Title: The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013)

Running head: EXPLICIT AND IMPLICIT KNOWLEDGE OF COLLOCATIONS

Authors:

Mark Toomer and Irina Elgort

School of Linguistics and Applied Language Studies, Victoria University of Wellington

Keywords: vocabulary, collocations, implicit knowledge, explicit knowledge, typographic enhancement, glossing

Author Note:

Acknowledgements

This research was supported by Victoria Doctoral Scholarship and the Faculty of Humanities and Social Sciences Research Grant #209320 to Mark Toomer. We are grateful to Dr. Lisa Woods, Statistical Consultant, School of Mathematics and Statistics, for her expert advice, as well as to three anonymous reviewers and Associate Editor Prof. Kara Morgan-Short whose comments and suggestions helped to make this a better article.

Address for correspondence:

Mark Toomer, School of Linguistics and Applied Language Studies, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand. E-mail: Mark.Toomer@vuw.ac.nz

Abstract

Sonbul and Schmitt (2013) have shown that exposure to L2 collocations in reading texts can produce gains in explicit knowledge, but they found no evidence of gains in implicit knowledge. The present study is a conceptual replication and an extension of Sonbul and Schmitt (2013). Sixty-two advanced ESL speakers read texts containing repeated occurrences of low-frequency medical collocations in three sessions over two days. Three incidental learning treatments were: reading-only (no typographic enhancement), bolding and bolding-plus-glossing. Collocational knowledge was assessed in two tests of explicit knowledge (cued-recall and form-recognition) and one test of implicit knowledge (primed lexical decisions). Repeated exposure to bolded collocations produced greater explicit knowledge than exposure to typographically-unenhanced collocations. Initial evidence of implicit knowledge development was observed in the unenhanced (reading-only) treatment, but not in the other treatments. In summary, our study replicated Sonbul and Schmitt's findings on explicit knowledge and extended their findings on implicit knowledge.

Introduction

A considerable proportion of language used in spoken and written discourse is formulaic; that is, words tend to co-occur in set configurations that can be described as multiword chunks or units (Sinclair, 1991). Access to knowledge of multiword units (MWUs) plays a central role in fluent comprehension and production. Research into formulaic language suggests that quality of knowledge and fluency of processing of MWUs are affected by the frequency of their occurrence in input, both in the first language (L1) and second language (L2) (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Wolter & Yamashita, 2018). However, more research is needed to trace the development of this frequency effect and clarify the role that learning and instructional approaches play in this development. Research into teaching and learning L2 MWUs has, so far, dealt primarily with the acquisition of explicit knowledge, with only a small number of studies investigating the development of implicit knowledge or comparing the developmental trajectories of explicit and implicit knowledge of MWUs. One such study, Sonbul and Schmitt (2013), found evidence for the development of explicit knowledge but not for the development of implicit knowledge of collocations. The present study was set up to corroborate their findings for explicit knowledge and extend the findings for implicit knowledge, for L2 speakers. Sonbul and Schmitt conjectured that a relatively short treatment duration and a limited number of exposures to the target collocations were possible reasons why the development of implicit collocational knowledge had not been observed. Our study puts this conjecture to the test by increasing the number of exposures and conducting the treatment over a two-day period.

Literature review

Factors in the acquisition and processing of L2 MWUs

Acquisition of MWUs in a non-native language is challenging (Laufer & Waldman, 2011; Li & Schmitt, 2010; Peters, 2014). This is because of the relatively low frequency of occurrence of MWUs in input compared with their composite individual words (Ellis, 2002; Siyanova-

Chanturia & Martinez, 2015) and because repeated exposure is needed for learning (e.g., Sonbul & Schmitt, 2013; Webb, Newton, & Chang, 2013). Durrant and Schmitt (2010), for example, found that repeated exposure to L2 adjective-noun collocations in sentence contexts resulted in better recall in a timed naming task, compared to a single exposure. The frequency effect has also been found in the visual processing of MWUs during reading. Siyanova-Chanturia, Conklin, and van Heuven (2011) reported a processing advantage for the reading of binomials (e.g., *bride and groom*), compared to their reversed forms. Studies that have examined the recognition and processing of MWUs that are either congruent or incongruent in the participants' L1 and L2 also show sensitivity to the frequencies of occurrence across the L1 and L2 (Carrol, Conklin, & Gyllstad, 2016; Wolter & Gyllstad, 2011, 2013). For example, Wolter and Gyllstad (2011) reported that advanced L2 learners were faster in making acceptability judgements about congruent collocations than about incongruent collocations and non-collocations.

Another factor contributing to the slow acquisition of MWUs is lack of salience in written discourse. This is because the MWUs are not marked as phrases within written texts and some are discontinuous (e.g., *as … as*) (Boers & Lindstromberg, 2009). As a result, learners often fail to detect MWUs (Bishop, 2004), instead focusing on single words (Henriksen, 2012). The amount of allocated attention can be increased through a brief externally-induced focus on target words in the context of meaning-focused activities, called *input enhancement* (Sharwood Smith, 1991, 1993; Szudarski & Carter, 2016). Techniques of input enhancement include typographic enhancement and glossing (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Ellis, 1999; Hulstijn, Hollander, & Greidanus, 1996). Eye-tracking studies show that typographic enhancement draws learners' attention to non-salient forms (Choi, 2017; Cintrón-Valentín & Ellis, 2015; Winke, 2013). The few studies that have investigated the effects of typographic enhancement on the learning of L2 MWUs (e.g., Boers et al., 2016; Choi, 2017; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016) suggest that participants remember unfamiliar MUWs better when they are typographically-enhanced in written input. The effect of glossing on the learning and retention of MWUs is still unclear: while glosses have been shown to enhance the learning of individual words (e.g., Abraham, 2008; Hulstiijn et al., 1996), few studies have examined the effect of glossing on the learning of L2 MWUs.

Explicit and implicit knowledge

Many studies which have measured vocabulary knowledge gained from reading, including knowledge of MWUs, have assessed gains in explicit knowledge only, with only a few studies investigating both explicit and implicit knowledge (e.g., Elgort & Warren, 2014; Sonbul & Schmitt, 2013). The investigation of the development of implicit knowledge, however, is needed for predicting whether learners can access formulaic language during meaning-focused language use in real time.

In memory research, the two type of knowledge (memory), explicit and implicit, are considered to be distinct and dissociable. Evidence from behavioural and neurological studies with patients with various lesions and deficits shows that these patients display "implicit knowledge of stimuli that they cannot explicitly perceive, identify, or process semantically" (Schacter, 1987, p. 513). Research with normal participants also suggesting that the explicit and implicit memory systems may be cognitively and neurologically distinct (e.g., Rugg et al., 1998; Voss & Paller, 2007; for reviews see Reber, 2013; Roediger, 1990). *Explicit* knowledge is described as conscious and analysable—the type of knowledge that can be retrieved with the help of meta-cognitive and task-related strategies (Ellis, 1993; Paradis, 1994). It is commonly argued that explicit knowledge is more likely to be acquired as a result of deliberate study (Ellis, 2004; Reber, 2013). *Implicit* knowledge is often described as intuitive, ballistic, and unanalysed (Anderson, 1983; Ellis, 1993, 1999)—the type of knowledge acquired implicitly through experience (Reber, 2008, 2013) and exposure to input (Paradis, 1994). However, Elgort (2011) showed that both explicit and implicit knowledge of novel words can develop over time in the course of certain types of deliberate study (see also Suzuki & DeKeyser, 2017).

