Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions

Suhad Sonbul

Umm Al-Qura University and University of Nottingham

Norbert Schmitt

University of Nottingham

To date, there has been little empirical research exploring the relationship between implicit and explicit lexical knowledge (of collocations). As a first step in addressing this gap, two laboratory experiments were conducted that evaluate different conditions (enriched, enhanced, and decontextualized) under which both adult native speakers (Experiment 1) and advanced nonnative speakers of English (Experiment 2) acquire collocations. Three different tests of collocational knowledge were used to assess gains after treatment: two traditional explicit tests (form recall and form recognition) and an innovative implicit test (priming). Results from mixed-effects modeling showed that all conditions led to significant long-term gains in explicit form recall and recognition both for natives and nonnatives, while no condition facilitated implicit collocational priming effects either for natives or nonnatives.

Keywords vocabulary; collocations; explicit knowledge; implicit knowledge; priming; mixed-effects modeling

Introduction

The relationship between explicit and implicit knowledge has long been the focus of research in psychology (for an overview, see Squire, 2004), and a similar interest has also emerged in applied linguistics and second language acquisition (SLA) circles (N. Ellis, 2008). To date, most SLA research into the explicit/implicit linguistic knowledge distinction has focused on second

language (L2) grammar acquisition (de Graaff, 1997; R. Ellis, 2005), although the nature of the relationship between explicit and implicit grammatical knowledge remains controversial. In contrast, there has been very little consideration of the distinction's implications for lexical knowledge. This might be due to the traditional dictionary metaphor which views the lexicon as nothing more than a list of forms and meanings that can be associated fairly easily through rote learning. However, the view that vocabulary knowledge is a much more complex construct is now gaining acceptance, with several lexical scholars claiming that it entails a number of different dimensions involving both explicit and implicit mental representations (Aitchison, 2003; N. Ellis, 1994; Nation, 2001; Schmitt, 2010). Still, the exact nature and interrelationship of explicit and implicit vocabulary knowledge has only begun to be investigated.

This study addresses this emerging lexical area by reporting on a two-experiment study (one with natives and one with nonnatives) exploring how well various direct and indirect input conditions promote both explicit and implicit knowledge of one aspect of vocabulary knowledge: collocations.

Explicit Versus Implicit Knowledge: Grammar Research

Although researchers disagree on the precise nature of explicit/implicit knowledge (Bialystok, 1981; Hulstijn, 2007; Paradis, 1994), the central difference lies in the presence or absence, respectively, of awareness. These two terms are defined here as follows:

> *Explicit knowledge* consists of the facts that speakers of a language have learned. These facts are often not clearly understood and may be in conflict with each other . . . . Explicit knowledge is held consciously, is learnable and verbalisable, and is typically accessed through controlled processing when learners experience some kind of linguistic difficulty in using the L2 . . . In contrast, *implicit knowledge* is procedural, is held unconsciously, and can only be verbalized if it is made explicit. It is accessed rapidly and easily and thus is available for use in rapid, fluent communication. (R. Ellis, 2006, p. 95, emphasis in original)

In essence, explicit knowledge seems related to declarative knowledge of language, while implicit knowledge is most closely related to speed and ease of access in the use of language. R. Ellis demonstrated that, while both dimensions relate to linguistic knowledge, they are separate constructs. In an influential validation study (R. Ellis, 2005), he found that measures of explicit knowledge

(metalinguistic knowledge test and untimed grammaticality judgment task) and measures of implicit knowledge (oral narration test, timed grammaticality judgment task, and oral imitation test) loaded onto two separate factors.

Most SLA researchers agree on the central role played by implicit knowledge in achieving proficiency (N. Ellis, 1993; R. Ellis, 2006). However, the key question is the relationship between explicit and implicit knowledge. This issue emerged toward the beginning of the 1980s when the interface debate was at the heart of L2 grammar instruction research. The interface debate was based on the conviction that developing implicit knowledge is the ultimate goal of language learning, and, thus, that direct instruction is only worthwhile if it can lead to (or at least aid) the growth of implicit knowledge (R. Ellis, 2007). The main question was, thus, whether direct instruction can or cannot influence the development of implicit knowledge. The noninterface position held that direct instruction is ineffective (Krashen, 1981; Paradis, 1994). The strong interface position maintained that explicit knowledge can be converted into implicit knowledge through practice (DeKeyser, 1994; B. McLaughlin, Rossman, & McLeod, 1983). The weak interface position is somewhere in-between, holding that explicit knowledge can feed into implicit knowledge either directly (if the learner is developmentally ready) or indirectly (through facilitating noticing or through helping learners notice gaps in their own output) (R. Ellis, 1994). In other words, although explicit and implicit knowledge are distinct and dissociated, they do interact substantially (N. Ellis, 2008; N. Ellis & Laporte, 1997).

Since the 1980s, many studies have been conducted to compare the effectiveness of various direct and indirect instructional conditions on the development of explicit and implicit grammatical knowledge (for reviews, see R. Ellis, 2002; Long, 1983). In a meta-analysis conducted at the turn of the millennium, Norris and Ortega (2000) reviewed 49 studies carried out between 1980 and 1998. The following conclusions were reached: (1) instruction results in large gains, (2) these gains are durable over time, and (3) direct instruction is more effective than indirect instruction. In spite of these important findings, Norris and Ortega point out a number of caveats, the most important being that "testing of learning outcomes usually favors explicit treatments by asking learners to engage in explicit memory tasks and/or in discrete de-contextualized L2 use" (p. 501) (see also Doughty, 2003, for a similar claim). However, Mackey and Goo (2007) observe that more studies are now applying measures of implicit grammatical knowledge. This is evident in Spada and Tomita's (2010) recent metaanalysis collapsing results of 41 studies that employed explicit (direct grammar instruction) and implicit (indirect manipulation of input with

no grammar instruction, e.g., enriched exposure) instructional techniques and utilized measures of either or both types of knowledge (*explicit*: controlled constructed tasks, e.g., grammaticality judgment tasks and multiple-choice items; *implicit*: free constructed tasks, e.g., free writing and picture description tasks). Results of the meta-analysis showed that direct instruction was more effective than indirect instruction not only in developing explicit knowledge but also in enhancing implicit knowledge. It should be noted, however, that not all free constructed tasks are necessarily tapping into implicit knowledge (Spada & Tomita, 2010; see also White & Ranta, 2002). This is supported by results of a recent study (Andringa, de Glopper, & Hacquebord, 2011) where a correlation between an explicit measure (grammaticality judgment task) and an implicit one (free written response task) was interpreted as evidence that the two are actually tapping into the same construct. Spada and Tomita (2010) concluded that there is a need for more validation studies to develop new measures and evaluate what they are assessing.

In fact, modern psycholinguistic techniques can be useful in exploring the explicit–implicit distinction. For example, Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, and Molinaro (2006) reported on a study conducted to gauge the development of English learners of L2 French in the rule of verbal person agreement over a period of 8 months. Results showed that, after 80 hours of instruction, participants performed very poorly in an explicit grammaticality judgment task, but showed Event-Related Potential (ERP) behavior (P600 component elicited for violations) that was not qualitatively different from natives. Similarly, Tokowicz and MacWhinney (2005) showed that performance of English learners of L2 Spanish in an explicit measure (grammaticality judgment task) was distinct from their ERP behavior for gender agreement violations. Although their explicit performance was near chance, they showed a clear P600 effect for these violations. Results from these two studies seem to suggest that, when using a highly sensitive psycholinguistic measure, implicit knowledge appears to develop before explicit knowledge, even after short instruction periods.

In sum, research has not yet clearly described the interrelationship between explicit and implicit grammatical knowledge and how it is acquired. Overall, direct instruction seems more effective than indirect instruction in the promotion of explicit knowledge. It may be more effective in promoting implicit knowledge as well, but this depends on what measures one accepts as tapping into implicit knowledge. If using psycholinguistic techniques that are undisputedly implicit in nature (e.g., ERP), then there is some evidence that the development of implicit knowledge may actually precede explicit knowledge.

Explicit and Implicit Lexical Knowledge

When considering the explicit–implicit knowledge distinction, there are reasons to believe that vocabulary is different from grammar. The main difference lies in what Hulstijn and de Graaff (1994, pp. 104–105) term *rule learning* of *nonlexical* structures versus *item learning* of *lexical* structures. In the area of grammar, this distinction is best exemplified in learning the regular form of the past tense *play–played* as opposed to the irregular form *become–became*, respectively. Along the same lines, it can be claimed that learning vocabulary is mostly *item learning*.

This item view has led to the assumption that vocabulary knowledge is declarative in nature and can never be implicit. Ullman (2001), for example, developed his declarative/procedural model of linguistic knowledge assuming, based on neurological evidence, that vocabulary knowledge is declarative while grammatical knowledge is procedural. Similarly, Hulstijn (2007) claims that vocabulary knowledge is symbolic and, thus, explicit in nature. These views seem to treat the lexicon as a simple dictionary containing only word forms connected to their meanings, which has led to the distinction between explicit and implicit lexical knowledge being largely ignored in research.

However, vocabulary knowledge is now viewed as a complex construct encompassing various aspects. Nation (1990, 2001) categorized a number of word knowledge aspects including form, meaning, grammatical characteristics, collocation, and several others. N. Ellis (1994), one of the first scholars to show an interest in the explicit–implicit contrast for lexis, considered Nation's aspects and distinguished two groups of components: (a) components related to the connection between the form of a word and its meaning and the various semantic relations between words which can be best learned explicitly and (b) components related to the form of the word and its receptive (input)/productive (output) aspects of usage which are best learned implicitly (but can also be learned explicitly).

N. Ellis's (1994) distinction does not stop at learning processes, but can also be applied to knowledge representations. Certain (meaning-related) aspects of word knowledge are inherently explicit whereas others (formal and usage-related aspects) might be either explicit or implicit depending on the presence or absence of awareness. R. Ellis (2004, p. 242) makes a similar claim:

> [W]hereas knowledge of the form of a word and of word collocations is largely implicit . . . Word meanings . . . probably constitute the largest single area in a learner's explicit knowledge. In addition, of course, individuals may develop a conscious awareness of the form of at least

some words—problematic spellings or the differences between American and British spellings, for example—and of some collocations.

