

RESEARCH

Incorrect inferences and contextual word learning in English as a second language

Irina Elgort

When readers encounter new words they may try to infer their meanings from context. Such contextual inferences may be correct or incorrect. This research considered the effect of incorrect meaning inferences on contextual word learning in English as a second language. Chinese speakers encountered 48 novel vocabulary items in informative single-sentence English contexts and were instructed to infer their meanings. They were able to verify their inferences by reviewing dictionary-type definitions at the end of the learning procedure. Participants' explicit knowledge of the critical vocabulary items was probed using a meaning generation task; their implicit knowledge was examined using a mixed-modality masked repetition priming lexical decision task. The results revealed a differential effect of incorrect inferences on the explicit and implicit knowledge of the vocabulary items. Explicit knowledge of meaning was less accurate after incorrect inferences than after correct inferences, but it was not worse than the knowledge gained when no explicit inference had been made. Implicit knowledge however was not affected by incorrect inferences. Pedagogical and research implications of the study findings are considered.

Keywords: Contextual word learning; Vocabulary; Meaning inferences; Implicit knowledge

1. Introduction

1.1. *Inferring word meanings from context in a second language*

Recently I interviewed fifty Chinese university students studying in an English-speaking country about their attitudes toward guessing word meanings from context when reading in English, their second language (L2). A common concern about this approach to learning new words expressed in these interviews was, 'What if I guess incorrectly?' Indeed, researchers have found that L2 readers make incorrect inferences about word meanings during reading (Bensoussan & Laufer, 1984; Frantzen, 2003; Kelly, 1990; Laufer, 1997; Mondria, 2003; Schouten-van Parreren, 1992). Incorrect inferences may arise when the context in which an unfamiliar word occurs provides insufficient support for inferring the meaning (is low-constraining, vague or ambiguous). Another reason for incorrect inferences is text difficulty for an individual reader; Hu and Nation (2000) found that about 98% of the running words in a text need to be known in order to achieve the level of understanding that supports the learning of new L2 words. Therefore, the same text may serve as informative context for one reader, but not another. Incorrect inferences may also arise when an unfamiliar L2 word is mistakenly identified as familiar (Laufer, 1997) or when a guess is made on the basis of

form similarity to another word, unrelated in meaning, without adequate attention to context (Frantzen, 2003).

The relationship between contextual inferencing and word learning is not straightforward. Contextual word learning is an incremental process influenced by text, word, learner and situational factors (Paribakht & Wesche, 1999). Correctly inferring a word's meaning from context is not always synonymous with learning (Frantzen, 2003; Haastруп, 1991; Hulstijn, 2001; Mondria, 2003; Mondria & Wit-de Boer, 1991; Pressley, Levin & McDaniel, 1987). The effect of incorrect inferences on word learning has been researched to a lesser extent, but it is generally assumed that incorrect inferences would interfere with the encoding of correct word meanings and have a negative effect on learning. Studies that looked into the effect of erroneous meaning inferences in reading on word learning tend to support this assumption (Hulstijn, 1992; Carpenter, Sachs, Martin, et al., 2012). Carpenter et al. (2012) provides a detailed insight into the nuanced effects of incorrect inferences vis-à-vis corrective feedback in L2 (German) word learning from reading. Importantly, they found that, once corrected, erroneous inferences were seldom repeated on post-tests. However, the majority of participants whose errors were not corrected (the no-feedback group) repeated their incorrect inferences on post-tests.

One of the limitations of previous studies into the effect of correct/incorrect inferences is exclusive use of off-line, explicit word knowledge measures. Since word learning through reading occurs incrementally, at early stages of

learning, an individual encounter with a word (even in an informative context) is often insufficient for the learner to be able to correctly articulate its core meaning, for example, in a meaning generation or translation task. In supportive single-sentence contexts, 3–4 encounters may be needed even for native (L1) speakers to learn word meanings (Bolger, Balass, Landen & Perfetti, 2008; Mestres-Missé, Rodrigues-Fornells & Munte, 2007). When reading in a second language, more encounters may be needed, especially when reading longer texts (e.g., Waring & Takaki, 2003). Nevertheless, readers may acquire some useful information about a new word each time they encounter it in context (L1: Borovsky, Kutas & Elman, 2010; L2: Webb, 2007). Therefore, it is important to go beyond traditional explicit, off-line meaning recognition and retrieval tests in order to evaluate partial word knowledge that may be below the threshold of awareness (Elgort & Warren, 2014; McLaughlin, Osterhout & Kim, 2004; Pellicer-Sánchez, 2015; see also Nation & Webb, 2011). Approaches that use online and offline, explicit and implicit, declarative and procedural word knowledge measures can shed further light on how different aspects of word knowledge are affected by making explicit correct and incorrect inferences during reading.

1.2. The effect of presence or absence of errors during vocabulary learning

Learning new words is underpinned by establishing new memory traces; therefore, learning procedures that avoid errors may result in more robust learning than those that include errors (Baddeley & Wilson, 1994). However, in L1 studies conducted with memory impaired (see Clare & Jones, 2008; Middleton & Schwartz, 2012 for critical reviews) and unimpaired participants, children (Warmington & Hitch, 2014) and adults (Bridger & Mecklinger, 2014; Warmington, Hitch & Gathercole, 2013), results are not clear-cut. Middleton and Schwartz (2012) point out that assumptions of errorless learning (EL) frameworks in cognitive rehabilitation are at odds with testing studies (of non-clinical populations) that show learning benefits associated with retrieval of information from long-term memory and a positive relationship between retrieval difficulty and learning (e.g., Karpicke & Roediger, 2007, 2008). The review (Middleton & Schwartz, 2012) suggests that, in errorless learning studies in amnesia, effectiveness of EL vs. trial-and-error (errorful) learning (EF) depends on the type of patient and training task, i.e., EF is better than EL for patients with mild to moderate memory impairments, while error avoidance may improve learning for patients with severe impairments when implicit learning tasks are used (p. 151).