Vocabulary knowledge is both explicit and implicit (Dóczi & Kormos, 2016; Hulstijn, 2002). In particular, the knowledge of meaning is thought to draw on explicit memory structures, while the ability to fluently process a word form in written or spoken input is thought to be mostly implicit (Ellis, 1994; Ellis, 2004; Paradis, 1994; Ullman, 2001). According to Ellis (1994, p. 226), memory studies show that "word identification operates according to implicit memory principles—it is affected by mere exposure and the frequency thereof. But explicit memory for words is clearly affected by the depth of processing and the degree to which subjects analyse their meaning..."

Knowledge of MWUs, especially collocations, may be mostly implicit (Durrant & Schmitt, 2010; Ellis, 2004). It has been suggested that collocations are learned inductively through repeated exposure (Hoey, 2005; Siyanova-Chanturia & Martinez, 2015), resulting in their faster processing, compared to free word combinations. Although L1 and L2 collocation priming studies exist (e.g., Durrant & Doherty, 2010; McKoon & Ratcliff, 1992; Wolter & Gyllstad, 2011), to our knowledge, Sonbul and Schmitt (2013) is the first study that investigated how implicit knowledge develops from exposure to novel collocations in written input (see below for more details).

Measures of implicit MWU knowledge

While measures of explicit memory (knowledge), such as free/cued recall and recognition tests, require conscious recollection of a specific learning episode, measures of implicit memory do not require explicit recollection. Instead, participants are instructed to perform a task in which implicit memory is "revealed by a facilitation or change in task performance that is attributable to information acquired during a previous study episode" (Schacter, 1987, p. 501).

Implicit knowledge of MWUs (gained through language experiences) has been measured using eye-tracking (Siyanova-Chanturia, Conklin, & van Heuven, 2011; Underwood, Schmitt & Galpin, 2004), self-paced reading (SPR) (Conklin & Schmitt, 2008; Millar, 2011; Schmitt & Underwood, 2004; Tremblay, Derwing, Libben, & Westbury, 2011) and primed lexical decisions (Sonbul & Schmitt, 2013). In one of the first eye-tracking studies on MWUs, Underwood, Schmitt and Galpin (2004) found that both native speakers (NSs) and non-native speakers (NNSs) of English fixated on fewer words when the words were in MWUs, the implication being that the participants more quickly predicted terminal words based on the first words in the MWUs. As noted above, Siyanova-Chanturia et al. (2011) reported in their eye-tracking study that NSs and NNSs of different English proficiencies read high-frequency phrases more quickly than low-frequency phrases.

A number of studies have used SPR to measure processing of MWUs. Schmitt and Underwood (2004) found no difference in the reading times of NSs and NNSs between the terminal words of the target MWUs and the control words. However, Tremblay et al. (2011) found, in three SPR experiments, that the sentences containing lexical bundles (e.g., *in the middle of the*) were processed more quickly than the control sentences (e.g., *in the front of the*). Millar (2011) reports that NS participants' reading times were significantly longer for non-native-like collocations (learner errors) (e.g., *best partner*) than for correct NS collocations (e.g., *ideal partner*). Millar (2011, p. 144) hypothesised that the incorrect learner collocations placed "an increased and sustained cognitive burden" on the participants.

Another technique commonly deployed in studies of cognitive processing of known lexical units is priming; priming can tap into language processing and use without participants' awareness of the priming procedure, that is, with little involvement of strategic, resource-heavy, cognitive processes (McNamara, 2005; Neely, 1991; Plaut & Booth, 2000; Posner & Snyder, 1975; Roediger, 1990). In its basic form, priming involves presentation of two successive stimuli: a prime and a target. A prime that has some relation to the target may alter the speed and accuracy of its processing. In priming studies, researchers observe how responses to the same targets vary as a function of their relationships with different types of prime. Priming methodology has been used extensively to study the organisation of lexical and semantic representations of monolinguals and bilinguals (Shelton & Martin, 1992). Collocation priming, as a form of associative priming, is assumed to be driven by the frequency of co-occurrence of the two stimuli (Plaut, 1995; Ratcliff & McKoon, 1988). In the next section, we provide a detailed account of what we believe is the first collocation learning study with L2 participants that uses the priming paradigm to measure implicit knowledge.

Sonbul and Schmitt (2013)

Sonbul and Schmitt (2013), henceforth S&S, used a primed lexical decision task (LDT) to evaluate the development of the implicit knowledge of collocations in the initial learning stages. The researchers conducted two experiments, one with adult NSs and the other with advanced adult NNSs from a variety of L1 backgrounds. In the latter experiment, 43 advanced-level students not majoring in medicine were exposed to 15 two-word, low-frequency medical collocations, in three learning conditions, in one learning session. In two incidental learning conditions, participants read a passage containing 10 collocations repeated three times; in one condition (enhanced), five collocations were bolded, while in the other condition (enriched) five collocations were typographically unenhanced. In the third, deliberate decontextualised learning condition, participants were instructed to remember five collocations that were presented onscreen for 10 seconds each.

A series of immediate and delayed (two weeks later) post-tests was administered. The first test was a perceptual collocation-priming lexical decision task (LDT) intended to measure implicit knowledge. In the task, the first word of each word pair was the prime and the second word the target, and participants were instructed to make lexical decisions (word or non-word) on the target. The word pairs in this task were the medical collocations (e.g., *cloud baby*) and their control pairs (e.g., *steam baby*)—two-word combinations containing the same second words as the collocations. For each target noun, each participant saw either the medical collocation or the control pair. The prime was displayed for 150ms and was immediately replaced by the target, which remained on-screen until response. The two remaining post-tests measured explicit collocational knowledge. The second post-test was a form-recall task: a summary sentence including a gap-fill and a definition in the margin; the first word of each collocation was missing but no answer options were provided. The third post-test was a form-recognition task: the summary sentence from the form-recall test was used again, but this time with four answer options and an "*I don't know*" answer option provided for each answer. No pre-test was given, but a control group was used as a baseline.

S&S found that both L1 and L2 participants gained explicit collocational knowledge from all three learning conditions. For native participants, highlighting collocations in context was not more effective at promoting explicit knowledge than no highlighting, but for L2 participants this difference was significant. No collocation priming was observed in S&S's study for either group; therefore, the authors concluded that no evidence for the development of implicit collocational knowledge was detected after one learning session containing three repetitions (regardless of the learning condition). Based on these findings, the authors conjectured that implicit knowledge is more difficult to facilitate than explicit knowledge (1) within a short one-off learning session, and (2) with three contextual encounters. We addressed these two points in our study by lengthening the treatment period to three sessions over two days and by allowing for more repetitions of the target items (nine instead of three contextual occurrences).

Aim of study and research questions

The present study is a conceptual replication and extension of Sonbul and Schmitt (2013). Cumming (2014) argues that "[a] study that keeps some features of the original and varies others can give a converging perspective, ideally both increasing confidence in the original finding and starting to explore variables that influence it" (Cumming, 2014, p.10; see also Lindstromberg & Eyckmans, 2017, and Marsden, Morgan-Short, Thompson, & Abugaber, 2018, on the rationale for replication studies in L2 research).

