be able to learn the novel form or form-meaning association incidentally without additional information or external support. Although Study 3 found a positive relationship between processing times and learning odds, items which received longer reading times (e.g., opaque items such as *guinea pig*) were not necessarily learned better. The contrary was often true, as indicated by lower learning gains for less compositional items.

In conclusion, the results of Study 3 showed that attention enhancement alters how learners process formulaic sequences, and that this may in some cases result in the kind of processing that enables retention of form. However, while typographic enhancement may cause learners to notice formulaic sequences, what learners do with the enhanced information depends on the degree of familiarity with the items, among other variables. Typographic enhancement by itself is by no means guaranteed to stimulate elaborate processing and retention of form (e.g., Leow & Martin, 2017; Winke, 2013). Further, increased attention may promote form recall of formulaic sequences, but is likely to only significantly affect learning when the meaning of a phrase is accessible, and when learners have sufficient time to rehearse a phrase in memory.

7.2.3 How does attention enhancement affect collocation learning when encounters are spaced?

The first three studies demonstrated the importance of item variables and engagement with formulaic sequences in incidental learning of formulaic sequences. A possible drawback of these studies is their focus on a single contextualized exposure. Because conscious engagement with L2 form has strong, instantaneous effects on learning (e.g., Ellis, 1993), its effects are more likely to be perceived in short incidental learning interventions, such as those examined in Studies 1-3. However, incidental learning is also determined by exposure frequency. What is retained from a single learning event needs to be consolidated and expanded in order to have meaningful effects on L2 processing and use.

Previous studies examining the effect of input enhancement on incidental learning of formulaic sequences have typically combined typographic enhancement with some form of input flooding (e.g., Northbrook et al., 2021; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016), showing that the combination of repetition and enhancement results in strong, durable gains that are likely to affect further processing. However, in non-flooded L2 input, for example in academic reading materials, online videos, etc., the occurrence rate of formulaic sequences is typically much lower (e.g., Durrant & Schmitt, 2010), and the positive, long-term effects of enhancement may not generalize to contexts where enhancement is only used with a single exposure.

The aim of the final study of this thesis was to examine the durability of the effect of enhancement on processing and learning of formulaic sequences during reading. The study focused on written text because this facilitated the experimental control of variables related to the target collocations, including the number of occurrences and collocate-node relationship. Further, in this study we examined the effects of incidental learning on explicit recall and recognition, as well as on processing fluency. Following recent studies (Conklin & Carrol, 2020; Norhbrook et al., 2021), the final study used eye-tracking to examine how repeated exposure during reading affected learners' processing times on novel L2 collocations.

The results of Study 4 confirmed the immediate attention-raising effect of typographic enhancement that was also found in previous studies (e.g., Choi, 2017). However, the results of the eye-tracking data, posttests, and post-experiment interviews clearly indicated that enhancement did not have a lasting effect on learners' memory for target collocations. In the second reading session, no clear differences emerged between enhanced and unenhanced items in terms of processing times, and for both item conditions, reading times on collocations decreased with each exposure at an equal rate.

Considering the strong effects of enhancement on collocation learning in previous studies using written texts (e.g., Szudarski & Carter, 2016; Toomer & Elgort, 2019), the minimal effect of typographic enhancement appears to follow from the fact that only a single exposure in the texts was enhanced. This may have led to a different interpretation of the function of enhancement, as suggested by the interview data. Unaware that they would be tested on the form of the L2 collocations, participants read the target items in function of the comprehension questions at the end of the experiment. Therefore, it seems that typographic enhancement in this experiment did not lead to "semi-incidental learning", because it did not cause learners to pay attention to the form of the target phrases (Pellicer-Sánchez & Boers, 2019).

This raises a few practical questions with regard to the use of typographic enhancement in L2 learning of formulaic sequences. Firstly, it could be argued that the combined effects of typographic enhancement and input flooding are only representative of learning contexts that provide massed exposure to formulaic sequences. However, in naturalistic L2 input, formulaic sequences tend to be widely dispersed and are unlikely to be encountered multiple times within the same text (Durrant & Schmitt, 2010). Of course, there are exceptions to this rule. For example, highly frequent lexical bundles and speech formulae may appear repeatedly in spoken conversations due to priming between interlocutors (e.g., Bybee, 1998). Technical and content-specific collocations may recur within or across texts of the same genre or texts that deal with the same general topics. Nevertheless, most formulaic sequences that are relevant to L2 acquisition are unlikely to be encountered repeatedly within close distance of each other.

This means that, in authentic, non-flooded L2 input, the repeated use of typographic enhancement may have minimal practical relevance.

Secondly, the use of typographic enhancement has typically been operationalized without giving learners explicit instructions to pay attention to target items. On the contrary, in most previous studies a comprehension test was announced to ensure incidental learning conditions, although there are a few notable exceptions (e.g., Montero Perez et al., 2015; Peters, 2012). For example, Montero Perez et al. (2015) found a significant interaction between the effects of enhanced captions and test announcement on learners' visual processing of novel words. When learners were forewarned of a vocabulary posttest, reading times on enhanced L2 words were significantly longer compared to a purely incidental learning condition. This finding shows that task demands may be at least as important in directing learners' attention to relevant linguistic features as visual, bottom-up attention enhancement (e.g., Laufer & Hulstijn, 2001; Schmidt, 1990). Like item variables and prior knowledge, task goals may affect how learners engage with word sequences in their L2 input. Therefore, we might have found entirely different results if typographic enhancement in Study 4 had been combined with the announcement of a collocation test, for instance.

The findings of Study 4 can also be understood in relation to the effect of processing efficiency and attention allocation when cognitive demands of a learning task are high. Learning novel collocations attached to complex scientific concepts may have been too cognitively demanding, and the difficulty of the scientific topics of the reading texts may have contributed to a strong focus on communicative meaning, and less attention for the lexical co-occurrences in the texts. A semi-incidental approach in which learners have to switch their conscious attention between the topic of the text and novel vocabulary or formulaic sequences may not be as effective when the cognitive demands of the reading task are already high. Giving learners an additional language-focused task may then result in attentional trade-off effects such as those observed by Choi (2017). However, because we employed a within-participants design, there is no way of finding out whether this was also the case in Study 4.

7.2.4 How fast might a processing advantage arise for L2 collocations encountered during reading?

The short answer is: "I don't know". Study 4 explored the use of a sentence-reading posttest, similar to that employed by Elgort et al. (2018) in their study on contextual vocabulary learning, to measure the effect of repeated exposure to L2 collocations during reading on implicit collocation knowledge. The premise of this test was the formulaic processing advantage often observed in L1 (e.g., Carrol & Conklin, 2020) and L2 reading (e.g., Siyanova-Chanturia et al., 2011), and the role of frequency and association strength underlying this processing advantage (e.g., Öksüz et al., 2020). Two recent studies found that the online processing of novel phrases is highly sensitive to repetition effects even after a few exposures during reading (Conklin & Carrol, 2020; Northbrook et al., 2021), in line with continuous accounts of multiword frequency effects such as lexical priming theory (Hoey, 2005). In contrast to these studies, we did not find a clear effect of reading on learners' processing times for L2 collocations in the sentence-reading posttest.

In the previous chapter, I offered a few explanations for why no significant interaction was found between item condition (target item, control item) and group (experimental group, comparison group). Aside from limitations associated with the posttest of reading fluency (see below), there are a few important differences between our study and the two previous studies that found a repetition effect on formulaic processing. One factor which I believe might help explain the different findings is the level of difficulty of the reading texts and of the target items. Conklin and Carrol (2021) examined the effect of repeated exposure to binomials on native speakers' online processing during reading. The target items in their study were highly transparent sequences that did not refer to novel concepts but had relatively straightforward form-meaning mappings (e.g., *pipes and wires, grass and leaves*). The same was true of the lexical bundles in Northbrook et al.'s (2021) study (e.g., *set off home, tired and hungry*), which appeared in simple short stories matched with the beginning learners' proficiency level.

Although most of the target items in Study 4 had a high degree of semantic transparency, they were associated with quite complex scientific concepts. Learning implicit associations between words may be impeded by high cognitive demands of a learning task (Conklin, 2020; Turk-browne, Junge, & Scholl, 2005). This may explain why no clear evidence of a processing advantage was found for the target collocations in Study 4. However, the faster reading times found for target collocations in the experimental and comparison groups suggest that the treatment did affect how learners processed the component words of the target collocations. Repeated exposure to words that appeared in the contexts of the reading materials primed their processing in short sentences, which presented the words in similar contexts. Although an additional control group not reading any experimental text would have strengthened this interpretation, it seems that the treatment did have some effect on learners' processing times in the sentence-reading task. Therefore, it is possible that a processing advantage would have arisen if learners had had more exposure to the technical collocations and their contexts. In academic reading materials, which also tend to have a high level of difficulty, fluent access to technical phrases may similarly only be achieved once learners are more familiar with the content of the academic texts the phrases appear in.

7.3 Pedagogical contribution the four studies

As explained above, the effects of typographic enhancement may vary considerably depending on how learners perceive the target items, and how they approach the learning activity. The results of Study 4 clearly show that learners do not necessarily interpret typographic enhancement as an invitation to attend to the association between two words, or to a word sequence as a whole, in the same way that it does not necessarily encourage them to find abstract grammatical patterns underlying enhanced elements of a text (e.g., Winke, 2013). Although enhancement certainly promotes noticing of the enhanced part of a text, as illustrated by the eye movement data in studies 3 and 4, rehearsal of the relevant form in memory depends on what learners interpret as relevant in relation to the task demands and their personal aims and background knowledge (Leow & Martin, 2017).

Assuming that typographic enhancement is a non-obtrusive method of drawing attention to relevant phrases, this form of input enhancement may ideally be combined with additional learning activities that promote learners' engagement with the target collocations and their contexts. This may be done simply by giving learners the explicit instruction to focus on enhanced collocations, or through deliberate learning activities. Such activities may include exercises that strengthen knowledge of the target collocations after reading or viewing, such as gap-fill exercises (e.g., Boers, Dang, & Strong, 2017; Stengers & Boers, 2015) or memorization tasks (e.g., Lindstromberg & Boers, 2008).

To support L2 learning of semantically opaque or idiomatic expressions, the use of activities that provide access to meaning, such as flashcards, may also be helpful (see Obermeier & Elgort, 2020). Further, data-driven learning approaches could be used to promote learning of meaning in a more contextualized manner. For example, concordances provide quick access to appropriate contextual uses of formulaic expressions, which may help learners familiarize themselves with the semantic and functional properties of word sequences that have low occurrence frequency in the input, without relying on input flooding (e.g., Cobb, 2019; Meunier, 2020). All of these types of learning activity have been found to boost learners' formulaic knowledge in their own right (Boers, 2020; Pellicer-Sánchez & Boers, 2019),and could be used in tandem with typographically enhanced L2 input.

Despite its potential benefits for learning, typographic enhancement has a number of potential disadvantages that need to be considered. First, researchers have argued that the attention-raising effects of typographic enhancement might hinder the implicit learning process underlying frequency effects in processing of formulaic sequences (Conklin, 2020; Toomer & Elgort, 2019). Although more evidence is needed to test this claim, it seems that negative effects of this kind may be more likely in learning activities that are cognitively demanding, for example in reading texts or videos that deal with complex topics.

Another limitation of typographic enhancement is that it relies on an external decision about which phrases an individual L2 learner should explicitly focus on. A clear trend that emerged from the studies in this thesis is that learners' prior knowledge strongly affected how they processed and learned formulaic sequences in the L2 input. Phrases that are relevant to one learner may be well-known or irrelevant to another learner, even among learners of the same L1 background and study program. Although teachers, corpus measures, and dictionaries could help identify potentially relevant phrases (e.g., Ellis et al., 2008; Martinez & Schmitt, 2012), it may prove difficult to enhance phrases that are useful to all learners without at the same time distracting from other relevant information in the input (Choi, 2017; Toomer & Elgort, 2019). More research into the potential trade-off effects of typographic enhancement is needed if we want to make sound recommendations about its use in L2 learning materials (Pellicer-Sánchez & Boers, 2019).

Based on the findings of the current thesis and the wider literature, it seems that purely incidental activities may already go a long way in expanding learners' formulaic repertoire (e.g., González Fernández & Schmitt, 2015). L2 learners appear to be just as sensitive to the sequential patterns in their input. In accordance with the principles of focus on form and assuming that both implicit and explicit processes can contribute to successful L2 acquisition, the best approach to L2 acquisition of formulaic sequences is likely to involve a good balance between exposure to authentic and representative L2 input, and external support where exposure alone falls short. The challenge lies in developing forms of support that work efficiently and do not detract from the natural communicative experiences that are offered by authentic L2 input.

7.4 Methodological contribution of the four studies

7.4.1 Pretesting effects

As in previous studies that used a pretest-posttest design, learning gains were compared between an experimental and a control condition (i.e., distractor items) to account for potential learning from the tests or outside the treatment. In addition to this statistical control, we also included questionnaires that allowed us to qualitatively examine the causes of learning effects for distractor items (and target items). For example, by asking students whether they had recognized formulaic sequences from the pretests during viewing, we could get an rough idea of how the pretest had affected their engagement with the target items.

The inclusion of questionnaire data was an invaluable addition to statistical control, because it revealed that pretesting effects existed even though these did not strongly affect gains for distractor items. In Study 2 (Chapter 4), it was found that the form recall test affected learners' responses in the meaning recall test, although a distractor task (i.e., the VLT) was inserted between the two tests. This reveals that using multiple tests which measure similar aspects of knowledge may lead to testing effects even when efforts are made to flush the learners' memory in between tests.

In addition, the questionnaire results, in particular those of Study 3 (Chapter 5), indicated that the use of a pretest may also have moderated the effect of the treatment, even though there was a one-week interval between the pretest and the eye-tracking experiment. This reveals an important limitation of the way in which pretesting effects are usually controlled in studies on incidental vocabulary learning. Using a traditional control group, or distractor items that only appear in the tests, do not account for the moderating effect of pretesting on the experimental treatment. Alternative approaches exist, such as Solomon's (1949) four-group design, which includes several control groups to account for learning from the pretest itself as well as the effect of the pretest on the experimental treatment. The use of questionnaires to account for testing effects may inspire future research to include similar verbal reports when adopting a pretest-posttest design, or to include additional control groups, in particular when different types of prior knowledge control (e.g., pseudowords) are deemed inappropriate or unfeasible.

7.4.2 Eye-tracking measures

In studies 3 and 4, learners' eye movements were measured in order to gain insight into the cognitive processes underlying incidental learning of formulaic sequences. In Study 3, eye-tracking was used to examine how learners visually processed formulaic sequences during caption reading, and to study the relationship between attention and learning gains. As discussed in Chapter 1, eye-tracking had been used in previous research to investigate incidental vocabulary learning from reading (e.g., Elgort et al., 2018), L2 processing of formulaic sequences (e.g., Siyanova-Chanturia et al., 2011), and multimodal processing (e.g., Montero Perez et al., 2015). Each of these topics has its own theoretical assumptions regarding the relationship between eye movements and cognitive processes. Because Study 3 focused on (a) caption reading in audiovisual input, (b) multiword unit processing, and (c) the relationship between attention and incidental learning, the set of eye-tracking indices used in this study was tailored to the conditions of all three topics.

Firstly, I included a binary measure of rereading to account for the short duration of captions. Binary rereading, or the odds of refixating a caption or a specific target item in the caption, was assumed to reflect re-analysis as a result of processing difficulty or increased attention, whereas rereading times reflected the degree of elaboration during re-analysis within the time constraints of the caption presentation. Both indices were associated with different independent variables, which confirmed that their inclusion provided a more complete account of learners' caption processing. Secondly, I included a measure of single word skipping. Unlike final-word skipping, a measure associated with predictability, for example in idiom processing (e.g., Carrol & Conklin, 2017), single word skipping may reflect the odds that learners read both words of the formulaic sequences instead of only one, due to increased visual salience. This measure was found to be associated with different variables than the other indices. Finally, total reading time was analyzed as an independent variable in relation to learners' posttest scores. The results of this analysis showed that reading times on target items were positively associated with learning odds, in line with Choi's (2017) findings for L2 collocations in reading.

In Study 4, eye-tracking also served a dual function, as a measure of collocation processing during the experimental treatment, and in a sentence-reading posttest. As a measure of collocation processing during reading, eye-tracking measures were included in a similar manner as in the study by Pellicer-Sánchez (2016), by tracking the relationship between exposure count (1-8) and reading times. In our study, we additionally analyzed the interaction between exposure count and typographic enhancement, to find out if reading times on target collocations changed between exposures in the same manner in the enhanced and unenhanced conditions. We additionally adopted a sentence-reading posttest similar to that used by Elgort et al. (2018), to compare reading times on target collocations and control items in short decontextualized sentences. As a first attempt to use this method with multiword expressions, the test proved useful for measuring the effect of the reading task on processing fluency. Unfortunately, lack of control of lexical components of the target and control items meant that we could not draw strong conclusions about collocation knowledge based on the results of this test. This test may be improved upon by (a) controlling the input frequency of the single word constituents, and (b) checking a priori differences in processing times between target items and control items, for example by giving the test to a control group (instead of, or in addition to the comparison group).