In studies with normal populations, similar issues arise. Warmington and Hitch (2014) reported some advantages for EL over EF in two deliberate word learning experiments with adult L1 participants. The first experiment involved learning novel word forms (i.e., non-words) for familiar concepts in a word-picture learning paradigm. The EF condition was operationalised by instructing participants to look at an object and provide (guess) its name. Participants were then given the correct object name and

asked to repeat it. In the EL condition, participants were provided with the name of each object upfront and asked to repeat it aloud. Word learning was measured using an object naming (production) task and a name-to-object recognition task. Compared to EF, EL resulted in a better performance on the naming task, but not on the name-to-object recognition task. In their second experiment, Warmington and Hitch (2014) used a word-definition learning paradigm and spaced repetitions for learning obscure L1 words (e.g., *dossil*), requiring the learning of both form and meaning. One familiarisation and two treatment blocks were administered. In the EL condition, participants were presented with definition-word pairs and then asked to retrieve the target word. In the second treatment block, an intervening definition was used to increase the practice interval, “e.g., *A coarse dark sugar is called jaggery. A paragraph mark is called pilcrow. What is the name for a coarse dark sugar? (response) What is the name for a paragraph mark? (response)*” (p. 589). In EF, the order was reversed; participants were first instructed to retrieve the word from its definition and then given the definition-word pair as feedback. In the second block, the feedback was delayed by including an intervening question-response sequence “e.g., *What is the name for a coarse dark sugar? (response) What is the name for a paragraph mark? (response) A coarse dark sugar is called jaggery. A paragraph mark is called pilcrow.*” Participants who learned via the EL method performed significantly better on a cued recall test, compared to the EF method. It needs to be pointed out, however, that the design of the two learning treatments in Experiment 2 may have differentially affected word learning outcomes. In the second cycle of EL, a delay in the target word retrieval would have likely been beneficial for learning because it increased the difficulty of the target word retrieval (Karpicke & Roediger, 2007). In the EF treatment, the delay in the provision of feedback could have interfered with learning (Shute, 2008), especially since an additional retrieval event (for a different target word) was introduced prior to the provision of feedback.

Rodriguez-Fornells, Kofidis and Munte (2004) tested word recognition and processing following ER and EF conditions with normal L1 (German) participants learning (memorising) 60 words, using behavioural and electrophysiological (ERP) measures. In the EF condition, participants were first instructed to come up with a number of words beginning with a specified set of letters (e.g., *B-R-U*), after which they were told which word was the correct answer (e.g., *Brust*, meaning *chest*) and asked to repeat it. In the EL condition, participants were given a word beginning with the same set of letters and asked to repeat it. In the recognition phase, participants were exposed to words beginning with the same letters that were either the target word, i.e., the word identified as correct in the learning treatment (e.g., *Brust*) or a foil, i.e., another high-probability candidate for the same letter-combination onset (e.g., *Bruder*, meaning *brother*). A better quality of memory performance (indexed by *d-prime*) was observed in the EL condition, but an overall number of detected targets was greater in the EF condition. Informed by the

finding of more positive ERPs (LPC component) for the correctly recognised target words in the EF (compared to the EL) condition, Rodriguez-Fornells et al., conjectured that the EF condition was associated with a deeper processing of the targets.

Research into error-free L1 word learning with memory impaired and normal participants has mostly been in the form of list or paired-associate learning out of context. Recently, however, Frishkoff, Collins-Thompson, Hodges and Crossley (2016) investigated whether, in contextual L1 word learning involving forced meaning generation (an approach that increases the risk of errors), the effect of erroneous guesses could be mitigated by providing immediate accuracy feedback. They found that the provision of feedback resulted in better word learning (measured by a meaning generation post-test), but only for mixed contexts (with some strong and some weak cues to the meaning of the target word). There was no significant effect of feedback when only high-constraining contexts (with very strong cues) were used, suggesting that words may be learned in such contexts, whether or not accuracy feedback is provided on readers' explicit meaning inferences (Frishkoff et al., 2016, p. 624). Presumably, because high-constraining sentences are more conducive to the encoding of correct meanings of novel words from context, opportunities to verify inference correctness through accuracy feedback have little to add to learning under such conditions.

Few studies have specifically investigated the effect of errors in L2 vocabulary learning (Boers, Dang & Strong, 2016; Boers, Demecheleer, Coxhead & Webb, 2014; Carpenter et al., 2012; Trenkic & Warmington, 2014). In a conference paper, Trenkic and Warmington (2014) reported an advantage for error-free over trial-and-error word learning, regardless of whether the participants' meaning guesses were correct or incorrect. In a collocation learning study, Boers et al. (2014) found that making a mistake in completing gap-fill exercises (trial-and-error learning) increased the likelihood of mistakes at post-test, even when corrective feedback was provided and the learners actively corrected their mistakes (cf. Carpenter et al., 2012; Frishkoff et al., 2016). Boers et al. (2016) partially replicated their previous results, but found that learning through trial-and-error was more successful under certain conditions (i.e., when exercises encouraged learners to think about collocations as chunks). Notwithstanding paucity of evidence on the topic, these investigations confirm that the effect of errors in L2 vocabulary learning is not straightforward, and further research is needed.

1.3. Present study

The present study contributes to our understanding of contextual word learning mediated by correct and incorrect meaning inferences, using measures of explicit and implicit knowledge. The key question in the study is *how incorrect inferences affect word learning in informative sentence contexts*. In the study, Chinese speakers of English read three stand-alone informative English sentences with embedded novel words. Participants were instructed to type inferences about the meanings of these words into a

text box (or leave it blank if they could not come up with an answer). After the inferencing task, the participants had a chance to review correct L2 definitions of the novel words; this ensured that all participants had an opportunity to verify their inferences and encode correct word meanings. This design allowed for a tight control over the presentation of the novel words (number and length of exposure; frequency of repetitions; manner in which the task was completed), in order to clarify the effect of inference correctness on learning. After the learning phase, participants' knowledge of the newly learned words was measured using (1) a meaning generation task that evaluated their ability to retrieve meaning from form (form-meaning mapping) and (2) a primed lexical decision task, in which the repetition-priming effect was used to evaluate their implicit vocabulary knowledge.