As a conceptual replication, our study aimed to verify the generalisability of S&S's finding that repeated encounters with low-frequency medical collocations produced explicit knowledge of these collocations and that typographic enhancement produced more explicit knowledge than no typographic enhancement (Polio & Gass, 1997). The present study also aimed to extend the findings of S&S by increasing the number of encounters with the collocations in context and extending the learning treatment from one to two days. In particular, we investigated whether a larger number of contextual exposures would facilitate the development of implicit knowledge of the collocations. Because the focus of our study was on learning MWUs from context, we did not include a deliberate decontextualized learning condition, replacing it with another type of input enhancement: glossing. As a conceptual replication and extension study, a number of the features of the original study were changed (for details, see Appendix A). The study addressed the following research questions:

1. *Does repeated exposure to lexical collocations in supportive contexts*[i] *lead to the development of explicit and/or implicit knowledge?*
2. *Does typographic enhancement affect the development of explicit and/or implicit knowledge of lexical collocations (compared to reading only)?*
3. *Does typographic enhancement combined with glossing affect the development of explicit and/or implicit knowledge of lexical collocations (compared to reading only)?*

Finally, we are extending the findings of S&S's study by providing estimated effect sizes for the comparisons of the learning observed in the typographically-enhanced and unenhanced exposure to unfamiliar collocations in context.

Method

Participants

Sixty-two L2 participants took part in the learning procedure and immediate post-tests (cf. 43 participants in S&S's Experiment 2); the delayed post-tests were taken by 46 participants. Volunteers were initially screened for English language proficiency based on their overall International English Language Testing System (IELTS) test scores (6.5 or higher), i.e., upper-intermediate to advanced proficiency. A further measure of lexical proficiency (for the accepted participants) was the Vocabulary Size Test (VST) (Nation & Beglar, 2007). The participants' mean VST score was 92.27 out of 140 (SD=23.44), i.e., around 9000 word-families.

The mean age of the participants was 30.13 (SD=7.04)[ii]; 60 out of 62 were university students. The participants were from different L1 backgrounds (see Appendix B). They reported being first exposed to English at the mean age of 9.20 and spending, on average, 2.42 years in English-speaking countries;. Because medical collocations were the learning, only those who had received no medical training were accepted into the study. Study participants received a $20 supermarket voucher and were offered up to 1½ hours of proof-reading by the first author. The participants who returned for the delayed post-test went into a draw for a $200 cash prize.

Materials

The learning targets were 15 lexical collocations—technical, medical phrases (e.g., *split hand*, *cloud baby)*, 13 of which had been used in the S&S study[iii]. Two collocations, *gene therapy* and *partial response,* replaced *specific diseases* and *principal cells* from the original study (see Appendix C)*.* The main problem with using *specific diseases* was the potential confusion between its technical meaning and its general meaning. *Principal cells* was rejected because its very specific meaning (one type of cells in the thyroid gland) made it very difficult to place in three texts on different topics. The replacement collocations were selected according to the criteria used by S&S: they were fairly transparent; they occurred with low frequency in the British National Corpus (BNC); and their first words had several synonyms which could be used as distractors in a multiple-choice test and used in control pairs.

The medical collocations were embedded in nine reading texts, each 500 words in length, with three occurrences of five collocations in each text (see Appendix D). Three of the texts were adaptations of 1000-word texts used by S&S; the remaining six texts were created for the present study. The first three texts contained an embedded explanation of each collocation on its first occurrence (e.g., *Stone heart occurs when heart muscles become stiff, which usually leads to the death of the patient*.). The texts were constructed to facilitate the learning of new items from context; 97% of the lexical coverage was provided by the most frequent 3000 words of the BNC, as calculated using the *Vocabprofile* tool (Cobb, n.d.).

Across all texts, each of the 15 collocation occurred nine times; the number of repetitions was based on the finding that at least eight repetitions are needed when learning new vocabulary from reading (Cobb, 2016; Elgort, Brysbaert, Stevens, & Van Assche, 2017; Pellicer-Sánchez, 2015; Pellicer-Sánchez & Schmitt, 2010). While there is little evidence regarding the number of encounters needed for the learning of collocations, Durrant and

Schmitt (2010) suggest that gaining initial receptive knowledge of collocations would require at least 8-10 exposures.

Marginal glosses (Jacobs, Dufon & Fong, 1994; Ko, 2005) used in the bolding-plus-glossing (bolding+glossing) learning condition contained definitions that were different from the contextual explanations embedded with the first encounter of the collocations in the texts. Each gloss was linked via a number to the corresponding collocation in the text (the collocation was not repeated in the gloss).

Treatment conditions

In the texts, the medical collocations appeared in one of the following three treatments: reading-only (no enhancement, also referred to as an *enriched condition* because of the boost in the frequency of occurrence across the three texts compared to the frequency of occurrence in the language); bolding (typographically-enhanced condition); and bolding+glossing (with typographic enhancement and glossed definitions). In the treatment group, each participant experienced five medical collocations in each treatment condition (i.e., reading-only, bolding and bolding+glossing), following a counterbalanced design. For each participant, the same five collocations were consistently presented in the same condition throughout the experiment. Thus, across all texts, participants experienced an equal number of items in all three treatments (i.e., a within-participant experimental design). The participants were assigned to one of the three presentation lists that determined which five collocations they saw in which treatment condition. For instance, *regional control* was presented in the reading-only condition in list A, in the bolding condition in list B and in the bolding+glossing condition in collocation-list C; *smooth diet* was presented in the reading-only condition in list B, in the bolding condition in list C and in the bolding+glossing condition in list A, and so on.

Measures

Collocational knowledge was assessed using the three tests from S&S. The same tests were administered as immediate and delayed post-tests, but the items were reordered in the delayed tests[iv]. Explicit knowledge of the collocations was measured in a pen-and-paper cued-recall test and a pen-and-paper form-recognition test. For both post-tests of explicit knowledge, a text of just over 500 words containing the 15 collocations was created from the first three treatment texts. Following S&S, the participants were given a shortened version of a text they had already read in a treatment session, in order to facilitate understanding. Only the second word of each collocation was provided in the text. In the explicit cued-recall test, participants were instructed to supply the first word of the collocation by writing in a space. Two aids were given: firstly, the context of the text, which indicated the collocation's meaning (e.g., "A baby who quickly spreads infectious diseases into its immediate environment is known as a ________ baby"); secondly, a gloss of the collocation, a short definition prompt which was different from the gloss for the reading texts was also given in the margin (e.g., for *cloud baby*: "baby carrying a disease"). The explicit form-recognition test comprised the same text but, instead of glossed definitions, participants were presented with multiple-choice options in the margin. The four choices for each missing collocate comprised the correct collocate, the control word (also used as a prime in the control pairs of the lexical decision task) and two further distractors (see Sonbul & Schmitt, 2013, p. 132, for a visual illustration of the two tests). An *I do not know option* was also provided.

Implicit knowledge was tested in a primed lexical decision task (LDT). The second word of the collocations was used as the target, to which lexical decisions were made. The target (e.g., *baby*) was preceded by the prime, which was either the first word of the collocation (e.g., *cloud*) or another word that had not been encountered by the participants as a collocate of the target (e.g., *steam*), making a *control pair*. The development of implicit knowledge of the collocations was operationalised as collocation priming, i.e., faster responses to the targets presented as part of the collocations than in the control pairs.

In addition to the medical collocations and control pairs, the stimuli in the LDT included pseudowords (e.g., *woup*, *flyst*). A list of pseudowords was created using Wuggy (Keuleers & Brysbaert, 2010). Seventy-six pseudowords were generated, two for each of the two real words in the 15 medical collocations (matched for the number of letters and syllables), plus 16 fillers for the word—nonword (e.g., *broad—iruax*) or nonword—word (e.g., *crorm—size*) filler trials.