This more sophisticated perspective highlights the value of the explicit–implicit distinction in understanding vocabulary knowledge. Unfortunately, the lack of conventional measures of implicit lexical knowledge has led to most previous lexical research being explicit in nature, assessing breadth, depth, and organization of vocabulary knowledge (see Read, 2000, for an overview). Implicit measures used in psychological research are just beginning to be used in investigating vocabulary acquisition issues (e.g., Elgort, 2011). One example is J. McLaughlin, Osterhout, and Kim (2004), who studied knowledge of the word form. They employed two measures of vocabulary form (a lexical decision task tapping explicit knowledge and an ERP measure gauging implicit knowledge) to observe the development of a group of English learners of French. Lexical decision and ERP behavior were obtained longitudinally in three sessions throughout a semester. Although French learners did not show much improvement in explicit performance, their ERP data reflected sensitivity to lexical status and semantic relatedness (similar to those found for native speakers in earlier research) after only 14 hours of instruction. In interpreting these results, the researchers claim that ERP is a more sensitive measure of linguistic improvement than explicit tests which might underestimate gains in knowledge.

Collocational Knowledge: Acquisition and Measurement

This section focuses on the other aspect of vocabulary knowledge besides word form which R. Ellis (2004) considers largely implicit in nature: collocational knowledge. Most research into collocations has been descriptive in nature, often analyzing either native speaker corpora or L2 output to see what phraseology has been produced (see, e.g., Siyanova & Schmitt, 2008). However, there has been little research into how knowledge about collocations has been acquired.

Durrant and Schmitt (2010) conducted an experiment in a controlled (laboratory) English-as-a-second-language (ESL) setting with advanced nonnatives. Participants were instructed to read aloud a number of sentences containing made-up, low-frequency, adjective-noun collocations (to ensure no preknowledge). All participants were tested with a timed cued recall (naming) task with instructions to remember combinations from the study phase. Results showed that L2 learners do retain memory of collocations they meet in input even after one or two exposures and even when these are not enhanced in any way. However, it should be noted that this study (and others conducted in the classroom context, e.g., Laufer & Girsai, 2008; Webb & Kagimoto, 2009) did not

employ any measure of implicit collocational knowledge. Although Durrant and Schmitt's test was timed and was thus tapping intuitions, we consider it a test of explicit knowledge in the sense that it elicited the target form (collocations) explicitly with a very clear reference to the study phase.

One early study using the priming paradigm did look at implicit knowledge of collocations. McKoon and Ratcliff (1992, Experiment 3) showed that priming can occur not only for made-up pairs but also for real-world collocations. In their study, they extracted noun-noun pairs (e.g., *hospital-baby*) from a small corpus. These pairs co-occurred in a six-word window higher than chance (as measured by high Mutual Information scores) but were not related according to association norms. Results of a priming/lexical decision test showed that there was a significant faciliation effect for these collocation pairs over control ones. This study seems to have opened the door for research into what is now called collocational priming.

This notion of collocational priming was first hypothesized by Hoey (1991, 2005), who claimed that collocations are acquired incrementally through exposure to text. Language users store concordances of every word encountered and its context, so each contextual encounter with a word might either create a new collocation or reinforce and modify an old one. According to Hoey (2005), there is little difference in the collocational priming process between adult natives and nonnatives; rather, the key issue is the amount and type of input. For him, natives generally thrive on the indirect input because they are likely to be surrounded by a massive amount of linguistic evidence, while nonnatives generally do better with direct exposure because they are not usually surrounded by such evidence. If there is enough exposure, then collocations will be learned inductively for both groups of learners, but if there is not enough exposure, then other more explicit techniques should be used to help native and nonnative learners. However, different activities raising learners' attention to collocational restrictions (e.g., exercises, illustrations, exposure to rich input) are likely to have differential effects on collocational priming.

Three recent studies examined this notion of collocational priming with native speakers of English using lexical decision tasks (Durrant & Doherty, 2010, Experiment 1; N. Ellis, Frey, & Jalkanen, 2009; Wolter & Gyllstad, 2011). Results of these studies showed that collocational priming is indeed possible and that native speakers are sensitive to collocational frequency. More specifically, a facilitation effect was found for high-frequency collocates over low-frequency collocates and control pairs. There is some question, however, of whether these studies were truly tapping into implicit knoweldge, as the priming techniques employed were not automatic in nature. Automatic priming

is contrasted with strategic priming in that the former operates rapidly without participants' conscious intention (implicit knowledge) while the latter requires intention and reflects slower processes (explicit knowledge) (Dagenbach, Horst, & Carr, 1990). After discussing different variables determining the nature of priming processes (stimulus onset asynchrony [SOA], the nature of task, the type of semantic relation, relatedness proportion [RP] measured as the number of related word-prime trials divided by all word-target trials, and nonword ratio [NR] measured as the ratio of nonword trials to all unrelated trials), McNamara (2005) concludes:

> An investigator interested in the automatic component of semantic priming would be well advised to use a short SOA (e.g., 200 ms or less), a low RP (e.g., <0.20?), and an NR of 0.50. Under these conditions, the task and the type of semantic relation do not seem to matter much. (p. 72)

Unfortunately, none of the three experiments mentioned above met these requirements. Thus, although they did show collocational priming, how this priming relates to implicit knowledge (vs. strategic explicit knowledge) is not clear.

Present Study: Aim, Operationalization of Terms, and Research Questions

To our knowledge, no research study to date has evaluated various learning conditions under which implicit knowledge along with explicit knowledge of collocations might be developed. The present study employs an established psycholinguistic measure in addition to traditional tests to investigate implicit/explicit lexical knowledge. Two experiments (with natives and nonnatives, respectively) are conducted to investigate the effectiveness of three typical learning conditions (enriched, enhanced, and decontextualized) under which explicit (form recall and recognition) and implicit (automatic priming) collocational knowledge might be acquired. The first two learning conditions are operationalized (following Reinders & R. Ellis, 2009) from an input perspective, as follows:

> Enriched input: "input that has been seeded with the target structure so that learners are exposed to high frequency over a period of time." (p. 282)
> Enhanced input: "input where the target feature has been emphasized in some way—glossing, bolding or underlining." (p. 283)

The decontextualized input condition, in turn, is defined as input where the target feature is presented out of context with direct instruction to commit it to memory. In accordance with Spada and Tomita's (2010) definitions (see earlier

discussion), the first condition is considered an indirect instructional condition while the second and third are considered direct teaching conditions.

The term collocation is operationalized here as a technical two-word pair that often co-occurs in medical contexts as evidenced in medical dictionaries. It should be noted here that the present study investigates initial stages of collocational learning, whereby a link is built between the two components of a collocation. Other aspects of acquisition, whereby a form–meaning link is established, are not assessed. Two testing sessions were included: one immediate and the other after a 2-week delay. The following research questions were addressed both for natives (Experiment 1) and nonnatives (Experiment 2):

1. Which type of exposure (enriched, enhanced, or decontextualized), if any, promotes explicit, form recall knowledge of collocations? Are any more effective than others?
2. Which type of exposure, if any, promotes explicit, form recognition knowledge of collocations? Are any more effective than others?
3. Which type of exposure, if any, promotes implicit knowledge (automatic priming effects) of collocations? Are any more effective than others?

Our two experiments included both indirect and direct input and measured both implicit and explicit lexical knowledge. Thus, it was possible to explore whether indirect input best facilitates implicit knowledge while direct input best facilitates explicit knowledge, or whether direct or indirect input is superior for both types of knowledge. Furthermore, we expected that comparison of results across the two experiments would enable us to determine if natives and nonnatives gain similar types of advantage from different input conditions in terms of learning implicit and explicit knowledge, as Hoey (2005) suggests.

Experiment 1: The Acquisition of Collocations by Adult Natives

Experiment 1 evaluates different conditions under which adult native speakers of English might acquire explicit and implicit knowledge of novel medical collocations. It consists of a study phase where participants are exposed to collocations, followed by immediate and delayed test phases where their learning gains are measured.

Method
Participants
The 35 native participants (15 males and 20 females) were undergraduates studying at a British university who received a payment (£8) for their

participation. They had no previous medical training and ranged in age from 18 to 27 ($M = 19.54$, $SD = 1.74$). Data for a further 8 participants was excluded from the experiment because of (1) low (less than 60%) scores in the reading comprehension test (see Procedure Section below) ($N = 1$), (2) having too many (more than 15%) incorrect or extreme responses in the lexical decision task ($N = 2$), or (3) not showing up in the delayed testing session ($N = 5$).

Stimuli

We intended to set up as natural a pedagogical setting as possible, and so wished to use real collocations as stimuli. However, this required finding collocations which educated native speakers would not already know. We decided to use medical collocations, as these are unlikely to be known by people without medical training. Choosing medical collocations for the present study involved two stages. First, five medical resources[1] were consulted in an attempt to identify candidates that fit the following criteria:

1. The collocation is transparent and is not technically loaded (i.e., no collocations comprising words of opaque Latin origin or names of drugs).
2. The British National Corpus (BNC) frequency of the collocation, checked using Davies's (2007) interface, is very low (0–8 occurrences in the whole corpus). This step indicated that the average native speaker would be unlikely to know the collocation.
3. The first word of the collocation has a number of synonyms that might look equally attractive to a person who has no knowledge of the collocation.

The candidate collocations were piloted with undergraduates similar to the experiment participants to ensure that they were all essentially unknown.

