HHS Public Access
Author manuscript
Lang Cogn Neurosci. Author manuscript; available in PMC 2016 June 01.

Published in final edited form as:
Lang Cogn Neurosci. 2015 June 1; 30(5): 506–528. doi:10.1080/23273798.2014.942673.

Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and ERP indicators of learning*

Irina Elgort,
Victoria University of Wellington, 6140, New Zealand

Charles A. Perfetti,
University of Pittsburgh, Pittsburgh, PA 15260

Ben Rickles, and
University of Pittsburgh, Pittsburgh, PA 15260

Joseph Z. Stafura
University of Pittsburgh, Pittsburgh, PA 15260

Abstract

New word learning occurs incidentally through exposure to language. Hypothesizing that effectiveness of contextual word learning in a second language (L2) depends on the quality of existing lexical semantic knowledge, we tested more and less proficient adult bilinguals in an incidental word learning task. One day after being exposed to rare words in an L2 (English) reading task, the bilinguals read sentences with the newly-learned words in the sentence-final position, followed by related or unrelated meaning probes. Both proficiency groups showed some learning through faster responses on related trials and a frontal N400 effect observed during probe word reading. However, word learning was more robust for the higher-proficiency group, who showed a larger semantic relatedness effect in unfamiliar contexts and a canonical N400 (central-parietal). The results suggest that the ability to learn the meanings of new words from context depends on the L2 lexical semantic knowledge of the reader.

Keywords

contextual word learning; L2 proficiency; N400; bilingualism; reading

Increasing vocabulary size and quality of word knowledge is an important goal of learning a second or foreign language (L2). For adult L2 learners, reading affords the opportunity for learning the meanings of words. How the reading processes actually lead to the development of memory traces representing new word knowledge is not completely clear, nor is the nature of the lexical semantic content of this knowledge. Do contextual encounters with an unfamiliar word in an adequately constraining sentence context lead to long-term

Address for correspondence: Dr Irina Elgort Victoria University of Wellington P.O. Box 600, Wellington, 6140, New Zealand. Phone: +64 4 4635970. Fax: +64 4 463 5284. irina.elgort@vuw.ac.nz.

modification of semantic memory? Does the trajectory of contextual learning depend on the knowledge and skill of the reader, such as the L2 proficiency? These questions are investigated in the present study that examines the effect of reader L2 (English) proficiency on the outcomes of early contextual L2 word learning.

Contextual word learning

Learning words from context involves language experiences that can establish memories for unfamiliar words and the context in which they occur. These processes are described in the instance-based framework of word learning (Bolger, Balass, Landen, & Perfetti, 2008), based on Reichle and Perfetti's (2003) adaption of Hintzman's (1986) memory model. (See Nadel & Moscovitch, 1997, for an alternative, multiple trace theory of memory consolidation). In the instance-based framework, an encounter with a word in various contexts results in episodic memories of the word learning events, i.e., the word plus its contexts. Each encounter results in its own episodic traces. Aspects of the episodic traces that are similar across encounters (e.g., modality, co-occurrence with other words, successful meaning inferences) are strengthened, while aspects that are specific to individual episodes (e.g., specific sentence and text contexts; incorrect meaning inferences) are not strengthened. Experiencing a new word in a range of informative contexts facilitates the establishment and consolidation of its lexical semantic representation, as a result of its co-occurrence with known words (Burgess & Lund, 1997; Landauer & Dumais, 1997). Eventually, with experience, the word's core meaning is abstracted from multiple episodic memories, and can be accessed independently from the original contexts.

Successful contextual learning depends on the quality of the linguistic context in which new words are encountered, with little or no contextual learning commonly observed for low-constraint or misleading (inconsistent) contexts (Batterink & Neville, 2011; Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2013; Frishkoff, Perfetti, & Collins-Thompson, 2010, 2011; Mestres-Missé, Rodríguez-Fornells, Münte, 2007). In supportive (informative) contexts, when the word meaning is constrained by the surrounding context, a learner may be able to quickly infer the meaning, sometimes from a single exposure. However, this initial knowledge is incomplete and fragile (susceptible to changes and adjustments brought about by additional contextual exposures). Multiple contextual exposures are needed for a robust lexical semantic representation to be established from reading.

Robust semantic learning appears to benefit more from exposures to a word in varied than in repeated (same) contexts (Bolger et al., 2008). Varied contexts promote richer semantic associations, help reject false inferences and encourage the establishment of new semantic features (Beck, McKeown, & Kucan, 2002; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009). Repeating previously-encountered contexts, on the other hand, reinforces memories of specific learning episodes, and is less likely to support the abstraction of meaning. In summary, contextual word learning is a slow, incremental process, viewed as a longitudinal progression towards context-independent lexical semantic memory representations.

The process of establishing robust lexical semantic representations is supported by inter-word connections. By sharing semantic features with other words, a new word becomes part of an existing lexical semantic network (Masson, 1995; McRae, de Sa, & Seidenberg, 1997; Plaut & Booth, 2000). Presumably, learners at lower L2 proficiencies have fewer and weaker L2 lexical semantic connections available and are, therefore, likely to be less effective in establishing L2 lexical semantic representations after initial incidental encounters with novel L2 words during reading.