Procedure

Each participant was allocated to either the treatment group (and, within that, to one of the three presentation lists as detailed in the *Treatment conditions* section) or to the control group, in the order of their volunteering for the study. There were 15 participants in the control group (n=14 on the delayed post-test as one participant withdrew from the study for personal reasons); these participants did not complete the reading procedure and only participated in the immediate and delayed tests. The treatment group initially comprised 47 participants (16 or 15 participants per presentation list), of whom 32 returned for the delayed post-test. In the treatment group, participants first completed the reading task, in which they were exposed to the medical collocations. They were instructed to focus on the meaning of the reading texts (incidental learning); in order to facilitate this approach to reading, they were asked to complete two reading-comprehension questions after each text. Participants were also instructed to read the marginal glosses at the points where references to glosses were included in the text. They completed three reading sessions over two days: two sessions on day one and one more session on the following day. A spaced repetition approach that incorporated an opportunity for an overnight memory consolidation (Lindsay & Gaskell, 2010; Wang et al., 2017) was used because it has been shown to facilitate vocabulary learning (Karpicke & Roediger, 2007).

Participants' knowledge of the collocations was measured after the third reading session on day two (immediate post-tests) and again after two weeks (delayed post-tests). In both immediate and delayed post-tests, the primed LDT was administered first, followed by the cued-recall task (gap-fill), then the form-recognition (multiple-choice) task. This test order was chosen (following S&S) to reduce any potential learning effect from one test to the next. The reading sessions were conducted with small groups of participants. All test sessions were conducted in a laboratory with no more than two participants at a time; a divider was placed between the two computers in the lab.

The primed LDT procedure was implemented in E-Prime 2.0 (Psychology Software Tools, Inc.). The participants were first presented with a fixation (+) displayed in the centre of the computer screen for two seconds. That was followed by the prime presented for 150 milliseconds, which was immediately replaced by the target, which remained on the screen until a response was registered. Participants were instructed to respond as quickly and accurately as possible to the target to indicate whether or not it was a word. To register their responses, participants pressed buttons labelled “Y” (for “yes”) or “N” (for “no”) on a response box connected to the computer.

After their the post-tests, each participant was asked about their prior knowledge of the medical collocations (addressing a limitation of S&S). Critical collocations self-reported as previously known were assigned a score of 1; otherwise it was 0. Overall, the participants' mean prior knowledge was estimated as 11% (SD=12%). Based on the prior knowledge self-report, the collocation, *gene therapy*, was excluded from the data analysis because it was familiar to more than a third of the participants (21 out of 62). Thus, 14 collocations were included in the data analysis.

Data analysis

The data analysis for all tests was conducted using mixed-effects modelling in R (version 3.4.4) (R Core Team, 2018). Participants and items (collocations) were entered in the models as crossed random effects. A minimally adequate statistical model was fitted to the data, using a stepwise variable selection and the likelihood ratio test for model comparisons (Baayen, Davidson, & Bates, 2008). Participants' lexical proficiency (VST scores) and self-reported prior knowledge of the collocations were treated as potential covariates. The immediate and delayed results were considered within the same statistical model for each of the three knowledge measures; a two-level predictor, session (immediate/delayed), was included in the data modelling procedure. The resulting statistical model contained only variables for which regression weights were different from zero ($p<0.1$), or they improved the model fit, or were involved in significant interactions. The models contained random slopes supported by the data (i.e., parsimonious mixed models) because this approach improves the balance between Type I error and power (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

For each test, two analyses were conducted. The first (preliminary) analysis was conducted to verify that the control group did not have the knowledge of the target collocations. The main analysis aimed to establish which treatments contributed to the development of the explicit and implicit knowledge of the collocations[v]. The treatment type predictor had three levels: reading-only (enriched), bolding, and bolding+glossing. The initial alpha level was set to .05. To control for the Type I error rate, the function *glht* in the R package *multcomp* (Hothorn, Bretz, & Westfall, 2008) was used to obtain multiplicity-adjusted *p*-values for the primary interest predictor (i.e., treatment) and the corresponding confidence intervals (CI).

In the two explicit knowledge tests (gap-fill and form-recognition), the accuracy of responses to the collocations encountered in the bolding and bolding+glossing treatments was compared with that to the collocations encountered in the reading-only condition. Responses were scored as either correct (1) or incorrect (0)[vi] and analysed using mixed logit models. For each explicit test, the odds ratios (OR) and standardised effect sizes[vii] (Chinn, 2000) are reported. The odds are defined as the probability of an event occurring divided by the probability of it not occurring (Field, 2013, p. 767). When the odds are more than 1.0, a "success" (in this case, a correct answer) is more likely than a "failure" (an incorrect answer) (Agresti, 2007, p. 28).

In the implicit knowledge analysis, inverse-transformed response times (-1000/RT) were used because the non-transformed RTs' distribution is positively skewed and does not fit the assumption of normal distribution. Implicit knowledge was operationalised as collocation priming; therefore, we compared RTs for each of the three treatments with those for the control pairs. Effect sizes were calculated following Brysbaert and Stevens (2018) (based on Westfall, Judd, & Kenny, 2014). Minimum a priori outlier removal was performed (i.e., only extreme outliers were removed) but the final regression models were subjected to model criticism; potentially harmful outliers (i.e., data points with standardized residuals exceeding 2.5 standard deviations) were removed and the model was refitted (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Brysbaert & Stevens, 2018). For the final models, see Appendix E.

Results

Post-tests of explicit knowledge: Preliminary analyses

In the preliminary analysis we compared the response accuracy of the treatment and control groups. The control group's mean accuracy rate was 1.62% on the cued-recall (gap-fill) test and 17.54% on the form-recognition (multiple-choice) test; the treatment group's mean accuracy was 66.86% on the cued-recall test and 95.37% on the form-recognition test. (*Note*: here and elsewhere in the body of the article, estimated means are based on the model predictions unless otherwise indicated; descriptive statistics calculated prior to the data analyses are reported in Appendix F).

In the explicit cued-recall test (control vs. treatment group), there was an interaction between Group and Session: the accuracy of the control group in cued-recall improved slightly (by 1.53%) on the delayed compared to the immediate post-test whereas, for the treatment group, it reduced slightly (by 6.44%). There was also a main effect of Group: the odds of obtaining a correct answer for the participants in the treatment group were 122 times the odds of obtaining a correct answer in the control group (see Table 1), with a large standardised mean difference (SMD) effect size of d=2.65. This means there is a 97% chance that a person picked at random from the treatment group will have a higher score than a person picked at random from the control group. The low gapfill accuracy recorded for the control group and a large effect size of Group gave us a clear mandate to proceed to the main analysis, which compared the three treatment conditions.

In the form-recognition test (control vs. treatment group), there was no statistically significant difference between the immediate and delayed post-test results; therefore, the variable Session was not included in the final model. For participants in the treatment group, the odds of obtaining a correct answer were 97 times the odds for participants in the control group (see Table 2), i.e., a large SMD effect size (*d*=2.65). Based on the control group's response accuracy of 17.54% in a four-choice test and a large effect size of Group, we proceeded with the main analysis.

Table 1 Response accuracy on the cued-recall test (control vs. treatment group): fixed effects

Parameter	Estimate	*SE*	*z*	*p*	Odds ratios	95% CI
(Intercept)	-4.07	0.79	-5.16	2.53E-07		
Group=Treatment	4.81	0.84	5.75	8.84E-09	122.14	23.75, 628.03
Session=Immediate	-2.84	1.09	-2.61	.009	0.06	0.01, 0.49
Prior.Knowledge=1	1.01	0.35	2.89	.004	2.73	1.38, 5.41
VSTc	0.02	0.01	2.23	.026	1.02	1.00, 1.04
Grp=Trt:Session=Imm	3.15	1.10	2.85	.004	23.33[viii]	2.68, 203.41

Notes. Intercept levels: Group=Control; Session=Delayed; Prior.Knowledge=0.

