Contextual Word Learning with Form-Focused and Meaning-Focused Elaboration

IRINA ELGORT, SARAH CANDRY, THOMAS J. BOUTORWICK, JUNE EYCKMANS, and MARC BRYSBAERT

To be published in *Applied Linguistics*

E-mail: irina.elgort@vuw.ac.nz

Abstract

Contextual L2 word learning may be facilitated by increasing readers' engagement with form and meaning of novel words. In the present study, two adult L2 populations, Chinese and Dutch speakers, read English sentences that contained novel vocabulary. These contextual exposures were accompanied either by form-focused elaboration (i.e., word-writing) or by meaning-focused elaboration (i.e., actively deriving word meaning from context). Immediate and delayed offline and online measures of word knowledge showed superior learning outcomes for the word-writing treatment. This finding is aligned with the predictions of the lexical quality hypothesis (Perfetti & Hart, 2002), highlighting the added value of more precise encoding of a word's form, in addition to learning its meaning. The key pedagogical implication of this study is that a simple act of copying novel words, while processing meaningful L2 input, may significantly boost quality of lexical knowledge.

Introduction

Contextual word learning in a second language may be facilitated by using deliberate learning procedures that direct readers' attention to the word (Laufer, 2005; Schmitt, 2008). The present study compares outcomes of two such procedures: (1) explicitly inferring word meanings from sentence context (meaning-focused elaboration) and (2) word-writing (form-focused elaboration). While inferring word meanings from context is common practice endorsed in the language education literature (L1: McKeown, 1985; L2: Nation & Webb, 2011; Schmitt, 2008), writing down an unfamiliar word is less common. Learning studies carried out using word-writing report mixed results, ranging from negative (Barcroft, 2006, 2007; Xu, Chang, Zhang, & Perfetti, 2013) to positive effects (Bogaards, 2001; Thomas & Dieter, 1987).

Contextual word learning (i.e., mapping a novel word form to meaning, as a result of encountering it in a meaningful context) is particularly challenging when the L2 writing system or orthography is markedly different from those of the learner's native language (L1). This is because visual word recognition and processing in an orthographically distant L2 is very challenging (Hamada & Koda, 2008; Koda, 1997). For this reason, deliberate learning procedures may be particularly beneficial for English language learners whose L1 is not alphabetic. The present study investigates the effects of meaning-focused and form-focused elaboration on contextual word learning with two learner groups: speakers of Chinese (an orthographically distant, non-alphabetic L1) and speakers of Dutch (an orthographically close, alphabetic L1).

Contextual word learning

According to the instance-based framework of word learning (Bolger, Balass, Landen, & Perfetti, 2008; Reichle & Perfetti, 2003), initial encounters with a new word in meaningful sentence contexts leave episodic traces of the learning event as a whole (i.e., word + context). After multiple contextual encounters with the same word, aspects of knowledge that are common to all encounters are reinforced (e.g., correctly inferred meaning senses, co-occurring words), while all other aspects are discarded, eventually leading to the development of a robust lexical-semantic representation of the new word (i.e., a stable form-meaning mapping).

Form-focused and meaning-focused elaboration in L2 studies

In an overview of L2 vocabulary learning studies, Schmitt (2008: 329) reports that "the overriding principle for maximising vocabulary learning is to increase the amount of engagement learners have with lexical items". At minimum, inferring meanings from context is described as a prerequisite for contextual word learning through reading (Nation & Webb, 2011). Other meaning-focused treatments advocated in conjunction with contextual word learning include the use of glosses and dictionaries, which tend to produce better contextual word learning than reading alone (Hulstijn, Hollander, & Greidanus, 1996; Knight, 1994; Luppescu & Day, 1993). Importantly, it is the combination of deriving word meanings from context and using additional meaning-focused treatments that create most favourable word leaning conditions (Ko, 1995). Fraser (1999), for example, observed that the combination of using contextual inferences and consulting a dictionary produced a 50% rate of word meaning recall, while only 30% recall was observed for each of the two approaches when used independently. Post-reading meaning-focused tasks that strengthen the weakly-established representations of contextually-learned words also improve meaning recall (Hill & Laufer, 2003; Mondria, 2003).

Schmitt (2008) also argues for attention to form in novel word learning, because the knowledge of form facilitates the learning of new meaning senses of L2 words

(Bogaards, 2001). Word-writing is an example of directing attention to form. Positive effects of word-writing (copying) on the English-speaking learners' ability to recall newly-learned L2 (French) words from English translations and in a free recall task were reported by Thomas and Dieter (1987; Experiments 1 & 3), after 3 cycles of paired-associate learning, with the L2 word presented with its L1 translation. Thomas and Dieter argued that when learning a word in a foreign language, copying brings learners' attention to the word form and structure, facilitating the creation of a more accurate lexical representation of the word in memory.

Schmitt (2008: 336) however warns against using form-focused learning as "an add-on" to other tasks citing Barcroft's (2002) limited processing resources proposition, according to which structural elaboration (such as word-writing) facilitates the learning of formal properties of a word and inhibits the encoding of its semantic properties, and vice versa, when processing demands are "sufficiently high". Barcroft (2006, 2007) reported that word-writing (copying) had a negative effect on the mapping of visual word form to its meaning in a word-picture association learning task with English-speaking learners of Spanish. However, the non-writing condition (i.e., word-picture rehearsal only) may have been privileged in these studies because it allotted more time to the rehearsal of the form-meaning connection measured in the posttests, while making the word form salient by presenting it out of context below the picture. Barcroft's claim that writing uses up the processing resources needed to create form-meaning associations are aligned with the findings reported in a Chinese L2 study with English-speaking learners (Xu et al., 2013). This study involved comparison of character writing and animation with reading-only as a means of character learning. In their study, writing resulted in a learning condition induced trade-off, i.e., better form recognition and cued recall but poorer meaning recall compared to the reading-only condition.

In the reviewed studies the target L2 lexical items appear in isolation in the learning phase, either in a word-picture association task (Barcroft, 2006, 2007) or with L1 translation equivalents (Thomas & Dieter, 1987; Xu et al., 2013). There is no L2 research (to our knowledge) that investigates the effect of word-writing on contextual word learning, while much attention has been devoted to meaning-focused treatments. To address this research gap, the present study is specifically designed to investigate whether word-writing (a deliberate learning procedure) facilitates contextual learning of L2 words to the same extent as deliberate focus on meaning. If the "resource depletion" (Barcroft, 2006) and the "learning condition trade-off" (Xu et al., 2013) hypotheses hold for contextual word learning, writing novel vocabulary in addition to its contextual learning may negatively affect form-meaning mapping.

Lexical Quality Hypothesis

Our prediction is, however, that fine-tuning the knowledge of form through word-writing will result in an overall improvement of lexical knowledge in contextual word learning. This prediction is based on the lexical quality hypothesis (Perfetti & Hart, 2002; Perfetti, 2007). Lexical quality refers to "the extent to which a mental representation of a word specifies its form and meaning components in a way that is both precise and flexible" (Perfetti, 2007: 359). High-quality lexical representations minimise confusion about word form and meaning as a result of fully specified component representations (e.g., orthographic representations that allow readers to distinguish between *pretty* and *petty*). Lexical quality also depends on the binding of orthographic, phonological, and semantic sources of information. A lexical representation can be described as stable if "word identity is reliably retrieved from an orthographic or phonological input" (Perfetti, 2007: 360). Lexical quality varies depending on the quantity and quality of the individual language user's experiences with a given word (e.g., for a builder the word *facia* will have a more stable lexical representation than for an average native English speaker).