Effects of second language proficiency

Previous L2 studies found better contextual learning of word meanings by learners with higher vocabularies (Elgort & Warren, 2014; Ferrel Tekmen & Daloğlu, 2006; Horst, Cobb, & Meara, 1998; Pulido & Hambrick, 2008). This learning advantage may be related to better text comprehension (Pulido, 2007), since vocabulary knowledge and reading comprehension are highly correlated (Jeon & Yamashita, 2014; see Ouellette, 2006, for a similar relationship between vocabulary knowledge and comprehension in L1).

L2 proficiency effects extend beyond learning, to words already "known", as revealed by asymmetries in the ways novice and advanced bilinguals access meanings of L2 words (Finkbeiner, Forster, Nicol, & Nakamura, 2004; Kroll & Dijkstra, 2002; Kroll, Michael, Tokowicz, & Dufour, 2002; Wang & Forster, 2010). Phillips, Segalowitz, O'Brien and Yamasakia (2004) also found that semantic categorization times for L2 (French) words were more variable for less proficient- than more-proficient bilinguals, suggesting that semantic processing becomes more automatic at higher L2 proficiency.

Neuro-linguistic studies (using fMRI and ERP) also show lexical proficiency effects in lexical semantic retrieval and processing. Ardal et al. (1990) compared an ERP signature of semantic processing, the N400 component, and an accompanying frontal negativity in the first and second language of adult bilinguals (and monolinguals), using congruous and incongruous sentence contexts. Longer N400 latencies in participants' L2 (compared to their L1) and a significant difference in amplitude between monolinguals and bilinguals at parietal locations were observed. The effect of incongruence at frontal locations was functionally similar to the parietal N400, but it was diminished for the L2. The researchers explained the differences in the latency of the N400 effect by less automatic processing in the L2 compared to the L1, suggesting that "N400 latency varies with the relative degree of automaticity achieved in a given language" (Ardal et al., 1990, p. 201). The reduced effect observed at frontal sites in the bilinguals' L2 was discussed with reference to their relative fluency of processing in the two languages (which corresponded to their self-reports of being more fluent in the L1 than in the L2). Furthermore, the subgroup of highly fluent bilinguals did not show any reduction in the late negativity in their L2. Later N400 peak latencies, longer durations, and smaller effects were also reported for the less proficient language of bilinguals by Kutas and Kluender (1994; see Rodríguez-Fornells et al., 2009, pp. 3717-3719, for an overview).

In a semantic categorization study, Phillips et al. (2004) found that the N400 distinguished lower- from higher-proficiency L2 learners when processing associatively related words. Kotz and Elston-Güttler (2004) found similar N400 proficiency differences in semantic category priming. Furthermore, differences between the ERP signature of L1 and L2 lexical processing are modulated by L2 proficiency (Midgley, Holcomb, & Grainger, 2009; Newman et al., 2012), and fMRI studies show that "additional brain activity, mostly in prefrontal areas" are involved at lower L2 proficiencies (but not higher L2 proficiencies) in L2 compared to L1 word retrieval (Abutalebi, 2008, p. 471).

The study

Here we focus on the outcomes of incidental learning exposures to new words in L2 contexts. We explore whether the establishment of lexical semantic representations is affected by the learner L2 lexical proficiency, when two known predictors of learning – the lexical difficulty (i.e., percentage of low frequency words in the learning context) and the contextual constraint (i.e., ease of deriving the meaning of novel words from context) – are controlled for. We test this hypothesis through a combination of behavioural and ERP indicators of word learning.

The initial learning phase exposed two groups of adult bilinguals (less and more proficient) to rare L2 (English) words embedded in three sentences that supported participants' ability to guess their meanings from context (e.g., "She hung her head in ignominy while the principal told her off for bad behavior The unacceptable actions of their youngest son brought ignominy to his family."; "*In the last 5 minutes the team scored*."; "an important goal, avoiding an almost certain ignominy."). One day later, these bilinguals read L2 sentences containing the newly-learned (critical) words in the sentence-final position. These sentences were either the same as in the learning phase (old) or different, i.e., not previously encountered by the participants (new). In the new sentence condition, critical words were either congruous (e.g., "After losing all his property, the man left his city in ignominy.") or incongruous with the sentence frame (e.g., "I thought it was curious that he wanted to look like a Mexican ignominy.") Each sentence was followed by a single-word meaning probe – a word either related or unrelated to the meaning of the critical newly-learned word in the sentence-final position (e.g., *ignominy – HUMILIATION; ignominy – INTRUSION*). Participants were instructed to perform a semantic relatedness task, i.e., decide as quickly and as accurately as possible if the sentence-final word and the probe word were related in meaning (Gernsbacher, Varner, & Faust, 1990, Exp. 4). This experimental paradigm created conditions to explore the quality of meaning representations established for the contextually-learned L2 words: 1) the semantic judgment task probed participants' ability to fluently access the meaning of the new words after an additional contextual exposure; 2) presenting these words in familiar vs. unfamiliar sentence contexts primed either their episodic or semantic representations; and 3) presenting these words in incongruous vs. congruous contexts provided a test of the stability of the newly-established lexical semantic representations, and of the involvement of cognitive control in the processing of word meanings.