Table 2 Response accuracy on the form-recognition test (control vs. treatment group): fixed effects

Parameter	Estimate	*SE*	*z*	*p*	Odds ratios	95% CI
(Intercept)	-1.55	0.33	-4.75	2.06E-06		
Group=Treatment	4.57	0.47	9.74	< 2e-16	96.83	46.99, 295.85
Prior.Knowledge=1	0.87	0.48	1.81	.070	2.39	0.84, 5.54

Notes. Intercept levels: Group=Control; Prior.Knowledge=0.

Post-tests of explicit knowledge: Main analyses

In the treatment group, the analysis of responses on the explicit knowledge tests showed that encountering a collocation in the bolding condition was significantly more likely to result in correct responses on the cued-recall and form-recognition tests compared to the reading-only condition (Table 3 & 4). In the cued-recall task, the bolding treatment resulted in the largest percentage of correct answers (75%), followed by the bolding+glossing treatment (68%) and the reading-only treatment (64%). In the form-recognition test, the bolding treatment was again the most effective, with a mean score of 98%, while the bolding+glossing and the reading-only treatments had scores of 96% and 94%, respectively.

In the cued-recall task, we found a significant difference in response accuracy between the bolding and the reading-only conditions (z=2.59, multiplicity-adjusted p=.026), but not between bolding+glossing and reading-only. The odds of obtaining a correct answer for the collocations learned in the bolding condition were 1.69 times the odds of those in the reading-only condition (a small effect size of d=0.29) (Table 3). Ellis and Sagarra (2011, p. 617) suggest that d≈0.25 might reflect educational significance (i.e., something was learned).

Table 3 Response accuracy on the cued-recall test (treatment group): fixed effects

Parameter	Estimate	*SE*	*z*	*p*	Odds ratios	95% CI
(Intercept)	0.56	0.41	1.35	.178		
Treatment=BO+GL	0.18	0.20	0.88	.379	1.19	0.81, 1.77
Treatment=BO	0.52	0.20	2.59	.010	1.69	1.14, 2.50
Session=Immediate	0.31	0.18	1.76	.079	1.37	0.96, 1.94
Prior.Know=1	1.05	0.35	2.99	.003	2.85	1.43, 5.65

Notes. Intercept levels: Treatment=Reading only; Session=Delayed; Prior.Knowledge=0. *BO*=bolding; *BO+GL*=bolding+glossing.

In the form-recognition task, there was a significant difference between the bolding and the reading-only conditions (z=3.08, adjusted p=.006) but not between bolding+glossing and reading-only (Table 4). The odds of obtaining a correct answer for the items learned in the bolding condition were 2.73 times the odds of those in the reading-only condition (a medium effect size, *d*=0.55).

Table 4 Response accuracy on the form-recognition test (treatment group): fixed effects

Parameter	Estimate	*SE*	*z*	*p*	Odds ratios	95% CI
(Intercept)	2.77	0.38	7.33	<.001		
Treatment=BO+GL	0.37	0.30	1.24	.214	1.44	0.81, 2.58
Treatment=BO	1.00	0.33	3.08	.002	2.73	1.44, 5.17

Notes. Intercept level: Treatment=Reading-only**.**

Reviewing results for the secondary-interest predictors, it is worth noting that, in the cued-recall test (Table 3), the odds of participants responding correctly to a collocation reported as known prior to the treatment was 2.84 times the odds for the unknown items (*d*=0.58). The participants were also less accurate on the delayed than on the immediate cued-recall test, but this difference did not reach statistical significance and the effect size was small (*d*=0.17). Neither Session nor the predictors reflecting individual differences (i.e., Prior knowledge and VST) affected form-recognition accuracy (Table 4).

Overall, the results of the explicit knowledge post-tests can be summarised as follows: Bolding > Reading-only; Bolding+Glossing ≈ Reading-only.

Implicit knowledge post-test: Preliminary analysis

Incorrect responses in the LDT (n=17, the accuracy rates of 99%) were removed prior to the data analysis. A preliminary analysis was conducted on the control group data separately. Three extreme outliers were removed after the data had been inverse-transformed. Experimental condition (intact collocations vs. control pairs) was the primary-interest predictor in this analysis. The results showed no priming, with estimated mean RTs to the control pairs 7 ms faster than RTs to the intact collocations (552 ms vs. 559 ms) (Table 5). The analysis also showed a statistically significant contribution of Session to the model, with estimated mean RTs on the delayed test about 66 ms faster than on the immediate test. There was a 125-ms difference between the RTs of the participants with the lowest and highest VST scores but the effect of VST did not reach statistical significance. Importantly, the absence of priming in this analysis suggests that participants in the control group did not have measurable implicit knowledge of the medical collocations.

Table 5 RTs of lexical decisions (control group only): fixed effects

Parameter	Estimate	*SE*	*df*	*t*	*p*	*d*
(Intercept)	-1.78	0.08	21.63	-22.82	< .001	
Exp.Cond=Intact.Collocation	0.02	0.03	243.29	0.78	.437	0.03
Session=Immediate	0.19	0.06	13.36	3.31	.006	0.30
VST(centered)	-0.01	0.00	12.84	-1.94	.074	0.01

Notes. Intercept level: Exp.Cond=Control.Pairs, Session=Delayed.

Implicit knowledge post-test: Main analysis

Two extreme outliers were removed from the analysis, leaving 1093 observations. In the main analysis of implicit knowledge, we looked for a priming effect for each of the treatments; therefore, RTs to the intact collocations encountered in each of the treatments were compared with the RTs to the control pairs.

Priming was present for items learned in the reading-only condition (t=-2.62, p<.01, multiplicity-adjusted p=.026), but not in the bolding+glossing or bolding condition (Table 6). For the collocations learned in the reading-only condition, RTs to the terminal word were significantly faster (by 18 ms) in the intact collocations than in the control pairs (538 ms vs 556 ms, respectively); however, the effect size was small (d=0.15). Across all treatments, RTs were 32 ms faster in the delayed than in the immediate test, and about 110 ms faster for participants who reported knowing more collocations. Descriptive statistics (Appendix F Table F2) show that collocation priming for the reading-only treatment was numerically greater in the immediate than in the delayed post-test, but there was no statistically significant interaction between Priming and Session in the data analysis.

Table 6 RTs of lexical decisions (experimental group only): fixed effects

	Estimate	*SE*	*df*	*t*	*p*	*d*
(Intercept)	-1.79	0.05	47.68	-34.39	<.001	
Treatment=Reading-only	-0.06	0.02	980.13	-2.62	.009	0.15
Treatment=BO	0.00	0.02	978.03	-0.18	.857	0.01
Treatment=BO+GL	-0.01	0.02	978.25	-0.53	.596	0.03
Session=Immediate	0.10	0.03	30.20	3.79	.001	0.25
Prior.Knowledge(centered)	-0.69	0.28	44.60	-2.44	.019	1.74

Notes. Intercept level: Treatment=Reading-only, Session=Delayed.

Discussion

Replication of S&S's findings for explicit knowledge

The results of our study replicate the finding of S&S's Experiment 2 (pp. 144-145) for the development of explicit knowledge of technical collocations. Both studies found that the treatment group was significantly more accurate on the cued-recall and form-recognition tests than the control group. This suggests that repeated exposure to technical lexical collocations in high-constraining written contexts can trigger their learning. Secondly, typographic enhancement (our bolding condition) resulted in more accurate responses compared to the enriched condition (our reading-only condition). Thirdly, the effect of Session on explicit collocational knowledge was not statistically significant in either post-test, for the treatment group. This suggests that participants mostly retained the explicit knowledge of the collocations gained from reading, regardless of the treatment, after two weeks. Fourthly, similar to S&S and learning studies in general, the participants' ability to recognise the collocations was superior to their ability to recall them (in our study, 95% vs. 67% respectively).