An important consequence of lexical quality is that a reliable and coherent representation of a word is retrieved easily and consistently, during offline and online processing[i]. This hypothesised consequence of lexical quality can be empirically tested. For example, the learner's offline ability to reliably retrieve a word's representation from a visual (orthographic) or auditory (phonological) input can be tested using dictation or meaning generation tasks, while its online processing and retrieval can be tested using the lexical decision task, in which participants are instructed to make word/non-word decisions under time pressure. Higher accuracy of lexical decisions would indicate more precise lexical representations, while faster responses would indicate higher ease/fluency of lexical processing.

Present study

The present study was conducted first with intermediate proficiency Chinese-speaking learners of English, and repeated with intermediate-to-high proficiency Dutch-speaking participants. Participants encountered novel L2 vocabulary items in informative proficiency-appropriate sentence contexts, and were instructed to perform either a word-writing or a meaning generation task with these items. Diverse informative contexts (i.e., contexts that support correct word meaning inferences) create favourable conditions for learning the meanings of novel words (Bolger et al., 2008). Deliberately deriving word meaning from context may be expected to further facilitate semantic encoding of novel words, as a result of additional conscious, effortful processes (McKeown, 1985). Word-writing practice, on the other hand, is predicted to result in more precise orthographic knowledge of these words, facilitating the quality of their lexical representations.

In line with the lexical quality hypothesis, we argue that precise knowledge of form (formal representation) is a key component of high-quality lexical knowledge. Adding word-writing to contextual learning enables learners to create a more precise orthographic representation of a novel word, over and above any initial lexical-semantic information gained as a by-product of its co-occurrence with other words in informative contexts (Kwantes, 2005; Landauer & Dumais, 1997; Reichle & Perfetti, 2003). Thus fine-tuning orthographic knowledge in conjunction with contextual word learning is predicted to facilitate form-meaning mapping, i.e., the binding of orthographic and semantic information. Deliberately deriving word meaning from context, on the other hand, may result in a better understanding of the meaning of a novel word (in informative contexts), but its contribution to form-meaning mapping (i.e., integration between formal and semantic representations) is predicated on the existence of a reasonably stable formal-lexical representation.

The following research question is posed in the present study: *Is word-writing a more beneficial additional form of engagement with unfamiliar L2 words in contextual word learning than a conscious effort to infer their meaning?* Specifically, the study investigates whether word-writing leads to a more accurate offline retrieval of word form and form-meaning mapping than deriving word meaning from context. It also evaluates whether word-writing boosts accuracy and fluency of visual processing of contextually-learned L2 words, compared to meaning-focused elaboration. Three post-tasks are used to address these questions: a dictation is used to evaluate productive retrieval of orthographic form (spelling), a decontextualised meaning generation task is used to probe form-meaning mapping (retrieval of meaning from form), and a speeded lexical decision task is used to examine accuracy and fluency of access to lexical representations of the newly-learned items under time pressure.

Study One: Chinese Speakers

Methodology

Participants

Study participants were 47 Chinese students (36 females), either pre-degree or in the first year of an undergraduate degree at a New Zealand university. Volunteers were accepted into the study if they had lived in an English speaking country for less than 12 months prior to the study, and if their *International English Language Testing System* (IELTS) scores were between 5.5 and 7.0 (intermediate to high intermediate)[ii].

Table 1: Study 1. Participants (L1 = Chinese)

Attribute	Mean	SD	Range
Age	24.5	4.0	19–36
Age of L2 (English) acquisition	11.8	2.9	6–20
Length of stay in English-speaking countries (months)	3.5	2.7	0.25–12
IELTS score	6.2	0.5	5.5–7.0
LexTALE score (receptive knowledge)	44.4	6.3	30.0–66.3
PVLT2k[a] (productive vocabulary knowledge)	13.5 (=75 per cent)	2.4	7–17
PVLT5k (productive vocabulary knowledge)	4.7 (=26 per cent)	2.0	1–12
O-Span score (working memory)	4.40	1.14	1–6

[a]The knowledge of productive L2 (English) vocabulary was measured at the 2,000 (PVLT2k) and 5,000 (PVLT5k) word frequency levels (see Laufer and Nation 1999: 46–51).

Learning materials

Critical items

Forty-eight critical vocabulary items (5-7 letters long) were used as learning targets in this study (hereafter, critical items). They were 24 low-frequency words (e.g., *egress*, *ramekin*) and 24 nonwords (e.g., *spiler, banity*) constructed by changing one letter in a real English word. The use of nonwords ensured that at least half of the critical items were unknown to all study participants. This was important in the absence of a vocabulary pretest[iii]. Moreover, low-frequency words often have irregular spelling and orthography-to-phonology mapping (because many of them are borrowings from other languages); while orthographically-legal nonwords are more representative of English orthography. On the other hand, the use of low-frequency real English words ensured that the participants gained some real value from the study.

The meaning of the critical items was related to one of two themes: cooking/food or building/housing (Appendix A). The two semantic fields of cooking and building were selected because they contain highly specialised words unlikely to be familiar to the participants, but good conceptual knowledge of these topics can be generally assumed. This is predicted to facilitate vocabulary learning because novel words can be grouped into thematic clusters that fit into the learners' existing schemata (Brewer & Nakamura, 1984; Mezynski, 1983; Tinkham, 1997). This means that learners can compare and contrast novel vocabulary with the words they already know, facilitating their integration into their existing semantic networks. To control for lexical characteristics that affect word learning, all critical items were concrete nouns.

Because the number of orthographic neighbours of a word form (i.e., words that can be created by changing a single letter in a given word, such as *function–junction*) affect word learning and processing (Andrews, 1997; Davis, 2012; Marinus, Nation, & de

Jong, 2015), the orthographic neighbourhood size was included in the data analyses as covariate (Table 2). Another item variable that was included in the analysis of the lexical decisions data was number of letters, because word length (in letters) may affect lexical decision times (Chumbley & Balota, 1984), especially for low-frequency words (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004).

Table 2: Critical vocabulary items

Attribute	Mean	SD	Range
Critical items (non-words)			
Frequency of base word (Subtlex-UK, Zipf scale)[a]	3.78	0.52	2.93–4.74
Number of orthographic neighbours (OrthN)[b]	1.58	0.83	1–4
Critical items (low-frequency words)			
Frequency (Subtlex-UK, Zipf scale)[c]	2.53	0.59	1.39–3.29
Number of orthographic neighbours (OrthN)	1.46	1.64	0–5

[a]Zipf scale frequency obtained from http://zipf.ugent.be/open-lexicons/interfaces/subtlex-uk/. Values 1–3 are low-frequency words, 4–7 are high-frequency words (Van Heuven *et al.* 2014).
[b]MCWord: An On-Line Orthographic Database of the English Language (Medler and Binder 2005).
[c]Note: Three items were not in the SUBTLEX-UK database.

Learning contexts

In the learning phase, each critical item was presented (in curly brackets) in three informative sentence contexts [e.g., (1) *A floor-to-ceiling door makes {egress} easy*. (2) *The mouse jumped down to the floor and ran around the room, trying to find an {egress}*. (3) *Beside the bed was a trap-door that permitted {egress} to the floor below*.]. The sentences were selected from the Corpus of Contemporary American English (corpus.byu.edu/coca). A small number of low-frequency words in the original sentences were replaced with higher frequency synonyms to facilitate reading comprehension and thus contextual word learning. In the resulting sentences, 91% of the words were within the first five thousand word frequency bands (Nation, 2006).

Tests and measures

Immediate measures

Post-tests of knowledge of form (dictation) and meaning (meaning generation) were conducted on the same day as the learning procedure. The tests were delivered using *Qualtrics* software (www.qualtrics.com). Participants were instructed to listen to the recordings of the critical items (presented in a random order), then type the item into one response field and its meaning in English into another field. A binary scoring (1-correct; 0-incorrect) was used. A score of 1 on the knowledge of form measure was assigned if the word was spelled correctly. On the knowledge of meaning measure, a score of 1 was given if the provided meaning was broadly aligned with the correct meaning of the critical item (e.g., for *egress*, a score of 1 was given in the following cases: "a way out such as small door or window", "a small way out", "entrance", "a small door or window", "access way", "a way of escape", "exit", "a small place to let thing go through", "doorway"). A score of 0 was assigned if there was no answer, or it was incorrect (e.g., for *egress*, "a kind of food", "a tool to the garden", "a container for individuals", "a plant"). Responses were scored by one of the authors and independently verified by another suitably qualified scorer (97% inter-rater agreement).