Learning measures and predictions

To allow fine grain views of the outcome of learning, we measured the decision times (RT) and the ERP responses (N400) when participants judged the semantic relatedness of newly-learned critical word and its meaning probe. Such judgments rely on semantic features (consistent patterns of activation on the neurological level) shared between the words being judged (Cree, McRae, & McNorgan, 1999; Hinton & Shallice, 1991; McRae et al., 1997), with faster responses observed on semantically related compared to unrelated word pairs (the *semantic relatedness effect*). This semantic relatedness effect results from the interactive semantic activation that spreads both forward (from the first to the second member of the pair) and backwards (from the second member to the first) (Balota, Boland, & Shields, 1989; Kiger & Glass, 1983; Koriat, 1981).

Because the amount of the semantic overlap between the two words depends on the quality of their lexical semantic representations, the behavioural effect (the difference in RTs between related and unrelated pairs) should depend on the extent to which contextual learning has established semantic features for the newly-learned word that overlap those of the probe word. When the relevant semantic features have been established, the process of semantic judgments reflects semantic memory. At the early stages of incidental contextual learning, memory representations for the new words are predicted to be contextually-bound (episodic), and only some initial lexical semantic features are likely to be extracted. The quality of the additional contextual exposure on day two, therefore, critically affects the trajectory of learning.

Encountering the critical word in a new supportive context facilitates the abstraction of meaning because it primes (and thus reinforces) already-established semantic features of the newly-learned word, and may add new features. Therefore, presenting critical words in new congruous contexts is expected to boost the semantic relatedness effect. When the context on day two matches that of the initial learning on day one, episodic traces of the prior exposure are reinforced, potentially delaying the meaning abstraction process (Bolger et al., 2008). For this (old) condition, the semantic judgment process is likely to mostly retrieve an episodic memory trace to compare. A comparison based primarily on episodic memory may produce no (or a reduced) semantic relatedness effect on the meaning judgement task.

When the context is misleading, as a result of being incongruous with the newly- learned word, contextually supported retrieval of the previously established semantic features becomes problematic. This exposure may also result in false inferences about the meaning of the critical word, sending learners on an erroneous learning path. Since the meaning of the newly-learned word presented in the incongruous condition is more difficult to retrieve, semantic judgments involving these words are also more difficult. A greater degree of cognitive control will be required to overcome the semantic interference exuded by the incongruous context, in order to make correct semantic judgements. Thus, no (or a reduced) semantic relatedness effect is predicted when the newly-learned words are embedded in incongruous contexts.

ERPs also provide evidence of semantic processing in the N400 amplitudes elicited by meaning probes. The N400 component (usually observed at central and parietal scalp regions) signals variation in meaning congruence: a word that is preceded by single-word or sentence contexts that are incongruous elicits a greater negativity about 400 ms after exposure than words that are more congruous with preceding contexts (Anderson & Holcomb, 1995; Kutas & Federmeier, 2000; Kutas & Hillyard, 1980; Nobre & McCarthy, 1994; Van Berkum, Hagoort, & Brown, 1999). The N400 has been used in L1 word learning studies as an index of semantic congruence of meaning probes and related newly-learned words, i.e., as a measure of word learning (e.g., Balass, Nelson, & Perfetti, 2010, Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2013; Frishkoff, Perfetti, & Collins-Thompson, 2010; Mestres-Missé, Rodríguez-Fornells, Münte, 2007; Perfetti, Wlotko, & Hart, 2005), In L2, N400 has been used in studies investigating L2 lexical semantic representations and processing (Kotz, 2001; Kotz & Elston—Güttler, 2004; Midgley et al., 2009; Newman et al., 2012; Phillips et al., 2004) and a few word learning studies (McLaughlin, Osterhout, & Kim, 2004; Ojima, Nakata, & Kakigi, 2005).

In the present semantic relatedness experiment, the N400 reduction on related trials depends on whether individual learners accessed the meaning of the newly-learned words. Furthermore, since the effects of learner L2 proficiency and linguistic and text difficulty on contextual word learning may be confounded, the effect of proficiency is studied under learning conditions where both lexical difficulty and contextual constraint are controlled for. If the N400 effect observed for higher- and lower-proficiency bilinguals is qualitatively and/or quantitatively different in this study for the two proficiency groups, we can be reasonably confident that contextual word learning is affected by the developmental state of the learner's L2 lexical knowledge. In other words, the study tests whether the process of semantic learning is restricted by L2 lexical proficiency.