In the present study, the participants' L2 vocabulary knowledge (measured by the VST) was not found to be a significant covariate in the analyses of explicit knowledge while, in S&S's explicit knowledge analysis, the participants' self-reported L2 proficiency was found to be a significant covariate that modulated the effect of test and session. A possible reason for this difference is that, in our cued-recall analysis, the self-reported knowledge of the collocations was a significant covariate (d=0.58); so, it may have explained most of the variance that had been explained by L2 proficiency in S&S's analysis. Notwithstanding this, it is noteworthy that the accuracy of form-recognition (measured by the multiple-choice test) seemed to be unaffected by the individual variables (vocabulary size or prior knowledge), and no attrition was found on the delayed test. Together, these results suggest that the participants in our study gained explicit receptive knowledge of the technical collocations and that typographic enhancement promoted the development of this knowledge.

Extension of S&S's study for implicit knowledge

A key contribution of the present research is that it provides initial evidence for the development of implicit knowledge of technical collocations, thus extending the findings of S&S's Experiment 2 (pp. 148-149). S&S point out that "the short, massed treatment … might have put the … enriched condition at a disadvantage because unenhanced exposure is said to aid development in knowledge gradually over time" (Sonbul & Schmitt, 2013, p. 153). Indeed, collocation priming was observed for the enriched, reading-only treatment in our study, suggesting that implicit knowledge of collocations can develop over time. The priming effect observed in the present experiment, where S&S found none, is likely due to a larger number of exposures to the collocations (nine vs. three), more learning sessions/texts (three vs. one), and the distribution of the learning sessions over two days (two reading sessions on day one and the third session on day two). The size of the recorded priming effect was small (d=0.15); this may reflect the nature of priming in a speeded lexical decision task, i.e., a difference of 18 ms is very small when compared to RTs that vary from 300 ms to 1500 ms. This difference is, however, similar to priming due to word frequency (see Keuleers, Lacey, Rastle, & Brysbaert, 2012), the effect of which is also small but consistent.

The study's research questions are now evaluated in light of the post-test results.

Research question 1: Does repeated exposure to lexical collocations in supportive contexts lead to their learning? The answer to the first research question is a cautious "yes". Participants in the present study gained explicit knowledge of the lexical collocations after nine contextual exposures across three sessions over two days; there was also some evidence of implicit knowledge in one (but not all) of the treatments (Table 6). Together with the relatively large 95% CIs (with the estimated mean inverse-transformed RT difference of -0.06; Lower=-0.11; Upper=-0.01) this small effect could indicate that the implicit collocational knowledge gained through nine contextual exposures over two days may still be fragile and that more encounters over a longer period of time may be needed for robust implicit collocational knowledge to develop. Alternatively, as elaborated in the *Limitations* section below, further replication studies could establish the stability of this small collocation priming effect using a larger sample size.

Research question 2: How does typographic enhancement affect the contextual learning of lexical collocations? The bolding of the collocations in the reading texts led to greater explicit knowledge in comparison with the reading-only condition, in both the cued-recall and form-recognition test. In terms of implicit knowledge, collocation priming was not observed in the bolding treatment but it was detected in the reading-only treatment. In summary, typographic enhancement may have promoted noticing of the medical collocations in context facilitating the acquisition of explicit knowledge, while the reading-only treatment that repeatedly exposed participants to the collocations in meaningful supportive contexts, with no typographic enhancement, was more effective in promoting the development of their implicit knowledge.

Research question 3: How does typographic enhancement combined with glossing affect the contextual learning of lexical collocations? Contrary to the conjecture that adding glossing to typographic enhancement might result in better learning (Hulstijn, et al., 1996), the bolding+glossing condition did not produce a learning advantage over the reading-only condition in the present study. It is possible that when the participants took their eyes away from the text to look at the marginal glosses their online contextual processing of the collocations was interrupted and the co-occurrence of the collocational components was not encoded.

Looking at the overall pattern of our results, we conclude that the different learning conditions appear to have promoted different types of processing of the unfamiliar collocations during reading and resulted in different types of knowledge. The enriched, reading-only condition repeatedly exposed the participants to the collocations in diverse supportive written contexts without explicitly drawing their attention to the constituent words, which seems to have triggered to the development of implicit knowledge. The bolding condition, presumably, made the collocations more salient and attracted readers' attention to them, thus improving accuracy of their recognition and recall. However, this treatment may have altered the process of word-to-text integration while reading (Perfetti, Yang, & Schmalhofer, 2008), thus interfering with the development of implicit knowledge.

Fluent reading comprehension involves rapid word identification, activation of word meanings and their integration into context (Perfetti & Stafura, 2014). The processing of a lexical collocation, such as *partial response* or *smooth diet*, in supportive contexts is unlikely to disrupt successful word-to-text integration, i.e., the integration of its component words with each other and the on-going context. If we accept that "a collocation is a sequence of words that co-occur more often than would be expected by chance" (Ellis & Ogden, 2017, p. 605) and that frequency of co-occurrence in processing is predictive of the strength of association in the memory (e.g., Ferrand & New, 2003; Günther, Dudschig, & Kaup, 2015; Kutas, 1993; see also Hutchison, 2003, for a review), increasing the frequency and distribution of contextual encounters in reading 'wires' the representations of the constituent words of the collocations together in the memory of the reader. This implicit learning leads to the development of implicit collocational knowledge.

When the words comprising a collocation are typographically-enhanced (bolded) in the reading text, their processing and integration into context changes as a result of additional attention allocated by L2 readers to the bolded items (see Choi, 2017, for an eye-tracking study of the effect of typographic enhancement on the processing of novel collocations during reading). Interestingly, Choi (2017) observed that, while increasing the L2 readers' recall of the target collocations, the use of typographic enhancement reduced their recall of non-enhanced lexical items in the same short passage. Choi concludes that typographic enhancement should be used with great care in L2 reading because "drawing learners' attention to target language items, such as collocations, is likely to undermine the creation of a coherent mental representation of the text" (Choi, 2017, p. 418). The reader's mental model of the text is critical in activating contextually relevant meaning representations of partially known collocations and, as a by-product, strengthening the mapping of their lexical and semantic representation. Perfetti et al. (2008, p. 314) explains that the lexico-semantic process of word-to-text integration "begins with the identification of a word from its orthographic input and, in overlapping phases, moves to the activation of associated meanings, both from the reader's lexicon and from the temporary memories that store words and referential situations from the text". Thus, implicit learning of L2 collocations during reading is supported by an accurate mental model of the text, the construction of which may be impeded when some words in the text are typographically-enhanced.

The bolding+glossing condition drew the participants' attention away from the text and may have disrupted the natural lexical processing of the collocations in reading, consequently, counteracting possible advantages of gloss-induced meaning elaboration on learning. In our study, the null hypothesis could not be rejected for the comparison of the bolding+glossing and reading-only treatments, in the explicit recognition and recall post-tests; there was also no evidence of implicit knowledge (operationalised as collocation priming) for the collocations encountered in the bolding+glossing condition. Using our findings we can now qualify S&S's provisional conclusion about a possible dissociation between explicit and implicit lexical knowledge in the learning of collocations: we submit that, although contextual learning of collocations can lead to the development of both explicit and implicit knowledge, the latter may take longer to develop and is less amenable to treatments promoting deliberate attention (noticing) through typographic enhancement.

Limitations, pedagogical implications and recommendations for future research

A possible limitation of this study was the limited number of items: 15 lexical collocations. For reasons outlined in the method section, we used the same number of collocations as S&S but increased the number of study participants. In future replications, it would be best to use power analysis (not currently common practice in SLA research) in order to estimate an optimal sample size (i.e., the number of participants and items) needed to investigate the research questions about the development of explicit and implicit collocational knowledge from reading, with and without typographic enhancements.