Delayed measures

The speeded visual lexical decision task was administered one day after the learning procedure. This task included 48 critical vocabulary items (from the learning phase), 48 higher-frequency English words (*Zipf BNC frequency* = 4.2; *SD* = 0.28) and 96 phonologically and orthographically legal nonwords (same length as the critical items) that had not been encountered by the participants prior to the lexical decision task. The processing of the contextually learned critical items was compared with that of the unfamiliar nonwords and higher-frequency L2 words. The effect of learning condition

(WW vs. ME) was examined in separate analyses of response accuracy and response time (RT) to the critical items.

Each lexical decision trial began with a plus sign presented in the middle of the screen for 300 ms. It was immediately replaced by the stimulus presented for 500 ms, which was replaced by a blank screen that stayed up until a response (button press) or the deadline of 3000 ms. The inter-trial interval was 200 ms. RTs were recorded from the stimulus onset. Participants were instructed to decide as quickly and as accurately as possible whether the visually presented stimulus was an English word. A practice block of 16 trials was used at the start of the procedure.

Additional measures of individual differences

Because L2 proficiency is an important predictor of contextual vocabulary learning (Author, 2015) and only limited control over participants' proficiency was possible at recruitment, their L2 lexical proficiency was further estimated using the following published instruments: LexTALE (Lemhöfer & Broersma, 2012 www.lextale.com) was used as a measure of receptive vocabulary knowledge, and Laufer and Nation's (1999) vocabulary levels test of controlled-productive ability (PVLT) was used to measure their productive vocabulary knowledge (Table 1). PVLT was measured at the 2,000 and 5,000 word frequency levels, and the average score was used in the data analyses.

Furthermore, because larger working memory tends to positively correlate with word learning in L1 (Cain, Lemmon, & Oakhill, 2004; Daneman & Green, 1986) and L2 (Juffs & Harrington, 2011), and because both word-writing and meaning deliberation may consume the limited processing resources needed to create form-meaning associations (Barcroft, 2006), participants' working memory was measured using an Operation Span (O-Span) task (Turner & Engle, 1989). Individual L2 vocabulary scores and working memory (O-Span) scores were included as covariates in the data analyses of the immediate and delayed tests.

Procedure

The learning procedure and lexical decision task were programmed and carried out using E-Prime software (Psychological Software Tools, Inc., Pittsburgh, PA). In the learning procedure, 48 critical items were presented in groups (blocks) of 12. Within each block the critical items were presented in sentential contexts one at a time, first in a familiarisation treatment, and then in two 'reading + elaboration' treatments. The lag between each presentation of the same critical item was 11 sentences. All 12 items were then presented with definitions, on the same screen, for meaning-verification. Overall, each critical item was presented four times; previous studies demonstrate that, in word learning from informative sentence contexts, three to four exposures are sufficient for some aspects of word knowledge to be established (Author, 2015; Bolger et al., 2008; Mestres-Missé, Rodriguez-Fornells & Munte, 2007). Participants took a 5-minute break after two learning blocks (24 items). They could also take short breaks after each block, and between cycles within blocks.

Familiarisation treatment

In the familiarisation treatment, sentences were presented for 60 seconds. The participants were instructed to read the sentence for meaning, paying attention to the word in brackets, and to press a designated key on the keyboard to listen to an audio-recording of that word. They were expected to gain some initial orthographic, phonological and semantic knowledge of the critical items in the first presentation (Chalmers & Burt, 2008; Share, 1995).

Learning conditions

In the second and third sentences, critical items were presented in one of the following two conditions: contextual learning with word-writing practice (WW), or contextual learning with explicit meaning-focused elaboration (ME). The participants were allotted 30 seconds per item. In the WW condition, the participants first read the whole sentence for meaning on the computer screen, then copied the critical item into a booklet and continued writing it until they heard a signal that marked the presentation of the next sentence (with a different item). Participants had to turn the page in the booklet and repeat the procedure. In the ME condition, participants first read the whole sentence for meaning, then typed an inferred meaning of the critical item into a text input box located below the sentence, on the same screen. The learning conditions were counterbalanced, so that each participant learned half of the critical items in the WW condition and half in the ME condition; across all learning trials, the critical items were learned in both conditions equal number of times. Participants were alternately assigned to the WW-first or ME-first learning set on their arrival to the learning session.

Definitions

In the final meaning-verification presentation, the items were accompanied by short definitions in English [e.g., *egress - a way out, such as a window or a small door required in every bedroom and basement*]. The participants were instructed to read the definitions and review their understanding of the items. Short definitions (that may be considered a special form of contextual exposure, Bolger et al., 2008) were used to ensure that participants had equal opportunities to encode the correct core meaning of the critical items, reducing the effect of incorrect contextual guessing for individual learners (Kelly, 1990; Mondria, 2003). Presenting dictionary-type definitions in post-initial encounters was considered to be environmentally plausible (Fraser, 1999)[iv].

Post-tests

After the learning procedure, participants completed an O-Span task (which also served as an intervening task), followed by the immediate post-tests of the knowledge of form and meaning. Participants were then instructed not to look up critical items before returning for session two, the next day. The lexical decision task was administered the following day, allowing for sleep-associated memory consolidation processes contributing to lexical-semantic integration (Lindsay & Gaskell, 2010). At the end of session two participants completed the L2 vocabulary tests, and were debriefed and interviewed about their contextual word learning practices in L2 and L1 reading. Both sessions were conducted with each participant individually.

Approach to data analysis

Linear mixed-effects (*lme*) modelling was used in the data analysis, using the *lmer* function (lme4 library package; Bates, 2011) of the interactive programming environment R. A *Generalized Linear Mixed Model* (mixed *logit* model) was used to analyse the binary data of the immediate knowledge of form and meaning tests, and the response accuracy data of the delayed lexical decision task (Jaeger, 2008). The *lme* approach was chosen because this statistical modelling procedure takes account of the combination of fixed and random effects associated with the learners (participants) and the items being learned (e.g., Baayen, Davidson, & Bates, 2008; Linck & Cunnings, 2015). In modelling our immediate and delayed test data, for example, we are able to account for both the participant and item characteristics known to affect L2 word learning (i.e., existing L2 vocabulary knowledge, individual differences in working memory, number of orthographic neighbours of the critical items), by including these characteristics in the model as covariates.

All analyses included participants and items as crossed random effects. Random slopes were fit for fixed effects, as appropriate. Non-dichotomous variables that were not normally distributed were transformed to bring them closer to normal distribution; they were also centered using the *scale()* function in R to avoid multicollinearity (Belsley, Kuh, & Welsch, 1980). The RT data were inverse transformed, scaled by 1000, and multiplied by −1 to normalise distribution (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010). For ease of reading, plots of results below are based on back-transformed estimates from the *lmer* models, with RTs expressed in milliseconds. The measure of statistical significance of the fixed effects is based on Markov Chain Monte Carlo (MCMC) sampling (10,000 iterations; Baayen et al., 2008).

A minimally adequate statistical model was fitted to the data, using a stepwise variable selection and the likelihood ratio test for model comparisons (Baayen et al., 2008). The resulting statistical model contained only variables that reached significance as predictors, improved the model fit, or were involved in interactions; all other predictors were excluded from further analysis. Following Baayen (2008) and Baayen & Milin (2010), initial minimal a-priori outlier screening was followed by model criticism, i.e., potentially harmful outliers (data points with absolute standardized residuals exceeding 2.5 standard deviations) were removed and the model was refit.