It was predicted that the effect of erroneous inferences in this study would not be detrimental for word learning. This prediction was motivated by the following considerations: (1) actively inferring meanings from context (meaning-focused elaboration) increases readers' engagement with novel words and the context (Hulstijn, Hollander & Greidanus, 1996; Laufer, 2005; Mondria & Wit-de Boer, 1991; Schmitt, 2008); (2) presenting novel words in informative contexts makes learning more likely, due to contextual clues and co-occurrence with known words (L1: Landauer & Dumais, 1997; Mestres-Missé, Rodrigues-Fornells & Munte, 2007; Chaffin, Morris & Seely, 2001; L2: Webb, 2008); (3) contextual word learning is facilitated by the use of definitions as a means of verifying contextual inferences (L1: Bolger et al., 2008; L2: Carpenter et al., 2012; Fraser, 1999; Kelly, 1990; Mondria, 2003).

Furthermore, it is predicted that explicit incorrect inferences are more likely to affect explicit word knowledge (as measured by the meaning generation task), but less likely to predict implicit knowledge that builds incrementally from individual encounters with the novel word in supportive contexts.

2. Methodology

2.1. Participants

Study participants (n = 47; female = 36) were either completing an English Proficiency Programme (n = 32) or enrolled in their first university course (n = 15) in New Zealand. Their mean age was 24.5 (St. Dev. = 4 years). All participants were Chinese speakers, whose English language proficiency ranged from intermediate to upper-intermediate, based on their overall IELTS scores of 5.5–7.0. Participants received grocery vouchers for their participation.

2.2. Materials and procedure

All data in the learning and testing phases of the study were collected individually from each participant, in a computer lab.

2.2.1. Vocabulary items

The learning targets in this study were 48 vocabulary items (5–7 letters) (henceforth, critical items), the meanings of which were related to two broad

themes: building/household (e.g., *pelmet*, *newel*) and cooking/food (e.g. *dollop*, *clabber*). These themes afford the use of technical vocabulary, while assuming that learners have enough background understanding of the topic and are able to use their pre-existing schemata to help them acquire new terms. Half of the items were made-up (nonce) words and the other half were low-frequency English words (Appendix A). The low-frequency words were used for the learners to derive some real learning value from their participation in the study. However, the orthographic and phonological characteristics of low-frequency words cannot be controlled, and such words may be more difficult to learn than nonce words constructed in accordance with the orthographic and phonological constraints of the target language. The use of nonce words also ensures that participants have no knowledge of the critical items prior to the study.

2.2.2. The learning phase

The critical items were embedded in three sentences each and presented on a computer screen, one at a time. The sentences were chosen to create reasonable opportunities for the participants to infer the meanings of the critical items from context (e.g., “A {pelmet} can frame the window space and conceal curtain rods.”; “Hard {pelmets} are usually made of wood; soft ones are in the same material as the curtains.”; “The main curtains should be topped with a {pelmet} and fall generously to the floor.”) During the first encounter with the critical item, the participants were instructed to read the sentence for meaning and then listen to the audio recording of the item presented in brackets. They could listen to the recording more than once by pressing a designated key on the keyboard.

On the second and third encounter, the participants were instructed to read sentences for meaning, then try to infer the meaning of the word in brackets and type it into the text box underneath the sentence. Thus, the participants had two attempts at explicitly articulating (in writing) their inferences about the meaning of each critical word. Once the meaning inference procedure was completed for each 12 critical items, the participants were presented with brief dictionary-type definitions (e.g., “pelmet – a narrow border of cloth or wood at the head of a window, to hide the fittings of curtains or blinds”). The vocabulary load of all learning materials was checked using VP-Compleat (www.lex tutor.ca/vp/): 96% of the words used in the sentences and 97.2% of the words used in the definitions were within the first 5000 word-frequency lists (based on Nation’s BNC/COCA word-family lists, available from www.victoria.ac.nz/lals/about/staff/paul-nation).

2.2.3. The testing phase

Testing was conducted over two days. The participants’ ability to retrieve meaning from form was tested using a meaning generation task. The participants were instructed to listen to recordings of the critical items and explain their meaning in English. This explicit knowledge test probed form-meaning mapping established for the newly learned items. The test was conducted on the same day as the learning phase, after an intervening task. It was scored by the researcher and checked by an independent

scorer with a postgraduate degree in linguistics. The same procedure was used to score first and second attempts to infer the meanings of critical items in the learning phase.

The following day, the participants completed mixed-modality masked repetition priming task (RPT) that measured implicit knowledge of the critical items by probing the quality of their recognition and processing. In repetition priming, the prime and target are the same item on related trials, but they are completely different on unrelated trials. In a masked priming procedure, the prime is presented for a short time and is preceded and followed by a mask (e.g., letters or symbols). This severely restricts information about the prime that can be consciously identified and makes it very difficult (if not impossible) for the participants to use consciously controlled processes and strategies. The priming manipulation is often combined with the lexical decision task, in which participants are instructed to decide whether the target is an English word or not, i.e., make a *lexical decision*. This decision requires participants to access lexical representations of words stored in their memory. Making a lexical decision involves accessing both formal-lexical and lexical semantic representations of the word (Joordens & Becker, 1997; Masson, 1995; Neely, 1991). In a primed lexical decision task, participants are expected to respond to the target faster (and/or more accurately) on related (repetition) trials than on unrelated trials. This priming effect occurs because the related prime activates the representation of the target word, making lexical decisions faster (Grainger & Ferrand, 1994). The locus of repetition priming is primarily lexical (Forster, 1998; Forster, Mohan & Hector, 2003). Hence, repetition priming occurs for known words but not for nonce words, because the latter have no lexical representations. In the present study, this design provides a means of checking whether the newly learned critical items were processed as words (i.e., whether their lexical representations had been established) and whether incorrect inferences negatively affect learning, i.e., reduce or eliminate the repetition priming effect.