Following a recommendation made in the manuscript review process (while aware of the limitations of this approach outlined by Ellis, 2010, pp. 58-60), we conducted post-hoc power simulations for the main analyses, using the R package *SIMR* (Green & MacLeod, 2016). We found that the form-recognition test had sufficient power (power=80.80%, 95% CI=78.22, 83.20), while the cued-recall (power=68.2%; 95% CI=65.21, 71.08) and primed LDT (power=61.00%; 95% CI=57.90, 64.04) analyses may have been somewhat underpowered. We must, however, caution against overinterpreting these power estimations because of their retrospective nature. We think that the SMD effect sizes that have been observed in our study for the comparison between the typographically-enhanced versus unenhanced condition on the two explicit knowledge post-tests (d=0.29 and d=0.55) are sufficiently convincing for the respective post-tests. We also think that the finding of an 18-ms facilitation priming in the speeded lexical decision task with an effect size of d=0.15 is about right for this experimental task. On the other hand, the SMD effect size for the comparison the bolding+glossing and reading-only treatments in the cued-recall test was small (d=0.10), and the priming effect size in the bolding+glossing treatments was very small (*d*= 0.03). Looking at these effect sizes for the bolding+glossing treatment, we suspect that any explicit or implicit knowledge gains resulting from this treatment (even if determined to be statistically significant using a larger sample size) are unlikely to be meaningful enough to justify the use of this treatment in practice. We suggest, therefore, that future replications focus on the comparison of the enriched (reading-only) and typographically-enhanced treatments in contextual learning of L2 collocations.

Using SIMR simulations as an a priori power analysis (based on the data from the present study and our models fitted in lme4) we estimated that, in order to achieve the power of 80% in the cued-recall and primed LDT analyses, a replication of our study would need 70-75 participants in the treatment group, with the same number of items. Alternatively, if the number of participants in the treatment group is kept the same as in our study, future replications would need to use 20-22 collocations to achieve 80% power.

We also acknowledge that the experimental treatments in this study may have been somewhat artificial. First, authentic texts are unlikely to have the same high density of occurrence of the lexical collocations as the present study. Multiple repetitions of the collocations in short texts may have impeded their coherence and flow. Our recommendation is to infuse longer texts (such as graded readers) with repeated occurrences of lexical collocations. Second, it is unusual to see repeated typographic enhancement, such as bolding, of key vocabulary items in authentic texts. In a typical university textbook, typographic enhancement of a technical term will usually occur only once (e.g., Clark & Randall, 2004; Field, 2013). Third, glosses were used in the present study in a somewhat unusual way: instead of supporting initial comprehension of the lexical items, they were used as a form of repeated meaning elaboration. This may have contributed to the weaker-than-expected effect of the bolding+glossing treatment on learning. Because the first time each collocation was presented in the text the sentence also included its embedded explanation, it is possible that the glosses added little further useful information. Finally, even though the participants had been instructed to read all the glosses, some may have not followed the instructions.

Notwithstanding these limitations, our finding that readers can develop both explicit and implicit knowledge of L2 lexical collocations when they encounter them multiple times in supportive contexts has important implications for language learners, teacher and materials developers. In particular, we recommend manipulating longer level-appropriate texts, such as graded readers or textbooks, to include multiple instances of collocations. Repetitions of single words in graded readers or other simplified materials have been found to produce gains in explicit knowledge of the items (Waring & Takaki, 2003), increase their processing speeds (Pellicer-Sánchez, 2015), and enhance their retention (Webb, Newton, & Chang, 2013). Northbrook & Conklin (2018) found that beginner English learners were sensitive to the frequency of occurrence of lexical bundles in their textbooks. However, further research is needed into the effect of repeated encounters with L2 collocations in graded readers (with and without typographic enhancement) on the development of explicit and implicit knowledge.

Our study also suggests that bringing L2 readers' attention to unfamiliar collocations (through typographic enhancement) facilitates the development of explicit collocational knowledge, while implicit learning may be more conducive to the development of implicit knowledge. Because the development of explicit and implicit collocational knowledge may have different developmental trajectories (see Sonbul & Schmitt, 2013, p. 151), we recommend that both explicit and non-explicit measures be used, when possible, in collocation learning research.

---

[i] Supportive (informative) contexts constrain the meaning of lexical items, making it possible for the reader to infer meanings of unfamiliar words or expressions from the surrounding text.

[ii] All but one of the participants were in the age range of 18-43; one participant was 61.

[iii] The collocations, treatment passages and explicit-knowledge tests were received directly from the first author of Sonbul and Schmitt (2013) and then adapted for our study. We used the same number of collocations as S&S for the reasons they outlined in their article (p. 153). Because our study is a replication, we investigated whether increasing the number and distribution of contextual encounters (while keeping the number of items constant) would create conditions needed for implicit knowledge to emerge. Since S&S's study may have been underpowered

(with 15 items and 43 participants across three treatment conditions), we recruited more participants (N=62). We will return to the issue of statistical power in the limitations section of the article.

[iv] No pre-test was included in this experiment since even one exposure to the target collocations could have affected the results by creating an initial memory trace and leading to unintentional learning (Nation & Webb, 2011; Sonbul & Schmitt, 2013). Instead, a control group was used to provide a baseline for the knowledge of the collocations prior to the treatment. The absence of (or negligible) knowledge of the collocations by participants in the control group (recruited from the same participant population) would suggest that these collocations were unfamiliar to the members of this population. Additionally, at the end of the testing phase, participants were asked to indicate their prior knowledge of the collocations, and this information was included in the data analysis.

[v] We did not combine the data of the control and experimental groups in one analysis because participants in the control group did not participate in the learning treatments and, therefore, the primary-interest variable—treatment— could not be applied to the control group.

[vi] Marking of the gap-fill test was straightforward for the vast majority of answers. Answers which were synonyms of the correct answers (e.g., *rest periods* for *gap periods*) were marked as incorrect. Answers which were misspellings (e.g., *cheif complaint* for *chief complaint*) or incorrect word forms (*silence areas* for *silent areas*; *gold hour* for *golden hour*) were marked as correct. One misspelling which was also another word—*ion lung* for *iron lung*—was marked as correct. The final questionable area was the inclusion of a compound word instead of a single word—*split-hand hand* for *split hand*—which was marked as correct.

[vii] Effect sizes in logit models were calculated by dividing the ln(OR) by 1.81, because the standard logistic distribution has variance $\pi^2/3 = 1.81$ (Chinn, 2000; Grissom & Kim, 2012, p. 273).

[viii] Note that the coefficient of an interaction is a ratio of odds ratios (and not a logged odds ratio).

References

Abraham, L. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199-226. DOI: 10.1080/09588220802090246

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken: New Jersey: John Wiley & Sons, Inc.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multiword phrases. *Journal of Memory and Language, 62*, 67–82.

Anderson, J. (1983). *The architecture of cognition.* Cambridge: Massachusetts: Harvard University Press.

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*, Cambridge: Cambridge University Press.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. DOI: 10.1016/j.jml.2007.005

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science, 19*, 241–248.

Bishop, H. (2004). Noticing formulaic sequences—a problem of measuring the subjective. *LSO Working Papers in Linguistics*, *4*, 15-19. Madison, Wisconsin: University of Wisconsin-Madison Department of Linguistics.

Boers, F., Demecheleer, M., He, L., Deconinick, J., Stengers, H., & Eyckmans, J. (2016). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *26*(2), 448-469. DOI: 10.1111/ijal.1214

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), 1-17. DOI: 10.1191/1362168806lr195oa

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition.* New York: Palgrave Macmillan.