7.4.3 Counterbalancing

Studies 3 and 4 used a counterbalanced within-participants design to measure the effects of typographic enhancement on reading times and learning gains. Most previous studies examining the effects of input enhancement on L2 vocabulary learning adopted a between-participants design (e.g., Choi, 2017; Montero Perez et al., 2015; Szudarski & Carter, 2016). In a counterbalanced design, each participant is exposed to both experimental conditions, so that participants act as their own control. An advantage of this approach is that individual differences have less impact on the comparison of the two conditions. This results in higher statistical power compared to a between-participants design with the same number of observations (Godfroid, 2020).

In a study on the effects on typographic enhancement, counterbalancing involves the creation of two versions or lists of the learning materials, each of which contains enhanced and unenhanced items. This results in an additional design choice regarding the distribution of enhanced and unenhanced items in the input. In Study 3, we split the experimental video in an enhanced and an unenhanced part. Because the two parts had to be viewed in chronological order, this meant that participants assigned to version 1 watched the first part of the video with enhancement, whereas participants assigned to version 2 watched the second part of the video with enhancement. At the risk of generating an order effect (i.e., an interaction between presentation list and the effect of enhancement), this approach was preferred because it minimized potential trade-off effects in which increased attention to enhanced items would bring about decreased attention to unenhanced items, which might lead to an overestimation of the enhancement effect. Because we used authentic viewing materials, we could not control the spacing between two target items.

In Study 4, care was taken to ensure enough space between the first occurrences of two target items, so that no attentional trade-offs would confound the enhancement effect. Therefore, in the first reading text of Study 4, enhanced and unenhanced items could be alternated. This meant that, in version 1 of the text, the even items (in terms of their order of occurrence) were enhanced, whereas in version 2, the odd items were enhanced. There may be no single best method of enhancing items in a counterbalanced design, and much will depend on properties of the input used, as our studies illustrate. The use of multiple short paragraphs or videos might give the highest degree of control over trade-off effects while avoiding order effects, but these may not always be most suitable if the goal is to reflect real-life reading or viewing situations. On the other hand, if the aim is to examine trade-off effects of typographic enhancement, then these two types of design (alternating vs. blocking) might be usefully combined.

7.5 Limitations and suggestions for future research

7.5.1 Pretesting effects

The three studies of the PhD used authentic audiovisual input, instead of, for example, manipulating the soundtrack or creating videos from scratch (see Cintrón-Valentín et al., 2019 for a good example). The gains found in these studies should thus be a fairly close reflection of how learning formulaic sequences might happen when EFL learners watch videos for leisure or during an English class. However, an important limitation of Studies 1-3 concerns the use of pretests to measure learning gains for formulaic sequences. Although this allowed us to track learning a wide range of (partially known) formulaic sequences without manipulating the audiovisual input, the learning gains found for the control items indicate that pretesting may have led to enhanced attention for target items during viewing.

One way in which pretests may contribute to the learning process is by identifying gaps in learners' knowledge of target items, which might then receive more attention during viewing one week later (e.g., Swanborn & de Glopper, 1999). This may be particularly problematic when formulaic sequences are partially known. For example, when learners can provide one of the single word constituents in a form recall pretest, the activation of the relevant knowledge structure, i.e., a partial form-meaning association, may carry over to the treatment. Further, the interview data of Study 3 suggest that the attention-raising effect of the pretest may have interacted with the effect of typographic enhancement. By highlighting phrases that appeared in the pretest, the enhanced items may have induced increased efforts to memorize the formulaic sequences in anticipation of a posttest. Although the inclusion of control items accounted for pretest effects to some extent, it is likely that pretesting made the learning conditions of the first three studies less incidental.

7.5.2 Auditory processing

Another limitation of this thesis is that it focused entirely on visual processing, without examining the contribution of the spoken modality. A few studies have shown that the simultaneous processing of spoken and written text tends towards asynchrony/misalignment (Conklin et al., 2020; Wisniewska & Mora, 2018). Both native and nonnative speakers seem to take advantage of the visual text by reading words ahead of their audio presentation. This trend has been found in processing of bimodal texts (Conklin et al., 2020) and video with captions (Wisniewska & Mora, 2018). Presumably, learners tend to read ahead of the audio in order to pre-process words orthographically before integrating the two modalities in memory, which may optimize processing efficiency (Bird & Williams, 2003). However, as Conklin et al. (2020) pointed out, this begs the question to what extent learners might use prosodic cues such as utterance boundaries to extract formulaic sequences from multimodal input (see Lin, 2012). Future studies might investigate how or to what extent learners make use of spoken and written modalities to learn formulaic phrases that are prosodically marked in speech. One way of doing this could be to use a paradigm similar to that developed by Bird and Williams (2003), to find out if the benefits associated with bimodal word processing extend to formulaic sequences.

7.5.3 Triangulation

Further, following studies such as Jung & Révész (2018), it may also be worthwhile to combine eye-tracking with stimulated recall, in order to gain a more complete account of the extent to which learners deliberately focus on formulaic forms during meaning-focused activities, while minimizing reactivity of the attention measures. The studies of the current thesis applied minimal triangulation through the combination of eye-tracking data and interview and questionnaire results. However, experiences gained while collecting eye-tracking data showed that much can be learned from participants' subjective recounts of their reading/viewing experience. For example, when learners were shown their eye movement data in the debriefing phase, many commented on the perceived level of difficulty of particular passages. A stimulated recall procedure focusing on learners' own eye-tracking results could be a worthwhile avenue for future research, for example to gain a better understanding on the degree of attention learners devote to formulaic sequences as compared to single word vocabulary.

7.5.4 Within-participants vs. between participants designs

In the studies focusing on typographic enhancement, different learning conditions were operationalized in a counterbalanced, within-participants design. This approach optimizes statistical power because each participant serves as their own control (Godfroid, 2020; Plonsky & Oswald, 2014). However, in a within-participants design, differences between the enhanced and unenhanced conditions can only be examined at the level of target items, but not at higher levels, e.g., effects on global processing strategies or comprehension of the wider discourse. It is possible that enhancement might cause learners to pay closer attention to formulaic language more generally, or, conversely, that it causes attentional trade-off effects (see e.g., Choi, 2017). For example, in the context of caption processing, Lee and Révész (2021) suggested that typographic enhancement might lead to increased caption reading. Although we did not observe any clear differences between the enhanced and unenhanced parts of the input in terms of global caption reading behavior, it may be interesting to test Lee and Révész' hypothesis with learners from different populations or subtitling traditions, to see if enhancement may indeed affect caption use. Future studies might also use a between-participants manipulation to examine the effect of enhancement on input comprehension.

7.5.5 Focus on form

Another limitation of Study 4 is that typographic enhancement was incorporated in an incidental learning activity without any explicit instruction to focus on form. As explained above, the effectiveness of typographic enhancement may strongly rely on the goals and demands of a reading task. Further, language teachers are unlikely to use enhancement without any form of explicit instruction or additional learning task to support L2 acquisition. Therefore, we must be careful in interpreting the lack of a significant effect of attention enhancement in Study 4. Future studies may either include an intentional learning condition to measure the interaction between task goals and input enhancement (e.g., Montero Perez et al., 2015), or they may move away from the traditional incidental approach by including explicit instructions to focus on meaning in all (enhanced) conditions.

7.5.6 Measuring the effect of a treatment on collocation processing fluency

The tests used to measure collocation knowledge in the four studies also come with a number of limitations. Firstly, the limitations of the sentence-reading posttest in Study 4 will need to be addressed in future research. I believe a reading fluency posttest can be a valuable measure in principle, but there is much room for improvement and elaboration. Firstly, as was discussed in the limitations section of Chapter 6, future studies should control the input frequency of single word constituents, as has been done in studies investigating collocation priming (e.g., Durrant & Schmitt, 2010; McKoon & Ratcliff, 1979).

Further, it may be interesting to also control the occurrence frequency of the full collocation, in a similar way as Conklin and Carrol (2020) did in their design. In addition, to control the effects of lexical variables future studies may counterbalance two alternative forms of each collocation in the reading texts, instead of only in the sentence-reading posttest. Conklin and Carrol (2020) counterbalanced two forms of each binomial (*wires and pipes* vs. *pipes and wires*) to control for prior order preferences. However, this type of control is more difficult with collocations than with binomials, since we cannot simply reverse the order of the lexical components to create an alternative collocation form. In line with the Firthian principle of meaning through collocation, achieving perfect semantic equivalence between two lexically varied forms is impossible. For that reason, it would be challenging to insert two lexically varied collocations into the same contexts. A solution could be to use target items containing one semantically empty or uninformative collocate, such as verb-noun collocations containing a delexical verb (e.g., *make* vs. *take*), or adjective-noun combinations containing an easily substituted adjective (e.g., evaluative adjectives such as *fantastic, lovely, disastrous*), although these may be more difficult to learn compared to more semantically informative collocations (e.g., Szudarski, 2012, see also Appendix 4).

7.5.7 Measuring productive use of formulaic sequences

In all but the final study, tests of declarative knowledge of formulaic sequences were used to measure the effect of incidental learning activities. Productive knowledge was measured using tests of form recall, which reflect learners' ability to retrieve a word sequence from memory when prompted by a definition or L1 translation. Producing the correct form in a controlled context is considerably less challenging than spontaneous production in speech or writing, which relies on more automatic access to target sequences. Because errors on the use of formulaic sequences are typically found in L2 production, more research is needed that investigates the effects of different learning activities on spontaneous productive use (e.g., Peters & Pauwels, 2015).

A challenge for future studies is finding ways to measure the effects of incidental learning activities on automatic or procedural knowledge at the productive level. This will be important to find out to what extent incidental activities with and without input enhancement may contribute to the free production of formulaic sequences, by affecting not just a form-meaning association, but also the efficient and automatic use required by L2 communication. Studies on incidental learning of grammatical constructions (e.g., Lee & Révész, 2020) have measured productive use by creating obligatory contexts for the constructions in speaking and writing tasks. However, eliciting production of specific formulaic sequences is more challenging (e.g., Duong, Montero Perez, Desmet, & Peters, 2021). A possible solution could be to study the effects of a longitudinal incidental learning intervention through the analysis of a spoken or written learner corpus that can be directly linked to the L2 input (see Candarli, 2020 for a good example in a formal learning context).

7.5.8 Longitudinal research

Finally, although the first three studies of this thesis explored the effects of viewing on L2 formulaic knowledge, the immense popularity of audiovisual input among EFL learners calls for more research into this type of input. For example, the effect of exposure frequency in audiovisual input has only been studied through repeated viewing (Majuddin et al., 2021). Although formulaic sequences do not typically recur often in authentic L2 input, a possible direction for future study might be to longitudinally examine the effects of narrow viewing on L2 acquisition of formulaic sequences. Related television programs contain many exposures the same, content-related single words (Rodgers & Webb, 2011), and the same may be true for technical or content-specific formulaic sequences. If formulaic sequences can be learned incidentally from contextual exposure without enhancement (e.g., González Fernández & Schmitt, 2015; Northbrook & Conklin, 2020), then this would be an easy and motivating way of increasing learners' formulaic knowledge. However, more data is needed to draw conclusions about the long-term effects of specific activities such as viewing.

7.6 Conclusion

I have presented the findings of four studies into the effects of incidental learning activities on the development of L2 formulaic knowledge. When learners engage in L2 viewing, they may acquire knowledge of formulaic sequences with relative ease, even from watching a single episode of L2 television. Enhancing formulaic sequences in a captioned video may cause learners to pay closer attention to formulaic sequences, and may thus indirectly contribute to learning gains. However, factors related to prior knowledge, and the semantic and lexical properties of formulaic sequences, are at least as important in determining the outcome of a viewing activity. Further, as shown in Study 4, increased visual attention does not always have the intended effect, and does not necessarily contribute to learning across exposures. Acquiring productive knowledge of formulaic sequences from minimal exposure comes down to a high degree of cognitive engagement, in line with Schmidt's (1990) noticing hypothesis. This engagement can only minimally be supported by bottom-up salience raising methods such as typographic enhancement.

In all likelihood, the learning effects found in the four studies of this thesis only scratched the surface of what can be achieved from engaging with L2 formulaic language in informal activities. What is now needed, is a more elaborate investigation of viewing, reading, and other activities a learner might engage in to expand their formulaic knowledge. To achieve a better understanding of the effects of variables such as frequency, salience, and learner engagement, there is a need for larger-scale and longitudinal research. In line with recent calls for tacit measures of vocabulary knowledge (Elgort & Warren, 2014), Study 4 explored the use of a sentence-reading posttest measuring the effect of reading on formulaic processing fluency. This study illustrated the difficulty of testing and interpreting the effects of a learning intervention on reading times. In future research, implicit measures will need to be expanded and improved upon, and more attention will need to be given to productive use of sequences in free or spontaneous production.

Researchers still have long road ahead in determining the benefits of different incidental learning activities, but for now, we may be carefully optimistic about the potential of reading and viewing for learning the formulaic sequences of the L2, with or without enhancement. The findings of this thesis are encouraging, given that informal activities may provide learners with an extensive amount of representative and authentic uses of formulaic language. In the long term, such activities may contribute generously to the development of proficient and idiomatic second language knowledge.

References

Aijmer, K. (2009). 'So er I just sort of I dunno I think it's just because...': A corpus study of 'I don't know' and "dunno" in learner spoken English. In A. H. Jucker, D. Schreier, & M. Hundt (Eds.), *Corpora: Pragmatics and discourse* (pp. 151–166).

Alotaibi, S., Pellicer-sánchez, A., & Conklin, K. (2021). The effect of input modes and number of exposures on the learning of L2 binomials. *ITL - International Journal of Applied Linguistics*, September 2021.

Alsadoon, R., & Heift, T. (2015). Textual input enhancement for vowel blindness : A study with Arabic ESL learners. *The Modern Language Journal*, *99*(1), 57–79. https://doi.org/10.1111/modl.12188

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101–102). Oxford: Oxford University Press.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of "make" in native and non-native student writing. *Applied Linguistics*, *22*, 173–194.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369– 406.

Arnon, I., & Priva, U. C. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371. https://doi.org/10.1177/0023830913484891

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*, 67–82. https://doi.org/10.1016/j.jml.2009.09.005

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. D. (1992). Working memory. *Science*, *255*, 556–559.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101–114. https://doi.org/10.1016/0346-251X(93)90010-E

Baltova, I. (1999). Multisensory language teaching in a multidimensional curriculum: The use of authentic Bimodal video in core French. *The Canadian Modern Language Review*, *56*(1), 31–48.

Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning of Four-Word Combinations. *Psychological Science*, *19*(3), 241–248.

Barsov, A. (1770). *Sobranie 4291 Drevnikh Rossiyskikh poslovits.* Moskva: Imperatorskiy Moskovskiy Universitet.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bednarek, M. (2012). "Get us the hell out of here": Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics*, *17*(1), 35–63.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, *14*(3), 275–311.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Johansson, J., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*, 509–533.

Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014a). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*, 399–418. https://doi.org/10.1017/S0142716412000434

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014b). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, *64*, 855–877. https://doi.org/10.1111/lang.12085

Bisson, M. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism: Language and Cognition*, *24*, 549–568.

Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 143–157). New York: Routledge.

Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, *21*(3), 362–380. https://doi.org/10.1177/1362168816651464

Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research*, *18*(1), 54–74. https://doi.org/10.1177/1362168813505389

Boers, F., Demecheleer, M., He, L., & Deconinck, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*(2), 448–469. https://doi.org/10.1111/ijal.12141

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Language Teaching Research*, *10*(3), 245–261.

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230245006

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83–110.

https://doi.org/10.1017/S0267190512000050

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics*, 41–62.

Boers, F., & Webb, S. (2018). Teaching and learning collocation in adult second and foreign language learning. *Language Teaching*, *51*(1), 77–89. https://doi.org/10.1017/S0261444817000301

Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across the lifespan*. Cambridge: Cambridge University Press.

Bolinger, D. (1967). Meaning and Memory. *Forum Linguisticum*, *1*, 1–14.

Bollansée, L., Puimège, E., & Peters, E. (2021). "Watch out! Behind you is the enemy!" An exploratory study into the relationship between extramural English and productive vocabulary knowledge. In V. Werner & F. Tegge (Eds.), *Pop culture and language learning* (pp. 199–214). London: Routledge.