Results

Immediate measures

The WW condition resulted in higher scores for both the knowledge of form (z=12.15 p<.001) and meaning (z=8.91 p<.001) than the ME condition (Appendix B Table B1a & b). On average, in the WW condition, the participants were able to retrieve the form for

56.9% and the meaning for 27.3% of the critical items; in the ME condition, it was 18.2% and 15.3%, respectively.

In the knowledge of form analysis, the effect of learning condition was modulated by the size of the items' orthographic neighbourhood and by the participants' working memory scores (O-Span) (Appendix B Table B1a). The knowledge of form was positively correlated with the size of the items' orthographic neighbourhood, and this effect was larger in the ME condition (Figure 1a; all plots are based on model predictions). It was about three times more difficult to learn the word-forms with no or few orthographic neighbours than with many neighbours, when the learning task required participant to focus mainly on the meaning (not on the form). The WW condition, on the other hand, resulted in lexical knowledge that was considerably less vulnerable to the size of the orthographic neighbourhood of the newly-learned word.

Furthermore, although participants with lower working memory were less successful in the retrieval of word forms on the immediate knowledge test regardless of the learning condition, those with larger working memory were able to correctly produce the form of 40% more words learned in the WW than in the ME condition (Figure 1b). Thus participants with larger working memory could realise this advantage better in the WW condition.

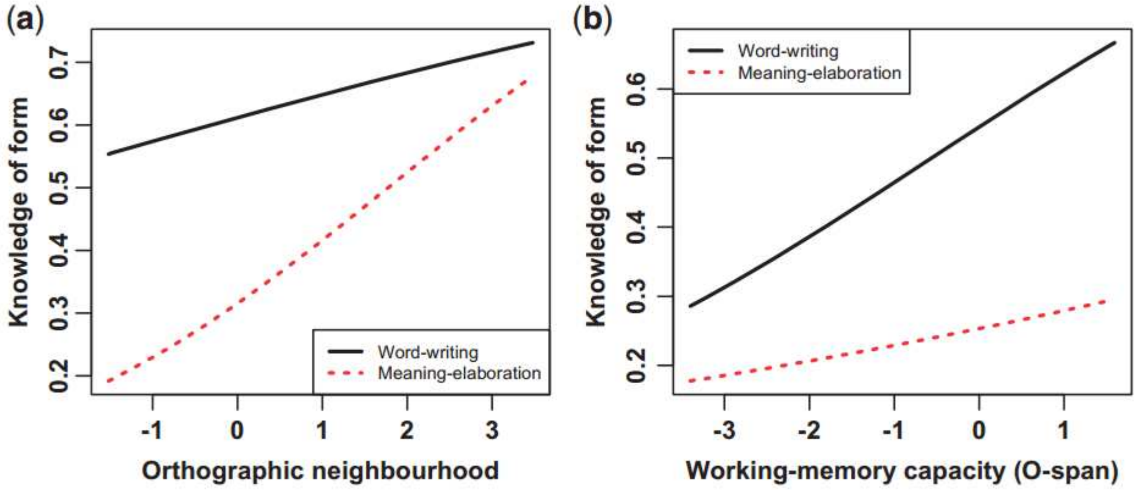

Figure 1: Study 1: The immediate knowledge of form test. Interactions between learning condition and (a) item orthographic neighbourhood; (b) participant working memory.

Finally, as expected, participant vocabulary knowledge (PVLT) was a reliable predictor of the learning of form and meaning.

Delayed measures

Lexical decisions on the higher-frequency English words were more accurate than those on the unfamiliar nonwords (z=4.79 p<.001), but there was no reliable difference in the accuracy of responses to the newly-learned critical items (z<1) and the unfamiliar nonwords (Appendix B Table B2). The accuracy of lexical decisions was modulated by participants' vocabulary knowledge (PVLT).

The RT analysis (Appendix B Table B3) showed an interaction between response accuracy and item type; responses to the higher-frequency words and the newly-learned items were faster than responses to the unfamiliar nonwords on correct trial, and slower on incorrect trials (Figure 2).

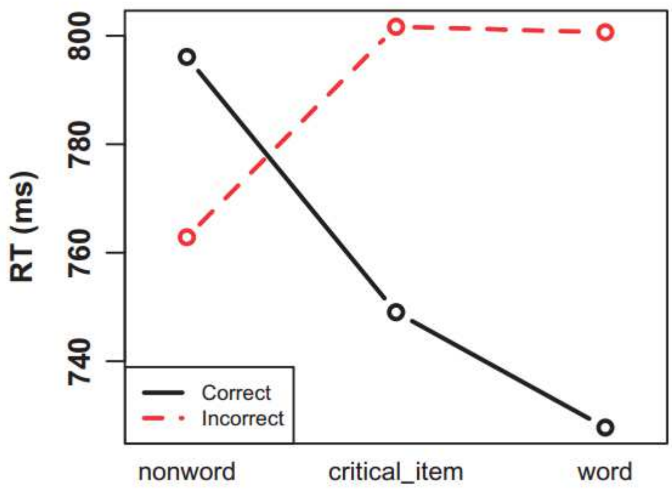

Figure 2: Study 1: Lexical decision task. RT analysis on the full data set. Interaction between target type and response accuracy

There was also a reliable interaction between RT and participants' lexical proficiency (LexTALE), such that higher-proficiency participants were slightly faster in responding to the words and newly-learned items than to the unfamiliar nonwords, while lower-proficiency participants were about 100 ms slower in responding to the words and newly-learned items than to the unfamiliar nonwords.

In order to evaluate the effect of learning condition on lexical quality, separate accuracy and RT analyses were conducted on the subset of the data representing lexical decisions to the critical items. Learning condition (WW vs. ME) was entered in the analyses as primary predictor. A significant effect of learning condition was observed in the accuracy analysis (z=6.45 p<.001) showing that participants were 14.4% more accurate in responding to the critical items learned in the WW condition (Appendix B Table B4). Similar to the immediate test of the knowledge of form, response accuracy to the critical items in the lexical decision task was positively correlated with item orthographic neighbourhood size and participant lexical proficiency (LexTALE). In the RT analysis, the effect of learning condition did not reach reliability (t=-1.18 p=.26), but

RTs to items learned in the WW condition were on average 7.5 ms faster than those to items learned in the ME condition.

Discussion of study one results

For the Chinese-speaking learners of English, handwriting (copying) novel L2 vocabulary encountered in informative sentence contexts resulted in superior learning outcomes than the same amount of time spent on deliberately deriving word meanings from context. The writing approach resulted in a reliably better knowledge of form and form-meaning mapping. Notably, word-writing, by virtue of focusing participant attention on spelling, was able to attenuate negative effects of small orthographic neighbourhood sizes (Andrews, 1997) on the ability to retrieve word forms.

The effect of learning condition on the knowledge of form was modulated by participants' working memory, i.e., the magnitude of advantage for contextual learning was positively correlated with participants' working memory capacity in the word-writing condition. This finding partially supports the limited processing capacity (resource depletion) hypothesis (Barcroft 2002), confirming that a resource-intensive activity, such as writing, differentially affects individual learners as a function of their available cognitive resources.