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition, 1*(1): 9, 1–20.

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*(3), 403-443. DOI: 10.1017/S0272263115000492

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19*, 3127—3131.

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403-426.

Cintrón-Valentín, M., & Ellis, N. (2015). Exploring the interface: Explicit Focus-on-Form Instruction and Learned Attentional Biases in L2 Latin*. Studies in Second Language Acquisition, 37(2)*, 197-235. doi:10.1017/S0272263115000029.

Clark, M., & Randal, J. (2004). *A first course in applied statistics—with applications in biology, business, and the social sciences.* Auckland: Pearson Education New Zealand.

Cobb, T. (2016). Numbers or numerology? A response to Nation (2014) and McQuillan (2016). *Reading in a Foreign Language*, *28*(2), 299-304.

Cobb, T. (n. d.). Compleat Lexical Tutor. Available at https://www.lextutor.ca/

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29(*1), 72-89.

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science, 25(1)*, 7–29.

Dóczi, B., & Kormos, J. (2016). Second Language Learning and Vocabulary Development. In B. Dóczi & J. Kormos, *Longitudinal Developments in Vocabulary Knowledge and Lexical Organization* (pp. 118-138). Oxford University Press.

Durrant, P. & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory, 6*(2), pp. 125-155. DOI:10.1515/cllt.2010.006

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

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367-413. DOI: 10.111/j.1467-9922.2010.00613.x

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2017). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*, 2017, 1-26. DOI: 10.1017/S0272263117000109

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: explicit and tacit knowledge and the role of learner and item variables. *Language Learning, 64*(2), 365-414. DOI: 10.111/lang.12052

Ellis, N. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 211-282). London: Academic Press.

Ellis, N. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *SSLA, 24*, 143-188.

Ellis, N., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, *33*(4), 589-624.

Ellis, P. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Ellis. R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly, 27*(1), 91-113.

Ellis, R. (1999). Factors in the incidental acquisition of second language vocabulary from oral input. In R. Ellis, *Learning a second language through interaction* (pp. 35-61). Amsterdam: John Benjamins.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227-275.

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

Green, P., & MacLeod, C. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493-498.

Grissom, R. & Kim, J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). Routledge.

Günther, F., Dudschig, C., & Kaup, B. (2015): Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology, 69(4), 626-653.*

Henriksen, B. (2012). Research on L2 learners' collocational competence and development – a progress report. In Bardel, C., Lindqvist, C., & Laufer, B. *EUROSLA Monographs Series 2, L2 vocabulary acquisition, knowledge and use* (pp. 29-56).

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

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal 50*(3), 346-363.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*, 193–223.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, *80*(3), 327-339.

Jacobs, G., Dufon, P., & Fong, C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*(1) (1994), pp. 19-28.

Karpicke, J., & Roediger, H. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *33*(4), 704-719. DOI: 10.1037/0278-7393.33.4.704

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627-633.

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287–304.

Ko, M. (2005). Glosses, comprehension, and strategy use. *Reading in a Foreign Language, 17*(2).

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: a corpus analysis of learners' English. *Language Learning*, *61*(2), 647-672. DOI: 10.111/j.1467-9922.2010.00621.x

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: a multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: acquisition and communication* (pp. 22-46). New York: Continuum.

Lindsay, S., & Gaskell, M. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, *60*(2), 45-63.

Lindstromberg, S., & Eyckmans, J. (2017). The particular need for replication in the quantitative study of SLA: A case study of the mnemonic effect of assonance in collocations. *Journal of the European Second Language Association*, *1*(1), 126-136.

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning, 00(0)*, 1–12. DOI: 10.1111/lang.12284.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94,* 305–315.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1155–1172.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Hove, UK: Psychology Press.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32*(2), 129-148. DOI: 10.1093/applin/amq035.

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

Nation, I., & Webb, S. (2011). *Researching and analysing vocabulary.* Boston: Heinle Cengage Learning.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 236-264). Hillsdale, NJ: Erlbaum.

Northbrook, J., & Conklin, K. (2018). Is what you put in what you get out? —Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics*, DOI: 10.1093/applin/amy027.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419)*.* London: Academic Press.

Pellicer-Sánchez, A. (2015). Incidental L2 vocabulary acquisition *from* and *while* reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*, 97-130. DOI: 10.1017/S027226315000224

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language*, *22*, 31-55.

Perfetti, C.A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*, 22–37.

Perfetti, C., Yang, C.-L., & Schmalhofer, F. (2008). Comprehension skills and word-to-text integration processes. *Applied Cognitive Psychology, 22*, 303-318.

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

Plaut, D.C. (1995). Semantic and associative priming in a distributed attractor network. In Cognitive Science Society (Ed.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society,* pp.37-42. Hillsdale, NJ: Erlbaum.

Plaut, D.C., & Booth, J.R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786-823.

Polio, C., & Gass, S. (1997). Replication and reporting: A commentary. *SSLA*, *19*, 499-508.

Posner, M.I., & Snyder, C.R.R. (1975). Facilitation and inhibition in the processing of signals. In Rabbitt, P.M.A., & Dornic, S. (Eds.), *Attention and performance V,* pp.669-682. New York: Academic Press.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*, 385-408.

Reber, P. J. (2008). Cognitive neuroscience of declarative and nondeclarative memory. In: S. Aaron, A. S. Benjamin, De BelleJ. S., B. Etnyre, & T. A. Polk (Eds.), *Advances in Psychology, 139* (pp. 113–123). North-Holland.

Reber, P. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia, 51*(10), 2026-2042.

Roediger, H. (1990). Implicit Memory: Retention Without Remembering. *American Psychologist, 45*, 1043.

Rugg, M. D., Mark, R. E. , Walla, P., Schloerscheidt, A. M., Birch, C. B., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392*(6676), 595-8.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 501-518.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced task. In N. Schmitt (Ed.), *Formulaic sequences: Processing, acquisition and use* (pp. 173-189). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (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. DOI: 10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549-669. DOI: 10.1093/applin/amt054

Sharwood Smith, M. (1991). Speaking the many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, *7*(2), 118-132.

Sharwood Smith, M. (1993). Input Enhancement in Instructed SLA: Theoretical Bases. *Studies in Second Language Acquisition, 15(2)*, 165-179.

Shelton, J.R., & Martin, R.C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1191-1210.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121-159. DOI: 10.1111/j.1467-9922.2012.00730.x

Szudarski, P. & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2). DOI: 10.1111/ijal.12092

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning* *67*(4), 747–790. DOI: 10.1111/lang.12241

Tremblay, A., Derwing, B., Libben, G., Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569-613. DOI: 10.1111/j.1467-9922.2010.0062.x

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105-122.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: an eye-movement study into the processing of formulaic sequences, in N. Schmitt (ed.), *Formulaic sequences* (pp. 73-95). Amsterdam: John Benjamins.

Voss, J.L., & Paller, K.A. (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. *Learning & Memory, 14,* 259-267. DOI: 10.1101/lm.529807

Wang, HC., Savage, G., Gaskell, M.G., Paulin, T., Robidoux, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review, 24*, 1186. https://doi.org/10.3758/s13423-016-1182-7

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.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learnin*g, *63*(1), 91-120. DOI: 10.1111/j.1467-9922.2012.00729.x

Westfall, J., Kenny, D., & Judd, C. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020-2045.

Winke, P. M. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition, 35(2)*, 323-352.

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.

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

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition, 40*(2), 395-416. DOI:10.1017/S0272263117000237.