Besides the canonical central-parietal N400 meaning congruence effect, some studies with newly-learned words reported a *frontal* N400 topography. In one contextual word learning study with L1 (Spanish) pseudowords (Mestres-Missé et al., 2007), participants made relatedness judgments on these pseudowords, as well as known words, immediately after the learning phase. While the topographic relatedness effect of known words conformed to canonical N400 central-parietal regions, the N400 relatedness effect of the newly-learned pseudowords was found over frontal sites. Mestres-Missé et al. (2007) suggested that at early learning stages "retrieval of the meaning of a novel word enlists a prefrontal network driven by retrieval effort and monitoring demands" (p. 1863). Effortful retrieval of newly-learned words was also associated with the frontal topography of semantic processing in Frishkoff et al. (2010). In the present study, a frontal N400 effect would thus indicate more effortful processing, either due to lower L2 proficiency or to an increased processing effort in some experimental conditions (e.g., in the incongruous condition) [i].

In summary, the study exposed higher- and lower-proficiency bilinguals to unfamiliar L2 words in constraining sentence contexts controlled for lexical difficulty, creating the possibility of incidental learning of word meanings. The learning outcomes were tested a day later in an L2 semantic relatedness judgment task with the critical words embedded in old, new or incongruous sentence contexts. This design allows tests of hypotheses that 1) contextual learning produces variable levels of abstraction of the word meaning from the

[i]In fMRI studies, the greater engagement of the pre-frontal cortex by lower proficiency bilinguals has been associate with their attempts to compensate for lower processing efficiency and with the involvement of executive control in accessing lexical representations of L2 words (Abutalebi, 2008).

context to a more stable semantic memory and 2) the extent to which semantic learning occurs depends in part on the L2 lexical proficiency of the bilingual reader.

Method

Participants

Twenty-six volunteers studying at the University of Pittsburgh were recruited into two groups: a higher-proficiency group (HPG, N=11) and a lower-proficiency group (LPG, N=15) (see Appendix A for details). All were late English bilinguals (i.e., did not start learning English in early childhood), right-handed, with no history of head injuries and with normal or corrected-to-normal vision. They were paid at an hourly rate of $10, with a possibility of earning a bonus for accurate and fast performance.

Proficiency assessment—LPG participants were recruited from students taking English proficiency courses at the English Language Institute of the University of Pittsburgh. Their TOEFL iBT scores were below 100. Participants for the HPG were recruited from students enrolled in education and international business courses at the same university. These students had high self- reported English language proficiency.

L2 lexical proficiency of each participant was assessed for quality and quantity of L2 word knowledge. The breadth of their vocabulary knowledge was measured using the Vocabulary Size Test (VST, Nation, 2006). The estimated mean score was 6307 word families (*SD* =1771) for the LPG and 9870 (*SD* =1610) for the HPG, a reliable difference of 3563 ($F_{(1,22)}$=25.42, p<.0001, η_p^2=.536). A speeded L2 (English) lexical decision task (LDT) containing 64 nonwords and 64 words (not used in the main study) provided a qualitative measure of L2 lexical proficiency: 1) the *coefficient of variation* (CV, Segalowitz & Segalowitz, 1993) and 2) *d-prime (d′)* accuracy scores. The CV (calculated as an individual's *SD/mean RT* ratio) provides a measure of variability corrected for the latency of responding. CV is an indicator of the relative deployment of controlled and automatic processes in such experimental tasks as lexical decision and semantic classification (Segalowitz, 2000; Phillips et al., 2004). Positive correlation between CV and RT (reflecting differential use of effortful processing) is a marker of higher lexical proficiency, while the absence of such a correlation (observed for less skilled language users) indicates heavy dependence on effortful processing (Harrington, 2006; Segalowitz & Segalowitz, 1993: 381). For the LPG, the mean CV was .401 (mean RT=1887 ms, *SD*=786 ms); for the HPG, it was .267 (mean RT=960 ms, *SD*=268 ms). The mean response accuracy index (*d′*) was .862 for LPG and 2.642 for HPG. The differences between groups were reliable for both CV and *d′*, ($F_{(1,22)}$=13.28, p<. 005, η_p^2=.376) and (F(1,22)=22.34,p<.0005, η_p^2=.504), respectively. Finally, there was a significant positive correlation between CV and RT for the HPG (r_s=.794, p<.01), but not for the LPG (r_s=.495, p=.072), further confirming the difference in L2 lexical proficiency between the two groups. Across all measures, the L2 lexical proficiency of the HPG was significantly higher than that of the LPG, both quantitatively and qualitatively.

Since word learning tends to be more successful for those with larger working memory capacity in both L1 (Cain, Lemmon, & Oakhill, 2004; Daneman & Green, 1986) and L2 (see Juffs & Harrington, 2011, for an overview), working memory scores for participants in the two groups were obtained using an Operation Span (O-Span) Task (Turner & Engle, 1989; Tokowicz, Michael & Kroll, 2004). The mean O-Span score for the LPG was 4.3 (*SE*=.29), and it was 3.9 (*SE*=.34) for the HPG, a non-significant difference (*p*=.399). This confirmed that the two groups were well-matched in regard to working memory.