The observed word-writing advantage is arguably related to the contextual nature of word learning in the present study and the type of tests used[v]. In studies that reported negative effects of writing on form-meaning mapping (reviewed earlier), novel words were usually presented in isolation in the learning phase, making their form more salient a priori. In our study, novel vocabulary was learned contextually and never presented in isolation. Therefore, the writing/copying procedure was the only learning condition that drew participants' attention specifically to the written form of the critical items. We argued that quality of formal-lexical representations (their specificity and completeness) is a key component of lexical quality. The writing condition provided the kind of engagement with the critical items that is needed to establish and fine-tune formal-lexical representations, while the meaning-focused elaboration directed readers' attention to the context in which the novel word was embedded rather than to the word-form itself. Secondly, the measure of form-meaning mapping in this study was aligned with real language use, i.e., to successfully complete the meaning generation task, participants had to recognize the spoken form of the critical item and retrieve other components of its representation from memory. This task probes a more advanced level of knowledge, compared to the previous studies, in which word form and meaning were provided (in multiple choice and matching tasks, e.g., Xu et al., 2013) and participants only had to match but not retrieve them.

In the lexical decision task, responses to the critical items were reliably more accurate when they were learned in the writing conditions. Accuracy of lexical decisions is indicative of stability of lexical representations, which signals lexical quality. The results of study one show that, at early stages of contextual word learning, word-writing contributes more to lexical quality than the common practice of deriving word meaning from context.

However, these findings may not generalise to native speakers of alphabetic languages. Developmental differences in young readers from distinct writing systems (English and Chinese) demonstrate that the brain adapts to "the special features of the orthography" (Cao, Brennan, & Booth, 2014). Chinese readers, for example, develop brain regions associated with holistic visual orthographic processing, while speakers of alphabetic languages develop brain regions associated with phonological processing and fine-grained visual form recognition (Perfetti et al., 2007). Word-writing may be particularly beneficial, therefore, for Chinese speakers, who are likely to need more targeted fine-tuning of orthographic and phonological representations of English words, compared to speakers of alphabetic languages who may enjoy less dramatic benefits from further developing precision of the word-form knowledge (Xu et al., 2013). In order to investigate whether differences between the two types of elaboration hold for native speakers of an alphabetic L1, a second study was conducted with a group of Dutch speakers. Dutch is an orthographic language that induces a high level of fine-tuning of orthographic representations from an early age (Marinus et al., 2015), providing a good case for testing whether the findings of study one can be generalised beyond L1-Chinese.

Study Two: Dutch Speakers

Study two followed the same design, learning and testing procedures, and used the same approach to data analysis.

Methodology

Participants

Study participants were 50 Dutch speakers (37 females) at a university in Belgium. Their English language proficiency was estimated between intermediate and advanced, based on the Common European Framework of Reference: B2 (n=22), C1 (n=14) and C2 (n=14), corresponding to IELTS scores of 5.5 – 9.0. All participants had formal English language instruction in secondary school; 21 were undergraduate students, 17 were graduate students and 12 were not students. Data from two participants were excluded from the analysis: one participant had a visual impairment and one participant was not a native speaker of Dutch (Table 3).

Table 3: Study 2. Participants (L1 = Dutch)

Attribute	Mean	SD	Range
Age	23	4.0	19–35
Age of L2 (English) acquisition	12.2	1.5	6–14
Length of stay in English-speaking countries (months)	0.4	1.5	0–9
LexTALE score (receptive knowledge)	74.9	11.7	50.0–97.5
PVLT2k (productive vocabulary knowledge)	15.8 (=88 per cent)	1.7	11–18
PVLT5k (productive vocabulary knowledge)	9.0 (=50 per cent)	3.3	4–18
O-Span score (working memory)	4.14	0.97	1–6

Results

Immediate measures

The WW condition resulted in higher scores for both the knowledge of form (z=10.66 p<.001) and meaning (z=3.76 p<.001) than the ME condition (Appendix B Table B5a & b). On average, in the WW condition, the participants were able to retrieve the form for 89% of the critical items and the meaning for 67.6% of the critical items; in the ME condition, it was 56.3% and 45.4%, respectively (96% inter-rater agreement).

The effects of participant and item variables were mostly aligned with those reported in study one. Regardless of the learning condition, participants' L2 vocabulary knowledge was a reliable predictor of the learning of form (z=4.02 p<.001) and meaning (z=4.44 p<.001). Participants' working memory (O-Span, z=1.95 p=.051) and item orthographic neighbourhood size (z=2.71 p<.01) were positively correlated with the knowledge of form, irrespective of learning condition (Appendix B Table B5a).

Delayed measures

Lexical decisions on the higher-frequency words were more accurate than those on the unfamiliar nonwords (z=7.37 p<.001). The accuracy of responses to the critical items was also reliably higher than that to the unfamiliar nonwords (z=4.13 p<.001)

(Appendix B Table B6). There was also an interaction between accuracy of responses to different stimulus types and participants' L2 lexical proficiency (LexTALE) in this analysis. Response accuracy to the newly-learned vocabulary items and higher-frequency words were unaffected by participants' proficiency, but their lexical proficiency was a strong predictor of response accuracy to the unfamiliar nonwords.

In the RT analysis (Appendix B Table B7), there was an interaction between response accuracy and item type; responses to the higher-frequency words and the newly-learned items were faster than responses to the unfamiliar nonwords on correct (but not incorrect) trials (Figure 3). RTs to the newly-learned items patterned with RTs to higher-frequency words: faster on correct and slower on incorrect trials; the pattern was reversed for the unfamiliar nonwords. RTs were also affected by individual differences in working memory capacity, with faster RTs observed for participants with larger O-Span scores.

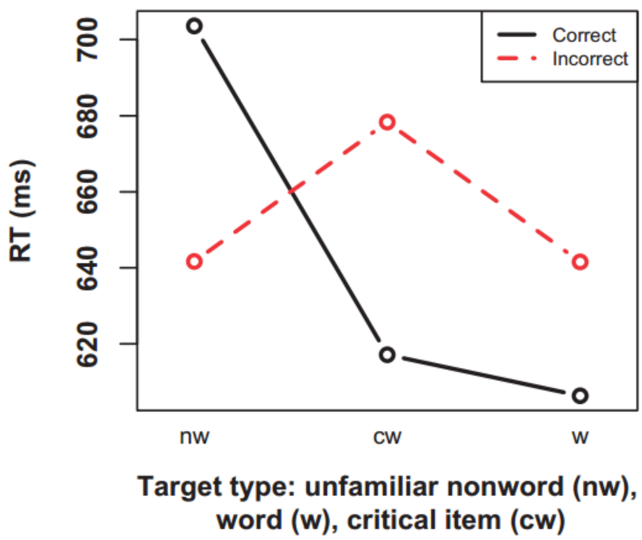

Figure 3: Study 2: Lexical decision task. RT analysis on the full data set. Interaction between target type and response accuracy

In order to evaluate the effect of learning condition on lexical quality, separate accuracy and RT analyses were conducted on the subset of the data representing lexical decisions to the critical items. Learning condition (WW vs. ME) was entered in the analyses as primary predictor. The effect of learning condition did not reach conventional reliability (z=1.94 p=.053) in the accuracy analysis, possibly due to the ceiling effect observed for the Dutch participants whose response accuracy for both learning conditions was over 99% (Appendix B Table B8). A reliable effect of learning condition was observed in the RT analysis (t=-3.54 p<.001), with the RTs to critical items learned in the WW condition being 15 ms faster than to those learned in the ME condition (Appendix B Table B9). Lexical decisions to the newly-learned critical items were also predicted by the O-Span scores; with those who had larger working memory registering faster response times (t=-1.99 p<.05).

Discussion of study two results

The key findings of study one were replicated in study two. Compared with the meaning-focused elaboration, using the word-writing procedure in contextual word learning resulted in a more accurate knowledge of form (spelling) and form-meaning mapping of the novel L2 words, and in faster lexical decisions. Numerically, benefits of word-writing were greater for the Chinese than the Dutch participants on both offline measures (Table 4), with the difference being greater on the knowledge of form test (over 17%). On the online measures, the locus of the word-writing advantage for the Chinese participants was in the accuracy of lexical processing, while for the Dutch participants it was in the speed of processing. Very high response accuracy of the Dutch participants irrespective of learning condition may be due to their fine-grained form recognition developed in learning their native Dutch (Marinus et al., 2015), as well as their higher overall L2 proficiency and superior knowledge of L2 vocabulary.