Brône, G., & Oben, B. (Eds.). (2018). *Eye-tracking in interaction: Studies on the role of eye gaze in dialogue: Volume 10*. New York: John Benjamins.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, *20*(2), 136–163.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior*, *46*(3), 904–911.

Bybee, J. (1998). The Emergent Lexicon. *Chicago Linguistic Society*, *34*, 421–435.

Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (Eds.), *From Pre-Language to Language* (pp. 109–134).

Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure: Typoglocial studies in language (TSL).*

Bygate, M. (1988). Units of oral expression and language learning. *Applied Linguistics*, *9*, 59–82.

Campoy, G. (2008). The effect of word length in short-term memory: Is rehearsal necessary? *Quarterly Journal of Experimental Psychology*, *61*(5), 724–734. https://doi.org/10.1080/17470210701402364

Candarli, D. (2021). A longitudinal study of multi - word constructions in L2 academic writing: The effects of frequency and dispersion. *Reading and Writing*, *34*, 1191–1223. https://doi.org/10.1007/s11145-020-10108-3

Carrol, G., & Conklin, K. (2014a). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research*, *7*(5), 1–11.

Carrol, G., & Conklin, K. (2014b). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, *17*(4), 784–797. https://doi.org/10.1017/S1366728913000795

Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs.' *Bilingualism: Language and Cognition*, *20*(2), 299–317. https://doi.org/10.1017/S1366728915000103

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95–122. https://doi.org/10.1177/0023830918823230

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition*, *38*, 403–443. https://doi.org/10.1017/S0272263115000492

Carrol, G., Littlemore, J., & Gillon, M. (2018). Of false friends and familiar foes: Comparing native and non-native understanding of figurative phrases. *Lingua*, *204*, 21–44. https://doi.org/10.1016/j.lingua.2017.11.001

Castelhano, M. S., & Rayner, K. (2008). Eye movements during reading, visual search, and scene perception: An overview. In K. Rayner, D. Shen, X. Bai, & G. Yan (Eds.), *Cognitive and cultural influences on eye movements* (pp. 3–33). Tianjin: Tianjin People's Publishing House.

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meaning of new words. *Memory and Cognition*, *25*, 203–226.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 225–235.

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

Chen, Y.-H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, *14*(2), 30–49.

Cheng, W., Greaves, C., Sinclair, J. M., & Warren, M. (2008). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics*, *30*(2), 236–252.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403–426. https://doi.org/10.1177/1362168816653271

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck : A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, 62–102. https://doi.org/10.1017/S0140525X1500031X

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, *16*(1), 22–29.

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

Cintrón-Valentín, M. C., & García-Amaya, L. (2021). Investigating Textual Enhancement and Captions in L2 Grammar and Vocabulary. *Studies in Second Language Acquisition*, *43*(5), 1068–1093. https://doi.org/10.1017/S0272263120000492

Cintrón-Valentín, M., García-Amaya, L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal*, *47*(4), 439–459. https://doi.org/10.1080/09571736.2019.1615978

Clifton, C., Staub, A., & Rayner, K. (2007). Eye Movements in Reading Words and Sentence. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341–372). Oxford: Elsevier.

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

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*(3), 38–63.

Cobb, T. (2019). From corpus to CALL. In A. Siyanova-chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 192–210). New York: Routledge.

Conklin, K. (2020). Processing single-word and multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 174–188). Routledge.

Conklin, K., Alotaibi, S., & Vilkaitė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257–276. https://doi.org/10.1177/0267658320921496

Conklin, K., & Carrol, G. (2021). Words go together like "bread and butter": The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*, (August 2020), 492–513. https://doi.org/10.1093/applin/amaa034

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge: Cambridge University Press.

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS One*, *10*, 1–38.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, *5*, 161–169.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, *2*(3), 223–235.

Cowie, A. P. (1994). Phraseology. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 3168–3171). Oxford: Pergamon.

Cowie, A. P. (2001). Speech formulae in English : problems of analysis and dictionary treatment. In G. van der Meer & A. G. B. ter Meulen (Eds.), *Making Senses: From Lexeme to Discourse. In Honor of Werner Abraham* (pp. 1–12). Groninger Arbeiten zur germanistishen Linguistik 44. Center for language and Cognition Groningen.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Teaching*, *29*(2), 243–263. https://doi.org/10.1177/0265532211419331

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied Linguistics*, *36*(5). https://doi.org/10.1093/applin/amt056

Crossley, S. A., Subtirelu, N., & Salsbury, T. (2013). Frequency effects or context effects in second language word learning. *Studies in Second Language Acquisition*, *35*, 727–755. https://doi.org/10.1017/S0272263113000375

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 560 million words 1990-present. https://corpus. byu.edu/coca/.

de Groot, A. M. B., & van Hell, J. G. (2005). The learning of foreign language vocabulary. In J. Kroll & A. de Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 9–29). Oxford: Oxford University Press.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure: Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, *23*(1), 171–185.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first and second language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*, 843–863.

Dijkstra, A., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197.

Divjak, D. (2019). *Frequency in Language: Memory, Attention and Learning*. Cambridge: Cambridge University Press.

Duong, T. P., Montero Perez, M., Desmet, P., & Peters, E. (2021). Learning vocabulary in spoken input- and output-based tasks. *TASK*, *1*(1), 100–126.

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*, 1–13. https://doi.org/10.1016/j.system.2019.102166

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations A meta-analysis. *International Journal of Corpus Linguistics*, *19194*(10), 443–477. https://doi.org/10.1075/ijcl.19.4.01dur

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real?

Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*(2), 125–155.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, *47*(2), 157–177.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188. https://doi.org/10.1177/0267658309349431

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, *40*, 341–366. https://doi.org/10.1017/S0272263117000109

Elgort, I., & Warren, P. (2014). L2 Vocabulary Learning From Reading: Explicit and Tacit Lexical Knowledge and the Role of Learner and Item Variables. *Language Learning*, *64*(2), 365–414. https://doi.org/10.1111/lang.12052

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers. *TESOL Quarterly*, *42*(3), 375–396.

Ellis, N.C. (2015). Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3–23).

Ellis, N. C. (1994). Implicit and explicit processes in language acquisition: An introduction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 1–32). San Diego: Academic Press.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*(1), 91–126.

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

Ellis, N. C. (2003). Constructions, chunking, and connectionism: the emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Oxford: Blackwell.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305–352.

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

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*, 164–194.

Ellis, N. C. (2007). The weak interface, consciousness, and form-focussed instruction: Mind the doors. In S. Fotos & H. Nassaji (Eds.), *Form focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 17–33). Oxford: Oxford University Press.

Ellis, N. C. (2018). Salience in usage-based SLA. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 21–40). New York: Routledge.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*(1), 29–62. https://doi.org/10.1515/text.1.2000.20.1.29

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

European Commission, D.-G. for C. (2018). *Media use in the European Union*. https://doi.org/doi/10.2775/116707

Evert, S. (2008). Corpora and collocations. *Corpus Linguistics. An International Handbook*, 1–53. https://doi.org/10.1515/9783110213881.2.1212

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*, 499–523. https://doi.org/10.1017/S0272263119000494

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage.

Firth, J. R. (1957). *Papers in linguistics, 1934–1951*. Oxford: Oxford University Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 170–210.

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

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*. https://doi.org/10.1111/lang.12225

Garner, J., & Crossley, S. (2018). A latent curve model approach to studying L2 N-gram development. *The Modern Language Journal*, *102*(3), 494–511. https://doi.org/10.1111/modl.12494

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

Gass, S., Winke, P., Isbell, D. R., & Ahn, J. (2019). How captions help people learn languages : A working-memory , eye-tracking study. *Language Learning & Technology*, *23*(2), 84–104.

Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In

A. P. Cowie (Ed.), *Phraseology. Theory, Analysis, and Applications* (pp. 125–143). Oxford: Oxford University Press.

Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. Farrar, Straus and Giroux.

Godfroid, A. (2020). *Eye Tracking in Second Language Acquisition and Bilingualism*. *Eye tracking in second language acquisition and bilingualism*. New York: Routledge. https://doi.org/10.4324/9781315775616

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483–517. https://doi.org/10.1017/S0272263113000119

Godfroid, A., & Hui, B. (2020). Five common pitfalls in eye-tracking research. *Second Language Research*, *36*(3), 277–305. https://doi.org/10.1177/0267658320921218

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness ? A triangulation of eye-movement data , verbal reports , and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Godfroid, A., & Spino, L. A. (2015). Reconceptualizing Reactivity of Think-Alouds and Eye Tracking : Absence of Evidence Is Not Evidence of Absence, (December), 896–928. https://doi.org/10.1111/lang.12136

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55–75.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Science*, *7*, 19–224.

Goldschneider, J. M., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787–814.

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, *41*(4), 481–505. https://doi.org/10.1093/applin/amy057

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*(1), 94–126. https://doi.org/10.1075/itl.166.1.03fer

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (Vol. 23, pp. 145–160). Oxford: Clarendon Press.

Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *IRAL - International Review of Applied Linguistics in Language Teaching*, *52*(3), 229–252. https://doi.org/10.1515/iral-2014-0011

Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology. An interdisciplinary perspective.* Amsterdam: John Benjamins.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In *Phraseology. An interdisciplinary perspective.* (pp. 27–49). Amsterdam: John Benjamins. https://doi.org/10.1075/z.139.07gra

Grant, L., & Bauer, L. (2004). Criteria for Re-defining Idioms: Are we Barking up the Wrong Tree? *Applied Linguistics*, *25*(1), 38–61. https://doi.org/10.1093/applin/25.1.38

Grice, J. W., & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. *Applied Multivariate Research*, *12*, 199–226.

Gries, S. T. (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, *18*(1), 137–165. https://doi.org/10.1075/ijcl.18.1.09gri

Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language* (pp. 21–33). Houndmills: Palgrave Macmillan.

Gyllstad, H. (2007). *Testing English collocations: developing receptive tests for use with advanced Swedish learners*. Lunds universitet.

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL - International Journal of Applied Linguistics*, *166*(2), 278–306. https://doi.org/10.1075/itl.166.2.04gyl

Gyllstad, H., & Wolter, B. (2016). Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter? *Language Learning*, *66*(2), 296–323. https://doi.org/10.1111/lang.12143

Harley, T. A. (2014). *The psychology of language: From data to theory*. London: Psychology Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.

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

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.

Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics*, *19*(1), 24–44.

Hsu, J. Y. (2007). Lexical collocations and their relation to the online writing ofTaiwanese college English majors and non-English majors. *Electronic Journal of Foreign Language Teaching*, *4*, 192–209.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.

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

Hulstijn, J. H. (2013). Incidental learning in second language acquisition. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 2632–2640). Chichester: Wiley-Blackwel.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558.

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions: An eye-tracking study. *Studies in Second Language Acquisition*, *39*, 401–430. https://doi.org/10.1017/S027226311600019X

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, *40*(6), 431–439.

Jones, S., & Sinclair, J. M. (1974). English lexical collocations. A study in computational linguistics. *Cahiers de Lexicologie*, *24*, 15–61.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*, 846–863.

Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*, 755–780.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, *68*, 1665–1692.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., … Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*, 7–36.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*, 262–284.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *The Modern Language Journal*, *78*, 285–

299.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah: Erlbaum.

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the Role of Phraseological Knowledge in Foreign Language Reading. *Applied Linguistics*, *38*(6), 848–870. https://doi.org/10.1093/applin/amv070

Kruger, J. L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the moving image: An overview of eye tracking studies. *Refractory: A Journal of Entertainment Media*, *25*, 1–14.

Kuiper, K., & Haggo, D. (1984). Livestock auctions, oral poetry, and ordinary language. *Language in Society*, *13*(2), 205–234.

Kuperman, V., & Dyke, J. A. Van. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *J Exp Psychol Hum Percept Perform.*, *39*(3), 802–823. https://doi.org/10.1037/a0030859

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*(1), 65–85.

Kurvers, J., & Uri, H. (2006). Metalexical awareness: Development, methodology or written language? *Journal of Psycholinguistic Research*, *35*, 353–367.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323.

Langacker, R. W. (1986). An Introduction to Cognitive Grammar. *Cognition*, *10*, 1–40. https://doi.org/10.1207/s15516709cog1001_1

Larsen-Freeman, D. (1983). The importance of input in second language acquisition. In R. Anderson (Ed.), *Pidginization and Creolization as Second Language Acquisition* (pp. 87–93). Rowley, MA: Newbury House.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning : A case for contrastive analysis and translation. *Applied Cognitive Psychology*, *29*(4), 694–716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness Vocabulary tests are contingent upon the test designer's definition of lexical knowledge. Lexical knowledge, in turn, has. *Language Learning*, *54*(3), 399–436.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvementLaufer, B. and J. Hulstijn. 2001. *Applied Linguistics*, *22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold revisited: Lexical Text Coverage, Learners' Vocabulary Size and reading Comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research*, *19*(6), 687–711. https://doi.org/10.1177/1362168814559797

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: The modulation of lexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews*, *5*(3), 107–127. https://doi.org/10.1177/1534582306289573

Lee, M., & Révész, A. (2018). Promoting Grammatical Development Through Textually Enhanced Captions : An Eye-Tracking Study. *The Modern Language Journal*, *102*(3), 557–577. https://doi.org/10.1111/modl.12503

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, Advance online publication. https://doi.org/10.1017/S0272263120000108

Lenth, R. (2018). 'emmeans: Estimated marginal means, aka least-squares means', R package. http://CRAN.R- project.org/package=emmeans.

Leow, R. P., & Martin, A. (2017). Enhancing the input to promote salience of the L2: A critical overview. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 167–186). Routledge.

Leveridge, A. N., & Yang, J. C. (2013). Testing learner reliance on caption supports in second language listening comprehension multimedia environments. *ReCALL*, *25*(2), 199–214.

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

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*, 85–102.

Lightbown, P. M., Spada, N., & White, L. (1993). The role of instruction in second language acquisition. *Studies in Second Language Acquisition*, *15*, 143–277 [Special Issue].

Lin, P. M. S. (2012). Sound evidence : The missing piece of the jigsaw in formulaic language. *Applied Linguistics*, *33*(3), 342–347. https://doi.org/10.1093/applin/ams017

Lin, P. M. S. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, *42*, 164–176.

Lin, P. M. S. (2019). Formulaic language and speech prosody. In A. Siyanova-chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 78–94). New York: Routledge.

Lin, P. M. S., & Siyanova-Chanturia, A. (2014). Internet television for L2 vocabulary learning. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 149–158). London: Routledge.

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*(1), 105–129. https://doi.org/10.1080/14790718.2012.679275

Lindstromberg, S., & Boers, F. (2008). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, *29*, 200–222.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Amsterdam: John Benjamins.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries and vocabulary learning. *Language Learning*, *43*, 263–287.

Macis, M. (2018). Incidental learning of duplex collocations from reading: Three case studies. *Reading in a Foreign Language*, *30*(1), 48–75.

Macis, M., & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, *21*(3), 321–340. https://doi.org/10.1177/1362168816645957

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, *43*(5), 985–1008. https://doi.org/10.1017/S0272263121000036

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, *33*(3), 299–320. https://doi.org/10.1093/applin/ams010

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York: Cambridge University Press.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A cognitive load perspective. *Applied Cognitive Psychology*, *28*(5), 653–660. https://doi.org/10.1002/acp.3050

McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, *35*, 51–73.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive in- ference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, *43*, 1735–1751.

McDonough, K., Trofimovich, P., Lu, L., & Abashidze, D. (2020). Visual cues during interaction: Are recasts different from noncorrective repetition? *Second Language Research*, *36*(3), 359–370.

Mckoon, G., & Ratcliff, R. (1979). Priming in Episodic and Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 463–480.

Meunier, F. (2020). Resources for learning multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 336–350). New York: Routledge.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, (63), 81–97.

Mirzaei, M. S., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning : A new tool to assist learners in developing second language listening skill. *ReCALL*, *29*(2), 178–199. https://doi.org/10.1017/S0958344017000039

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS One*, *4*.

Mohamed, A. A. (2017). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*, 269–293.

Mondria, J.-A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, *12*, 249–267.

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching,* Advance online publication. https://doi.org/10.1017/S0261444821000501

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(181), 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing Vocabulary Learning Through Captioned Video: An Eye-Tracking Study. *Modern Language Journal*, *99*(2), 308–328. https://doi.org/10.1111/modl.12215

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

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

Muñoz, C., Pujadas, G., & Pattemore, A. (2021). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*, Advance online publication. https://doi.org/10.1177/02676583211015797

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, *48*(3), 323–363.

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

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *63*(1), 59–81. https://doi.org/10.1353/cml.2006.0049

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

Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston, MA: Heinle, Cengage Learning.