Learning session: Materials and procedures

Critical words—Participants were presented with 90 rare English words (*critical words*) embedded in three high-constraining sentences. The critical words were 45 nouns and 45 adjectives used in two previous studies, Balass (2011) and Bolger et al. (2008). All words were at the low-end of the scale for orthographic and meaning familiarity for native English speakers in the database of over 500 rare English words Balass (2011), which made them highly unlikely to be known by L2 participants [ii]. Their average frequency in CELEX (Baayen, Piepenbrock, & Gulikers, 1995) was .86 opm (*SD*=.77; seven words were not in the corpus) and the average length was eight letters (*SD*=1.7).

Sentences—For each critical word, three high-constraining sentences were used in the learning session (270 sentences in total) because high-constraining contexts generate more robust learning than low-constraining contexts (Borovsky et al., 2012, 2013; Frishkoff et al., 2010). Constraint levels were based on the results of the cloze procedure used in Balass (2011, pp. 29-31), with semantic constraints being scored between 0 (lowest constraint) to 1 (highest constraint). The sentences used in this study had scores that were higher than 0.6.

The non-critical words in the sentences were examined through the BNC (British National Corpus) version of the online VP tool (http://www.lextutor.ca). Lower frequency words were replaced with their higher-frequency synonyms, resulting in 95.7% of the words being within the first 3000 most frequent words of English, and 98.1% being within the first 5000 words. This simplification procedure resulted in all sentences having vocabulary levels that were appropriate for the L2 participants (with the LPG group's mean vocabulary size estimated as 6307). Thus contextual word learning for both groups was facilitated by using lexically-appropriate high-constraining contexts allowing participants to infer meanings of unknown words.

Procedure—Participants engaged in a self-paced reading task in a computer lab, with no more than two participants in the same room at a time. Participants read sentences presented in the middle of a computer screen and pressed the space bar to move to the next sentence. The same pseudorandom sentence order was used with all participants. They were instructed to read for general understanding, and were not informed that their word knowledge would be tested in the second session. The learning task was organized in three blocks of 90 sentences each. After each block participants took a break of up to five minutes. The task took about one and a half hours to complete.

[ii]It was not possible to check participants' knowledge of the critical words prior to the incidental learning session without drawing their attention to them and avoiding pre-learning. However, after the cloze test on day two, participants were informally asked if they knew these words prior to the study. Only one participant in the HPG group mentioned that she may have encountered a few of the words prior to the experiment, but she did not know their meanings.

Participants encountered each critical word in three different sentences, with the word repeated every fifth sentence. The task was organized in 15-sentence cycles (six cycles per block), with two short true-false comprehension questions administered at the end of each cycle (Appendix B, Table 1). Critical words were not used in these questions nor were they needed to answer the questions. Participants used the response box to submit their answers; their time on task was limited to five seconds per question. To ensure that participants read the sentences for understanding, they received a bonus if they answered over 80% of the questions correctly. The resulting average accuracy for both groups was over 80%, indicating that the learning materials were well-understood.

Participants also self-rated their understanding of each sentence on a five-point scale, from "did not understand" (one) to "fully understood" (five), using a response box connected to the computer. The rating screen was terminated by the button press, or after three seconds. The resulting high mean ratings (HPG: M=4.24, SD=.35; LPG: M =3.94, SD =.44) indicate that the sentences used in the learning session were at an appropriate difficulty level for both groups. The small (.30) difference in the groups' self-rated understanding of the sentences was not reliable ($t(22)$=1.76, p=.093).

At the end of the session, participants completed two proficiency tests (described above) and an O-Span task measuring their working memory capacity. Some lower-proficiency participants who could not complete all three supplementary tasks in the first session completed the remaining tasks at the end of the second session, on day two.

Testing session: Materials and procedures

The testing session was performed on the next day to allow for overnight consolidation (Davis, Di Betta, Macdonald, & Gaskell, 2009; Lindsay & Gaskell, 2010). Individual participants read a sentence with the critical word embedded in one of the three context types described below and then made semantic relatedness decisions while ERPs were recorded. The sentences were displayed one word at a time, with critical words always appearing in the sentence-final position. Participants made decisions on whether the last word of each sentence and the following word (semantic probe) were related in meaning. At the end of the testing session, participants completed a pen and paper four- choice cloze test of critical word knowledge, using the sentences from the learning session that had not been used in the semantic relatedness experiment.

Higher-frequency words—In addition to the 90 critical words, the semantic relatedness task included 30 higher- frequency words (15 nouns and 15 adjectives; mean CELEX frequency = 22.3 opm) that did not occur in the learning session, in order to compare participants' performance on known and newly-learned words. Each of the higher-frequency words was matched in meaning with a critical word; for example, *vague* (higher-frequency word) was matched with *abstruse* (critical word) and paired with the meaning probe *OBSCURE* in the related condition. All higher-frequency words were within the first six thousand words in the BNC corpus (with 90% in the five thousand and 87% in the four thousand bands) and, therefore, were expected to be known to the bilinguals in both proficiency groups.

Semantic probes—The *semantic probe* (the second stimulus in the semantic relatedness task) was selected using the *WordNet WordNet* database (WordNet, 2010). Semantic probes were in the same synset (group of related lexical items) as the corresponding related critical words [iii]. Semantic probes were feature-rich basic category words (mean CELEX frequency=12.9 opm; 90% in the first six thousand words; mean length=8.7 letters) that shared semantic features with related critical words (Brown, 1958; Murphy, 2004).