Nevertheless, word-writing was able to give a boost to the fluency of lexical processing of the contextually-learned critical items by the Dutch participants, compared to meaning-focused elaboration.

Table 4: Mean results and differences between the WW and ME learning conditions in the two studies, based on lme model predictions

Measure	Study 1: Chinese			Study 2: Dutch		
	WW	ME	Difference	WW	ME	Difference
Knowledge of form (per cent)	56.9	18.2	38.6	89.0	67.6	21.4
Knowledge of meaning (per cent)	27.3	15.3	12.0	56.3	45.4	10.9
Accuracy of lexical decisions (per cent)	76.7	62.2	14.4	99.6	99.4	0.2
RT on lexical decisions (ms)	792.8	800.3	−7.5	644.7	659.7	−15.0

Similar to study one, the participants' lexical proficiency and working memory predicted learning outcomes (on the offline and online measures). The presence of the working memory effect suggests that both elaboration procedures used with contextual word learning created high processing demands on the participants' cognitive resources. However, the Dutch participants with larger working memory were able to utilise this advantage regardless of the learning condition, while the Chinese participants could only do so if the critical item was learned in the word-writing condition.

General discussion

This study investigated two types of elaboration, both of which increase the amount of engagement learners have with lexical items in contextual word learning (Schmitt, 2008): deliberately deriving word meaning and handwriting/copying novel vocabulary.

While inferring word meanings from context is a commonly used strategy (Ender, 2014; Fraser, 1999; Paribakht & Wesche, 1999), it is far less common to write down novel words. In fact, copying L2 words during early stages of their learning has been explicitly discouraged, based on the results of previous studies, in which word-writing was used with deliberate, de-contextualised word learning (Barcroft, 2007: 724).

The findings of the present study are straightforward; the studies with Chinese and Dutch participants show that better lexical quality was observed when contextual L2 word learning was accompanied by word-writing, compared to deliberately deriving word meanings, regardless of whether the first language of participants was logographic (Chinese) or alphabetic (Dutch). This finding is aligned with the prediction of the lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) that more precise (specified) orthographic representations, resulting from word-writing, would mediate the development of more complete phonological representations and facilitate the integration of formal-lexical and lexical-semantic representations (form-meaning mapping).

We argue that, when novel vocabulary occurs in diverse informative contexts, conducive to the learning of meaning, the role of word-writing is particularly pertinent. This is because deliberate meaning inferencing represents the type of encoding already in place, when processing words in meaningful contexts, but word-writing engages the learner in a new type of encoding - the encoding of word-form. Therefore, the study shows the added value of word-writing in conjunction with contextual word learning, rather than superiority of form-focused elaboration per se.

This last point is important, because the present study was designed to maximise word learning opportunities by presenting novel vocabulary in informative sentence contexts, as well as providing the phonology (pronunciation) in first exposure, and supplying definitions at the end of the learning phase. In particular, reviewing the definitions of the critical items immediately after contextual exposure, has likely contributed to the establishment of more robust meaning representations, in both learning conditions (Fraser, 1999; Ko, 1995). This is because "definitions … can interact with contexts to communicate core meanings" and are more effective in establishing high-quality meaning representations than contexts alone (Bolger et al., 2008: 122).

Learning and teaching implications

The study shows that a simple technique of copying, when combined with a meaning-focused activity of sentence reading, contributes significantly to the lexical quality of newly-learned words. It may even help attenuate negative effects of orthographic neighbours and individual working memory capacity (for lower-proficiency learners) on the formal-lexical representations established through contextual word learning. These findings highlight the value of providing learners with form-focused encoding opportunities in meaning-focused language learning. We have argued that it is the combination of different types of encoding that makes word-writing effective; thus word-writing out of context is less likely to be an effective learning approach (cf. Barcroft, 2006, 2007; Xu et al., 2013).

In practical terms, L2 learners should be encouraged to copy unfamiliar vocabulary encountered in meaningful contexts (e.g., as a vocabulary notebook entry). This practice would be particularly fruitful when combined with the use of digital referencing tools (e.g., electronic dictionaries and dynamic glosses) to listen to novel words and bring up context-appropriate definitions.

Limitations and suggestions for future research

The present study represents a supportive contextual word learning scenario, and further research is needed to verify whether the word-writing advantage holds under less favourable learning conditions, without additional pronunciation and definitional support, and for longer texts. The learning targets in this study were concrete nouns; and it would be prudent to check if the word-writing treatment is equally effective in contextual learning of other word-types. Furthermore, learning targets were pre-identified for the participants, using brackets; outside the laboratory, this would correspond to learning words that have been explicitly noticed and recognised as unfamiliar by the reader (which is not always the case in L2). Another disclaimer is that we only measured receptive word knowledge, and no claims are made about the effect of word-copying on productive word use. Research is also needed to examine longer-term learning; studies of L2 word learning from reading emphasise the importance of multiple, repeated exposures, as well as opportunities to interact with new words. It is possible that meaning-focused elaboration would have a better chance at enhancing contextual word learning overtime, once the orthographic representations of the new lexical items have been established.

Notwithstanding these limitations, to our knowledge, this is the first carefully controlled study that compares the effects of adding word-writing and meaning-focused elaboration to contextual word learning in L2 (English), across two populations with dramatically different L1 writing systems; and the findings are very encouraging for the word-writing procedure.

---

[i] Online processing occurs in real-time, and is usually tested under time pressure to prevent the use of conscious strategies. Offline processing is time-free, and allows the use of conscious

strategies. Word processing in fluent reading is an example of online processing; using a guessing strategy involves offline processing.

[ii] An overall IELTS score is an average of four individual band scores for reading, listening, writing, and speaking, calculated on a 9-band scale.

[iii] No pretest of word knowledge was administered because pretests, at minimum, expose participants to target word forms in isolation, making their form more salient a priori. To exclude any potential effects of five French cognates (*armoire, fontina, ottoman, ramekin,* and *vitrine*) on the study findings, we repeated the analyses of the Dutch data after excluding these critical items. This exclusion did not change the pattern of results in any of the analyses.

[iv] In a post-study interview about their contextual word learning practices outside the laboratory, the participants reported a tendency not to use dictionary look-ups at first encounter with a novel word, treating repeated encounters as an indication of the word's usefulness/importance and a reason to verify its meaning in the dictionary.

[v] The meaning-verification cycle in the learning procedure may have also contributed to the findings. We will return to this point in the general discussion.

References

Andrews, S. (1997). The role of orthographic similarity in lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review, 4*, 439-461.

Author (2015).

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

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*, 283–316.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*, 323-363.

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research, 22*, 487-497.

Barcroft, J. (2007). Effects of word and fragment writing during L2 vocabulary learning. *Foreign Language Annals, 40*, 713-726.

Bates, D. M. (2011). *Linear mixed model implementation in lme4*. Retrieved from http://cran.r-project.org/web/packages/lme4/vignettes/Implementation.pdf

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: Wiley.

Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition, 23*, 321–343.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Contextual variation and definitions in learning the meaning of words. *Discourse Processes, 45*, 122-159.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition, Vol.1* (pp. 119–160). Hillsdale, NJ: Erlbaum.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96,* 671–681.

Cao, F., Brennan, C., & Booth, J. R. (2014). The brain adapts to orthography with experience: Evidence from English and Chinese. *Developmental Science*. doi: 10.1111/desc.12245.

Chalmers, K. A., & Burt, J. S. (2008). Phonological and semantic information in adults' orthographic learning. *Acta Psychologica, 128*, 162–175.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory and Cognition*, *12*, 590–606.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25,* 1–18.