Nattinger, J., & DeCarrico, J. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nattinger, J. R. (1980). A lexical phrase grammar for ESL. *TESOL Quarterly*, *14*, 337–344.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English. *Applied Linguistics*, *24*, 223–242.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: 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*(1), 94. https://doi.org/10.2307/747835

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298–320. https://doi.org/10.1177/1362168816639619

Nicklin, C., & Vitta, J. P. (2021). Effect-driven sample sizes in second language instructed vocabulary acquisition research. *The Modern Language Journal*, *21*, 218–236. https://doi.org/10.1111/modl.12692

Noreillie, A.-S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages. *ITL - International Journal of Applied Linguistics*, *169*(1), 212–231. https://doi.org/10.1075/itl.00013.nor

Northbrook, J., & Conklin, K. (2019). Is What You Put in What You Get Out ? — Textbook-derived Lexical Bundle Processing in Beginner English Learners. *Applied Linguistics*, *40*(5), 816–833. https://doi.org/10.1093/applin/amy027

Northbrook, J., Conklin, K., & Allen, D. (2021). 'Did you see that?' — The role of repetition and enhancement on lexical bundle processing in English learning materials, October 2021. https://doi.org/10.1093/applin/amab063

Nunberg, G., Sag, I., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491–538.

Obermeier, A., & Elgort, I. (2021). Deliberate and contextual learning of L2 idioms : The effect of learning conditions on online processing. *System*, April 2021. https://doi.org/10.1016/j.system.2020.102428

Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic Language Volume 2: Acquisition, Loss, Psychological Reality, and Functional Explanations* (pp. 375–386). John Benjamins.

Öksüz, D., Brezina, V., & Rebuschat, P. (2020). Collocational processing in L1 and L2 : The effects of word frequency, collocational frequency, and association. *Language Learning*, *71*(1), 55–98. https://doi.org/10.1111/lang.12427

Paivio, A. (1986). *Mental representations, a dual coding approach*. Oxford: Oxford University Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25.

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

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, *35*(1). https://doi.org/10.1177/0267658317694221

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130–149. https://doi.org/10.1017/S0267190512000098

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge: Cambridge University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191–225). London: Longman. https://doi.org/10.4324/9781315836027-12

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381–402. https://doi.org/10.1177/1362168815618428

Pellicer-Sánchez, A., & Boers, F. (2019). Teaching and learning of formulaic language. In A. Siyanova-chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 153–173). New York: Routledge.

Pellicer-Sánchez, A., & Conklin, K. (2019). Eye tracking as a data collection method. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 371–382). New York.

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. (2012). Learning German formulaic sequences: the effect of two attention-drawing techniques. *The Language Learning Journal*, *40*(1), 65–79. https://doi.org/10.1080/09571736.2012.658224

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

https://doi.org/10.1177/1362168813505384

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138. https://doi.org/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*(1), 142–168.

Peters, E. (2019). The Effect of Imagery and On-Screen Learning From Audiovisual Input. *TESOL Quarterly*, *53*(4), 1008–1032. https://doi.org/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. 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*(3), 747–782. https://doi.org/10.1111/lang.12351

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes*, *20*, 28–39. https://doi.org/10.1016/j.jeap.2015.04.002

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

Pichette, F., De Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, *33*(1), 66–82. https://doi.org/10.1093/applin/amr037

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–29.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language.* Harper Collins.

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–86). Hillsdale, NJ: Erlbaum.

Price, K. (1983). Closed-captioned TV: An untapped resource. *MATSOL Newsletter*, *12*(2), 1–8.

Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input : An eye-tracking study. *Second Language Research*, October 2021. https://doi.org/10.1177/02676583211049741

Puimège, E., & Peters, E. (2019a). Learners ' English Vocabulary Knowledge Prior to Formal Instruction : The Role of Learner-Related and Word-Related Variables. *Language Learning*, *69*(4), 943–977. https://doi.org/10.1111/lang.12364

Puimège, E., & Peters, E. (2019b). Learning L2 vocabulary from audiovisual input: an exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal*, *47*(4), 424–438. https://doi.org/10.1080/09571736.2019.1638630

Puimège, E., & Peters, E. (2020). Learning formulaic sequences through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *42*(3), 525–549. https://doi.org/10.1017/S027226311900055X

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*(4), 479–496. https://doi.org/10.1080/09571736.2019.1616806

Pujola, J.-T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, *14*(2), 235–262.

Pulido, D. (2007). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *57*, 155–199.

Pulido, M. F. (2021). Individual chunking ability predicts efficient or shallow L2 processing: Eye-tracking evidence from multiword units in relative clauses. *Frontiers in Human Neuroscience*, *11*, January 2021. https://doi.org/10.3389/fpsyg.2020.607621

Rayner, K. (1998). Eye Movements in Reading and Information Processing : 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191–201.

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9–36). Hillsdale, NJ: Erlbaum.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning : Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Révész, A., Michel, M., & Lee, M. (2019). Exploring second language writers' pausing and revision behaviors. *Studies in Second Language Acquisition*, *41*, 605–631.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*, 213–235. https://doi.org/10.1017/S0272263112000861

Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning*, *45*, 283–331.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.

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/10.4018/jksr.2013010103

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. (2020). Incidental vocabulary learning from viewing television. *ITL – International Journal of Applied Linguistics*, *171*(2), 191–220.

Rychlý, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN* (pp. 6–9). Brno: Masaryk University.

Saragi, T., Nation, I. S. P., & Meister, F. (1978). Vocabulary learning and reading. *System*, *6*, 72–78.

Schmidt, R. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137–174). Rowley, MA: Newbury House.

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

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam: John Benjamins.

Schmitt, N. (Ed.). (2010). *Researching vocabulary: a vocabulary research manual.* Basingstoke: Palgrave.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 1–22). John Benjamins.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *Modern Language Journal*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the bahviour of two versions of the vocabulary levels test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Schneider, U. (2018). ΔP as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory*, *16*: 249–274.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*, 165–179.

Simard, D., & Foucambert, D. (2013). Observing noticing while reading in L2. In J. M.

Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 207–226). Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i at Manoa.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140–160). New York: Longman.

Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, *53*, 148–160.

Siyanova-Chanturia, A., Conklin, K., Cafarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Brain and Language Representation and processing of multi-word expressions in the brain. *Brain and Language*, *175*, 111–122. https://doi.org/10.1016/j.bandl.2017.10.004

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776–784. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The Idiom Principle Revisited. *Applied Linguistics*, *36*(5), 549–569. https://doi.org/10.1093/applin/amt054

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2019). *Understanding formulaic language: A second language acquisition perspective.* New York: Routledge.

Siyanova-Chanturia, A., & Spina, S. (2020). Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study. *Language Learning*, *70*(2), 420–463. https://doi.org/10.1111/lang.12383

Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 38–61). New York: Routledge.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429–458. https://doi.org/10.3138/cmlr.64.3.429

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*, 510–532.

Slattery, T. J., Pollatsek, A., & Rayner, K. (2007). The effect of the frequencies of three consecutive content words on eye movements during reading. *Memory & Cognition*, *35*, 1283–1292.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, *46*, 137–150.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on offline/on-line collocational processing. *Bilingualism*, *18*(3), 419–437. https://doi.org/10.1017/S1366728914000674

Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning*, *63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Spätgens, T., & Schoonen, R. (2020). Single, but not unrelated: Key issues in researching single-word items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 511–528). New York: Routledge.

Spina, S. (2019). The development of phraseological errors in Chinese learner Italian: A longitudinal study. In A. Abel, A. Glaznieks, V. Lyding, & L. Nicolas (Eds.), *Widening the scope of learner corpus research. Selected papers from the fourth Learner Corpus Research Conference* (pp. 95–119). Louvain: Presses Universitaires de Louvain.

Spinner, P., Gass, S. M., & Behney, J. (2013). Ecological validity in eye-tracking: An empirical study. *Studies in Second Language Acquisition*, *35*, 389–415.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407. https://doi.org/10.1598/RRQ.21.4.1

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, *8*(2), 209–243. https://doi.org/10.1075/ijcl.8.2.03ste

Stengers, H., & Boers, F. (2015). Exercises on collocations: A comparison of trial-and-error and exemplar-guided procedures. *Journal of Spanish Language Teaching*, (2), 152–164.

Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does ‘chunking’ foster chunk-uptake? In S. De Knop, F. Boers, & A. De Rycker (Eds.), *Fostering language teaching efficiency through Cognitive Linguistics* (pp. 99–117). Berlin: Mouton de Gruyter.

Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *International Review of Applied Linguistics in Language Teaching*, *49*, 321–343.

Suárez, M. del M., Gilabert, R., & Moskvina, N. (2021). The mediating role of vocabulary size, working memory, attention and inhibition in early vocabulary learning under different TV genres: An exploratory study. *TESOL Journal*, *12*(4), 1–29. https://doi.org/10.1002/tesj.637

Suárez, M., & Gesa, F. (2019). Learning vocabulary with the support of sustained exposure to captioned video: do proficiency and aptitude make a difference ? captioned video : do pro fi ciency and aptitude make a di ff erence ? *The Language Learning Journal*, *47*(4), 497–517. https://doi.org/10.1080/09571736.2019.1617768

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-46048-6

Swanborn, M. S. L., & de Glopper, K. (1999). Incidental Word Learning While Reading: A

Meta- Analysis. *Review of Educational Research*, *69*(3), 261–285.

Szudarski, P. (2012). Effects of meaning- and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching and Research*, *1*(2), 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 (United Kingdom)*, *26*(2), 245–265. https://doi.org/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.

Tinker, M. A. (1936). Reliability and validity of eye-movement measures of reading. *Journal of Experimental Psychology*, (19), 732–746.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In R. R. Heredia & A. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 171–207). Cambridge: Cambridge University Press.

Tiv, M., Gonnerman, L., Whitford, V., Friesen, D., Jared, D., & Titone, D. (2019). Figuring Out How Verb-Particle Constructions Are Understood During L1 and L2 Reading. *Frontiers in Psychology*, *10*, 1–18. https://doi.org/10.3389/fpsyg.2019.01733

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations : A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, *69*(2), 405–439. https://doi.org/10.1111/lang.12335

Turk-browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

Underwood, G., Schmitt, N., & Galpin, N. (2004). The eyes have it: an eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 153–173). John Benjamins.

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

Van Zeeland, H., & Schmitt, N. (2013). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*(3), 191–210.

https://doi.org/10.1017/S0261444807004338

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

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Ablex.

Verspoor, M. H., de Bot, K., & van Rein, E. (2011). English as a foreign language. The role of out-of-school language input. In A. De Houwer & A. Wilton (Eds.), *English in Europe today. Sociocultural and educational perspectives* (pp. 147–166). Amsterdam: John Benjamins.

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

Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(10), 1632–1642. https://doi.org/10.1037/xlm0000259

Vilkaitė, L. (2017). Incidental acquisition of collocations in L2 Effects of adjacency and prior vocabulary knowledge. *ITL - International Journal of Applied Linguistics*, *168*(2), 248–277. https://doi.org/10.1075/itl.17005.vil

Vilkaitė, L. (2016). *Formulaic language: Distribution, processing, and acquisition*.

Vilkaitė, L., & Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations? *Applied Linguistics*, *40*, 329–354.

Vinogradov, V. V. (1938). *Sovremennyy russkiy yazyk. Grammaticheskoe uchenie o slove.* Moskva: Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo.

von der Malsburg, T., & Angele, B. (2018). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language*, *94*, 119–133. https://doi.org/10.1016/j.jml.2016.10.003.False

Vu, D. Van, & Peters, E. (2020). Learning Vocabulary from Reading-only, Reading-while-listening, and Reading with Textual Input Enhancement: Insights from Vietnamese EFL Learners. *RELC Journal*, Advance online publication. https://doi.org/10.1177/0033688220911485

Vu, D. Van, & Peters, E. (2021). Incidental learning of collocations from meaningful input: A longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, Advance online publication. https://doi.org/10.1017/S0272263121000462

Vu, D. Van, & Peters, E. (2022). A longitudinal study on the effect of mode of reading on incidental collocation learning and predictors of learning gains. *TESOL Quarterly*, Advance online publication. https://doi.org/10.1002/tesq.3111

Waibel, B. (2008). *Phrasal verbs: German and Italian learners of English compared*. Saarbrücken, Germany: VDM.

Wang, A., & Pellicer-Sánchez, A. (2022). Incidental Vocabulary Learning From Bilingual Subtitled Viewing : An Eye-Tracking Study. *Language Learning*, Advance online publication. https://doi.org/10.1111/lang.12495

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*(2), 130–163. https://doi.org/10.1177/003368828501600214

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

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232–245.

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. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 225–239). New York: Routledge.

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support : How do frequency and distribution of occurrence affect learning ? https://doi.org/10.1177/1362168814559800

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, *43*(1), 55–77.

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

Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford: Oxford University Press.

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

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

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 205–223.

Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*, 312–339.

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.

Winke, P., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue. Eye-movement recordings in second language research. *Studies in Second Language Acquisition*, *35*(2), 205–212.

Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, *35*, 323–352.

Winke, P., Sydorenko, T., & Gass, S. (2013). Factors Influencing the Use of Captions by Foreign Language Learners: An Eye-Tracking Study. *The Modern Language Journal*, *97*(1), 254–275. https://doi.org/10.1111/j.1540-4781.2012.01432.x

Wisniewska, N., & Mora, J. C. (2018). Pronunciation learning trough captioned videos. In J. Levis (Ed.), *Proceedings of the 9 Pronunciation in Second Language Learning and Teaching conference* (Vol. ISSN 2380-, pp. 204–215). Ames, IA: Iowa State University.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*(4), 430–449. https://doi.org/10.1093/applin/amr011

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*, 1193–1221. https://doi.org/10.1017/S0142716414000113

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing. *Studies in Second Language Acquisition*, *40*(2), 395–416. https://doi.org/10.1017/S0272263117000237

Wood, D. (2006). Uses and functions of formulaic sequences in second-language speech: An exploration of the foundations of fluency. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *63*(1), 13–33. https://doi.org/10.1353/cml.2006.0051

Wood, D. (Ed.). (2010). *Perspectives on formulaic language: Acquisition and communication*. London: Continuum.

Wood, D. (2015). *Fundamentals of formulaic language: An introduction.* London: Bloomsbury.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, *32*, 213–231. https://doi.org/10.1017/S027226310000629X

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 123–148). Amsterdam: John Benjamins.

Wulff, S. (2009). Converging evidence from corpus and experimental data to capture idiomaticity. *Corpus Linguistics and Linguistic Theory*, *5*(1), 131–159. https://doi.org/10.1515/CLLT.2009.006

Wulff, S. (2010). Marrying cognitive-linguistic theory and corpus-based methods : On the compositionality of English V NP-idioms. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 223–238.

Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 37–56). John Benjamins. https://doi.org/10.1075/sibil.54.03wul

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*(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*, 831–856.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *57*(4), 541–572.

Zykova, I. V. (2016). Linguo-cultural studies of phraseologisms in Russia : past and present1 from the history of research within the Russian phraseological tradition. *Yearbook of Phraseology*, *7*, 127–148. https://doi.org/10.1515/phras-2016-0007

Appendices

Appendix 1. Instruments of Study 1

Form recall test

Translate the following words/expressions into English.

levensduur = l _ _ _ _ _ _ _ y
bedrijfsevenement = c _ _ _ _ _ _ _ _   e _ _ _ _
schaalbaar = s _ _ _ _ _ _ _
schaalvoordelen = e _ _ _ _ _ _ _ _   o _   s _ _ _ _
ondervragen = i _ _ _ _ _ _ _ _ _ _
basismateriaal/grondstof = r _ _   m _ _ _ _ _ _ _ _
kleinhandel = r _ _ _ _ _
onuitgesproken boodschap = s _ _ _ _ _ _ _ _ _   m _ _ _ _ _ _
jezelf verdedigen = h _ _ _   your   o _ _
aandacht trekken/krijgen = c _ _ _ _   the   e _ _
jong/pril/beginnend = f _ _ _ _ _ _ _ _
vroegtijdige boeking = a _ _ _ _ _ _   b _ _ _ _ _ _
inleven = e _ _ _ _ _ _ _ _
verleiden = e _ _ _ _ _
grote aantallen = v _ _ _   n _ _ _ _ _ _
een gok wagen = t _ _ _   a   p _ _ _
iets onder de knie krijgen = m _ _ _ _ _   the   a _ _
voorspellen = f _ _ _ _ _ _ _
tot de verbeelding spreken = c _ _ _ _ _ _   the   i _ _ _ _ _ _ _ _ _ _
eigenzinnig = q _ _ _ _ _
waanzin = l _ _ _ _ _
op lange termijn = d _ _ _   the   l _ _ _
draaiende/in beweging = u _   a _ _   r _ _ _ _ _ _
innemend = e _ _ _ _ _ _ _
smakeloos = i _ _ _ _ _ _
opbrengen = t _ _ _   o _ _ _
teruggaan = h _ _ _   b _ _ _
een cursus organiseren = r _ _   a   c _ _ _ _ _
terugdeinzen = s _ _   a _ _ _
iets in detail onderzoeken = d _ _ _ _   d _ _ _   into something
ijver = a _ _ _ _
verduidelijking = c _ _ _ _ _ _ _ _ _ _ _ _
onderneming = v _ _ _ _ _ _
benutten = t _ _   i _ _ _
tot stand komen = c _ _ _   a _ _ _ _
welbespraakt = a _ _ _ _ _ _ _ _ _
gloeien = a _ _ _ _ _
balken = b _ _ _
bedroefd = b _ _ _ _ _
dilemma = q _ _ _ _ _ _ _

Form recognition test

Circle or highlight the missing words/expressions.