The design of the RPT was based on the mixed-modality repetition priming paradigm developed by Grainger, Diependaele, Spinelli, Ferrand and Farioli (2003). The masked prime was presented visually (in the written form) while the target was presented in the auditory modality, as a sound recording. A mixed-modality priming paradigm engages both orthographic and phonological representations and processing. Used as a measure of contextual learning, this paradigm is, therefore, more informative than within-modality repetition priming. In reading, for example, phonological (as well as orthographic) representations of known words are activated automatically by the visual input, contributing to word identification and lexical access.

In the RPT, the 48 critical items from the learning phase and 48 phonologically and orthographically plausible nonce words (*nonwords*, not encountered by the participants prior to completing the task) were used as auditory targets presented for lexical decisions. On repeated trials, the visual primes were the same items as the auditory targets (*egress* – *EGRESS*); on unrelated trials, they were not

(*dollop* – *EGRESS*). In addition, 48 unrelated prime-target pairs were used as fillers (24 word targets with nonword primes and 24 nonword targets with word primes), in order to reduce the use of task-related strategies. A practice block of 16 trials was used at the start of the procedure. Each trial began with the presentation of the forward mask (a row of # signs) together with two vertical lines (above and below the centre of the mask). After 500 ms, the forward mask was replaced by the prime presented in the lower-case letters in the middle of the screen for 67 milliseconds. It was immediately replaced by the backward mask consisting of pseudorandom strings of consonants (XPWKXHKPWK) not used in the corresponding prime or target. The auditory target was presented 16 milliseconds after the onset of the backward mask which remained on the screen until the end of the trial.

Participants were instructed to decide as quickly and as accurately as possible whether the spoken stimulus (the target) was an English word by pressing the Yes or No button on an electronic response box connected to the computer. After the experiment, participants were debriefed to evaluate prime visibility. Most of the participants were either not aware of the presence of the prime (57%) or could not derive any useful information about the prime (30%), with only 13% indicating that they were able to see “a word” in the prime position at least sometimes. This confirms that any priming effect observed in this experiment should be due to the processing of the critical items, and not to task-related strategies or deliberate decision making.

2.2.4. Participants' vocabulary knowledge

Because readers' existing L2 vocabulary knowledge plays a role in the learning of new words (e.g., Elgort, Perfetti, Rickles, & Stafura, 2014), the participants' vocabulary test scores were included in the analyses of their explicit and implicit word knowledge as secondary interest predictors. The participants' vocabulary knowledge in English was measured using *LexTALE* (Lemhöfer & Broersma, 2012; www.lextale.com) and a vocabulary levels test (*VLT*) of controlled productive ability (Laufer & Nation, 1999). *LexTALE* is a Yes-No vocabulary test that includes real and nonce words. It was included in the modelling of response accuracy in the lexical decision (see section 2.3). Productive *VLT* is a cloze test; test-takers fill in missing words in high-constraining sentences. This test measures participants' ability to retrieve a known word form from memory using strong contextual cues about its meaning (i.e., meaning-form mapping). Participants' *VLT* scores on the 2000 word-frequency level were included in the analysis of the meaning generation task. The tests were administered at the end of the data collection procedure. The participants' average productive *VST* score on the 2000 level was 75% (13.5 out of 18) and their average *LexTALE* score was 44%.

2.3. Analysis

Linear mixed-effects modelling was used in the data analysis (Bates & Sarkar, 2010; Bates, 2011). All analyses included participants and items as crossed random effects. Random slopes were fitted for fixed effects, as appropriate. A *Generalized Linear Mixed Model* was fitted to the

response accuracy data from the meaning generation task (Jaeger, 2008). The RPT data analysis was conducted after excluding the filler trials and trials with nonword targets. Non-dichotomous variables that were not normally distributed were transformed to bring them closer to normal distribution; they were also centred using the *scale* function in R to avoid multicollinearity (Belsley, Kuh & Welsch, 1980). A minimally adequate statistical model was fitted to the data, using a stepwise variable selection and the likelihood ratio test for model comparisons (Baayen, Davidson & Bates, 2008). The resulting statistical models 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.¹

In the analysis of the meaning generated task, response accuracy (correct/incorrect) was used as the dependent variable and inference correctness was entered as a primary predictor with three levels (correct, incorrect and no inference entered). The participants' *VLT*(2k) score was used in the analysis as a secondary-interest predictor.

In the analysis of the primed lexical decisions, the following questions were examined: (1) whether mixed-modality repetition priming was observed for the newly learned items (but not for the nonwords) and whether inference correctness affected this priming; and (2) whether incorrect inferences negatively affect the accuracy and latency of lexical decisions on the critical items used as targets (across repeated and unrelated trials). The accuracy of responses and response times (RT) were the dependent variables in the corresponding analyses of the RPT data. In both analyses, inference correctness was used as a primary-interest predictor; vocabulary knowledge (the *LexTALE* score) was used as a secondary-interest predictor in the response accuracy analysis. In addition to being an index of participants' vocabulary knowledge, *LexTALE* score acts as a covariate, accounting for the individual variability in their approaches to making word-nonword decisions. The RT analysis also included number of letters (item length), RT on the preceding trial and response accuracy, as additional variables.²

3. Results

On the first attempt to infer meanings of the critical items during the learning phase, 61.5% of inferences were correct, 29.2% were incorrect and, for 9.3%, no inference was entered. On the second attempt, 63.4% were correct, 28.6% were incorrect and, for 8%, no inference was entered.