Second, for each of the 15 collocations which fit the above criteria, a control pair was devised for the priming task (see Materials section below). This pair included the second word of the intact collocation along with a synonym of the first (e.g., *decaying lung* as a control for *vanishing lung*), with the pair not occurring in the BNC. The first word of the control pair was closely matched with the first word of the intact collocation item-by-item in terms of the following factors: part of speech, semantic plausibility, BNC frequency (the frequency per million of the control word is within the range of ±30 of the intact word frequency), and length (there is no more than one letter difference between the control and the intact words). Appendix S1 in the Supporting Information online presents the intact collocations and their controls for Experiment 1.
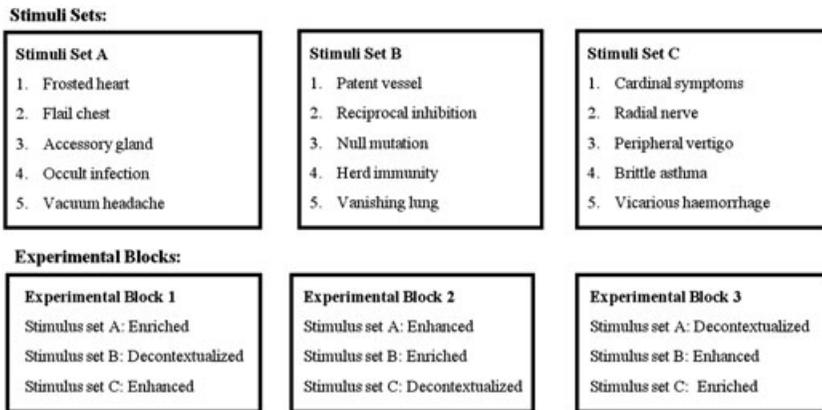
Stimuli Sets:

Stimuli Set A	Stimuli Set B	Stimuli Set C
1. Frosted heart	1. Patent vessel	1. Cardinal symptoms
2. Flail chest	2. Reciprocal inhibition	2. Radial nerve
3. Accessory gland	3. Null mutation	3. Peripheral vertigo
4. Occult infection	4. Herd immunity	4. Brittle asthma
5. Vacuum headache	5. Vanishing lung	5. Vicarious haemorrhage

Experimental Blocks:

Experimental Block 1	Experimental Block 2	Experimental Block 3
Stimulus set A: Enriched	Stimulus set A: Enhanced	Stimulus set A: Decontextualized
Stimulus set B: Decontextualized	Stimulus set B: Enriched	Stimulus set B: Enhanced
Stimulus set C: Enhanced	Stimulus set C: Decontextualized	Stimulus set C: Enriched

Figure 1 Target collocations counterbalanced across treatment conditions (Experiment 1).

Materials

We then chose a reading passage from *The Language of Medicine in English* (Tiersky & Tiersky, 1992) which covered a wide range of medical conditions, thus facilitating the inclusion of the intact collocations. According to Rott (2009), a single exposure can result in an initial memory trace (see also Durrant & Schmitt, 2010) which should later be strengthened through subsequent exposures. Thus, the passage was adopted and shortened to include three occurrences of 10 intact collocations, varying according to our two contextualized experimental conditions (i.e., enriched and enhanced) as follows. Five collocations were embedded in the passage (the enriched condition) while the other five were embedded and made salient with red, bold font (the enhanced condition). In the third learning condition, five collocations were taught in isolation in a PowerPoint presentation (the decontextualized condition), where each collocation was presented individually on a different slide (no meaning was provided). Each collocation was presented in red font, flashing on the screen three times for a total of 10 seconds.

The 15 collocations and the 3 learning conditions were placed into a counterbalanced design consisting of three experimental blocks, as shown in Figure 1. Because it was not possible to administer a pretest in order not to attract participants' attention to the critical stimuli before the experiment, a control group was included as a baseline for comparison. The 35 participants were randomly assigned to the experimental groups (Block 1, $N = 10$; Block 2, $N = 8$; Block 3, $N = 8$) or to the control group ($N = 9$).

Measures

Because the focus of the present experiment is on the distinction between explicit and implicit collocational knowledge, three measures were developed: two explicit and one implicit. The two explicit measures were based on Schmitt, Dornyei, Adolphs, and Durow (2004). The first (form recall) employed a cloze format to measure participants' ability to recall the form of the collocation. Participants were presented with a summary of the reading passage with target collocations in parentheses. A brief definition of each collocation was included in the margin as a clue. To avoid any potential effects of this test on the subsequent form recognition one, the first letter of each missing word was not included. Alternatively, the number of letters in each space was represented by dashes to restrict participants' responses. The following example is the form recall test item for *vanishing lung*:

Lung infections might lead to different conditions such as the (_ _ _ _ _ _ _ _ _ **lung**) syndrome.

> Decrease in the lung's size as shown in an X-ray

The second explicit test measured form recognition. It used the same passage as in the form recall test, but with multiple-choice items given in the margin. Each item included five options: the correct first word (the key), three plausible distracters (one is the control and two other synonyms with the same part of speech), and one "I don't know" option to reduce guessing. Below is the form recognition test item for *vanishing lung*:

Lung infections might lead to different conditions such as the (_____ **lung**) syndrome.

> **a.** fading
> **b.** declining
> **c.** decaying
> **d.** vanishing
> **e.** I DON'T KNOW

While the explicit measures are fairly traditional, the measurement of implicit collocation knowledge required a more innovative approach. We opted to use the priming paradigm to capture implicit knowledge. In our priming task (designed on E-Prime[®], Psychology Software Tools, Inc., www.pstnet.com), participants were presented with the first word of the collocation as the prime and the second as the target. They had to decide whether the second string of letters is a real English word or not by pressing the appropriate key (YES or NO) on a response box. The priming task was designed in accordance with McNamara's (2005) guidelines to ensure automatic, nonstrategic priming.

Critical stimuli for this task included the 15 intact (exposed) collocations and their 15 control pairs. The intact and control primes were combined into

two counterbalanced lists such that seven or eight intact pairs and seven or eight control pairs were included in each list. Targets which were matched with their intact primes in one list were matched with their control primes in the other. No prime or target word was used more than once in either list. In addition, 53 filler items were included, in which the second item of the pair was either a word ($K = 22$) or a nonword ($K = 31$).[2] All nonwords were selected from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002) and were pronounceable and orthographically legal. In order to reduce the effect of the implicit test on the explicit form recognition test, 30 fillers included distracters from the multiple-choice test as primes. Thus, in each stimuli list, participants were presented with 68 trials (7 or 8 intact pairs, 7 or 8 control pairs, and 53 fillers). Half of the participants ($N = 18$) were tested with the first stimuli set, while the other half ($N = 17$) were tested with the second set. Twenty practice trials consisting of nonstudy items were constructed to allow participants to get used to the lexical decision task.

In each trial, a fixation point was presented (2,000 milliseconds) followed by the first word of each pair (the prime), which remained on the screen for 150 milliseconds. This was immediately replaced by the second string (the target), which remained on the screen until the participant had made a lexical decision. The order of the trials was randomized across subjects to avoid potential order effects. E-Prime recorded accuracy and response latency for each participant in all trials.

Procedure

Upon arrival at the lab for the single-person session, the participant signed the consent form in which he or she was briefed on the purpose of the study. Then, the study phase began. While participants in the experimental blocks took part in both study and test phases, those in the control group only took part in the test phase. The participant was instructed to read the whole passage once within 10 minutes and to try to focus on the message as there would be some compre-hension questions at the end. Upon finishing, the first author took the reading passage and participants were instructed to watch a PowerPoint presentation that included five collocations initiated with the following instruction:

> Now, you will be presented with five different phrases that usually occur in medical contexts. Please study them carefully and try to remember the combination.

After the presentation, a general comprehension test (eight multiple-choice items) was administered for a dual purpose. First, it was intended to ensure

that participants were reading for meaning; the mean score gained by the 26 participants in the experimental group was acceptable (7.26 out of 8.00, 90.87%, $SD = 0.96$). Second, the test functioned as an interval between the study phase and the test phase.

To avoid earlier measures affecting later ones, the tests were carefully sequenced. The first was the computerized priming task, which took 6 minutes. The participants then performed a distracter task (a digit-span memory test) that was intended to minimize any effect of the implicit test on the subsequent explicit tests. Next the participants completed the paper-and-pencil form recall test within 10 minutes, followed by the form recognition test (5 minutes). Finally, participants filled out a short background questionnaire.

At the end of the session, participants were invited to a follow-up experiment to be conducted two weeks later. This entailed a delayed testing session where the participants took the explicit and implicit tests without being exposed to any treatment. To minimize any possible transfer from the immediate to the delayed testing sessions, the choices were reordered in the delayed multiple-choice test.

Scoring and Data Analysis

When scoring and coding the explicit form recall test, a response that was correct in spelling and morphology was coded as CORRECT. Because this test aimed at assessing knowledge of collocations rather than knowledge of the word form, a response was also coded as CORRECT if the word was misspelled but recognizable (*vacume* for *vacuum*) or if the correct stem was provided but a wrong affix was used (e.g., *reciprocate* for *reciprocal*) (partially correct answers constituted only 0.66% of all data points in the recall test). A response was coded as INCORRECT if no response or a wrong response was given. Scoring the explicit form recognition test was straightforward; a correct answer was coded as CORRECT and an incorrect, missing, or "I don't know" response was coded as INCORRECT.

We conducted the analysis with R version 2.12.1 (R Development Core Team, 2010) using mixed-effects models, that is, models that include both fixed effects (the independent variables) and random effects (participants and items). Mixed-effects modeling is now becoming a popular method of analysis in psycholinguistic research for a number of reasons. First, it can treat participants and items as random variables in one model allowing for "the simultaneous generalization of the results on new items and new participants" (Gagné & Spalding, 2009, p. 25). Second, it allows testing main independent variables in addition to other psycholinguistic (item-related and participant-related) variables.

Third, it can handle interval-scale (e.g., reaction time) measures through linear mixed-effects (LME; see Baayen, 2008) models as well as categorical (e.g., correct versus incorrect) measures using mixed logit models (the LME alternative for categorical data; see Jaeger, 2008). Lastly, it can cope with missing values and imbalanced designs. Because the present study involved both a correct/incorrect categorical measure along with an interval-scale reaction time measure and because we aimed at testing various item-related and participant-related factors, we opted to use mixed-effects modeling as the analysis method.

The method starts with the simplest (null) model, which includes only the dependent measure and the random variables. Fixed effects are then added incrementally and χ^2 (likelihood ratio) tests are used to check whether the inclusion of additional predictors contributes significantly to the model. Once the interim best-fit model is reached, variables are excluded one by one to check whether any predictor is redundant (i.e., a variable whose presence/absence in the model does not lead to any significant difference) and can thus be excluded. Once these are excluded, we are left with the final best-fit model, which best describes the overall variance in the research design. As mixed-effects modeling is likely to be new for many readers, we provide additional detailed description of the procedure we followed in Appendix S3 of the Supporting Information online.