1. With debts of $4 million, the _ is clearly heading for trouble.
   venture, oxter, waily , carlin
2. The cost of _ was going up.
   callow materials, rude materials, raw materials, crass materials
3. For _ in car design, you really need to keep it simple.
   reduct, scimtone, longevity, shiverly
4. How did the problem _ in the first place?
   come ago, come onto, come away, come about
5. I've had two job offers, and I'm in a real _ about which one to accept.
   claricle, quandary, micrenity, combuzzle
6. I admire the way you _ in a discussion.
   hold your own, bear your own, stay your own, keep your own
7. Would you please give a of your last statement?
   residation, clarification, sollumination, deravelment
8. The company _ more than $3.5 million every year.
   turns across, turns over, turns through, turns for
9. It's very easy to _ with the characters in her books.
   empathize, comprease, relevate, reamish
10. He tried to give the _ that he was a man of the people.
    astute message, subliminal message, circumspect message, contemplative message
11. This strategy can lead to problems further _ .
    up the line, down the line, on the line, around the line
12. He can _ victory or defeat.
    upsurp, prosurvy, forecast, premitten
13. The adverts _ the customer into buying things they don't really want.
    barlone, delead, entice, insuade
14. It's always useful to _ to the mechanics of a market.
    drill down, drill on, drill in, drill up
15. She tried to _ by waving her arms.
    grab his eye, bust his eye, catch his eye, clasp his eye
16. The teacher explained the theory in an _ way.
    elurable, insome, overrigging, engaging
17. It’s getting dark, time to _ .
    head away, head back, head over, head under
18. I’m not sure if the idea is _ from school level to state level.
    extendic, ascendile, scalable, elastive
19. The hotel has a facility for _ .
    onward booking, progressive booking, advanced booking, leading booking
20. _ of people have signed up for this.
    strong numbers, vast numbers, bulky numbers, colossal numbers
21. The police captain _ the suspect.
    interrogated, demunted, perilled, crossmeasured
22. We have been working for 6 months to _ of bread making.

cram the art, grasp the art, master the art, excel the art

23. The story of the young wizard has _ of the world's children.
    grabbed the imagination, secured the imagination, captured the imagination, apprehended the imagination
24. You need two years' experience in _ to get the job.
    upjoint, grendin, retail, circuary
25. Her _ for basketball impressed me.
    vervotion, fiercity, ardor, wascal
26. At a _ you can meet interesting people.
    corporate event, corporate juncture, corporate exploit, corporate episode
27. We _ in cooperation with other organisations.
    break courses, drop courses, pace courses, run courses
28. She gave an entertaining, and _ speech.
    articulate, distoken, defillible, cohoring
29. Big public companies enjoy _ .
    economies of extent, economies of range, economies of scale, economies of scope
30. If only we could _ all that energy and creativity.
    tap into, tap on, tap for, tap from
31. It might be a good idea to _ on a few technology stocks.
    win a punt, take a punt, find a punt, earn a punt
32. The current economic climate is particularly difficult for _ businesses.
    marxish, fledgling, incloin, smaltering
33. The farm can be _ in the autumn of 2005.
    up and running, up and moving, up and working, up and trotting
34. He was tall and had a _ sense of humour.
    orning, catent, quirky, vullable
35. I never _ from hard work.
    shy away, shy off, shy about, shy around
36. The mules suddenly started _ .
    windering, neiding, balkening, braying
37. After their children had left home the couple felt _ .
    laffered, berunted, demullied, bereft
38. Why anyone buys music with such _ lyrics is a mystery.
    honum, insipid, dramby, beigic
39. It was pure _ spending all that money.
    hoodelly, insillity, lunacy, becility
40. Some steels can be _ at lower temperatures than others.
    lampered, annealed, scoughened, bemended

Meaning recall test

Translate the following words/expressions into Dutch or describe their meaning in Dutch.

quirky =
to bray =
economies of scale =

to entice =
bereft =
to interrogate =
engaging =
scalable =
corporate event =
vast numbers =
advanced booking =
to empathize =
to master the art =
quandary =
to capture someone's imagination =
to hold your own =
to catch someone's eye =
insipid =
venture =
down the line =
to anneal =
to take a punt =
fledgling =
up and running =
to tap into =
to come about =
lunacy =
longevity =
to forecast =
to head back =
to run courses =
to shy away =
to turn over =
to drill down (into) =
articulate =
ardor =
clarification =
retail =
subliminal message =
raw materials =

Appendix 2. Target items and instruments of Study 2

Target items

The tables below contain the target and distractor items with values for raw frequency in OpenSubtitles, MI score in OpenSubtitles, congruency, transparency, and collocate-node relationship.

Target items with frequency of occurrence, raw frequency, MI score, congruency, semantic transparency, and collocate-node relationship

Item	FoO	Frequency	MI score	Congruency	Transparency	Collocate-node
abusive language	1	21	10.25	1	free combination	adj.-noun
beyond the pale	1	98	11.29	0	figurative idiom	other
bodily fluids	1	425	17.87	1	free combination	adj.-noun
breeding ground	1	336	10.02	1	figurative idiom	adj.-noun
building site	1	292	5.98	1	free combination	adj.-noun
cause offence	2	10	5.42	0	free combination	verb-noun
cell phone	1	23300	11.55	0	free combination	noun-noun
collateral damage	1	1106	15.52	1	free combination	adj.-noun
common denominator	1	284	15.44	1	free combination	adj.-noun
day care	1	828	8.87	1	free combination	noun-noun
depend on	2	19375	7.94	0	free combination	verb-prep.
draw from	2	610	5.25	0	figurative idiom	verb-prep.
drive off	1	1874	4.86	0	free combination	verb-particle
evolutionary advantage	1	21	19.53	1	free combination	adj.-noun
fight or flight response	2	114	7.77	1	free combination	noun-noun
find out	3	117839	6.52	0	restr. collocation	verb-prep.
foul language	1	251	11.36	1	free combination	adj.-noun
free speech	1	295	7.13	0	free combination	adj.-noun
get away with	3	11358	4.31	1	figurative idiom	verb-particle
give rise to	3	354	7.44	0	figurative idiom	verb-noun
grist to the mill	1	3	16.66	0	pure idiom	noun-noun
guinea pig	1	1637	15.11	0	figurative idiom	noun-noun
heck of a lot	1	201	7	0	free combination	noun-noun
highest echelons	1	10	13.43	0	free combination	adj.-noun
ice cubes	1	978	13.27	1	free combination	noun-noun
in the end	1	12001	6.42	0	figurative idiom	other
jazz up	1	77	6.4	0	pure idiom	verb-prep.

keep a diary	1	303	9.3	1	restr. collocation	verb-noun
laugh like hell	1	16	2.25	0	restr. collocation	verb-noun
locker room	1	2103	10.34	0	free combination	noun-noun
make sense	1	20458	10.63	0	restr. collocation	verb-noun
meet your maker	1	380	9.44	0	pure idiom	verb-noun
melting pot	1	113	8.19	0	figurative idiom	adj.-noun
mince words	1	264	12.71	0	restr. collocation	verb-noun
mutual pleasure	1	10	6.33	1	free combination	adj.-noun
pain relievers	1	46	13.26	1	free combination	noun-noun
pass over into	1	4	4.54	1	free combination	verb-particle
private parts	1	315	9.05	0	free combination	adj.-noun
pushing up daisies	1	63	11.55	0	pure idiom	noun-noun
sexual depravity	1	8	10.81	1	free combination	adj.-noun
sheer coincidence	1	33	10.57	1	free combination	adj.-noun
shift the blame	1	38	8.27	0	free combination	verb-noun
spark off	1	34	5.66	0	figurative idiom	verb-particle
spend a penny	3	59	7.02	0	pure idiom	verb-noun
spread the word	1	1003	8.98	0	restr. collocation	verb-noun
steady pace	4	56	10.77	1	free combination	adj.-noun
supernatural powers	1	246	12.76	1	free combination	adj.-noun
take into account	1	773	6.38	0	restr. collocation	verb-noun
take umbrage	1	60	10.14	0	free combination	verb-noun
tap into	1	1716	8.83	0	figurative idiom	verb-prep.
tell off	1	162	2.25	0	restr. collocation	verb-particle
turning point	1	640	4.56	1	free combination	adj.-noun
unleash a torrent of	1	1	13.26	1	restr. collocation	verb-noun
well versed in	1	108	13.79	0	free combination	other
white lie	1	366	6.88	0	restr. collocation	adj.-noun
win the right	1	47	1.3	0	restr. collocation	verb-noun

Note: FoO = Frequency of occurrence in the input.

Distractor items with frequency, MI score, congruency, semantic transparency, and collocate-node relationship.

Item	Frequency	MI	Congruency	Transparency	Collocate-node
break the mold	119	8.26	0	figurative idiom	verb-noun
chip in	1506	6.15	0	figurative idiom	verb-part
cut back on	424	3.41	0	figurative idiom	verb-part
get round to	170	1.03	0	pure idiom	verb-noun
hit the roof	156	7.01	0	figurative idiom	verb-noun
joint effort	56	8.48	1	free combination	adjective-noun
once in a blue	103	8.33	0	Restr. collocation	noun-noun
piece of advice	907	9.23	0	free combination	noun-noun
pull a muscle	143	8.53	0	free combination	verb-noun

Form recall test

Give the English expression for each of the meanings described below. The first letter of each word is given.

dood zijn	p _ _ _ _ _ _ u _ d _ _ _ _ _ _
leugentje om bestwil	w _ _ _ _ l _ _
hoogste rangen, hoogste niveaus	h _ _ _ _ _ _ e _ _ _ _ _ _ _
lichaamssappen (bloed, zweet, enz.)	b _ _ _ _ _ f _ _ _ _ _
pijnstillers	p _ _ _ r _ _ _ _ _ _ _ _
bovennatuurlijke krachten	s _ _ _ _ _ _ _ _ _ _ _ p _ _ _ _ _
gsm, mobiele telefoon	c _ _ _ p _ _ _ _
opwekken (bv. emoties)	s _ _ _ _ o _ _
geslachtsdelen, edele delen	p _ _ _ _ _ _ p _ _ _ _
iemand die gebruikt wordt voor wetenschappelijke tests, proefkonijn	g _ _ _ _ _ p _ _
aanhoudend tempo, gestaag tempo	s _ _ _ _ _ p _ _ _
op je woorden letten, een blad voor de mond nemen	m _ _ _ _ your w _ _ _ _
bijdragen, meebetalen	c _ _ _ i _
voedingsbodem	b _ _ _ _ _ _ _ g _ _ _ _ _
goed vertrouwd met, goed op de hoogte van	w _ _ _ v _ _ _ _ _ i _

gezamenlijke inspanning	j _ _ _ _ e _ _ _ _ _
plots heel boos worden	h _ _ t _ _ r _ _ _
een stroom van iets (bijv. verontwaardiging) veroorzaken	u _ _ _ _ _ _ a t _ _ _ _ _ _ o _
dagelijks in een dagboek schrijven, een dagboek bijhouden	k _ _ _ a d _ _ _ _
reactie op een gevaarlijke situatie, vecht- of vluchtsrespons	f _ _ _ _ o _ f _ _ _ _ _ r _ _ _ _ _ _ _
upgraden, "pimpen"	j _ _ _ u _
bouwwerf	b _ _ _ _ _ _ _ s _ _ _
overgaan van iets naar iets anders	p _ _ _ o _ _ _ i _ _ _
aanstoot geven, beledigen	c _ _ _ _ o _ _ _ _ _ _
gebruik maken van, benutten, profiteren van	t _ _ i _ _ _
perversie, seksueel verwerpelijk gedrag	s _ _ _ _ _ d _ _ _ _ _ _ _ _
een spier verrekken	p _ _ _ a m _ _ _ _ _
iets op een andere, vernieuwende manier doen, het stramien doorbreken	b _ _ _ _ t _ _ m _ _ _
kleedkamer	l _ _ _ _ _ r _ _ _
eindelijk tijd vinden voor iets, ergens aan toe komen	g _ _ r _ _ _ _ t _
veroorzaken, aanleiding geven tot	g_ _ _ r_ _ _ to
beledigende taal	a _ _ _ _ _ _ l _ _ _ _ _ _ _
dagverblijf	d _ _ c _ _ _
iets is duidelijk, gemakkelijk te begrijpen en logisch	m _ _ _ s _ _ _ _
zeldzaam, zelden of nooit	o _ _ _ i _ a b _ _ _ m _ _ _
moment waarop een belangrijke verandering plaatsvindt, keerpunt	t _ _ _ _ _ _ p _ _ _ _
grof taalgebruik	f _ _ _ l _ _ _ _ _ _ _
plaats of cultuur waar verschillende mensen, stijlen, theorieën, enz. samenkomen, smeltkroes	m _ _ _ _ _ _ p _ _
de boodschap of het nieuws verspreiden	s _ _ _ _ _ t _ _ w _ _ _

sterven	m _ _ _ your m _ _ _ _
te weten komen, ontdekken	f _ _ _ o _ _
uiteindelijk	i _ t _ _ e _ _
ijsblokjes	i _ _ c _ _ _ _
niet betrapt worden, er mee wegkomen	g _ _ a _ _ _ w _ _ _
afhangen van	d _ _ _ _ _ o _
wederzijds genoegen, wederzijds genot	m _ _ _ _ _ p _ _ _ _ _ _ _
rekening houden met	t _ _ _ i _ _ _ a _ _ _ _ _ _
naar de wc gaan	s _ _ _ a p _ _ _ _
wegrijden	d _ _ _ _ o _ _
vrijheid van meningsuiting	f _ _ _ s _ _ _ _ _
berispen, de les lezen, terechtwijzen	t _ _ _ o _ _
onaanvaardbaar, over de schreef	b _ _ _ _ _ t _ _ p _ _ _
eigenschap die een (dier)soort helpt voortbestaan in haar concurrentie met andere soorten, evolutionair voordeel	e _ _ _ _ _ _ _ _ _ _ _ a _ _ _ _ _ _ _ _
bezuinigen op	c _ _ b _ _ _ o _
leren uit, putten uit	d _ _ _ f _ _ _
onbedoelde schade die bij een aanval tegen iets of iemand anders wordt aangericht, nevenschade	c _ _ _ _ _ _ _ _ _ d _ _ _ _ _
puur toeval	s _ _ _ _ c _ _ _ _ _ _ _ _ _ _
gemeenschappelijke basis, gemene deler	c _ _ _ _ _ d _ _ _ _ _ _ _ _ _ _
kennis of informatie die iemand goed uitkomt, koren op de molen	g _ _ _ _ t _ t _ _ m _ _ _
heel hard lachen, je doodlachen	l _ _ _ _ l _ _ _ h _ _ _
een recht verkrijgen	w _ _ t _ _ r _ _ _ _
de schuld afschuiven op iemand anders	s _ _ _ _ t _ _ b _ _ _ _
aanstoot nemen, beledigd zijn door iets	t _ _ _ u _ _ _ _ _ _
een tip, advies	p _ _ _ _ o _ a _ _ _ _ _

heel veel a h _ _ _ o _ a l _ _

Questionnaire 1

1 = strongly DISAGREE – 5 = strongly AGREE

1. The topic of the video was interesting. 1 2 3 4 5
2. The topic of the video is relevant to this English course. 1 2 3 4 5
3. The length of the video was appropriate/OK. 1 2 3 4 5
4. The video was easy to understand. 1 2 3 4 5
5. I mainly focused on the content of the video. 1 2 3 4 5
6. Watching a video is a good way to improve your English. 1 2 3 4 5
7. Watching a video is a good way to improve your listening skills in English.
   1 2 3 4 5
8. Watching a video is a good way to improve your English vocabulary.
   1 2 3 4 5
9. Watching a video is a good way to improve your grammar skills in English.
   1 2 3 4 5
10. I paid attention to the words that were used in the video. 1 2 3 4 5
11. What have you learned in terms of content?
12. What have you learned in terms of language/vocabulary (new words)/grammar? Write as many items as you can remember.