3.1. The meaning generation task

On average, the participants were able to correctly retrieve meanings of about 15% of the critical items in the meaning generation task. There was no statistically significant effect of inference correctness on the first attempt on the accuracy of meaning retrieval. However, there was a reliable effect of inference correctness on the second attempt (**Table 1**), such that when the meaning of the critical word was inferred incorrectly or no attempt was entered, the resulting knowledge was inferior to that when the meaning was inferred correctly ($z = -2.74$,

$p = 0.006$ and $z = -1.96$, $p = 0.051$, respectively). The meaning generation scores were, on average, about 7% less accurate if the inference entered was incorrect and about 10% less accurate if an inference was not entered (**Figure 1A**). There was also a reliable effect of existing vocabulary knowledge, such that the participants with higher VLT scores were more accurate in generating the meaning of the critical words ($z = 3.62$, $p = 0.0003$).

Finally, there was a significant interaction between inference correctness and vocabulary knowledge ($z = 2.16$, $p = 0.031$). For the participants with smaller vocabularies, making a correct meaning inference resulted in a higher probability of retrieving correct meanings at test, compared to when the inference was incorrect or no inference was entered. Interestingly, making an incorrect inference was no worse than making no inference (**Figure 1B**). However, the negative effect of incorrect inferences diminished with an increase in the L2 vocabulary knowledge, and no difference

in the meaning generation score was observed for participants with highest vocabulary scores whether their recorded meaning inferences were correct or incorrect. Conversely, the negative effect of not entering an inference remained even for the most advanced participants (**Figure 1B**).

3.2. The mixed-modality masked repetition priming task (RPT)

The main finding in the RPT was that repetition priming was not affected by incorrect inferences on either the first or second attempt (**Table 2** and **3**). A robust priming effect was observed in the response accuracy ($z = 4.04$, $p = 5.3e-05$) and latency ($t = -2.42$, $p = 0.022$) analyses (**Figure 2**), and there were no reliable interactions between priming and inference correctness in either analysis. This suggests that contextual word learning progressed whether participants' explicit meaning inferences were correct or incorrect.³

Table 1: Analysis of responses in the meaning generation task (fixed effects).

	Coef. β	SE(β)	z	p
(Intercept)	-1.80	0.22	-8.09	6.0e-16
Inference2Att=Incorrect	-0.79	0.29	-2.74	.006
Inference2Att=No	-0.91	0.46	-1.96	.051
Vocabulary*	3.73	1.03	3.61	3.0e-04
Inf2Att=Incorrect:Vocabulary	3.45	1.60	2.16	.031
Inf2Att=No:Vocabulary	-0.15	2.58	-0.06	.954

Intercept levels: Inference on 2nd attempt=Correct.

*Vocabulary: Log-transformed, centred productive VLT scores.

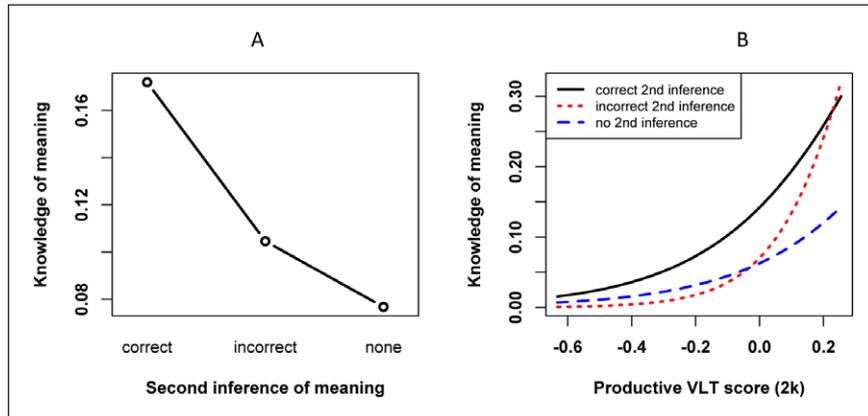


Figure 1: Analysis of responses in the meaning generation task. Panel A: effect of inference correctness; Panel B: interaction between inference correctness and vocabulary knowledge.

Table 2: RPT: Response accuracy analysis (fixed effects).

	Coef. β	SE(β)	z	p
(Intercept)	1.21	0.25	4.93	8.2e-07
Condition=Related	0.66	0.16	4.04	5.3e-05
Inference2Att=Incorrect	-0.19	0.20	-0.94	.345
Inference2Att=No	-0.52	0.31	-1.70	.089
Vocabulary*	4.87	1.45	3.36	.001

Intercept levels: Condition=Unrelated; Inference on 2nd attempt=Correct.

*Vocabulary: log-transformed, centred LexTALE scores.

Meaning inferences on the second attempt had a weak effect on the accuracy of lexical decisions to the critical items (across related and unrelated trials); the responses were somewhat less accurate when no inference was entered compared to the correct inferences ($z = -1.70$, $p = 0.089$), although this difference did not reach statistical significance. No response accuracy effect was observed for correct versus incorrect inferences (Table 2, Figure 3A).

In the RT analysis (across related and unrelated trials), there was a weak effect of meaning inference correctness on the first (but not second) attempt, but this effect was not statistically significant. Lexical decisions to the critical items were somewhat slower when the item's meaning was inferred incorrectly ($t = 1.55$, $p = 0.106$) or was not inferred ($t = 1.68$, $p = 0.135$), compared to when it was inferred correctly (Table 3, Figure 3B).

4. Discussion and Conclusions

The present study examined the effect of meaning inferences on contextual learning of novel L2 words. Critical vocabulary items were presented in single-sentence contexts that provided reasonable opportunities for the learner to generate correct meaning inferences. After the contextual exposure, participants reviewed correct dictionary-style definitions of the critical items. Time-on-task, number of encounters with critical items, and the spacing of repetition were controlled.

The results show that students were able to infer meanings of over half of the critical items correctly during reading but, on post-test, the meanings of only for 15% of the items were retrieved correctly. Although some previous contextual L2 word learning studies show similarly low gains (e.g., Waring & Takaki, 2003;

Table 3: RPT: RT analysis (fixed effects).