Davis, C. J. (2012). The orthographic similarity of printed words. In J. S. Adelman (Ed.), *Visual Word Recognition: Models and Methods, Orthography and Phonology, Volume 1* (pp. 120-206). Hove, England: Psychology Press.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, R. H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language, 62*, 284-301.

Ender, A. (2014). Implicit and explicit cognitive processes in incidental vocabulary acquisition. *Applied Linguistics*, doi: 10.1093/applin/amu051.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21*, 225–241.

Hamada, M., & Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word learning. *Language Learning, 58*, 1–31.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL, 41*, 87–106.

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

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.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*, 137–166.

Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System, 18*, 199–207.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78*, 285–299.

Ko, M. H. (1995). Glossing in incidental and intentional learning of foreign language vocabulary and reading. *University of Hawaii Working Papers in ESL, 13,* 49–94.

Koda, K. (1997). Orthographic knowledge in L2 lexical processing: A crosslinguistic perspective. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 35–52). Cambridge: Cambridge University Press.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review, 12*, 703–710.

Landauer, T. K., & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*, 211–240.

Laufer, B. (2005). Focus on form in second language vocabulary acquisition. In S. H. Foster-Cohen, M. P. Garcia-Mayo & J. Cenoz (eds.), *EUROSLA Yearbook* 5. Amsterdam: John Benjamins, 223–250.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*, 33–51.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*, 325–343, doi:10.3758/s13428-011-0146-0.

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning, 65*, Suppl. 1, 185–207.

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

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

Marinus, E., Nation, K., & de Jong, P. F. (2015). Density and length in the neighborhood: Explaining cross-linguistic differences in learning to read in English and Dutch. *Journal of Experimental Child Psychology, 139*, 127–147.

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20,* 482–496.

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language [Internet]. Available from: www.neuro.mcw.edu/mcword/.

Mestres-Missé, A., Rodríguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex, 17*, 1858–1866.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research, 53*, 253-279.

Mondria , J. (2003). The effects of inferring, verifying, and memorizing on the retention of L2 word meanings: An experimental comparison of the "meaning-inferred method" and the "meaning-given method." *Studies in Second Language Acquisition, 25*, 473–499.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59–82.

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

Paribakht, S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, *21*, 195–224.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam, the Netherlands: Benjamins.

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., & Tan, L-H. (2007). Reading in two writing systems: Accommodation and assimilation in the brain's reading network. *Bilingualism: Language and Cognition, 10*, 131–146.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading, 7*, 219–237.

Share, D. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.

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

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13*, 138-163.

Thomas, M. H. *&* Dieter, J. N. (1987). The positive effects of writing practice on integration of foreign words in memory*. Journal of Educational Psychology,* 79*,* 249–253*.*

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176-1190.

Xu, Y., Chang, L. Y., Zhang, J., & Perfetti, C. A. (2013). Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Language Annals, 46*, 423–444.

Appendix A: List of stimuli and their definitions

Topic	Lexicality	Critical word	Definition
building	low-freq. word	egress	a way out, such as a window or a small door, that is required in every bedroom and basement.
building	low-freq. word	lathe	a machine for shaping or carving wood, metal or other material, by rotating it against cutting tools.
building	low-freq. word	vitrine	a glass display case.
building	low-freq. word	anvil	a heavy steel or iron block with a flat top, on which metal is hammered and shaped.
building	low-freq. word	dibble	a pointed hand tool for making holes in the ground for seeds or young plants.
building	low-freq. word	ottoman	a low stool, usually covered with fabric, designed to go at the foot of a chair.
building	low-freq. word	scriber	a sharp instrument used for making marks on wood, bricks or metal.
building	low-freq. word	gable	a triangular area of a house beneath the sloping roof.
building	low-freq. word	pelmet	a narrow border of cloth or wood at the head of a window, to hide the fittings of curtains or blinds.
building	low-freq. word	armoire	a wardrobe or cabinet for storing clothing.
building	low-freq. word	bodger	a skilled craftsman, who makes wooden chair legs, poles and beams.
building	low-freq. word	newel	a post supporting a handrail of a staircase.
building	nonword	parsage	a woodcutting saw with a removable blade.
building	nonword	torsh	a type of nonmetallic electrical cable that is used for indoor wiring.
building	nonword	surmit	a construction vehicle with tracks or large wheels and a wide blade, used for moving earth or debris.
building	nonword	shottle	paving material, such as small stones, used for garden paths and sidewalks.

building	nonword	spiler	a person who lays gas, water and sewer pipes.
building	nonword	opean	a mechanical tool for lifting people or heavy objects.
building	nonword	creptor	a device for measuring and recording the use of electricity or water by a household or business.
building	nonword	emback	a small simple building used as a temporary shelter, in a forest or mountain area.
building	nonword	tylon	a brick cleaning substance used after brick laying work is completed.
building	nonword	banity	a painted pattern or design on a wall or ceiling.
building	nonword	ferch	a horizontal piece of wood or stone that forms the bottom of an entrance or doorway.
building	nonword	proster	a large strong steel or iron beam forming a main supporting element in a framework of buildings or bridges.
cooking	low-freq. word	offal	the internal organs of butchered animals (such as cow or pig) including the liver, heart, and stomach that are safe to eat.
cooking	low-freq. word	dollop	a shapeless mass or lump of soft food, such as cream.
cooking	low-freq. word	ramekin	a small dish for baking and serving an individual portion of food.
cooking	low-freq. word	pepita	flat seeds of vegetables of the squash family, such as a pumpkin.
cooking	low-freq. word	saran	plastic food wrap.
cooking	low-freq. word	fontina	semi-soft cheese, with a mild and slightly nutty flavour.
cooking	low-freq. word	busser	a person who clears tables in a restaurant or cafe.
cooking	low-freq. word	gourd	a plant with a hard skin and shell, that is used as a storage container, musical instrument or a decoration.
cooking	low-freq. word	griddle	a piece of cooking equipment with a flat cooking surface and a heat source underneath.
cooking	low-freq. word	ladle	a large long-handled spoon with a cup-shaped bowl, used for serving soup, stew or sauce.
cooking	low-freq. word	kipper	fish preserved by splitting, salting, and drying or smoking.

cooking	low-freq. word	clabber	raw milk that has soured due to natural fermentation. It is similar to yoghurt.
cooking	nonword	troppy	a person who takes particular pleasure in fine food and drink.
cooking	nonword	tragger	a bowl with small holes used for draining substances cooked in water.
cooking	nonword	rocot	a finely chopped cooked mixture of diced meet and vegetables (mostly potatoes and onions).
cooking	nonword	gastle	a tool similar to an eye dropper, used during cooking to cover meat in its own juices or with a sauce.
cooking	nonword	quist	a wine merchant or wine maker.
cooking	nonword	recresh	small bubbles formed in or on a liquid that rise to the surface (e.g., during boiling or pouring).
cooking	nonword	trince	the chef who is second in authority in a restaurant or kitchen, below the head chef.
cooking	nonword	capsale	a little bite of food served before dinner. It is usually carefully decorated and has intense flavours.
cooking	nonword	prend	a small kitchen tool with a broad, flexible blade that is used to mix, spread, or lift food.
cooking	nonword	bondit	a type of dessert. A square doughnut served very hot topped with powdered sugar.
cooking	nonword	dybasty	a procedure of taking a small amount of food into the mouth to test its quality.
cooking	nonword	cheis	the flat middle section of chicken wings used to make chicken nibbles and snacks.

Appendix B

Table B1. Study one: Analysis of the immediate test data.