Questionnaire 2

In the video you watched last week, there were some expressions that also occurred in the vocabulary tests two weeks ago. Did you notice this while watching the video? *Yes/no*

If yes, did you pay extra attention to the expressions while watching the video? *Yes/no*

Did you think you would get vocabulary tests afterwards, with the same expressions? *Yes/no*

When you received the first vocabulary test today, did you notice that the same expressions were tested in the previous vocabulary tests, a few weeks before? *Yes/no*

Did you notice that the expressions in the test also occurred in the video? *Yes/no*

Did you expect a second test, like last time? *Yes/no*

Did you try to memorize expressions from the first test for the second test? *Yes/no*

Do you think the first test helped you get a higher score on the second test? *Yes/no*

Appendix 3. Target items, instruments, and results of Study 3

Item variables

Target items with values for corpus frequency, mutual information, semantic decomposability, and single word frequency

Item	Frequency	MI	Semantic	SW	SW (lemma)
highest echelons	0.07	14.26	2.42	1172	echelon
foul language	0.28	8.66	5.14	5649	foul
bodily fluids	0.34	14.4	6.23	5633	bodily
common denominator	1.17	12.91	4.00	1969	denominator
supernatural powers	0.24	14.72	6.24	6256	supernatural
sexual depravity	0.02	13.31	4.36	1008	depravity
mutual pleasure	0.02	14.61	6.12	20624	mutual
heck of a lot	0.68	3.69	3.78	12946	heck
pass over (into)	0.01	8.24	4.24	203017	pass
tell off	0.03	10.44	3.47	1119451	tell
tap into	2.71	9.24	3.21	25294	tap
fair description	0.08	4.87	5.34	42081	description
end up	73.19	7.19	4.76	212155	end
evolutionary	0.10	9.27	5.5	10939	evolutionary
abusive language	0.08	7.15	6.06	7681	abusive
guinea pig	0.68	16.39	2.45	5284	guinea
spark your interest	0.42	6.84	5.25	12393	spark
unleash a torrent	0.02	3.69	3.7	2798	torrent
subliminal effect	0.01	5.56	3.48	805	subliminal
take into account	3.96	8.24	3.91	101626	account
pain relievers	0.27	13.36	6.36	2821	reliever
jab line	0.01	3.38	2.44	2002	jab
turning point	5.61	11.33	5.3	8607	turning
win the right	0.10	3.21	4.61	288384	win
beyond the pale	0.52	3.21	2.2	3	pale
sheer coincidence	0.06	16.03	4.03	10667	coincidence
racial epithet	0.19	12.76	2.72	1507	epithet
take umbrage	0.10	10.44	2.84	382	umbrage

Notes. Frequencies and mutual information scores were obtained from COCA (Davies, 2012). MI = mutual information. SW = least frequent word in the MWU.

Form recall test

Give the English expression for each of the meanings described below. The first letter of each word is given.

Nederlandse omschrijving	**Engelse vertaling**
dood (zijn)	p _ _ _ _ _ _ u _ d _ _ _ _ _ _
vrije meningsuiting	f _ _ _ s _ _ _ _ _
je interesse opwekken	s _ _ _ _ your i _ _ _ _ _ _ _
leugentje om bestwil	w _ _ _ _ l _ _
hoogste rangen, hoogste niveaus	h _ _ _ _ _ _ e _ _ _ _ _ _ _
lichaamssappen (bloed, zweet, enz.)	b _ _ _ _ _ f _ _ _ _ _
een spier verrekken	p _ _ _ a m _ _ _ _ _
pijnstillers	p _ _ _ r _ _ _ _ _ _ _ _
bovennatuurlijke krachten	s _ _ _ _ _ _ _ _ _ _ _ p _ _ _ _ _
geslachtsdelen, edele delen	p _ _ _ _ _ _ p _ _ _ _
iemand die deelneemt aan wetenschappelijke tests, proefkonijn	g _ _ _ _ _ p _ _
op je woorden letten, een blad voor de mond nemen	m _ _ _ _ your w _ _ _ _
bijdragen, meebetalen	c _ _ _ i _
verborgen of nauwelijks waarneembare invloed	s _ _ _ _ _ _ _ _ _ e _ _ _ _ _
voedingsbodem	b _ _ _ _ _ _ _ g _ _ _ _ _
goed vertrouwd met, goed op de hoogte van	w _ _ _ v _ _ _ _ _ i _
de kern of het hoogtepunt van de grap, de clou	j _ _ l _ _ _
gezamenlijke inspanning	j _ _ _ _ e _ _ _ _ _
plots heel boos worden	h _ _ t _ _ r _ _ _
een stroom van iets (bijv. verontwaardiging) veroorzaken	u _ _ _ _ _ _ a t _ _ _ _ _ _ o _

upgraden, "pimpen"	j _ _ _ u _
bouwwerf	b _ _ _ _ _ _ _ s _ _ _
overgaan van iets naar iets anders	p _ _ _ o _ _ _ i _ _ _
aanstoot geven, beledigen	c _ _ _ _ o _ _ _ _ _ _
gebruik maken van, benutten, profiteren van	t _ _ i _ _ _
perversie, seksueel verwerpelijk gedrag	s _ _ _ _ _ d _ _ _ _ _ _ _ _
iets op een andere, vernieuwende manier doen, het stramien doorbreken	b _ _ _ _ t _ _ m _ (_) _ _
kleedkamer	l _ _ _ _ _ r _ _ _
eindelijk tijd vinden voor iets, ergens aan toe komen	g _ _ r _ _ _ _ t _
terechtkomen	e _ _ u _
beledigende taal	a _ _ _ _ _ _ l _ _ _ _ _ _ _
zeldzaam, zelden of nooit	o _ _ _ i _ a b _ _ _ m _ _ _
moment waarop een belangrijke verandering plaatsvindt, keerpunt	t _ _ _ _ _ _ p _ _ _ _
grof taalgebruik	f _ _ _ l _ _ _ _ _ _ _
plaats of cultuur waar verschillende mensen, stijlen, theorieën, enz. samenkomen, smeltkroes	m _ _ _ _ _ _ p _ _
de boodschap of het nieuws verspreiden	s _ _ _ _ _ t _ _ w _ _ _
sterven	m _ _ _ your m _ _ _ _
wederzijds genoegen, wederzijds genot	m _ _ _ _ _ p _ _ _ _ _ _ _
rekening houden met	t _ _ _ i _ _ _ a _ _ _ _ _ _
naar de wc gaan	s _ _ _ _ a p _ _ _ _
wegrijden	d _ _ _ _ o _ _
berispen, de les lezen, terechtwijzen	t _ _ _ o _ _
onaanvaardbaar, over de schreef	b _ _ _ _ _ t _ _ p _ _ _

eigenschap die een (dier)soort helpt voortbestaan in haar concurrentie met andere soorten, evolutionair voordeel	e _ _ _ _ _ _ _ _ _ _ _ _ a _ _ _ _ _ _ _ _
bezuinigen op	c _ _ b _ _ _ o _
racistische opmerking	r _ _ _ _ _ e _ _ _ _ _ _
onbedoelde schade die bij een aanval tegen iets of iemand anders wordt aangericht, nevenschade	c _ _ _ _ _ _ _ _ _ d _ _ _ _ _
puur toeval	s _ _ _ _ c _ _ _ _ _ _ _ _ _ _
gemeenschappelijke basis, gemene deler	c _ _ _ _ _ d _ _ _ _ _ _ _ _ _ _
kennis of informatie die iemand goed uitkomt, koren op de molen	g _ _ _ _ t _ t _ _ m _ _ _
heel hard lachen, je doodlachen	l _ _ _ _ l _ _ _ h _ _ _
een recht verkrijgen	w _ _ a r _ _ _ _
de schuld afschuiven op iemand anders	s _ _ _ _ t _ _ b _ _ _ _
aanstoot nemen, beledigd zijn door iets	t _ _ _ u _ _ _ _ _ _
een tip, advies	p _ _ _ _ o _ a _ _ _ _ _
heel veel	a h _ _ _ o _ a l _ _
getrouwe beschrijving	f _ _ _ d _ _ _ _ _ _ _ _ _ _
opscheppen	b _ _ _ your o _ _ t _ _ _ _ _ _

Questionnaire

1 = helemaal mee oneens – 5 = helemaal mee eens

13. Het onderwerp van de documentaire was interessant. 1 2 3 4 5
14. De documentaire duurde niet te lang. 1 2 3 4 5
15. De documentaire was gemakkelijk te begrijpen. 1 2 3 4 5
16. Ik lette vooral op de inhoud van de documentaire. 1 2 3 4 5
17. Ik lette op de woorden die in de documentaire voorkwamen. 1 2 3 4 5
18. Tv kijken is een goede manier om je talenkennis te verbeteren. 1 2 3 4 5
19. Tv kijken is een goede manier om je luistervaardigheid te verbeteren.1 2 3 4 5
20. Tv kijken is een goede manier om je woordenschatkennis te verbeteren.1 2 3 4 5
21. Tv kijken is een goede manier om je kennis van grammatica te verbeteren.1 2 3 4 5

22. Hoe vaak (ongeveer) kijk je Engelstalige tv/film? Omcirkel.
    a. Met Nederlandse ondertitels
    (bijna) nooit / elke maand / elke week / een paar keer per week / elke dag

    b. Met Engelse ondertitels
    (bijna) nooit / elke maand / elke week / een paar keer per week / elke dag

    c. Zonder ondertitels
    (bijna) nooit / elke maand / elke week / een paar keer per week / elke dag

    d. Met ondertitels in een andere taal (niet NL of EN)
    (bijna) nooit / elke maand / elke week / een paar keer per week / elke dag

23. Hoeveel jaar heb je Engelse les gekregen ((lagere school) + middelbaar + (hoger onderwijs))?
24. Had je al eens een aflevering van Fry's Planet Word bekeken? **Ja/nee**
    Had je de aflevering die vandaag werd getoond al eens bekeken? **Ja/nee**
25. Wat heb je geleerd i.v.m. de inhoud van de documentaire?
26. Wat heb je geleerd i.v.m. Engelse woordenschat/grammatica? Schrijf alle items op die je je herinnert.

Analysis of learning gains

The tables below present the results of the models comparing learning gains of target items and distractors. The first model included main effects for time (pretest vs. posttest) and item type (target item vs. distractor), as well as an interaction term between these two variables. The second and third models show the effect of time (pretest vs. posttest) for the target items and distractors separately.

Best-fitting Model for Learning Gains of Target Items vs. Distractors (n = 3,016)

	B	*SE*	*z*	*p*
Fixed effects				
Intercept	-1.45**	0.52	-2.81	0.005
Time	1.19***	0.16	7.51	< 0.001
Item type	0.15	0.68	0.22	0.82
Time*Item type	-0.68**	0.21	-3.22	0.001
Random effects	*Variance*		*SD*	
(1|item)	6.18		2.49	
(1|subject)	0.71		0.84	
Adjusted ICC /		0.68/0.66		
AIC		2532.16		

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. The reference level for time was "pretest", and the

Best-fitting Model for Learning Gains of Target Items (n = *1,456)*

	B	*SE*	*z*	*p*
Fixed effects				
Intercept	-1.45**	0.52	-2.77	0.006
Time	1.2***	0.16	7.48	< 0.001
Random effects	*Variance*		*SD*	
(1 | item)	6.33		2.52	
(1 | subject)	0.75		0.87	
Adjusted ICC /		0.68/0.66		
AIC		1194.84		

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. The reference level for time was "pretest".

Best-fitting Model for Learning Gains of Distractors (n = *1,560)*

	B	*SE*	*z*	*p*
Fixed effects				
Intercept	-1.38*	0.55	-2.52	0.01
Time	0.55***	0.15	3.75	< 0.001
Random effects	*Variance*		*SD*	
(1 | item)	6.87		2.62	
(1 | subject)	1.37		1.17	
Adjusted ICC /		0.72/0.71		
AIC		1300.37		

Note. *** $p < .001$, ** $p < .01$, * $p < .05$. The reference level for time was "pretest".

Appendix 4. Pilot to Study 4

In preparation of Study 4 (Chapter 6), a pilot experiment was conducted which used a different design from that adopted in the main experiment. Analysis of the pilot data revealed methodological limitations which led to considerable changes in the study design. Below, I outline the original design piloted for Study 4 and discuss the motivations for methodological changes.

The pilot experiment focused on the effect of repeated exposure during reading on processing fluency, form recall, and form recognition of collocations. The effect of typographic enhancement was not included in the pilot, because the primary aim was to test elements of the design related to the reading materials (e.g., difficulty of the content, lexical profile), the eye-tracking experiment (e.g., text layout, delineation and labelling of interest areas, etc.) and the posttests (e.g., guessing in multiple-choice test of form recognition). Typographic enhancement would be added to the main experiment in the same counterbalanced fashion as reported in Chapter 6.

The pilot focused on two research questions:

1. What is the effect of repeated exposure on L2 processing fluency of collocations during reading?
2. What is the effect of repeated exposure during reading on form recall and form recognition of collocations?

Method

The experiment adopted a within-participants, counterbalanced design. Participants read texts containing familiar nouns which appeared either with (experimental condition) or without (control condition) a modifying adjective. This approach was similar to that adopted by Durrant and Schmitt's (2010), who used a sentence-reading activity to measure the effect of repetition on learners' explicit memory of novel adjective-noun collocations (e.g., *powerful neck, cheap ball*). Durrant and Schmitt (2010) used two counterbalanced lists of items, each containing different nouns with (e.g., *powerful neck*) and without (e.g., *ball*) a modifying adjective. The pilot experiment was inspired by their design, but incorporated target collocations into larger reading texts.

Participants

For this pilot, 20 Flemish university students were recruited who matched the profile of participants in the main experiment of Study 4. Participants completed the Vocabulary Levels Test in order to get a rough idea of their lexical proficiency level. Their average score was 139/150 (*SD* = 8).

Reading materials

Twenty-four short texts were prepared for the experiment, each approximately 300 words in length. The initial aim was to spread the reading experiment across four weeks, in order to measure the effect of repetition over a longer period of time and with spaced exposures (to reflect the low occurrence rate of collocations in L2 input). The texts were adapted from TED-ed videos in the same way as the reading materials in Study 4: target collocations were inserted and low-frequency words were replaced with higher-frequency words or deleted in order to lower the lexical difficulty of the texts.

Target collocations were 32 semantically transparent, low-frequency (=< 0.1/million in COCA and SUBTLEX-US) adjective-noun collocations, each of which appeared eight times in total across the 24 texts. Because of their low frequency, the collocations were unlikely to be familiar or conventional to learners. However, care was taken to ensure that target collocations were plausible in the contexts of the reading materials. The following collocations were used as target items: *unique signal, talented species, special chance, wild predator, handy ability, exact difference, interesting creature, careful experiment, simple evidence, curious quality, complex mind, straight movement, familiar effect, exciting skill, original behavior, invisible damage, weird process, rare power, useful spot, unusual trick, strange condition, precise study, clever scientist, complicated activity, amazing beast, active function, valuable sense, alert response, dangerous habitat, smart mammal, typical trait, common habit.*

Instruments

Eye-tracking

Participants' eye movements were recorded using the Eyelink Portable Duo (SR Research). Interest areas were created for each of the target collocations and for the single word components. To determine the effect of repeated exposure on participants' processing of the collocations, reading times on the nouns in the experimental and control conditions were compared. A significant difference in reading times between these two conditions would be taken as evidence for learning or increased familiarity as a result of repeatedly encountering the collocations during the treatment.

Posttests of form recall and form recognition

We also included two posttests similar to the ones used in Study 4. The first was a form recall test in which participants had to provide the correct noun for each adjective in a short, non-constraining sentence. The second test was a form recognition test in which participants selected the noun from four options.

Example item of the form recall test:

These wolves are **wild** _ .

Example item of the form recognition test:

It was a **precise** _ .

a. study
b. process
c. system

d. investigation
e. I don't know

Treatment conditions

The collocations occurred either in the control condition (noun-only) or in the experimental condition (adjective-noun). The two conditions are illustrated in the table below. Two presentation lists were created so that each participant would encounter each item in only one of the conditions throughout the experiment. This means that for each participant, 16 collocations would appear in full, and 16 items would appear as nouns only (without the adjectives). In addition, the 24 short texts were grouped into four blocks of six, one block for each session of the experiment. In each block, the 32 items would be presented twice, or eight times in total.