Context Conditions—Three sentence conditions were used to present critical words in the semantic judgment task: sentences used in the learning session on day one (the *old* condition); sentences not used in the learning session, in which the critical word was congruous with the meaning of the sentence frame (the *new* condition); sentences not used the learning session, in which the critical word was incongruous with the sentence frame (the *incongruous* condition) (Appendix B, Table 2). The additional 30 higher-frequency words were always used in congruous contexts.

Three counterbalanced experimental sets (120 trials each) were created so that each critical word occurred in one of the three conditions (i.e., old, new or incongruous) in each set, with all 90 critical words presented on an equal number of trials in the three conditions across all sets. The same 30 sentences with higher-frequency words were used in all three sets. In each set, half of the trials contained related critical-probe word pairs, and half unrelated pairs. Participants were assigned to the next available set on their arrival to the testing session.

Testing procedure—The experimental procedure (Figure 1) was programmed and carried out on E-Prime software (Psychological Software Tools, Inc., Pittsburgh, PA), which sent event information to the EGI NetStation EEG recording system. Instructions and trials were presented on a 15-in CRT monitor with a 60 Hz refresh rate. Each trial started with a row of crosses in the middle of the screen (a fixation point) displayed for 500 ms, followed by a blank screen, the duration of which randomly varied between 50 and 250 ms to reduce timing dependent oscillatory EEG patterns. A sentence was then presented one word at a time, with each word (except last) displayed for 350 ms and followed by a blank screen displayed for 250 ms. The last word of the sentence appeared with a period and was displayed for 1000 ms, followed by a blank screen for 200 ms. After this, a semantic probe in capital letters was displayed for 1000 ms. Following the probe, participants saw a screen instructing them to press the "1" on the response box if the two words were related/similar in meaning, and the "2" if they were not. This screen was terminated by a button press or after 2000 ms. Participants used their right hand to register a response. To start the next trial they pressed both buttons together. Response latencies were recorded from the onset of the task screen that signalled a semantic judgment. Participants were instructed to keep still and unblinking during the trial, but were encouraged to blink and rest between trials. They were given five practice trials to acclimate to the task. The experiment was conducted in three blocks of 40 trials, with an opportunity for a longer break between blocks. Participants were

[iii]Due to the very low frequency of the critical words it was not possible to estimate their relatedness using such semantic similarity tools as *LSA* (http://lsa.colorado.edu/) or word association norms, as most of these words were not available for analysis. However, the mean LSA pairwise comparison index calculated for higher-frequency words (matched with the critical words) and related probes was .31 (SD=.17), while this index calculated for the unrelated condition was .09 (SD=.05). Although there are no direct interpretations of LSA scores, the LSA website gives the following as example input: cat/mouse .34; house/dog .02.

instructed to silently read sentences displayed one word at a time and make decisions on whether the last word of each sentence (displayed with a period) and the following word in capital letters were related in meaning. They were told to consider just the sentence-final word in their decisions, and not the whole sentence. They were not asked to judge whether the sentence-final word agreed with the rest of the sentence, in order to encourage natural processing of the critical word.

ERP Recordings

Participants were fitted with a 128 electrode Geodesic sensor net (Tucker, 1993) with Ag/AgCl electrodes (Electrical Geodesics, Inc., Eugene, OR). Impedances were kept below 50 kΩ, an acceptable level with this system (Ferree, Luu, Russell, & Tucker, 2001). During the recording, a vertex reference was used, and the data were later rereferenced offline using an average reference. Six eye channels were monitored for artefacts related to eye-blinks and eye-movements. The EEG signal was digitally sampled at a rate of 500 Hz, and was hardware filtered between .1 and 200 Hz during recording. After recording, a 30 Hz lowpass FIR filter was applied to the data. Next, epochs of 1200 ms were created relative to the onset of meaning probes (200 ms before onset of stimulus, 1000 ms after). Within a segment, differential voltages greater than ±75 μV, and ±140μV on two separate sets of eye channels were considered eye movements and eye blinks, respectively. Any channels displaying voltages exceeding ±200 μV within the segment were considered “bad” for that segment. Channels that contained artefacts on 20% or more trials were removed and interpolated later. Segments containing either eye artifacts or more than 10 bad channels were rejected. Of the original 26 volunteers, two (one from each proficiency group) were excluded from the analyses; one due to a technical error during recording, and the other due to having too few usable channels after artifact detection. Next, visual analysis of averaged segments was used to reject up to 12 scalp channels per subject, inclusive of those rejected through automatic artefact detection. The removed channels (M=7, SD=2.87) were replaced by spherical spline interpolation using data from neighbouring channels (Ferree, 2006). Channels were then average referenced, and corrected for the polar average referencing effect (PARE; Junghöfer, Elbert, Tucker, & Braun, 1999). The average of the 200 ms pre-stimulus baseline was subtracted from the data, and the trials were averaged per condition for each participant.