(a) Analysis of the knowledge of form					
	Coef.β	**SE(β)**	***z***	***p***	
(Intercept)	-1.25	.25	-5.07	3.9E-07	***
LearningCondition[a]=WW	1.41	.12	12.15	< 2.0E-16	***
cOrthN	.61	.16	3.87	1.1E-04	***
cO-Span	.17	.15	1.13	.260	
cPVLT	.42	.08	5.54	3.0E-08	***
LearningCondition:cOrthN[b]	-.40	.09	-4.49	7.1E-06	***
LearningCondition:cO-Span	.26	.10	2.61	.009	**
(b) Analysis of the knowledge of meaning					
	Coef.β	**SE(β)**	***z***	***p***	
(Intercept)	-2.78	.28	-9.75	< 2.0E-16	***
LearningCondition=WW	1.20	.13	8.91	< 2.0E-16	***
cPVLT	.63	.10	5.99	2.1E-09	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Intercept levels: LearningCondition=meaning-focused elaboration.

[a] WW=word-writing; ME=meaning-focused elaboration; mall “c” (e.g., cOrthN) – indicates that a variable is centered; PVLT=the average score of PVLT2k and PVLT5k.

[b] A colon sign (:) between variables indicates an interaction between them, e.g., LearningCondition:cOrthN stands for an interaction between learning condition and number of item’s orthographic neighbours.

Table B2. Study one: Lexical decision task. Response accuracy analysis on the full data set.

	Coef.β	**SE(β)**	***z***	***p***	
(Intercept)	1.22	.14	8.76	< 2.0E-16	***
TargetType=critical_item	-.19	.20	-.93	.355	
TargetType=word	1.02	.21	4.79	1.6E-06	***
cPVLT	.10	.04	2.34	.020	*
TargetType=critical_item:cPVLT	.09	.03	3.01	.003	*
TargetType=word:cPVLT	.26	.04	6.65	3.0E-11	***

Intercept levels: TargetType=unfamiliar nonword.

Table B3. Study one: Lexical decision task. RT analysis on the full data set.

	Coef.*β*	*t* value	MCMC mean	HPD95 lower	HPD95 upper	pMCMC	
(Intercept)	6.640	228.56	6.641	6.600	6.681	1.0E-04	***
TargetType=critical_item	.049	4.25	.050	.027	.071	1.0E-04	***
TargetType=word	.048	3.69	.049	.024	.075	4.0E-04	***
Target.ACC=1 (correct)	.043	6.23	.043	.030	.056	1.0E-04	***
cNoL	.030	7.38	.030	.022	.038	1.0E-04	***
cLgLexTALE	-.069	-.33	-.067	-.352	.219	.643	
cLgPrev.RT[a]	.147	8.23	.153	.122	.184	1.0E-04	***
TargetType=cr_item:cLgLexTALE	-.175	-4.89	-.174	-.246	-.104	1.0E-04	***
TargetType=word:cLgLexTALE	-.145	-4.10	-.145	-.216	-.078	1.0E-04	***
TargetType=cr_item:Target.ACC=1	-.110	-9.50	-.111	-.134	-.089	1.0E-04	***
TargetType=word:Target.ACC=1	-.138	-10.73	-.139	-.164	-.113	1.0E-04	***

Intercept levels: TargetType=unfamiliar nonword, Target accuracy(Target.ACC)=0 (incorrect).

[a] Prev.RT stands for an RT on the preceding trial. It is included in the regression model to control for longitudinal effects observed in lexical decisions (Baayen et al., 2008).

Table B4. Study one: Lexical decision task. Response accuracy analysis on the critical items.

	Coef.*β*	**SE(*β*)**	***z***	***p***	
(Intercept)	.60	.21	2.88	.004	**
LearningCondition=WW	.69	.11	6.45	1.1E-10	***
cLgLexTALE	3.87	1.13	3.43	.001	***
cOrthN	.26	.11	2.35	.019	*

Intercept levels: Learning mode=meaning-focused elaboration.

Table B5. Study two: Analysis of the immediate test data.

(a) Analysis of the knowledge of form

	Coef.*β*	**SE(*β*)**	***z***	***p***	
(Intercept)	1.16	.29	3.94	8.2E-05	***
LearningCondition=WW	1.36	.13	10.66	< 2.0E-16	***
cPVLT	.32	.08	4.02	5.8E-05	***
cO-Span	.41	.21	1.95	.051	
cOrthN	.53	.20	2.71	.007	**

(b) Analysis of the knowledge of meaning

	Coef.*β*	**SE(*β*)**	***z***	***p***	
(Intercept)	-0.23	0.37	-0.63	0.526	
LearningCondition=WW	0.44	0.12	3.76	1.7E-04	***
cPVLT	0.45	0.10	4.44	9.1E-06	***

Intercept levels: LearningCondition=meaning-focused elaboration.

Table B6. Study two: Lexical decision task. Response accuracy analysis on the full data set.

	Coef.*β*	**SE(*β*)**	***z***	***p***	
(Intercept)	3.141	.279	11.24	< 2.0E-16	***
TargetType=critical_item	1.596	.386	4.13	3.7E-05	***
TargetType=word	3.644	.495	7.37	1.8E-13	***
cLexTale	.091	.017	5.39	6.9E-08	***
TargetType=critical_item:cLexTALE	-.120	.012	-9.77	< 2.0E-16	***
TargetType=word:cLexTALE	.009	.018	.49	.627	

Intercept levels: TargetType=unfamiliar nonword.

Table B7. Study two: Lexical decision task. RT analysis on the full data set.

	Coef.*β*	***t* value**	**MCMC mean**	**HPD95 lower**	**HPD95 upper**	**pMCMC**	
(Intercept)	-1.570	-53.80	-1.569	-1.618	-1.522	1.0E-04	***
TargetType=critical_item	.084	3.66	.084	.039	.128	1.0E-04	***
TargetType=word	-2.7E-04	-.01	.001	-.050	.052	.970	
Target.ACC=1 (correct)	.137	12.61	.137	.115	.158	1.0E-04	***
cOspan	-.061	-2.23	-.043	-.096	.010	.115	
cLgPrev.RT	.103	6.01	-.157	-.280	-.032	.016	*
cTrial	-3.7E-04	-2.28	.105	.072	.136	1.0E-04	***
TargetType=cr_item:Target.ACC=1	-.283	-13.34	-4.0E-04	-.001	-1.0E-04	.026	*
TargetType=word:Target.ACC=1	-.228	-9.17	-.283	-.327	-.243	1.0E-04	***

Intercept levels: TargetType=unfamiliar nonword, Target accuracy (Target.ACC)=0 (incorrect).

Table B8. Study two: Lexical decision task. Response accuracy analysis on the critical items.

	Coef.*β*	**SE(*β*)**	***z***	***p***	
(Intercept)	5.173	.485	10.66	< 2.0E-16	***
LearningCondition=WW	.468	.242	1.94	.053	

Intercept levels: Learning mode=meaning-focused elaboration.

Table B9. Study two: Lexical decision task. RT analysis on the critical items.

	Coef.*β*	***t* value**	**MCMC mean**	**HPD95 lower**	**HPD95 upper**	**pMCMC**	
(Intercept)	-1.523	-46.06	-1.519	-1.576	-1.462	1.0E-04	***
LearningCondition=WW	-.035	-3.54	-.036	-.056	-.016	2.0E-04	***
Target.ACC=1 (correct)	-.090	-4.87	-.092	-.128	-.056	1.0E-04	***
cOspan	-.055	-1.99	-.055	-.097	-.011	.013	*
cInvPrev.RT	.106	5.27	.113	.077	.150	1.0E-04	***
cTrial	-4.9E-04	-2.14	-.001	-.001	-1.0E-04	.034	*

Intercept levels: LearningCondition=meaning-focused elaboration, Target accuracy=0 (incorrect).

Note: cInvPrev.RT and cTrial are included in the regression model to control for longitudinal effects. “Inv” (e.g., cInvPrev.RT) stands for inverse transform (InvPrev.RT= –1,000/Prev.RT).