We developed two mixed-effects models, one for explicit knowledge and one for implicit knowledge. For explicit knowledge, the analysis (mixed logit modeling) started with a null model including our binomial dependent variable (CORRECT versus INCORRECT responses) and participants and items as random effects. We then added predictor variables incrementally in the following order: main effects (TEST TYPE: form recall, form recognition; CONDITION: enriched, enhanced, decontextualized, control; and SESSION: immediate, delayed[3]) and then item-related effects (length of Word 1, log frequency of Word 1, length of Word 2, log frequency of Word 2, frequency of the collocation). Finally, variables were excluded one by one to check for redundancy to arrive at the final best-fit model. The procedure for implicit knowledge (using LME modeling) was the same except that the dependent variable was log response time[4] (reaction times were log-transformed to reduce skewness in the distribution) instead of CORRECT/INCORRECT responses, and the first main effect was ITEM TYPE (intact, control) instead of TEST TYPE.

Results
Explicit Form Recall and Form Recognition Tests
An assumption of the experiment is that the medical collocations were unknown to the participants. This assumption is supported by the control group results

Table 1 Responses and percentile scores in explicit tests (form recall and recognition) under all conditions (natives)

Condition	Session	Form Recall (Cloze-test)				Form Recognition (Multiple-choice)			
		Correct	*%*	*Incorrect*	*%*	*Correct*	*%*	*Incorrect*	*%*
Control[a]	Immediate	2	1.5	133	98.5	43	31.8	92	68.2
	Delayed	13	9.6	122	90.4	46	34.1	89	65.9
Decontextualized[b]	Immediate	44	33.8	86	66.2	97	74.6	33	25.4
	Delayed	37	28.5	93	71.5	92	70.8	38	29.2
Enhanced[b]	Immediate	36	27.7	94	72.3	93	71.5	37	28.5
	Delayed	34	26.2	96	73.8	89	68.5	41	31.5
Enriched[b]	Immediate	29	22.3	101	77.7	90	69.2	40	30.8
	Delayed	34	26.2	96	73.8	84	64.6	46	35.4

[a]Max score $= 135$ ($K = 15 \times N = 9$).
[b]Max score $= 130$ ($K = 5 \times N = 26$).

(Table 1), which showed very low form recall collocation knowledge (1.5%). However, it should be noted that the recall percentage score for the delayed test (9.6%) was higher, which suggests that the implicit priming test might have had a direct effect on explicit test results (see Limitations section). As opposed to the lack of form recall knowledge by the control group, this group answered one third of the items in the recognition test correctly (around 32% immediate and 34% delayed). This implies that participants were merely guessing on the four-option multiple-choice test. Thus, if participants under the treatment conditions score higher than the 9.6% level (recall) and 34.1% level (recognition), gains should be attributed to the treatment.

We modeled CORRECT answers likelihood using a mixed logit model. Main effects were added to the null model incrementally. The addition of TEST TYPE significantly improved the null model ($\chi^2(1) = 429.80$, $p < .001$, $N = 2,100$). Adding CONDITION further improved the model ($\chi^2(3) = 26.68$, $p < .001$, $N = 2,100$). Then, a TEST TYPE $\times$ CONDITION interaction was added to the model but this did not lead to any significant improvement ($\chi^2(3) = 0.15$, $p = 0.99$, $N = 2,100$). Moreover, adding SESSION as a main effect did not improve the model ($\chi^2(1) = 0.08$, $p = .78$, $N = 2,100$), neither did adding it as an interacting variable with CONDITION ($\chi^2(4) = 5.15$, $p = .27$, $N = 2,100$) or TEST TYPE ($\chi^2(2) = 1.14$, $p = .56$, $N = 2100$). SESSION was thus removed from the model. Item-related variables were then added incrementally, but none

Table 2 Summary of the best fit mixed logit model for variables predicting natives' correct explicit responses

Predictor		*Estimate*	*SE*	*Wald-Z*	*p*
(Intercept)		−3.15	0.38	−8.38	<.001
TEST TYPE:	*recognition*	2.31	0.12	18.81	<.001
CONDITION:	*decontextualized*	2.11	0.38	5.53	<.001
	enhanced	1.90	0.38	4.99	<.001
	enriched	1.72	0.38	4.52	<.001

Note. The model has random intercepts for participants and items. $N = 2,100$, log-likelihood = -1,034.08.

improved the model. Finally, significant predictors were removed backward but none was found to be redundant. The final best-fit model is presented in Table 2.

The best-fit mixed logit model indicates that more correct responses were produced in the recognition test (634/863) than in the recall test (229/863). Also, more correct responses (237, 252, 270/863) were produced under the three experimental conditions (enriched, enhanced, decontextualized, respectively) than those produced by the control group (104/863). In order to assess differences across experimental conditions, a multiple comparison analysis (see Bretz, Hothorn, & Westfall, 2010) was conducted on the CONDITION variable. Results showed that more correct responses were produced under the decontextualized condition than the enriched condition (*Wald Z* = 2.55, *p* = .046) while no significant difference was found between the enhanced condition and the other two conditions (enriched: *Wald Z* = -1.17, *p* = .63, decontextualized: *Wald Z* = 1.38, *p* = .49). Third, lack of interaction between TEST TYPE and CONDITION implies that (1) the advantage of the recognition gains over recall gains was present under all conditions and (2) significant differences across conditions were present both for recall and recognition gains. Finally, the fact that SESSION was not a significant predictor implies that explicit gains did not drop significantly after the 2-week gap and that the effects reported above are true for immediate and delayed results.

Overall, immediate and delayed results (for both recall and recognition tests) can be summarized as follows:

Enhanced = Enriched > Control

Enhanced = Decontextualized > Control
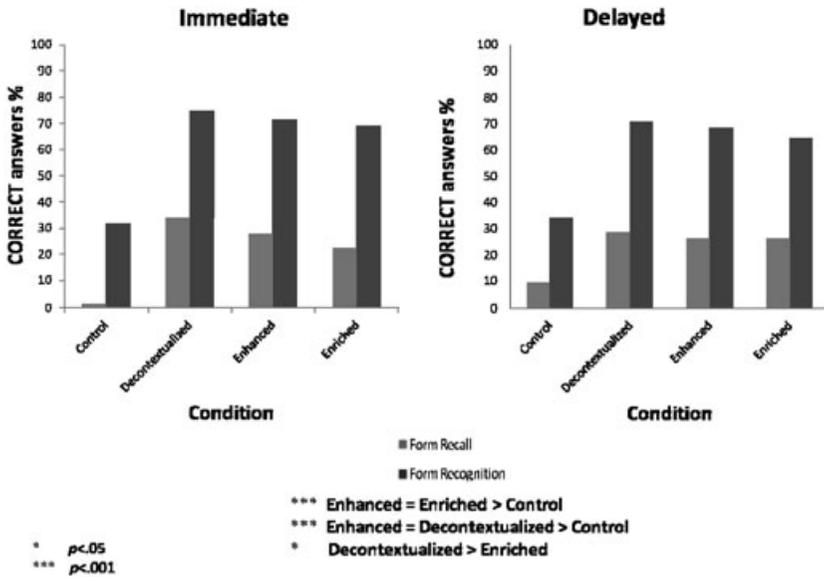
Decontextualized > Enriched

Figure 2 Percentages and comparisons of explicit (form recall and recognition) scores under all conditions in both sessions (natives).

These results are also graphically illustrated in Figure 2. The three experimental conditions led to higher explicit gains than the control group (around 16–19 percentage points for recall and 30–37 percentage points for recognition both on the delayed test[5]). Also, the decontextualized condition led to more correct answers than the enriched condition on the delayed test (two percentage point difference in recall and six in recognition).

Implicit Priming Test
If implicit memory for the intact collocations was formed during the study phase, the experimental conditions should produce faster (shorter) reaction times to these collocations than to their control counterparts, but this should not be the case for the control group. Table 3 presents mean reaction times under all conditions.

First we checked the overall accuracy in the lexical decision task, and it was high (97.43%).[6] We then fit a LME model with log reaction time as the dependent variable and main effects (ITEM TYPE, CONDITION, and SESSION) added incrementally as predictors to the null model. ITEM TYPE did not improve the null model ($\chi^2(1) = 0.00$, $p = 1.00$, $N = 1,001$) and was thus removed. Then, CONDITION was added as a predictor but did not

Table 3 Mean reaction time (RT) in milliseconds for intact and control items under all conditions (natives)

Condition	Session	Intact Items		Control Items		Difference in
		RT Mean	SE[a]	RT Mean	SE[a]	RT Mean
Control	Immediate	695.81	26.31	702.02	27.59	6.21
	Delayed	651.89	17.93	639.00	21.67	−12.89
Decontextualized	Immediate	644.59	26.16	650.11	23.48	5.52
	Delayed	605.10	19.81	637.27	23.47	32.17
Enhanced	Immediate	656.00	23.77	679.15	28.61	23.15
	Delayed	611.84	19.04	661.43	23.38	49.59
Enriched	Immediate	674.00	25.95	648.18	25.91	−25.82
	Delayed	644.32	24.44	648.66	24.37	4.34

[a]Standard error of mean.

lead to any significant improvement in the null model ($\chi^2(3) = 4.00$, $p = .26$, $N = 1,001$), and so was also removed. SESSION was added as a predictor and was found to significantly improve the model ($\chi^2(1) = 10.11$, $p = .001$, $N = 1,001$) with shorter (faster) reaction time in the delayed session. After testing all main effects, item-related variables were added incrementally to assess their effects on log reaction time. Only Word 1 length ($\chi^2(1) = 5.27$, $p = .02$, $N = 1,001$), Word 2 length ($\chi^2(1) = 18.13$, $p < .001$, $N = 1,001$), and Word 2 log frequency ($\chi^2(1) = 15.04$, $p < .001$, $N = 1,001$) led to significant improvement in the model. Thus, the shorter the first word and the shorter/more frequent the second word, the shorter the log reaction time is. We then tested interactions between significant predictors (SESSION, W1 length, W2 length, and W2 log Frequency). Only the interaction between W2 length and W2 log Frequency led to a significant improvement in the model ($\chi^2(1) = 5.71$, $p = .02$, $N = 1,001$). In this model, W2 log frequency ceased to be a significant predictor. Thus, the effect of the second word frequency on reaction times (RT) does not seem to be direct but is rather modulated by length (i.e., more frequent second words lead to faster RT only if they are short). Finally, removing the significant predictors backward did not show any redundant effects, and so all were retained. Table 4 presents the final best-fit model.