Results and Analyses

In the pen and paper four-choice cloze task administered at the end the testing session, the HPG participants were better able to select the correct critical word for a previously seen sentence (86%) than were the LPG participants (58%), $F_{(1,22)}$ =13.50, p<.01, η_p^2=.380.

Behavioural results

The mean accuracy of semantic decisions was .81 (SD=.39) for the HPG and .66 (SD=.48) for the LPG. The HPG achieved reliably higher d' across conditions (M=2.57) than the LPG (M=1.46), $F_{(1,22)}$=19.24, p<.001, η_p^2= .466.

Summary preview of the RT results—The RT analysis (see below for details) showed the semantic relatedness effect, with reliably faster responses on related than on unrelated trials. This effect was significantly greater for the HPG compared with the LPG. In the analyses by condition, responses were faster in the old than in the new condition, but a reliable semantic relatedness effect (faster responses on related trials) was observed only when critical words were presented in the new (not in the old) condition, for the HPG. For the LPG, there was no difference in the semantic relatedness effect between the two conditions, but participant response latencies on related trials in the new (but not old) condition were modulated by the probe word length (with longer RTs observed for longer probes). This interaction suggests that semantic judgments were less automatic (more effortful) for the LPG learners when they had to access the meanings of critical words in new contexts.

RT analyses—The final data set included 2059 observations after removing incorrect responses and "no response" data points. The mean response latency was 560 ms (*SD*=274) for the HPG, and 616 ms (*SD*=365) for the LPG (see Table 1 for mean RTs and accuracy by condition).

Inspection of the distribution of RTs revealed a marked non-normality. The RT data were log-transformed to normalize the distribution. A *Linear Mixed-Effects* (*lme*) model was fitted to the RT data. Participants and items (semantic probes) were included in the models as crossed random effects. The model included a by-participant random slope for the trial number, in order to account for the varying effect of earlier and later trials on individual participants (Baayen et al., 2008). Statistical significance of the fixed effects was based on Markov Chain Monte Carlo (MCMC) sampling (10,000 iterations; Baayen, Davidson, & Bates, 2008). Full results are reported in the Appendices. For ease of reading, all plots of results are based on back-transformed estimates from the *lmer* models, i.e., with RTs expressed in milliseconds.

First, an *lme* model was fitted to the RT data using the primary-interest predictors: proficiency (HPG/LPG), relatedness (related/unrelated) and condition (new/old/incongruous/higher-frequency). The model fit was significantly improved by adding probe length (in letters) as a secondary-interest *item* variable, consistent with length effects reported in other research (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004). The model fit was improved by a three-way interaction among proficiency, relatedness and condition (new vs. old) (Appendix C Table 1). A reliable simple effect of relatedness in this analysis (t=−4.19, p<.001) indicates that, overall, judgments were faster on related than on unrelated trials. An interaction between relatedness and proficiency (t=2.17; p<.05) reflected differences in the magnitude of the semantic relatedness effect, with a larger effect observed for the HPG than for the LPG (Figure 2, note - all figures are based on model predictions). Decisions were 139 ms faster for related than unrelated pairs for HPG, but only 43 ms faster for LPG participants. For further analyses, the latency data were split by proficiency, and a new model was fitted to the response latency data in each proficiency group (Appendix C Table 2a & b).

Analysis of the HPG data—An *lme* model was fitted to the HPG RT data, with relatedness and condition as primary predictors. Probe length did not reach significance as predictor and was not used in this model. The model fit was improved by a significant interaction between condition and relatedness (Appendix C Table 2a; Figure 2a). Participants in the HPG were reliably faster on related than unrelated trials (t=–4.55, p<.001), and their judgments were reliably faster in the old than in the new condition (t=-2.06,p<.05). The semantic relatedness effect (i.e., faster responses on related than on unrelated trials), however, was observed only when critical words were embedded in new (not in old) sentences (Figure 2a).

Analysis of the LPG data—An *lme* model was fitted to the LPG RT data, with the two primary predictors (relatedness and condition) and probe length (as a secondary-interest variable). Although the LPG participants tended to be faster on related than on unrelated trials (t =–2.14, p<.05) (Appendix C Table 2b; Figure 2b), there was a reliable three-way interaction between the three predictors (t=–2.26, p<.05). On related trials the difference in RTs between the new and old conditions was modulated by probe length: when critical words were presented in new (and incongruous) contexts latencies of responses increased as probe length increased – an indication of the effortful (less automatic) nature of lexical semantic processing – while no such effect occurred in the old condition or with higher-frequency words (Figure 2c).

ERP results

ERPs were analyzed to examine the effect of critical words on the semantic processing of the meaning probes [iv]. The amplitude of the N400 component observed for the probe is an indicator of degree of its semantic congruence with the preceding stimulus (newly-learned word), as perceived by the reader (Balass et al., 2010; Holcomb, 1993; Kutas & Hillyard, 1989; Mestres-Missé et al., 2007; Nobre & McCarthy, 1994).