Example sentences from reading materials in the two counterbalanced lists of the pilot

Example sentence from reading text	List
Along with our other great ape cousins, these *interesting creatures* belong to our Hominidae family tree.	1
In fact, few animals could survive the *conditions* in which naked mole rats have to thrive.	1
Along with our other great ape cousins, these *creatures* belong to our Hominidae family tree.	2
In fact, few animals could survive the *strange conditions* in which naked mole rats have to thrive.	2

Procedure

Participants in the pilot came to the eye-tracking lab for a single reading session. Only 12 out of 24 texts were used, to limit the duration of the experiment to a maximum of two hours. Participants were assigned to one of the two presentation lists presented in the table above. To account for the full range of exposures (1-8), and to test all reading materials, different participants read different texts. This meant that the occurrence frequency of target items ranged between 1 and 8 exposures. The presentation of the screens and calibration procedures were similar to those in the experiment of Study 4. After each short text, participants responded to a true-or-false comprehension question using the keyboard. After reading, participants completed the form recall and form recognition tests, as well as a short questionnaire on the difficulty level of the reading texts (similar to the post-viewing questionnaires in Studies 1-3).

Results

I only report descriptive results of the posttest scores, because the sample size was too small to conduct inferential statistical analysis. The results of the form recall test showed a floor effect: no learning gains were found for any of the participants at this level of knowledge. In the form recognition test, participants recognized a small number of target items, but most of the participants' correct responses were guesses, as suggested in the table below. There was, however, a small difference between the experimental and control conditions, which suggests that learners did pick up a few of the target collocations.

Results of the form recognition test (number of correct responses and number of guesses

	Control	Experimental
Total score (max = 16)	4.5 (2.8)	6.4 (2)
"I don't know" (max = 16)	4 (4.2)	2.25 (1.8)

The eye-tracking results showed no obvious differences between the control and experimental conditions in first pass times and total reading times. Differences in reading times across exposures are not reported, because of the small number of observations for each level of this variable.

Results of eye-tracking data per item condition

		First pass time	Total reading time
Control	M	376.91	476.71
	SD	182.70	313.32
Experimental	M	367.66	434.27
	SD	216.67	305.72

Limitations of the pilot and motivation for design changes

There are a few obvious differences between the pilot experiment and the experiment of Study 4 in terms of design, materials, and instruments. In the pilot, the effect of repeated exposure during reading on processing fluency of target collocations was measured by comparing reading times of nouns in the experimental condition with reading times of nouns in the control condition. This approach was chosen because it allowed for a within-participant comparison of experimental and control conditions while controlling for general repetition effects (or decreases in reading times due to increased topic familiarity, see Study 4). However, this design made control of other variables more difficult.

Firstly, the target nouns in the control condition appeared in a slightly different location on the screen than target nouns in the experimental condition, because the removal of the adjective meant that the target noun shifted to the left of the screen. The position of target items is a low-level visual variable that may affect reading times. For example, reading studies typically only measure eye-tracking data for items that do not appear at the end or beginning of a line (e.g., Conklin et al., 2020). Therefore, the use of a counterbalanced approach as operationalized in the pilot may not be ideal when using eye-tracking measures.

Secondly, the decision to counterbalance the experimental and control conditions in the way described above meant that the target items in the pilot study were all adjective-noun collocations containing a contextually redundant adjective (e.g., evaluative adjectives such as *amazing, interesting, handy, useful, …*). Further, unlike Durrant and Schmitt's (2010) sentence-reading task, the target collocations were incorporated repeatedly into continuous texts, which made the selection of meaningful, contextually relevant target items even more difficult. The results of the form recall and form recognition tests showed that, in spite of the tests being administered immediately after reading, participants could not recall any of the target collocations, and only correctly recognized a very small number of items. Participants informally reported that they found the tests very difficult, and had resorted to guessing most of the time. I believe that the communicative redundancy of the adjectives in the target collocations may have made the target items very difficult to recall (e.g., Cintrón-Valentín & Ellis, 2016), but also irrelevant as targets for learning.

In order to increase reliability of the eye-tracking data, and to allow for the selection of more learnable and pedagogically relevant collocations, the counterbalanced design as described here was abandoned. Instead, Study 4 included a sentence-reading posttest to measure processing fluency of target collocations from the reading texts and control items that did not appear in the reading texts. Instead of the low-redundant adjective-noun collocations from the pilot, more meaningful collocations with technical/contextual meanings were included in the experiment of Study 4.

The format of the posttests of form recall and form recognition in the pilot experiment were also changed. In the cued recall test of the pilot, participants had to recall the correct noun based on the target item. In the experiment of Study 4, the noun was provided, and participants had to recall the correct adjective. This format was more comparable to that used in previous studies (e.g., Sonbul & Schmitt, 2013). In addition, the low-constraining contexts of the pilot tests were replaced with meaningful, constraining sentences in the posttests of Study 4. This was done to make the tests slightly easier and more sensitive to learning of the form-meaning associations of the target collocations.

Finally, a small but important change to the eye-tracking experiment involved the use of the keyboard during the experiment. In the pilot, participants used three different keys on the keyboard to select a response to the comprehension questions, and moved from one screen to the next by pressing the spacebar. However, participants could not press the correct keys without moving their heads, which led to considerable drift in the eye movement registration. Therefore, in the experiment of Study 4, all keyboard commands were replaced by a mouse click, which enabled participants to proceed through the experiment with minimal need for recalibration.

Although the design of Study 4 still had its limitations (see Section 7.5), the pilot experiment led to considerable improvements in the study design, and made an important contribution to the preparation of the reading materials, target items, and eye-tracking experiments.

Appendix 5. Target items, instruments and results of Study 4

Target collocations

Target collocations and their frequency per million words in the COCA

	Frequency per million - collocation	Frequency per million - word 1	Frequency per million - word 2
action potential	0.12	137.26	39.72
automatic recognition	0.01	15.24	28.54
capturing surface	0.00	57.41	80.70
central brain	0.02	104.51	100.03
clever arm	0	14.10	160.92
compound eye	0.08	12.63	346.62
conscious pain	0.00	15.88	95.94
dead zone	0.64	178.41	48.32
global workspace	0.01	99.69	0.81
human stress	0.00	257.30	46.32
inner access	0	27.79	95.18
internal model	0.05	45.19	191.13
light cup	0.00	214.50	96.68
nervous activity	0.01	33.72	162.67
novel task	0.01	53.40	82.70
pain behavior	0.04	95.94	138.18
physical color	0	109.44	159.76
pinhole effect	0	0.29	217.63
reflecting skin	0.00	10.23	82.46
reflex response	0.02	2.29	135.72
sensitive cell	0.02	28.08	127.57
shared control	0.03	52.32	200.54
simple spot	0	120.53	73.89
visual sensitivity	0.01	37.29	12.72

Form recall test

De zinnen hieronder bevatten vetgedrukte woordcombinaties waarvan één woord ontbreekt. Lees de context en probeer het juiste woord in te vullen. Als je het antwoord niet kent, zet dan "x".

All animals, even insects, show a(n) _ **response** in harmful or painful situations: this is a direct reaction of the nervous sytem.	_ response
A(n) _ **baby** is a young child that spreads infectious diseases very efficiently.	_ baby
The human eye contains two types of _ **cells** that can detect light.	_ cells
The _ **complaint** or most important sign of measles is a high temperature.	_ complaint
The bodies of octopuses have _ **arms** that can think all by themselves.	_ arms
The different parts of the nervous system have _ **control**: they decide over the body's movements together.	_ control
Insects have _ **eyes**, which consist of many tiny units that can detect fast movements.	_ eyes
Tree rings and carbon dating are examples of _ **clocks** used by evolutionary biologists.	_ clocks
The vertical space through which deep-sea divers go up and down is called the _ **column**.	_ column
The human brain is capable of advanced _ **activity** or interaction between neurons.	_ activity
Very few insects fly like a(n) _ **bullet** straight from the flower where they have picked up pollen to another flower.	_ bullet
Neurons communicate by means of _ **potentials**, which are electrical impulses.	_ potentials
A(n) _ **shadow** is an area on the side of a mountain that is sheltered from the wind and rain.	_ shadow
Our minds can form _ **models** of information from the senses, which makes us capable of complex thought.	_ models
A mechanical device that helps polio patients breathe is called a(n) _ **lung**.	_ lung
In the octopus, the _ **brain** works together with the rest of the nervous system.	_ brain
_ **interactions** happen when the combination of two substances, like caffeine or alcohol, causes different effects than either would individually.	_ interactions
The environment is threatened by many forms of _ **stress**, such as CO emissions and farm waste.	_ stress
The human brain contains _ **areas** where injuries cause no clear symptoms.	_ areas
The second process of pain is called _ **pain**, and this is only felt by some animals, like birds and mammals.	_ pain
In the deep ocean, animals are adapted to catch _ **snow**, which are particles of dead animals and plants that drift down from above.	_ snow
One symptom of epilepsy is _ **absence**, during which the patient briefly loses consciousness.	_ absence

After a mental trauma, a person may suffer a psychological condition called _ **shock**.	_ shock
In the evolution of the eye, the _ **cup** had a smaller opening, which enabled the eye to detect the direction of movement.	_ cup
Sometimes a patient has to follow a(n) _ **diet**, during which they eat only food containing little fiber.	_ diet
During heart surgery, a serious complication called _ **heart** may occur, during which the heart stops beating.	_ heart
The adaptations of the human eye lead to incredible _ **sensitivity**, with great clarity and resolution.	_ sensitivity
Scientists have observed the _ **behavior** of animals in the wild and in the lab, to understand how they experience pain.	_ behavior
An ice age is an example of a change in the _ **climate** which drives evolution, along with changes in the genetic climate.	_ climate
Critically injured patients need to receive treatment straight away, preferably during the _ **hour** immediately after the accident.	_ hour
Cancer drugs are often given to patients in cycles, which consist of drug treatments followed by _ **periods** during which no treatment is given.	_ periods
As a result of a lack of oxygen, many parts of the ocean have formed _ **zones**, where nothing can survive.	_ zones
What we can't do unconsciously, are _ **tasks**, like driving a car for the first time.	_ tasks
The octopus has a _ **skin** that makes camouflage much easier.	_ skin
A potential cure for aids is _ **therapy**, which puts biological material into a patient's body.	_ therapy
At the back of the eyeball, humans have a(n) _ **surface**, or retina, which detects the incoming light.	_ surface
The evolution of the eye started from a _ **spot** that could only detect light.	_ spot
The _ **colors** we can see are not created by the brain but are a property of the light.	_ colors
Some fractures occur between one end of the bone and the _ **end**, which is the end that doesn't move.	_ end
When an electric eel attacks, it curls up and creates a(n) _ **storm** which exhausts and immobilizes the prey.	_ storm
The first process of pain is the _ **recognition** of harm, which all animals experience.	_ recognition
When a person suffers from _ **hand** syndrome, they can only move the muscles in one part of the hand.	_ hand
Consciousness gives us _ **access**, allowing us to interact with our thoughts and emotions.	_ access
The _ **number** of an atom is equal to the combined number of protons and neutrons.	_ number
A snake sometimes uses a _ **bite** without any venom, which serves as a warning but doesn't kill.	_ bite
In the evolution of the eye, the _ **effect** improved the sharpness of vision by capturing only a thin beam of light.	_ effect

Evolutionary biologists study _ **trees**, or family trees that only split into two separate branches.	_ trees
All input from the senses is brought together in the _ **workspace**, where we form coherent thoughts.	_ workspace

Form recognition test

De zinnen hieronder bevatten vetgedrukte woordcombinaties waarvan één woord ontbreekt.

Lees de context en probeer het juiste woord in te vullen. Kies altijd één van de vier opties,

ook als je het antwoord niet kent.

All animals, even insects, show a(n) _ **response** in harmful or painful situations: this is a direct reaction of the nervous sytem.

1. trigger 2. impulse 3. instinct 4. reflex

A _ **baby** is a young child that spreads infectious diseases very efficiently.

1. steam 2. mist 3. cloud 4. fog

The human eye contains two types of _ **cells** that can detect light.

1. responsive 2. sensitive 3. receptive 4. precise

The _ **complaint** or most important sign of measles is a high temperature.

1. chief 2. prime 3. first 4. key

The bodies of octopuses have _ **arms** that can think all by themselves.

1. bright 2. smart 3. intelligent 4. clever

The different parts of the nervous system have _ **control**: they decide over the body's movements together.

1. mutual 2. joint 3. common 4. shared

Insects have _ **eyes**, which consist of many tiny units that can detect fast movements.

1. multiple 2. camera 3. round 4. compound

Tree rings and carbon dating are examples of _ **clocks** used by evolutionary biologists.

1. organic 2. primitive 3. natural 4. cellular

The vertical space through which deep-sea divers go up and down is called the _ **column**.

1. blue 2. ocean 3. water 4. sea

The human brain is capable of advanced _ **activity** or interaction between neurons.

1. nervous 2. organized 3. sensory 4. united

Very few insects fly like a(n) _ **bullet** straight from the flower where they have picked up pollen to another flower.

1. magic 2. direct 3. magnetic 4. electric

Neurons communicate by means of _ **potentials**, which are electrical impulses.

1. action 2. force 3. energy 4. power

A _ **shadow** is an area on the side of a mountain that is sheltered from the wind and rain.

1. dry 2. wind 3. weather 4. rain

Our minds can form _ **models** of information from the senses, which makes us capable of complex thought.

1. structural 2. organic 3. internal 4. subjective

A mechanical device that helps polio patients breathe is called a(n) _ **lung**.

1. metal 2. iron 3. steel 4. stone

In the octopus, the _ **brain** works together with the rest of the nervous system.

1. chief 2. master 3. central 4. main

_ **interactions** happen when the combination of two substances, like caffeine or alcohol, causes different effects than either would individually.

1. drug 2. influence 3. dose 4. effect

The environment is threatened by many forms of _ **stress**, such as CO emissions and farm waste.

1. cultural 2. modern 3. industrial 4. human

The human brain contains _ **areas** where injuries cause no clear symptoms.

1. invisible 2. mute 3. silent 4. dark

The second process of pain is called _ **pain**, and this is only felt by some animals, like birds and mammals.

1. conscious 2. subjective 3. alert 4. rational

In the deep ocean, animals are adapted to catch _ **snow**, which are particles of dead animals and plants that drift down from above.

1. floating 2. sinking 3. aquatic 4. marine

One symptom of epilepsy is _ **absence**, during which the patient briefly loses consciousness.

1. wide 2. open 3. direct 4. pure

After a mental trauma, a person may suffer a psychological condition called _ **shock**.

1. skin 2. head 3. shell 4. spine

In the evolution of the eye, the _ **cup** had a smaller opening, which enabled the eye to detect the direction of movement.

1. ray 2. window 3. sense 4. light

Sometimes a patient has to follow a _ **diet**, during which they eat only food containing little fiber.

1. smooth 2. strong 3. stable 4. plain

During heart surgery, a serious complication called _ **heart** may occur, during which the heart stops beating.

1. solid 2. stone 3. stiff 4. metal

The adaptations of the human eye lead to incredible _ **sensitivity**, with great clarity and resolution.

1. optic 2. graphic 3. visual 4. dynamic

Scientists have observed the _ **behavior** of animals in the wild and in the lab, to understand how they experience pain.

1. injury 2. harm 3. pain 4. damage

An ice age is an example of a change in the _ **climate** which drives evolution, along with changes in the genetic climate.

1. incidental 2. secondary 3. outer 4. external

Critically injured patients need to receive treatment straight away, preferably during the _ **hour** immediately after the accident.

1. golden 2. rich 3. capital 4. critical

Cancer drugs are often given to patients in cycles, which consist of drug treatments followed by _ **periods** during which no treatment is given.

1. break 2. interval 3. rest 4. gap

As a result of a lack of oxygen, many parts of the ocean have formed _ **zones**, where nothing can survive.

1. dry 2. void 3. cold 4. dead

What we can't do unconsciously, are _ **tasks**, like driving a car for the first time.

1. initial 2. original 3. novel 4. fresh

The octopus has a _ **skin** that makes camouflage much easier.

1. wavering 2. coloring 3. shading 4. reflecting

A potential cure for aids is _ **therapy**, which puts biological material into a patient's body.

1. growth 2. cell 3. organic 4. gene

At the back of the eyeball, humans have **a(n) _ surface**, or retina, which detects the incoming light.

1. perceiving 2. observing 3. viewing 4. capturing

The evolution of the eye started from a **_ spot** that could only detect light.

1. plain 2. simple 3. basic 4. pure

The **_ colors** we can see are not created by the brain but are a property of the light.

1. material 2. physical 3. objective 4. visible

Some fractures occur between one end of the bone and the **_ end**, which is the end that doesn't move.

1. still 2. locked 3. fixed 4. firm

When an electric eel attacks, it curls up and creates **a(n) _ storm** which exhausts and immobilizes the prey.

1. magnetic 2. electrical 3. dynamic 4. charged

The first process of pain is the **_ recognition** of harm, which all animals experience.