The nonsignificant interaction between CONDITION and ITEM TYPE indicates that natives' RT for intact versus control primes was not predicted by the type of input (enriched, enhanced, decontextualized, or no input [control]) either in the immediate or the delayed sessions. This result provides a clear answer to the third research question; namely, natives did not develop implicit

Table 4 Summary of the best fit LME model for variables predicting natives' log reaction time

Predictor	Estimate	MCMC	SE	t	pMCMC	Pr(>|t|)
(Intercept)	6.15	6.15	0.05	114.05	0.001	<.001
SESSION: immediate	0.04	0.04	0.01	3.16	0.006	0.002
Word 1 Length	0.02	0.02	0.00	3.62	0.001	0.004
Word 2 Length	0.03	0.03	0.00	6.03	0.001	<.001
Word 2 Log Frequency	0.05	0.05	0.04	1.30	0.15	0.20
W2 Length × W2 Log Frequency	−0.01	−0.01	0.01	−2.36	0.03	0.02

Note. The model has random intercepts for participants and items. $N = 1,001$, $R^2 = 0.54$. MCMC = Monte Carlo Markov chain; pMCMC = p values estimated by the MCMC chain method using 10,000 simulations; $Pr(>|t|) = p$ values obtained with the t test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom.
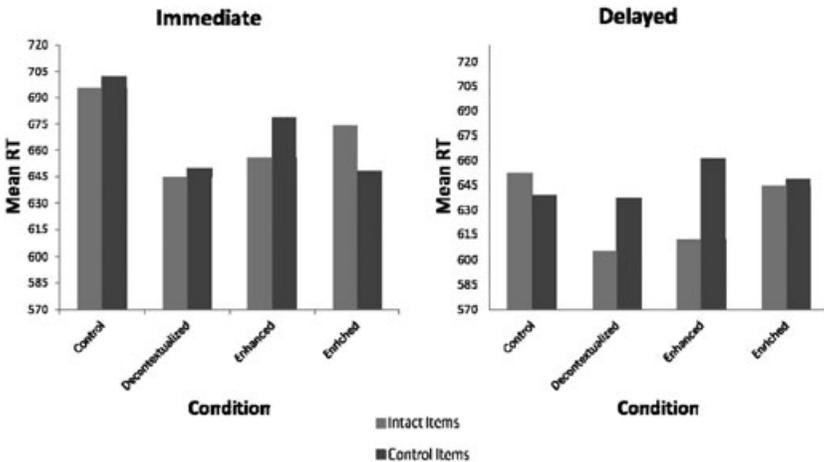

Figure 3 Mean reaction times for intact and control items under all conditions in both sessions (natives).

knowledge of collocations under any condition, at least as measured by our priming measure. This can be summarized as follows (and is graphically illustrated in Figure 3):

Immediate: Enriched condition = no effect
Enhanced condition = no effect

Decontextualized condition = no effect
Control group = no effect

Delayed: Enriched condition = no learning
Enhanced condition = no learning
Decontextualized condition = no learning
Control group = no learning

Discussion

Experiment 1 compared three treatment conditions for their ability to promote collocation learning. On the explicit side of knowledge (form recall and form recognition), all treatment conditions led to learning. Moreover, the decontextualized condition led to higher gains than the enriched condition for both types of knowledge and in both testing sessions. However, the enhanced condition seems to stand in the middle between the other two treatment conditions, in that it did not lead to any significant difference. Furthermore, just as with single words, form recognition knowledge is stronger than form recall knowledge.

In terms of implicit knowledge, one may have speculated that explicit learning conditions would lead to explicit knowledge, while implicit knowledge would accrue best from an indirect learning condition without an explicit component. However, we found that natives did not develop either short-term or long-term implicit memory traces for intact collocations under any treatment condition. To conclude, it seems that promoting explicit knowledge in native speakers is possible with a wide range of treatment conditions. Implicit knowledge, on the other hand, seems to be difficult to develop from such a short treatment period.

Experiment 2: The Acquisition of Collocations by Adult NonNatives

Experiment 2 builds upon Experiment 1 and aims at investigating the conditions under which advanced adult nonnatives might acquire English medical collocations. The design mirrors that of Experiment 1 except for one modification: collocations were chosen carefully to ensure a suitable level of difficulty for nonnatives.

Method
Participants
The 43 participants (13 males and 30 females) were postgraduate students (master's = 25, doctoral = 18) majoring in nonmedical fields at a British

university, who had all met the university entry requirement (minimum IELTS score of 6.00 or TOEFL score of 550). They ranged in age from 21 to 41 ($M =$ 28.65, $SD = 4.91$) and came from a variety of first language (L1) backgrounds (Arabic = 6, Bengali = 1, Bulgarian = 1, Chinese = 9, Dutch = 2, Farsi = 1, French = 2, Hindi = 1, Italian = 1, Kurdish = 1, Luxemburgish = 1, Malay = 2, Polish = 2, Portuguese = 2, Romanian = 2, Russian = 1, Spanish = 3, Swedish = 1, Swiss = 1, Thai = 1, Turkish = 1, and Yoruba = 1). The participants had spent a mean of 25.59 months in the United Kingdom ($SD =$ 26.91, $Min = 6$, $Max = 132$), but were first exposed to English at an average age of 10.78 years ($SD = 4.41$, $Min = 3$, $Max = 22$). Their self-rated proficiency scores (on a scale from 1 = very poor to 5 = excellent) were: reading $M =$ 4.14, $SD = 0.74$; writing $M = 3.97$, $SD = 0.89$; speaking $M = 3.86$, $SD = 0.64$, and listening $M = 3.95$, $SD = 0.69$. Their overall self-rated proficiency score (averaged across skills) was 3.98 ($SD = 0.62$). The participants were offered a £10 payment for their participation. Data for a further 5 participants were excluded from the experiment for having too many (more than 15%) incorrect or extreme responses in the lexical decision task ($N = 3$) or for not showing up for the delayed test ($N = 2$).

Stimuli

The procedures from the previous experiment were employed in choosing medical intact collocations and their controls, except for the following additional criterion:

> Component words of each collocation should belong to the most frequent 3,000 lemmas in the BNC (Leech, Rayson, & Wilson, 2001) or to the General Service List (West, 1953).

This step was intended to ensure that the nonnatives knew the individual words comprising each collocation. We ended up with 15 viable items (see Appendix S2 in the online Supporting Information). These were divided into three balanced sets of five collocations, which were then counterbalanced across the three treatment conditions, as shown in Figure 4.

Materials

The passage from Experiment 1 was adapted to include three occurrences of each collocation and simplified somewhat to be of a suitable level of difficulty for the nonnative participants.[7] Other than this, the treatment conditions were the same. The 43 participants were randomly assigned to the experimental groups (Block 1, $N = 11$; Block 2, $N = 11$; Block 3, $N = 12$) or to the

Stimuli Sets:

Stimuli Set A	Stimuli Set B	Stimuli Set C
1. Smooth diet	1. Cloud baby	1. Shell shock
2. Silent areas	2. Fixed end	2. Split hand
3. Gap periods	3. Chief complaint	3. Golden hour
4. Iron lung	4. Specific diseases	4. Principal cells
5. Stone heart	5. Pure absence	5. Regional control

Experimental Blocks:

Experimental Block 1	Experimental Block 2	Experimental Block 3
Stimulus set A: Enriched	Stimulus set A: Enhanced	Stimulus set A: Decontextualized
Stimulus set B: Decontextualized	Stimulus set B: Enriched	Stimulus set B: Enhanced
Stimulus set C: Enhanced	Stimulus set C: Decontextualized	Stimulus set C: Enriched

Figure 4 Target collocations counterbalanced across treatment conditions (Experiment 2).

control group ($N = 9$). Participants in these blocks did not differ in terms of the time they spent in the United Kingdom, age they were exposed to English, or self-rated proficiency (all $ps > .20$).

Measures, Procedures, Scoring, and Data Analysis
The designs of the two explicit tests (cloze and multiple-choice) and the implicit, priming test followed guidelines from Experiment 1. Likewise, the procedures and scoring[8] of this experiment were the same as those followed in Experiment 1, with two exceptions. First, because nonnatives are obviously slower than natives in lexical decision, we decided to extend the long cutoff time in the implicit measure to 1,800 milliseconds. Responses shorter than 150 milliseconds or longer than 1,800 milliseconds (constituting 1.97% of all data points) were excluded from further analysis. Second, one participant-related factor (overall self-rated proficiency score averaged across skills) was included to check its effect on explicit and implicit gains.

Results
Explicit Form Recall and Form Recognition Tests
The experiment assumed that medical collocations were unknown to the non-native participants, and the results of the control group on the immediate tests seem to confirm this (Table 5), with only a 2.2% score on the form recall test, and a 29.6% on the form recognition test, which is close to chance. However, it should be noted that, similar to results of Experiment 1, the percentage scores for the delayed tests (10.4% recall, 45.9% recognition) are far higher.

Table 5 Responses and percentile scores in explicit tests (form recall and recognition) under all conditions (nonnatives)

Condition	Session	Form Recall (Cloze-test)				Form Recognition (Multiple-choice)			
		Correct	*%*	*Incorrect*	*%*	*Correct*	*%*	*Incorrect*	*%*
Control[a]	Immediate	3	2.2	132	97.8	40	29.6	95	70.4
	Delayed	14	10.4	121	89.6	62	45.9	73	54.1
Decontextualized[b]	Immediate	72	42.4	98	57.6	142	83.5	28	16.5
	Delayed	63	37.1	107	62.9	125	73.5	45	26.5
Enhanced[b]	Immediate	88	51.8	82	48.2	148	87.1	22	12.9
	Delayed	80	47.1	90	52.9	133	78.2	37	21.8
Enriched[b]	Immediate	60	35.3	110	64.7	127	74.7	43	25.3
	Delayed	63	37.1	107	62.9	114	67.1	56	32.9

[a]Max score = 135 ($K = 15 \times N = 9$).
[b]Max score = 170 ($K = 5 \times N = 34$).