	Coef. β	SE(β)	t	p
(Intercept)	-0.58	0.02	-35.80	1.0E-04
Condition=Related	-0.02	0.01	-2.42	0.022
Inference1Att=Incorrect	0.01	0.01	1.55	0.106
Inference1Att=No	0.02	0.01	1.68	0.135
NoL*	0.03	0.01	2.58	0.006
Response accuracy=correct	-0.05	0.01	-5.81	1.0E-04
RT on preceding trial**	0.20	0.04	5.43	1.0E-04

Intercept levels: Condition=Unrelated; Inference on 1st attempt=Correct; Response accuracy=incorrect.
 *NoL: number of letters, centred; **Inverse transformed, centred RT on the preceding trial.

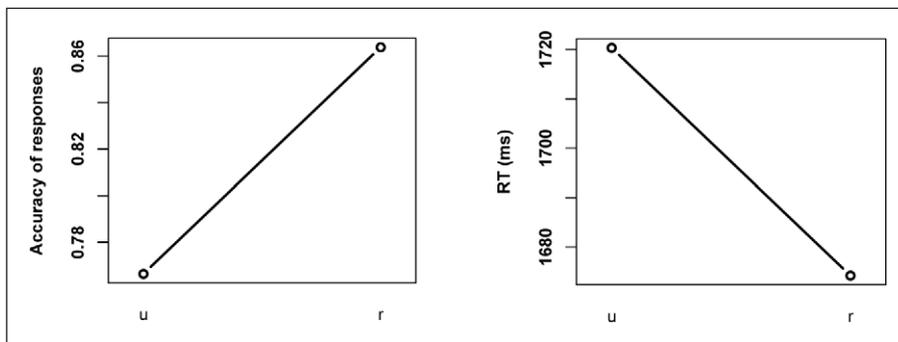


Figure 2: RPT analysis. The priming effect is the difference between the result in the related (r) and unrelated (u) conditions.

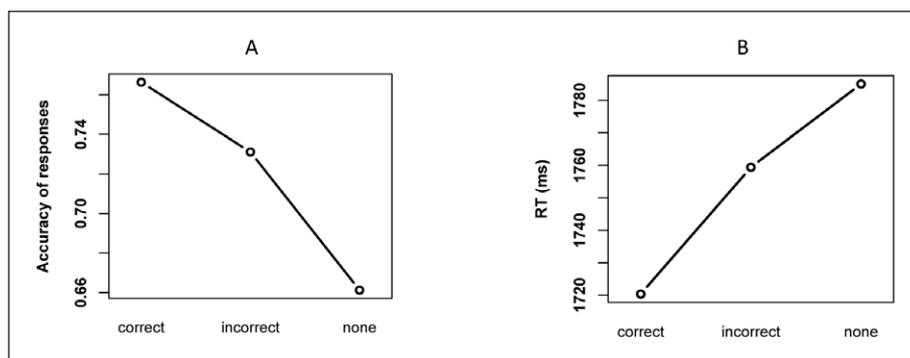


Figure 3: RPT. Panel A: effect of inference correctness on 2nd attempt on response accuracy; Panel B: effect of inference correctness on 1st attempt on RTs.

Zahar, Cobb & Spada, 2001), the following features of the present study may have contributed to this finding: (1) participants' proficiency level and their L1 (Chinese) that is considered distant from the target language; (2) the burden of learning new word forms and concepts, rather than mapping new L2 labels to familiar concepts; and (3) the nature of the meaning generation test, in which critical items were presented in neutral context that did not provide clues about their meanings.

Although the participants tended to be less accurate in the meaning generation task when their second contextual inference was incorrect (compared to when it was correct), the negative effect of incorrect inferences was no worse than that of not providing an inference. Furthermore, meaning generation scores of participants with larger L2 vocabularies were not negatively affected by erroneous inferences; conversely, producing no inference resulted in lower scores. This suggests that negative effects of making incorrect inferences on word learning from reading diminish as L2 readers' proficiency increases.

Auditory lexical decisions to the contextually learned items were marginally less accurate and slower (compared with the correctly inferred items) when no inference was registered; incorrect inferences had no effect on the accuracy of lexical decisions but made responses marginally slower. Importantly, accuracy of meaning inferences during reading had no effect on the mixed-modality masked repetition priming, operationalised in this study as a measure of implicit word knowledge. This suggests that lexical representations were established for the critical L2 items irrespective of the accuracy of explicit attempts to infer their meanings. Thus, when unfamiliar L2 words occur in informative sentence contexts, explicit incorrect meaning inferences during reading have some negative effect on the establishment of explicit form-meaning mapping for lower proficiency participants, but they appear to be benign as far as the development of implicit knowledge and establishment of lexical representations are concerned.

These results have important implications for vocabulary research; they show that the choice of measures affects findings in word learning studies, especially at early stages of learning. This is because different aspects of word knowledge may have different learning trajectories. The ability to explicitly articulate an accurate core meaning for a novel word may take longer to develop in contextual learning (even after consulting a dictionary), but the development of its lexical representation can be underway from the first contextual encounter, whether or not the reader is able to explicitly articulate an accurate meaning inference.

Taken together, the results of the present study confirm the hypothesis that explicit meaning inferences during L2 reading do not necessarily predict the development of implicit word knowledge. Implicit lexical knowledge is likely to develop with each informative contextual encounter, as a by-product of the co-occurrence of the new word with known words and by virtue of the new word assuming a specific grammatical and thematic role in a sentence (Ferretti, McRae & Hatherell, 2001; Landauer & Dumais, 1997). Nevertheless, making an effort to infer

word meanings from context appears to be beneficial, compared to not doing so, at least for the development of explicit word knowledge. A possible reason for the boost provided by an attempt to infer meanings from context is that it brings readers' attention to the immediate and larger sentence context, facilitating their engagement with contextual cues.

Based on the study findings, a recommendation can be made that L2 readers attempt to infer meanings of unfamiliar words from context, without being overly concerned about making explicit incorrect guesses. Even when initial guesses are not fully on target, the act of guessing the meaning from context seems to contribute to the incremental establishment of lexical representation which can be fine-tuned with future encounters.