1. routine 2. automatic 3. spontaneous 4. natural

When a person suffers from **_ hand** syndrome, they can only move the muscles in one part of the hand.

1. split 2. cleft 3. sliced 4. torn

Consciousness gives us **_ access**, allowing us to interact with our thoughts and emotions.

1. inner 2. psychological 3. mental 4. private

The **_ number** of an atom is equal to the combined number of protons and neutrons.

1. energy 2. mass 3. core 4. volume

A snake sometimes uses a **_ bite** without any venom, which serves as a warning but doesn't kill.

1. dull 2. fake 3. raw 4. dry

In the evolution of the eye, the **_ effect** improved the sharpness of vision by capturing only a thin beam of light.

1. pocket 2. pinhole 3. opening 4. keyhole

Evolutionary biologists study **_ trees**, or family trees that only split into two separate branches.

1. double 2. binary 3. forked 4. split

All input from the senses is brought together in the **_ workspace**, where we form coherent thoughts.

1. global 2. vital 3. interior 4. primary

Sentence-reading posttest

In this test, each of the 24 collocations from the reading texts (e.g., *dead zone*) is matched with a control phrase containing a different final word (e.g., *dead belt*). The control collocations have very low or zero frequency in the COCA. They are matched with the target items in single word frequency and length. The control items are embedded in the same sentences as the target collocations. All critical sentences are matched for the number of words that appear before the critical phrase (3-5 words). The stimuli also include 24 distractor collocations (some of which were borrowed from Toomer and Elgort, 2019) which are later tested in the delayed posttests of cued recall and form recognition (to control for learning from the reading posttest), and 24 filler sentences not containing technical collocations.

Condition	**Context before item**	***Collocation***	**Context after item**
Target	A great number of	*dead zones*	have appeared, causing some fish to leave the area.
Control	A great number of	*dead belts*	have appeared, causing some fish to leave the area.
Distractor	Research shows that	*silent areas*	may not really exist, but some scientists insist they do.
Filler	Dolphins recognize themselves in mirrors, which shows they are aware of their own bodies.		

Each collocation is only encountered in one of the conditions by each participant. Participants are assigned to one of two presentation lists. Collocations appearing in the target condition in list 1, appear in the control condition in list 2, and vice versa. The distractor items are split between the two presentation lists, so that 12 distractors appear in each list. The 24 filler sentences appear in both lists. This results in 60 sentences per participant, in four conditions: target (n = 12), control (n = 12), distractor (n = 12) and filler (n = 24). The order of presentation of the sentences is randomized for each participant. To make sure that participants read the sentences attentively, one third of the sentences is followed by a true-or-false comprehension question.

List	Collocation	Condition	Sentence
1	compound heads	control	We learned that the compound heads of lobsters are able to detect very fast motions.
1	nervous ability	control	Scientists study the nervous ability of this monkey by using a special type of scanner.
1	simple cover	control	Some worms have a simple cover that can detect contrasts when there is enough light.

1	physical images	control	We can see the physical images on our TV with only 3 different wave frequencies.
1	capturing screen	control	In some people, the capturing screen does not have active cones, causing color-blindness.
1	novel trial	control	Most types of novel trial cannot be done without focus, even by very clever people.
1	reflex movement	control	Apart from the reflex movement of the animal, we can't tell if the spider feels anything.
1	conscious loss	control	It's unlikely that conscious loss can be felt by insects, but we shouldn't harm them.
1	capacity in	control	Because of the action capacity in our nerve system, neurons can respond very quickly.
1	global workshop	control	Information in the global workshop can stay there for a short time, about 30 seconds.
1	light ball	control	The insects have a light ball that makes it much easier to hide from enemies.
1	inner network	control	We have a kind of inner network that clearly separates our thinking from that of fish.
1	shared control	target	The body parts use their shared control to make navigation through tunnels possible.
1	dead zones	target	A great number of dead zones have appeared, causing some fish to leave the area.
1	central brain	target	In these animals the central brain is very important, but a dog's head has more neurons.
1	automatic recognition	target	This process is the automatic recognition that makes worms react when you touch them.
1	reflecting skin	target	Look at the reflecting skin of the animal: it can mirror the background perfectly.
1	internal models	target	Humans can build internal models of many things, but fish only achieve simple reactions.
1	clever arms	target	To test how the clever arms work together, scientists set up an experiment with a tunnel.
1	human stress	target	To limit the amount of human stress in the future, we need to change our farming methods.
1	visual sensitivity	target	We can learn about their visual sensitivity by studying the eye structure very closely.
1	sensitive cells	target	On top of the sensitive cells for normal vision, birds can also see ultraviolet light.
1	pain behavior	target	By observing the pain behavior of these crabs, we found out that they leave their shells.
1	pinhole effect	target	Thanks to the pinhole effect and other features, human eyes are like a camera.
1	silent areas	distractor	Research shows that silent areas may not really exist, but some scientists insist they do.
1	fixed end	distractor	In one type, the fixed end of the bone remains still while the other end bends.
1	cloud baby	distractor	The medical term cloud baby was first used to refer to these children in the 1960s.

1	iron lung	distractor	They invented the iron lung in the 1920s to make breathing easier for polio patients.
1	chief complaint	distractor	The most typical chief complaint of a person suffering from a heart attack is chest pain.
1	magic bullet	distractor	They fly like a magic bullet from one flower to the next of only one orchid species.
1	drug interaction	distractor	This can create a drug interaction that could lead to kidney failure or liver damage.
1	electrical storm	distractor	The result of the electrical storm is that the eel can capture and swallow its prey alive.
1	rain shadow	distractor	The desert lies in the rain shadow and receives only a few inches of water per year.
1	smooth diet	distractor	A person on a smooth diet eats food containing little fibre, such as soft cheese.
1	golden hour	distractor	The essence of the golden hour is that an injured patient needs treatment immediately.
1	shell shock	distractor	Common symptoms of shell shock include extreme tiredness and the inability to sleep.
1	NA	filler	On our planet, we have two polar regions, which are called the Arctic and the Antarctic.
1	NA	filler	We should accept that wildfires are vital to the existence of healthy forest ecosystems.
1	NA	filler	To save the killer whales from extinction, we need to restore salmon populations.
1	NA	filler	Fruits and vegetables such as berries, cucumbers and broccoli are over 90% water.
1	NA	filler	Because viruses can reproduce by the millions, they can very quickly develop mutations.
1	NA	filler	Supported by water, sea animals can ignore gravity, which means they can grow much larger.
1	NA	filler	Unlike plants, animals spend energy on all sorts of things, like eating and staying warm.
1	NA	filler	Sugar is a general term used to describe a class of molecules called carbohydrates.
1	NA	filler	The garbage patch can shift from the size of Texas to the size of the United States.
1	NA	filler	The DNA of ants changes from generation to generation and differences become larger.
1	NA	filler	The largest species of jellyfish has tentacles that can extend more than 100 feet.
1	NA	filler	Most insects have smell detectors on their antennae, similar to those in human noses.
1	NA	filler	Studies found that some kinds of pesticide can damage the honeybees' navigation skills.
1	NA	filler	As glaciers shrink because of climate change, the amount of regional water decreases.
1	NA	filler	We can build fish reserves closed to all fishing to help ecosystems restore themselves.

1	NA	filler	Shrimp are caught by dragging nets the size of a football field along the ocean bottom.
1	NA	filler	It may be that electric fish can withstand their own shocks because they are so large.
1	NA	filler	Dolphins recognize themselves in mirrors, which shows they are aware of their own bodies.
1	NA	filler	Ocean currents are driven by the wind, tides, and the rotation of the Earth.
1	NA	filler	Claws are ideal for piercing and hooking, but their points make grabbing difficult.
1	NA	filler	All matter is made up of atoms, which consist of three types of smaller particles.
1	NA	filler	So far, scientists have only been able to explore one per cent of the deep ocean floor.
1	NA	filler	Evidence of our influence is seen all over the ocean, no matter how deep or distant.
1	NA	filler	The best fishing grounds are found by the grandmother, who leads the whales in the hunt.
2	compound eyes	target	We learned that the compound eyes of lobsters are able to detect very fast motions.
2	nervous activity	target	Scientists study the nervous activity of this monkey by using a special type of scanner.
2	simple spot	target	Some worms have a simple spot that can detect contrasts when there is enough light.
2	physical colors	target	We can see the physical colors on our TV with only 3 different wave frequencies.
2	capturing surface	target	In some people, the capturing surface does not have active cones, causing color-blindness.
2	novel task	target	Most types of novel task cannot be done without focus, even by very clever people.
2	reflex response	target	Apart from the reflex response of the animal, we can't tell if the spider feels anything.
2	conscious pain	target	It's unlikely that conscious pain can be felt by insects, but we shouldn't harm them.
2	action potentials	target	Because of the action potentials in our nerve system, neurons can respond very quickly.
2	global workspace	target	Information in the global workspace can stay there for a short time, about 30 seconds.
2	light cup	target	The insects have a light cup that makes it much easier to hide from enemies.
2	inner access	target	We have a kind of inner access that clearly separates our thinking from that of fish.
2	shared forces	control	The body parts use their shared forces to make navigation through tunnels possible.
2	dead belts	control	A great number of dead belts have appeared, causing some fish to leave the area.
2	central memory	control	In these animals the central memory is very important, but a dog's head has more neurons.

2	automatic perception	control	This process is the automatic perception that makes worms react when you touch them.
2	reflecting edge	control	Look at the reflecting edge of the animal: it can mirror the background perfectly.
2	internal figures	control	Humans can build internal figures of many things, but fish only achieve simple reactions.
2	clever legs	control	To test how the clever legs work together, scientists set up an experiment with a tunnel.
2	human weight	control	To limit the amount of human weight in the future, we need to change our farming methods.
2	visual impression	control	We can learn about their visual impression by studying the eye structure very closely.
2	sensitive base	control	On top of the sensitive base for normal vision, birds can also see ultraviolet light.
2	pain language	control	By observing the pain language of these crabs, we found out that they leave their shells.
2	pinhole sense	control	Thanks to the pinhole sense and other features, human eyes are like a camera.
2	pure absence	distractor	When an attack of pure absence occurs, a patient may suddenly feel weak or numb.
2	split hand	distractor	A patient with a split hand can have trouble communicating, leading to social rejection.
2	gene therapy	distractor	One type of gene therapy may in the future help the immune system control the virus.
2	external climate	distractor	Changes to the external climate and to genetics may have an effect on evolution.
2	dry bite	distractor	By means of a dry bite the snakes can send a message to warn their victim.
2	mass number	distractor	This means the mass number goes down, and the atom turns into a different element.
2	water column	distractor	Going down in the water column we see much larger fish than at the sea surface.
2	binary trees	distractor	The study of binary trees took several years to complete, because there were so many.
2	marine snow	distractor	Survival depends on marine snow and other curious phenomena found in the deep ocean.
2	natural clocks	distractor	The diversity of natural clocks is useful for archeologists, but also for dating fossils.
2	gap periods	distractor	In most cases the gap periods tend to be much longer than the treatment periods.
2	stone heart	distractor	In recent years stone heart has been avoided through the use of modern medical techniques.
2	NA	filler	On our planet, we have two polar regions, which are called the Arctic and the Antarctic.
2	NA	filler	We should accept that wildfires are vital to the existence of healthy forest ecosystems.
2	NA	filler	To save the killer whales from extinction, we need to restore salmon populations.

2	NA	filler	Fruits and vegetables such as berries, cucumbers and broccoli are over 90% water.
2	NA	filler	Because viruses can reproduce by the millions, they can very quickly develop mutations.
2	NA	filler	Supported by water, sea animals can ignore gravity, which means they can grow much larger.
2	NA	filler	Unlike plants, animals spend energy on all sorts of things, like eating and staying warm.
2	NA	filler	Sugar is a general term used to describe a class of molecules called carbohydrates.
2	NA	filler	The garbage patch can shift from the size of Texas to the size of the United States.
2	NA	filler	The DNA of ants changes from generation to generation and differences become larger.
2	NA	filler	The largest species of jellyfish has tentacles that can extend more than 100 feet.
2	NA	filler	Most insects have smell detectors on their antennae, similar to those in human noses.
2	NA	filler	Studies found that some kinds of pesticide can damage the honeybees' navigation skills.
2	NA	filler	As glaciers shrink because of climate change, the amount of regional water decreases.
2	NA	filler	We can build fish reserves closed to all fishing to help ecosystems restore themselves.
2	NA	filler	Shrimp are caught by dragging nets the size of a football field along the ocean bottom.
2	NA	filler	It may be that electric fish can withstand their own shocks because they are so large.
2	NA	filler	Dolphins recognize themselves in mirrors, which shows they are aware of their own bodies.
2	NA	filler	Ocean currents are driven by the wind, tides, and the rotation of the Earth.
2	NA	filler	Claws are ideal for piercing and hooking, but their points make grabbing difficult.
2	NA	filler	All matter is made up of atoms, which consist of three types of smaller particles.
2	NA	filler	So far, scientists have only been able to explore one per cent of the deep ocean floor.
2	NA	filler	Evidence of our influence is seen all over the ocean, no matter how deep or distant.
2	NA	filler	The best fishing grounds are found by the grandmother, who leads the whales in the hunt.

Results

Results of the best fitting models for eye-tracking measures during reading (n = *7,216)*

	Estimate	SE	*T*	*p*
Fixed effects				
Intercept	6.87	0.05	130.17	< .001
TE	-0.22	0.03	-6.39	< .001
Exposure 2	-0.32	0.03	-9.61	< .001
Exposure 3	-0.29	0.03	-8.71	< .001
Exposure 4	-0.37	0.03	-11.13	< .001
Exposure 5	-0.40	0.03	-12.13	< .001
Exposure 6	-0.38	0.03	-11.47	< .001
Exposure 7	-0.39	0.03	-11.60	< .001
Exposure 8	-0.48	0.03	-14.32	< .001
Item lengt	0.92	0.11	8.69	< .001
TE*Exposure 2	0.23	0.05	4.82	< .001
TE*Exposure 3	0.27	0.05	5.75	< .001
TE*Exposure 4	0.29	0.05	6.06	< .001
TE*Exposure 5	0.24	0.05	5.13	< .001
TE*Exposure 6	0.18	0.05	3.86	< .001
TE*Exposure 7	0.18	0.05	3.86	< .001
TE*Exposure 8	0.28	0.05	5.84	< .001
Random effects	Variance	SD		
(1 ǀ Item)	0.07	0.27		
(1 ǀ Participant)	0.01	0.10		
Residual	0.25	0.50		
Marginal R2 / conditional R2	0.12 / 0.34			
AIC	10772.18			

The reference level of TE is 'enhanced', and the reference level of Exposure count is 'exposure 1'.

Best fitting models for reading times in the sentence-reading posttest in the experimental group (total reading time and first pass time)

	Total reading time (*n* = 940)				First pass time (*n* = 953)			
	Estimate	SE	*T*	*p*	Estimate	SE	*T*	*p*
Fixed effects								
Intercept	6.80	0.06	116.41	< 0.001	5.63	0.08	73.412	< 0.001
Item = control	0.19	0.05	3.40	0.001	0.08	0.07	1.156	0.251
Item = enhanced	-0.002	0.04	-0.08	0.939	-0.06	0.07	-0.793	0.428
Length (log transformed)	0.78	0.14	5.54	< 0.001	-0.01	0.16	-0.043	0.966
Random effects	Variance	SD			Variance	SD		
(1 ǀ Item)	0.02	0.15			0.01	0.09		
(1 ǀ Participant)	0.07	0.26			0.11	0.33		
Residual	0.16	0.40			0.65	0.80		
Marginal R2 / conditional R2	0.10 / 0.43				0.004 / 0.16			
AIC	1137.80				2384.85			

Best fitting models for reading times in the sentence-reading posttest in the experimental group (rereading time and total reading time of the final word)

	Rereading time (*n* = 773)				Total reading time final word (*n* = 903)			
	Estimate	SE	*T*	*p*	Estimate	SE	*T*	*p*
Fixed effects								
Intercept	6.43	0.07	92.54	< 0.001	6.08	0.06	103.07	< 0.001
Item = control	0.22	0.07	3.29	0.002	0.21	0.06	3.70	< 0.001
Item = enhanced	0.07	0.06	1.22	0.223	-0.05	0.05	-1.01	0.311
Length (log transformed)	0.73	0.17	4.40	< 0.001	0.58	0.08	7.06	< 0.001
Random effects	Variance	SD			Variance	SD		
(1 ǀ Item)	0.02	0.15			0.02	0.14		
(1 ǀ Participant)	0.09	0.30			0.06	0.25		
Residual	0.31	0.56			0.22	0.47		
Marginal R2 / conditional R2	0.06 / 0.30				0.13 / 0.37			
AIC	1423.08				1345.56			