We modeled CORRECT answers likelihood using a mixed logit model. The addition of TEST TYPE significantly improved the null model ($\chi^2(1) = 432.52$, $p < .001$, $N = 2,580$). Adding CONDITION further improved the model ($\chi^2(3) = 65.56$, $p < .001$, $N = 2,580$). Then, a TEST TYPE × CONDITION interaction was added to the model not leading to any significant improvement ($\chi^2(3) = 4.92$, $p = .18$, $N = 2,580$). Moreover, adding SESSION as a main effect did not improve the model ($\chi^2(1) = 1.70, p = .19, N = 2,580$). However, adding it as an interacting variable with CONDITION improved the model ($\chi^2(4) = 28.14$, $p < .001$, $N = 2,580$) and its interaction with TEST TYPE further improved the model ($\chi^2(1) = 5.30$, $p = .02$, $N = 2,580$). Item-related variables were then added incrementally but none improved the model. Finally, adding the single participant-related variable (overall self-rated proficiency score) improved the model ($\chi^2(1) = 12.65, p < .001, N = 2,580$). Furthermore, this variable interacted with TEST TYPE ($\chi^2(1) = 4.88$ $p = .03$, $N = 2,580$) and SESSION ($\chi^2(1) = 10.76$, $p = .001$, $N = 2,580$), significantly improving the model. As a final step in the analysis, significant predictors were removed backward. Only the TEST TYPE × SESSION interaction was found to be redundant ($\chi^2(1) = 3.28$, $p = .07$, $N = 2,580$) and was thus removed from the model. The final best-fit model is presented in Table 6.

Significant main effects in Table 6 can be summarized as follows: (1) more correct responses were produced in the recognition test (891/1,334) than in the recall test (443/1,334); (2) more correct responses were produced (364,

Table 6 Summary of the best fit mixed logit model for variables predicting nonnatives' correct explicit responses

Predictor			Estimate	SE	Wald-Z	p
(Intercept)			−7.63	1.05	−7.25	<.001
TEST TYPE:	*recognition*		3.87	0.72	5.36	<.001
CONDITION:	*decontextualized*		1.60	0.35	4.54	<.001
	enhanced		2.12	0.36	5.96	<.001
	enriched		1.42	0.35	4.04	<.001
SESSION:	*immediate*		1.42	0.72	1.98	0.047
Overall Proficiency			1.32	0.25	5.37	<.001
CONDITION:	*(decontextualized)* X SESSION: *(immediate)*		1.48	0.31	4.70	<.001
	(enhanced)		1.42	0.32	4.48	<.001
	(enriched)		1.17	0.31	3.76	<.001
TEST TYPE:	*(recognition)*	X *Overall Proficiency*	−0.44	0.18	−2.45	0.01
SESSION:	*(immediate)*	X *Overall Proficiency*	−0.61	0.17	−3.55	<.001

Note. The model has random intercepts for participants and items. $N = 2,580$, log-likelihood $= -1,249$.

449, 402/1,334) under the three experimental conditions (enriched, enhanced, decontextualized, respectively) than those produced by the control group (119/1,334); (3) SESSION led to a significant drop in explicit gains over the 2-week delay with more correct responses in the immediate than the delayed session (680, 654/1,334, respectively); and (4) as the proficiency of L2 learners increased, they produced more correct answers in explicit measures. It should be noted, however, that these main effects were modulated by a number of interactions. First, the TEST TYPE × overall proficiency interaction seems to suggest that the difference between gains achieved in the two explicit tests (recall versus recognition) is modulated by overall proficiency with a larger difference between recall and recognition knowledge as proficiency increases (see Figure 5, left panel). Second, the normal direction of the SESSION effect (showing attrition in knowledge) changed as it interacted with overall proficiency. More proficient L2 learners showed a clear improvement in explicit knowledge in the delayed session (Figure 5, right panel).

Third, because the CONDITION × SESSION interaction was significant, we defined a contrast matrix to explore all possible comparisons for this specific interaction through multiple comparison analysis (see Table 7)[9], and three points are worth noticing: (1) the difference between the control group and the three experimental conditions was present in both sessions; (2) among the three experimental conditions, only one difference was significant in both sessions (i.e., the difference between the enhanced and the enriched conditions); and
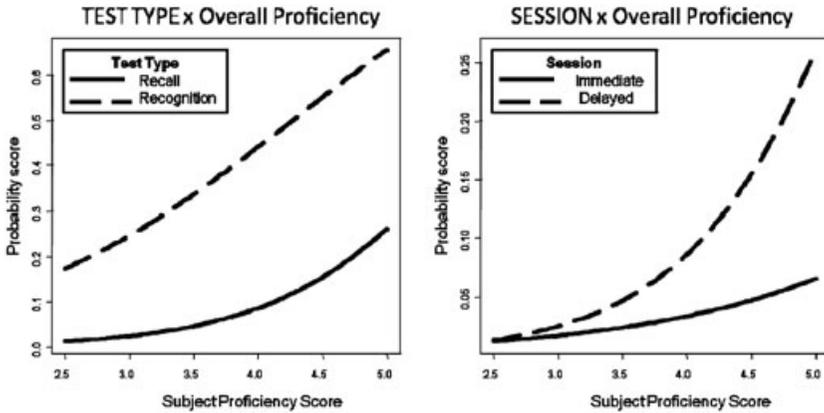
Figure 5 Interaction between proficiency and TEST TYPE (left panel) and between proficiency and SESSION (right panel) in the logit model for the nonnative explicit data (*Note*. The vertical axis depicts the probability of CORRECT answers transformed from log odds).

(3) a significant difference between the immediate and the delayed sessions was only present for the control group (with fewer correct responses in the immediate session). Finally, CONDITION did not interact with TEST TYPE suggesting similar results for both recall and recognition tests.

To conclude, the three treatment conditions led to delayed recall gains of 27–37 percentage points and to delayed recognition gains of 21–32 percentage points over the control group. Also, the enhanced condition led to more correct answers than the enriched condition in the delayed test (10 and 11 percentage point difference in recall and recognition, respectively). These results can be summarized as follows (and are illustrated in Figure 6):

$$\text{Decontextualized} = \text{Enriched} > \text{Control}$$

$$\text{Decontextualized} = \text{Enhanced} > \text{Control}$$

$$\text{Enhanced} > \text{Enriched}$$

Implicit Priming Test

Overall accuracy in the lexical decision task was high (98.53%). Mean reaction time under all conditions is presented in Table 8.

We fit a LME model with log reaction time as the dependent variable and main effects (ITEM TYPE, CONDITION, and SESSION) added incrementally

Table 7 Multiple comparisons for the CONDITION × SESSION interaction (nonnatives' explicit results)

Predictor	*Estimate*	*SE*	*Wald-Z*	*p*
Immediate Session Contrasts				
Decontextualized–Control	2.57	0.35	7.35	<.001
Enhanced–Control	2.95	0.35	8.37	<.001
Enriched–Control	2.18	0.35	6.26	<.001
Enhanced–Decontextualized	0.38	0.18	2.13	0.37
Enriched–Decontextualized	−0.39	0.17	−2.30	0.27
Enriched–Enhanced	−0.77	0.17	−4.39	<.001
Delayed Session Contrasts				
Decontextualized–Control	1.36	0.33	4.10	<.001
Enhanced–Control	1.76	0.33	5.26	<.001
Enriched–Control	1.21	0.33	3.64	<.001
Enhanced–Decontextualized	0.40	0.17	2.32	0.26
Enriched–Decontextualized	−0.15	0.17	−0.90	0.98
Enriched–Enhanced	−0.55	0.17	−3.21	0.02
Immediate Versus Delayed[a]				
Control	−0.83	0.23	−3.64	0.01
Decontextualized	0.38	0.17	2.25	0.30
Enhanced	0.36	0.18	2.05	0.41
Enriched	1.42	0.17	0.85	0.99

[a]With delayed as the reference level.

Table 8 Mean reaction rime (RT) in milliseconds for intact and control items under all conditions (nonnatives)

| Condition	Session	Intact Items		Control Items		Difference in
		RT Mean	*SE*[a]	RT Mean	*SE*[a]	RT Mean
Control	Immediate	711.93	25.81	712.78	28.03	0.85
	Delayed	644.23	24.20	627.32	27.47	−16.91
Decontextualized	Immediate	715.29	22.22	673.01	24.74	−42.28
	Delayed	705.85	28.96	664.99	22.21	−40.86
Enhanced	Immediate	657.58	20.08	761.04	29.20	103.46
	Delayed	677.56	22.59	726.80	29.03	49.24
Enriched	Immediate	783.68	30.50	712.72	25.64	−70.96
	Delayed	688.96	24.86	693.73	24.58	4.77

[a]Standard error of mean.
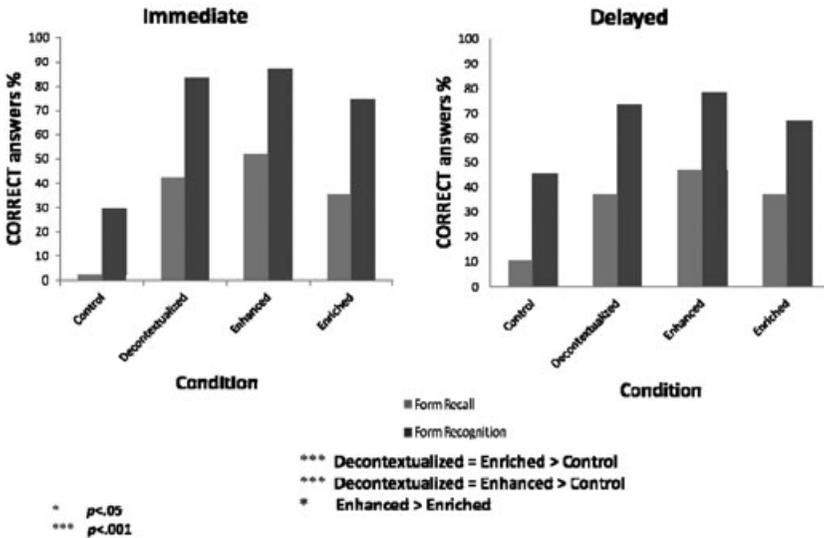
Figure 6 Percentages and comparisons of explicit (form recall and recognition) scores under all conditions in both sessions (nonnatives).

to the null model. Similar to Experiment 1, ITEM TYPE did not improve the null model ($\chi^2(1) = 0.00$, $p = 1.00$, $N = 1,246$), neither did CONDITION ($\chi^2(3) = 6.77$, $p = .08$, $N = 1,246$). However, SESSION did improve the model ($\chi^2(1) = 17.63$, $p < .001$, $N = 1,246$) with shorter (faster) reaction time in the delayed session. Only two-item related variables further improved the model: Word 2 length ($\chi^2(1) = 27.92$, $p < .001$, $N = 1,246$) and Word 2 log frequency ($\chi^2(1) = 13.42$, $p < .001$, $N = 1,246$). The shorter and more frequent the second word, the shorter the log reaction time is. Finally, overall proficiency was added to the model and was found to significantly improve it ($\chi^2(1) = 7.05$, $p = .008$, $N = 1,246$). The more proficient the participant, the shorter his/her log reaction time is. We then tested interactions between significant predictors (SESSION, W2 length, W2 Frequency, and overall proficiency), but none led to any significant improvement in the model. Finally, when removing significant predictors backward, none was found to be redundant. Table 9 presents the final best-fit model.