4.1. Limitations

This study examined a contextual word-learning scenario that creates favourable learning conditions: the critical items were pre-identified in the sentences and presented in informative contexts, an option to listen to each item was provided during the first contextual encounter, and dictionary-type definitions were presented after the contextual exposure. It is possible that, under less favourable conditions, incorrect inferences may have a larger negative effect on learning. Future research is needed to evaluate the effect of incorrect inferences on contextual word learning outside of the laboratory, under less favourable conditions. Importantly, measures of implicit knowledge should be used alongside more traditional measures of explicit knowledge, in order to charter incremental development of word knowledge from reading. For example, semantic priming can be used to further investigate the development of lexical semantic representations of new words from reading.

Additional File

The Additional file for this article can be found as follows:

- **Appendix A.** List of Stimuli and Their Definitions.
DOI: <https://doi.org/10.22599/jesla.3.s1>

Notes

- ¹ In the original modelling of the data, word type (nonce word vs. low-frequency word) had been entered into the models, but it was removed from the final (reported) models because it had no reliable effect and did not improve model fit.
- ² Incorrect responses were not removed from the RT analysis; instead, response accuracy was included in the analysis as an independent variable.
- ³ In a separate analysis conducted with nonword targets, no priming was found in the accuracy or RT analyses.

References

- 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. DOI: <https://doi.org/10.1016/j.jml.2007.12.005>

- Baddeley, A., & Wilson, B. A.** (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*, *32*, 53–68. DOI: [https://doi.org/10.1016/0028-3932\(94\)90068-X](https://doi.org/10.1016/0028-3932(94)90068-X)
- Bates, D. M.** (2011). Linear mixed model implementation in lme4. Retrieved from: <http://cran.r-project.org/web/packages/lme4/vignettes/Implementation.pdf>.
- Bates, D. M., & Sarkar, D.** (2010). lme4: Linear mixed-effects models using S4 classes. Retrieved from: <http://CRAN.R-project.org/package=lme4>, R package version 0.999375–33.
- Belsley, D. A., Kuh, E., & Welsch, R. E.** (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley. DOI: <https://doi.org/10.1002/0471725153>
- Bensoussan, M., & Laufer, B.** (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, *7*, 15–32. DOI: <https://doi.org/10.1111/j.1467-9817.1984.tb00252.x>
- Boers, F., Dang, T. C., & Strong, B.** (2016). Comparing the effectiveness of phrase-focused exercises. A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*. DOI: <https://doi.org/10.1177/1362168816651464>
- Boers, F., Demecheleer, M., Coxhead, A., & Webb, S.** (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research*, *18*, 54–74. DOI: <https://doi.org/10.1177/1362168813505389>
- 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. DOI: <https://doi.org/10.1080/01638530701792826>
- Borovsky, A., Kutas, M., & Elman, J.** (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*, 289–296. DOI: <https://doi.org/10.1016/j.cognition.2010.05.004>
- Bridger, E. K., & Mecklinger, A.** (2014). Errorful and errorless learning: The impact of cue-target constraint in learning from errors. *Memory & Cognition*, *42*, 898–911. DOI: <https://doi.org/10.3758/s13421-014-0408-z>
- Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R.** (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin & Review*, *19*, 81–86. Retrieved from: <http://search.proquest.com/helicon.vuw.ac.nz/docview/1021964881?accountid=14782>. DOI: <https://doi.org/10.3758/s13423-011-0185-7>
- Chaffin, R., Morris, R. K., & Seely, R. E.** (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 225–235. DOI: <https://doi.org/10.1037/0278-7393.27.1.225>
- Clare, L., & Jones, R. S. P.** (2008). Errorless learning in the rehabilitation of memory impairment: A critical review. *Neuropsychology Review*, *18*, 1–23. DOI: <https://doi.org/10.1007/s11065-008-9051-4>
- Elgort, I., Perfetti, C., Rickles, B., & Stafura, J.** (2014). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, *30*, 506–528. DOI: <https://doi.org/10.1080/23273798.2014.942673>
- Elgort, I., & Warren, P.** (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414. DOI: <https://doi.org/10.1111/lang.12052>
- Ferretti, T. R., McRae, K., & Hatherell, A.** (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516–547. DOI: <https://doi.org/10.1006/jmla.2000.2728>
- Forster, K. I.** (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research*, *27*, 203–233. DOI: <https://doi.org/10.1023/A:1023202116609>
- Forster, K. I., Mohan, K., & Hector, J.** (2003). The mechanics of masked priming. In: Kinoshita, S., & Lupker, S. J. (Eds.), *Masked Priming: The State of the Art*, 3–37. Hove, UK: Psychology Press.
- Frantzen, D.** (2003). Factors affecting how second language Spanish students derive meaning from context. *Modern Language Journal*, *87*, 168–199. Retrieved from: <http://www.jstor.org/helicon.vuw.ac.nz/stable/1193032>. DOI: <https://doi.org/10.1111/1540-4781.00185>
- Friskhoff, G. A., Collins-Thompson, K., Hodges, L., & Crossley, S.** (2016). Accuracy feedback improves word learning from context: evidence from a meaning-generation task. *Reading and Writing*, *29*, 609–632. DOI: <https://doi.org/10.1007/s11145-015-9615-7>
- Grainger, J., Diependaele, K., Spinelli, E., Ferrand, L., & Farioli, F.** (2003). Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *29*, 1256–1269. DOI: <https://doi.org/10.1037/0278-7393.29.6.1256>
- Grainger, J., & Ferrand, L.** (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory & Language*, *33*, 218–233. DOI: <https://doi.org/10.1006/jmla.1994.1011>
- Haastруп, K.** (1991). *Lexical Inferencing Procedures or Talking about Words*. Tübingen: Gunter Narr Verlag.
- Hu, H. M., & Nation, I. S. P.** (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403–430. Retrieved from: <http://nflrc.hawaii.edu/rfl/PastIssues/rfl131hsuehchao.pdf>.
- Hulstijn, J. H.** (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In: Arnaud, P. J. L., & Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*, 113–125. London: Macmillan. DOI: https://doi.org/10.1007/978-1-349-12396-4_11
- Hulstijn, J. H.** (2001). Intentional and incidental second-language vocabulary learning: Are appraisal of elaboration, rehearsal and automaticity. In: Robinson, P. (Ed.), *Cognition and Second Language Instruction*, 258–286. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/cbo9781139524780.011>