Similar to Experiment 1, the best fit LME model seems to suggest that nonnatives' RT performance for intact as opposed to control items was not predicted by condition (i.e., no interaction between ITEM TYPE and CONDITION). It can be concluded, thus, that implicit knowledge was not

Table 9 Summary of the best fit LME model for variables predicting nonnatives' log reaction time

Predictor	Estimate	MCMCmean	SE	t	pMCMC	Pr(>|t|)
(Intercept)	6.62	6.61	0.16	40.60	0.001	<.001
SESSION: immediate	0.05	0.05	0.01	4.19	0.001	<.001
Word 2 Length	0.05	0.05	0.00	8.46	0.001	<.001
Word 2 Log Frequency	−0.04	−0.04	0.01	−4.01	0.001	<.001
Overall Proficiency	−0.11	−0.11	0.04	−2.71	0.007	0.007

Note. The model has random intercepts for participants and items. $N = 1,246$, $R^2 = 0.48$. MCMC = Monte Carlo Markov chain; pMCMC = p values estimated by the MCMC chain method using 10,000 simulations; $Pr(>|t|) = p$ values obtained with the t test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom.
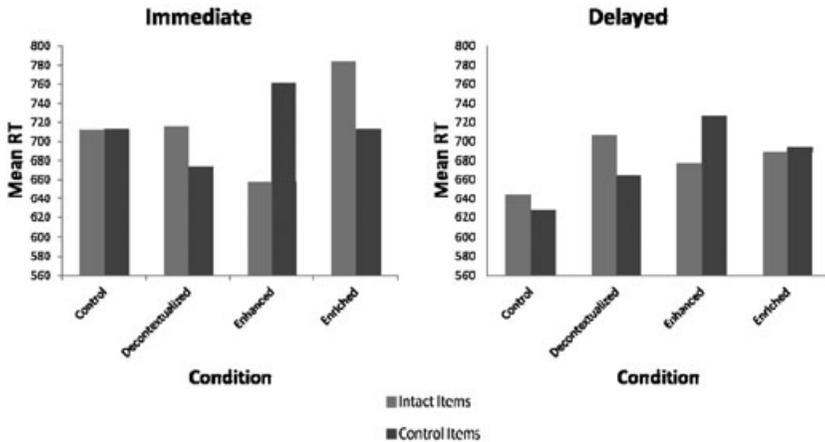

Figure 7 Mean reaction times for intact and control items under all conditions in both sessions (nonnatives).

developed under any condition. The results can be summarized as follows (see Figure 7):

Immediate: Enhanced condition = no effect
Decontextualized condition = no effect
Enriched condition = no effect
Control group = no effect
Delayed: Enhanced condition = no learning
Decontextualized condition = no learning

        Enriched condition = no learning
        Control group = no learning

Discussion

As opposed to native speakers, explicit learning plays a much more prominent role in L2 vocabulary acquisition (e.g., Nation, 2001; Schmitt, 2008), although little is known about the learning of collocations. Experiment 2 explored three learning conditions and found that they all led to significant gains in collocation knowledge at both the form recall and form recognition levels of mastery. This was true for the immediate test, but more importantly, it was also the case of the delayed test. Therefore, all three of the treatments were effective in facilitating durable learning. Moreover, the enhanced condition was more effective than the enriched condition in enhancing both types of explicit knowledge (though not more effective than the decontextualized condition). Finally, recognition gains were always higher than recall gains and this difference increased as the proficiency of the L2 learners increased.

What about implicit knowledge? The immediate and delayed priming tests showed that there was no improvement of this knowledge. It seems that the types and amounts of treatment exposure in this experiment were not sufficient to establish long-term implicit knowledge, at least not that the priming paradigm could capture. Overall, although the treatments all facilitated the learning of explicit collocational knowledge, they did not facilitate measurable learning of implicit knowledge.

General Discussion

The Relationship Between Explicit and Implicit Lexical Knowledge

Previous research has shown that direct conditions are generally more effective than indirect conditions in promoting lexical learning, but it must be said that this research has a number of important limitations: (1) the lexical items researched have almost always been individual words, (2) the aspect of word knowledge addressed has almost always been the form-meaning link, and (3) the research measurements have almost always been explicit tests. It is therefore perhaps not surprising that direct approaches have usually been shown to have an advantage, because the form-meaning link of individual words is relatively amenable to explicit introspection, and the explicit measurements have privileged explicit knowledge. Thus, it is unclear whether nonexplicit indirect learning conditions can be as, or more, effective in promoting implicit knowledge, simply because implicit knowledge has seldom been tested before.

Our study design allows such a comparison between direct/indirect input and explicit/implicit knowledge. The high-proficiency educated natives benefited from both indirect and direct (enhanced and decontextualized) exposure to collocations, in terms of enhancing explicit (recall and recognition) knowledge. Moreover, the decontextualized condition showed a clear advantage over the enriched condition in developing both recall and recognition knowledge although it did not hold this advantage over the other direct (enhanced) condition. These results are in line with previous research which shows that explicit input best leads to explicit knowledge. But does explicit input still hold this advantage when it comes to enhancing implicit knowledge? The native results suggest that it does not. None of the three treatment conditions led to significant implicit knowledge gains. Thus, with very proficient speakers, it seems that the rather small amounts of input in our various treatment conditions may lead to explicit lexical knowledge, but not implicit priming effects.

What about advanced nonnatives? First, similar to natives, all of the three treatments led to durable form recognition and form recall collocation learning. Second, unlike natives, the enhanced condition seems to lead to higher explicit gains than the enriched condition although it did not hold this advantage over the decontextualized condition. Also, the decontextualized condition did not lead to any higher gains than the enriched condition. This is incongruent with research with individual words, where more direct treatments almost always hold an advantage. However, it fits with Bishop's (2004) finding that formulaic sequences that were enhanced (underlined and in red font) were clicked upon more often for glosses than the same sequences when appearing in plain font. This certainly suggests the benefits of making formulaic language more salient in texts for L2 learners through typographical enhancement. Third, nonnatives did not develop any durable implicit memory traces of collocations.

In both experiments, there was a clear dissociation between explicit and implicit lexical knowledge, and it looks as if one does not imply the other. Implicit knowledge (as measured using the priming paradigm) seems to be harder to enhance than explicit knowledge not only for nonnative speakers, but also for native users of the language. The fact that implicit knowledge was harder to enhance than explicit knowledge seems to contradict recent ERP evidence from grammar (Osterhout et al., 2006; Tokowicz & MacWhinney, 2005) and vocabulary (J. McLaughlin et al., 2004) research where implicit knowledge was found to precede explicit knowledge. It should be noted, however, that the automatic priming paradigm might not be sensitive enough in capturing the very initial implicit memory traces built after such a short exposure period. This short treatment period (with no built-in recycling) may have disadvantaged

the learning of implicit knowledge more than that of explicit knowledge. That is, it may be that recycling is more important for the learning of implicit knowledge than it is for explicit knowledge. R. Ellis (2007) observes that studies that did show growth in implicit grammatical knowledge involved prolonged and substantial instruction.

Natives Versus Nonnatives: Absolute Learning of Explicit Collocation Knowledge

Hoey (2005) suggests that native speakers and nonnative learners of a language do not differ in how they acquire collocations, and that it mainly depends on the type and amount of input they receive. It is thus interesting to compare the native results with the advanced-proficiency nonnatives to explore this claim. As neither group developed implicit collocational knowledge, we will only discuss results of the two explicit measures. Both natives and nonnatives showed clear evidence of durable learning of explicit knowledge under all treatment conditions with one direct condition (decontextualized for natives and enhanced for nonnatives) showing an advantage over the indirect, enriched, condition. Thus, our results seem to support Hoey's (2005) claim. Adult natives and nonnatives do not seem to gain greatly different advantages from direct versus indirect input when acquiring collocations: both natives and nonnatives benefited more from direct types of input. This result is also in line with Spada and Tomita's (2010) conclusion that direct instruction leads to higher grammatical explicit knowledge gains than indirect instruction. It seems that explicit knowledge gains after treatment is one area where vocabulary and grammar findings overlap.

It is also useful to consider the amount of absolute long-term (delayed) learning that accrued from the input conditions. After the focused instruction (decontextualized) treatment, natives were able to recognize the collocations form about 37 percentage points better than the nonstudy control condition; after the enhanced treatment, about 34 percentage points better; and after the enriched treatment, about 30 percentage points better. Typically form recall mastery is more difficult to achieve, but even here, the decontextualized treatment led to about 19 percentage point gains and the enhanced/enriched conditions led to about 16 percentage point gains. One might expect that the nonnatives would have less uptake from the treatments. This was the case for the easier form recognition test (about 21–32 percentage points better than the controls), but not for the higher-level form recall knowledge (about 27–37 percentage points better than the controls). This result may seem surprising as both natives and nonnatives were exposed to the same amount of input. One might speculate

that while natives tend to focus mainly on the communicative meaning, non-natives inevitably focus relatively more on the language form, and this may be a factor in enhancing their learning of collocational word combinations from input (we did not test the learning of the collocation's meaning) at least on the explicit side. This conclusion is sustained by the finding that, in comparison with the enriched condition, natives benefited more from decontextualized exposure where their attention was explicitly focused on the collocation's form, while advanced nonnatives benefited more from having their attention focused on the collocations through contextualized, salient exposure.

Limitations and Suggestions for Future Research

Inevitably, there are a number of limitations that need to be noted in this study. First, it is possible that some participants had some previous knowledge of the collocations targeted in this study. However, although we did not administer a pretest in order not to direct participants' attention to these collocations, our piloting indicated that learners similar to our participants had no previous knowledge of the specialist medical collocations. Second, all participants took all tests in the same order and in a short period of time. In spite of the procedures taken to minimize the effect of earlier tests on later ones, the more proficient L2 learners in the control group showed a significant improvement in the delayed explicit tests. This might be explained as a direct effect from the priming test to the explicit tests. It should be noted, however, that despite this potential undesirable effect of the priming test, the statistical analysis dealt carefully with that limitation as only gains that were significantly higher than those achieved by the control group were considered to be a reliable evidence of learning. Third, the study used a relatively small number of stimuli, but the use of the priming test made this unavoidable, as it was necessary to find collocations for which suitable control counterparts (controlled for frequency, length, and semantic plausibility) could be found. Finally, many factors known to affect L2 learning were not considered in the present study (e.g., time on task, repetition, and recycling). The short, massed treatment, for example, might have put the indirect approach (enriched condition) at a disadvantage because unenhanced exposure is said to aid development in knowledge gradually over time (N. Ellis, 2002; R. Ellis, 2009; Spada & Tomita, 2010). Thus, results of the present study should be interpreted carefully and may not yet have clear implications for language teaching practice.

While the area of vocabulary studies is now establishing an understanding of the acquisition of explicit knowledge (e.g., Schmitt, 2008), it is only

just beginning to consider how to facilitate fluent and automatic (i.e., implicit) knowledge. Future research should attempt to develop and validate other tests of implicit lexical knowledge (eye tracking, ERP, among other behavioral, free constructed, tasks) to inform the explicit/implicit debate. Another interesting line for future research concerns how beginning and intermediate English-as-a-foreign-language learners (as opposed to the advanced ESL learners investigated here) develop implicit, in addition to explicit, collocational knowledge in language classrooms. Our results, along with Webb and Kagimoto (2009), have shown proficiency as a major predictor of performance on explicit tests of collocational knowledge. It is interesting to see what would happen when implicit collocational knowledge is assessed.

The nature of explicit and implicit lexical knowledge and the interrelationship between them is an emerging lexical issue that is likely to attract increased attention in the future. This study is a start in exploring this topic, but can only be considered a first step. It will surely raise more questions than it has answered, but hopefully it has at least begun the discussion.

Revised version accepted 20 December 2011

Notes

1 (i) *Blakiston's Gould medical dictionary* (Gould & Gennaro, 1979), (ii) *Collins dictionary [of] medicine* (Youngson, 2004), (iii) *Mosby's medical, nursing, and allied health dictionary* (Anderson, Keith, & Novak, 2002), (iv) *Stedman's medical terms and phrases* (Stedman, 2004), and (v) *Stedman's medical dictionary for the health professions and nursing* (Stedman, 2005).

2 We tried our best to achieve McNamara's (2005) recommended ratios for automatic priming. Relatedness proportion (RP) was close to the 0.20 value (the number of related trials was 7 or 8 out of all 37 word-word trials (7 or 8/37 ≈ 0.20). As for nonword ratio (NR), there were 31 nonword trials out of all the 60 or 61 unrelated trials resulting in the acceptable NR of ≈ 0.50 (31/60 or 61).

3 When categorical variables are entered as predictors in a logit mixed model or a LME model, the levels are ordered alphabetically and the first is treated as the reference level (e.g., control is the reference level for the CONDITION variable here). Thus, when any level is found to be significant, this would mean that it is significant in comparison to the reference level.

4 For the implicit, priming test, responses shorter than 150 ms or longer than 1300 ms (constituting only 2.15% of all data points) were excluded from further analysis. The long cut-off time in the present priming task is far higher than that used in previous research. However, it should be noted that words used in the present lexical decision

task are rather low-frequency (see Appendix S1 in the Supporting Information) and would be processed slower than those included in previous research.

5 Because immediate tests can only inform about short-term effects, we prefer to interpret immediate gains as short-term enhancement effects. We reserve the notion of "learning" for gains on delayed tests, as only they can demonstrate acquisition that is durable (see Schmitt, 2010, for more on this).

6 For both experiments, only correct responses were included in the final reaction time analysis.

7 The mean reading-comprehension score gained by the 34 participants in the experimental group was acceptable. All participants scored higher than 60% in this test (*Mean* = 6.88 out of 8.00, 86.03%, *SD* = 0.84).

8 Only 1.09% of the data points coded as CORRECT in the explicit recall test represent partial, rather than complete, knowledge in Experiment 2.

9 This analysis was run in a model including the CONDITION × SESSION interaction only (excluding all other significant predictors) in order to catch all possible contrasts for this interaction.

References

Aitchison, J. (2003). *Words in the mind: An introduction to the mental lexicon* (3rd ed.). Oxford, UK: Blackwell.

Anderson, D. M., Keith, J., & Novak, P. D. (2002). *Mosby's medical, nursing, and allied health dictionary* (6th ed.). London: Mosby.

Andringa, S., de Glopper, K., & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning*, *61*, 868–903.

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

Bialystok, E. (1981). Some evidence for the integrity and interaction of two knowledge sources. In R. Anderson (Ed.), *New dimensions in research on the acquisition and use of a second language* (pp. 62–74). Rowley, MA: Newbury House.

Bishop, H. (2004). Noticing formulaic sequences—A problem of measuring the subjective. *LSO Working Papers in Linguistics*, *4*, 15–19.

Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple comparisons using R*. Boca Raton, FL: Chapman & Hall/CRC.

Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 581–591.

Davies, M. (2007). *VIEW: Variation in English words and phrases*. Retrieved July 22, 2011, from http://corpus.byu.edu/bnc/

de Graaff, R. (1997). The eXperanto experiments: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*, 249–276.

DeKeyser, R. (1994). Implicit and explicit learning of L2 grammar: A pilot study. *TESOL Quarterly*, *28*, 188–194.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Malden, MA: Blackwell.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*, 125–155.

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

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*, 367–413.

Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit. *European Journal of Cognitive Psychology*, *5*, 289–318.

Ellis, N. C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*, *11*, 37–56.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N. C. (2008). Implicit and explicit knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.). *Volume 6: Knowledge about language* (pp. 119–132). New York: Springer.

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). Psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In U. Römer & R. Schulze (Eds.), *Exploring the lexis-grammar interface* (pp. 89–114). Amsterdam: John Benjamins.

Ellis, N. C., & Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 89–113). Mahwah, NJ: Erlbaum.

Ellis, R. (1994). Implicit/explicit knowledge and language pedagogy. *TESOL Quarterly*, *28*, 166–172.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition*, *24*, 223–236.

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

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, *40*, 83–107.

Ellis, R. (2007). Explicit form-focused instruction and second language acquisition. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 437–455). Malden, MA: Wiley-Blackwell.

Ellis, R. (2009). Implicit and explicit learning, knowledge, and instruction. In R. Ellis et al. (Eds.), *Explicit and implicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters.

Gagné, C. L., & Spalding, T. L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language*, *60*, 20–35.

Gould, G. M., & Gennaro, A. R. (1979). *Blakiston's Gould medical dictionary: A modern comprehensive dictionary of the terms used in all branches of medicine and allied sciences, with illustrations and tables* (4th ed.). New York/London: McGraw-Hill.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford, UK: Oxford University Press.

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

Hulstijn, J. H. (2007). Psycholinguistic perspectives on language and its acquisition. In J. Cummins & C. Davison (Eds.), *The international handbook of English language teaching* (pp. 783–796). Norwell, MA: Springer.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? *AILA Review*, *11*, 97–112.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. London: Pergamon Press.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*, 694–716.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Long, M. H. (1983). Does second language instruction make a difference? A review of the research. *TESOL Quarterly*, *17*, 359–382.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 407–452). Oxford, UK: Oxford University Press.

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.

McLaughlin, B., Rossman, T., & McLeod, B. (1983). Second language learning: An information-processing perspective. *Language Learning*, *33*, 135–158.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.

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

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

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

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56, Supplement 1*, 199–230.

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

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 Nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.

Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.

Reinders, H., & Ellis, R. (2009). The effects of two types of input on intake and the acquisition of implicit and explicit knowledge. In R. Ellis et al. (Eds.), *Explicit and implicit knowledge in second language learning, testing and teaching* (pp. 281–302). Bristol, UK: Multilingual Matters.

Rott, S. (2009). The effect of awareness-raising on the use of formulaic constructions. In R. Corrigan, E. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic language* (pp. 405–422). Amsterdam: John Benjamins.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Press.

Schmitt, N., Dornyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 55–86). Amsterdam: John Benjamins.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*, 429–458.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta analysis. *Language Learning*, *60*, 263–308.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning & Memory*, *82*, 171–177.

Stedman, T. L. (2004). *Stedman's medical terms and phrases*. London: Lippincott Williams & Wilkins.

Stedman, T. L. (2005). *Stedman's medical dictionary for the health professions and nursing* (Illustrated 5th ed.). London: Lippincott Williams & Wilkins.

Tiersky, E., & Tiersky, M. (1992). *The language of medicine in English* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall Regents.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*, 173–204.

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

Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, *43*, 55–77.

West, M. (1953). *A general service list of English words*. London: Longman.

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in a second language. *Language Awareness*, *11*, 259–290.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, *32*, 430–449.

Youngson, R. M. (2004). *Collins dictionary [of] medicine* (3rd ed.). Glasgow: HarperCollins.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Intact Medical Collocations and their Controls (Experiment 1).
Appendix S2. Intact Medical Collocations and their Controls (Experiment 2).
Appendix S3. Mixed-Effects Modeling Procedures in the Study.