- 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. DOI: <https://doi.org/10.2307/329439>
- 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. DOI: <https://doi.org/10.1016/j.jml.2007.11.007>
- Joordens, S., & Becker, S.** (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *2–1*, 1083–1105. DOI: <https://doi.org/10.1037/0278-7393.23.5.1083>
- Karpicke, J. D., & Roediger, H. L.** (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *33*, 704–719. DOI: <https://doi.org/10.1037/0278-7393.33.4.704>
- Karpicke, J. D., & Roediger, H. L.** (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968. DOI: <https://doi.org/10.1126/science.1152408>
- Kelly, P.** (1990). Guessing: No substitute for systematic learning of lexis. *System*, *18*, 199–207. DOI: [https://doi.org/10.1016/0346-251X\(90\)90054-9](https://doi.org/10.1016/0346-251X(90)90054-9)
- 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. Retrieved from: <http://search.proquest.com/helicon.vuw.ac.nz/docview/614361064?accountid=14782>. DOI: <https://doi.org/10.1037/0033-295X.104.2.211>
- Laufer, B.** (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In: Coady, J., & Huckin, T. (Eds.) *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, 20–34. Cambridge University Press.
- Laufer, B.** (2005). Focus on form in second language vocabulary acquisition. In: Foster-Cohen, S. H., Garcia-Mayo, M. P., & Cenoz, J. (Eds.), *EUROSLA Yearbook*, *5*, 223–250. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/eurosla.5.11lau>
- Laufer, B., & Nation, P.** (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*, 33–51. DOI: <https://doi.org/10.1177/026553229901600103>
- 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: <https://doi.org/10.3758/s13428-011-0146-0>
- Masson, M. E. J.** (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *21*, 3–23. DOI: <https://doi.org/10.1037/0278-7393.21.1.3>
- McLaughlin, J., Osterhout, L., & Kim, A.** (2004). Neural correlates of second language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704. DOI: <https://doi.org/10.1038/nn1264>
- Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F.** (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1858–1866. DOI: <https://doi.org/10.1093/cercor/bhl094>
- Middleton, E. L., & Schwartz, M. F.** (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation*, *22*, 138–168. DOI: <https://doi.org/10.1080/09602011.2011.639619>
- Mondria, J.-A.** (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. DOI: <https://doi.org/10.1017/S0272263103000202>
- Mondria, J.-A., & Wit-de Boer, M.** (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics*, *12*, 249–267. DOI: <https://doi.org/10.1093/applin/12.3.249>
- Nation, I. S. P., & Webb, S.** (2011). *Researching and Analyzing Vocabulary*. Boston, MA: Heinle.
- Neely, J. H.** (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: Besner, D., & Humphreys, G. W. (Eds.), *Basic Processes in Reading: Visual Word Recognition*, 264–336. Hillsdale, NJ: Erlbaum.
- 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. Retrieved from: <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/reading-and-incidental-l2-vocabulary-acquisition/D53AF9CDB587AC544BA9C09330C45766>. DOI: <https://doi.org/10.1017/S027226319900203X>
- Pellicer-Sánchez, A.** (2015). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, 1–34. DOI: <https://doi.org/10.1017/S0272263115000224>
- Pressley, M., Levin, J. R., & McDaniel, M. A.** (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. In: McKeown, M. G., & Curtis, M. E. (Eds.), *The Nature of Vocabulary Acquisition*, 107–127. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rodriguez-Fornells, A., Kofidis, C., & Munte, T. F.** (2004). An electrophysiological study of errorless learning. *Cognitive Brain Research*, *19*, 160–173. DOI: <https://doi.org/10.1016/j.cogbrainres.2003.11.009>
- Schmitt, N.** (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363. DOI: <https://doi.org/10.1177/1362168808089921>
- Schouten-van Parreren, C.** (1992). Individual differences in vocabulary acquisition: A qualitative experiment in the first phase of secondary education. In: Arnaud, P. J. L., & Bejoint, H. (Eds.), *Vocabulary and Applied Linguistics*, 94–101. London: Macmillan.

- Shute, V. J.** (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189. DOI: <https://doi.org/10.3102/0034654307313795>
- Trenkic, D., & Warmington, M.** (2014, September). Improving Vocabulary Learning in Second Language Speakers: An Errorless Learning Approach. *Paper presented at EUROSLA 24 Conference*. York England.
- Waring, R., & Takaki, M.** (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15, 130–163. Retrieved from: <http://nflrc.hawaii.edu/rfl/October2003/waring/waring.html>.
- Warmington, M., & Hitch, G. J.** (2014). Enhancing the learning of new words using an errorless learning procedure: Evidence from typical adults. *Memory*, 22, 582–94. DOI: <https://doi.org/10.1080/09658211.2013.807841>
- Warmington, M., Hitch, G. J., & Gathercole, S. E.** (2013). Improving word learning in children using an errorless technique. *Journal of Experimental Child Psychology*, 114, 456–465. DOI: <https://doi.org/10.1016/j.jecp.2012.10.007>
- Webb, S.** (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46–65. DOI: <https://doi.org/10.1093/applin/aml048>
- Webb, S.** (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20, 232–245. Retrieved from: <http://nflrc.hawaii.edu/rfl/October2008/webb/webb.html>.

How to cite this article: Elgort, I. (2017). Incorrect inferences and contextual word learning in English as a second language. *Journal of the European Second Language Association*, 1(1), 1–11, DOI: <https://doi.org/10.22599/jesla.3>

Submitted: 24 December 2016

Accepted: 19 June 2017

Published: 01 August 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.