In order to cover a larger portion of the scalp and to account for within and between-group variability in the locations of effects, we used nine clusters, each centred on an electrode of the 10-20 system (F3,Fz,F4,C3,Cz,C4,P3,Pz,P4) and extended to the nearest 5-7 electrodes (Figure 3). Averaged portions of the waveform used in amplitude analysis were chosen based on visual inspection of the peak latency of the N400 component, respective of the previously established time window (300-500 ms after the stimulus onset), and for the sustained effect of meaning congruence in the 500-700 ms time window. Repeated-measures analysis of variance was used to examine the ERP component of interest.

Visual inspection of ERPs—The ERPs time locked to semantic probes for related vs. unrelated trials are plotted in Figure 4; and those by the sentence-final word condition and relatedness in Figure 4a. Figure 5 shows voltage maps computed by subtracting the waveform responses to related probes from the responses to unrelated probes (these plots reveal the scalp distribution of N400 differences for the HPG and the LPG). A visual inspection showed differences in the N400 effect for the two proficiency groups: a clear N400 effect distinguishing related and unrelated trials observed at central (Cz and C4) and

[iv]All trials (both correct and incorrect responses in the behavioural task) are included in the ERP analyses presented in the paper. The ERP recordings taken on the meaning probe before the explicit task responses are made are interpreted as an immediate brain response to the word in relation to the prior context (i.e., the newly-learned critical word that ends the carrier sentence). In some cases, the ERP is a more sensitive indicator of what the participant knows than the following behavioural (yes/no) decision (McLaughlin, Osterhout, & Kim, 2004).

parietal (P3 and P4) sites in the canonical time window (300-500 ms) for the HPG, but not for the LPG. The parietal N400 observed for the HPG had a less defined peak and a longer peak latency, compared to the central N400. The effect of meaning relatedness at frontal sites occurred for both proficiency groups within the expected time window (peaking just prior to 400 ms). Although the morphology of the frontal effect was similar for the two groups, it was somewhat smaller for the LPG than for the HPG. In addition, both LPG and HPG participants showed a sustained late effect of meaning congruence on the probe word around midline (across parietal, central and frontal sites) between 500 and 700 ms. For the HPG, this effect was observed broadly across both right hemisphere and midline sites, whereas for the LPG participants this late effect could be observed more clearly at midline sites.

Analysis of the canonical N400 effect—The initial analysis focused on the mean amplitudes within the canonical time window (from 300 to 500 ms after probe exposure) over a slightly right-lateralized central- parietal electrodes - an N400 topography commonly associated with meaning integration in semantic priming or semantic categorization tasks (Kutas & Federmeier, 2011; Kutas & Van Petten 1994; Swaab, Baynes, & Knight, 2002). A repeated-measures ANOVA was performed on the N400 amplitude data (measured in *microvolts*) at four sites (Cz, C4, Pz, P4) . The within-participant factors were meaning relatedness (related/unrelated), sentence-final word condition (new/old/incongruous/higher-frequency) and scalp topography (anteriority and laterality), while the between-participant factor was proficiency group (HPG/LPG). The analysis revealed a simple effect of meaning congruence ($F_{(1,22)}$=16.55, p=.001, η_p^2=.43), with the mean amplitude less negative on related (M=.52 μV, SE=.26) than on unrelated (M=−.19 μV, SE=.27) trials. However, there was a statistically significant interaction between relatedness and proficiency ($F_{(1,22)}$=13.01, p<.01, η_p^2=.37) (Figure 6), which revealed that the N400 integration effect was reliably observed only for the HPG. There was also a significant effect of scalp anteriority (central vs. parietal) ($F_{(1,22)}$=25.29, p<.001, η_p^2=.54) and an interaction between anteriority and relatedness ($F_{(1,22)}$=6.84, p<.05, η_p^2=.24). The mean N400 amplitude was significantly less negative over posterior (M=.90 μV, SE=.31) than over central (M=-.57 μV, SE=.28) sites (Figure 4). There were no reliable interactions between sentence-final word condition and meaning relatedness in this analysis.

An additional repeated-measures ANOVA was performed on the ERPs at the parietal midline electrode cluster (Pz), which is associated with automatic semantic processing. The results showed a reliable interaction between relatedness and proficiency $F_{(1,22)}$=4.93, p<0.05, η_p^2=.18, with the N400 semantic congruence effect observed only for the HPG, but not the LPG. Importantly, there was also a reliable interaction between relatedness and sentence-final word condition $F_{(3,66)}$=3.22, p<0.05, η_p^2=.13, reflecting a larger semantic congruence effect in the new context condition compared with both old and incongruous conditions (Figure 7 and 4a).

Analysis of the frontal N400 effect—The N400 amplitudes were also analyzed at frontal sites since negativities in this time-window and region have previously been implicated in more effortful lexical semantic processing (Frishkoff, et al., 2010; Mestres-

Missé et al., 2007). The analysis was performed on the mean amplitudes within the 300 to 500 ms time window after probe exposure. A repeated-measures ANOVA was performed on the N400 amplitude data (at Fz and F4 sites), using the within-subject factors of meaning relatedness (related/ unrelated), sentence-final word condition (new/old/incongruous/higher-frequency), and scalp topography (central/right), and the between-subject factor of proficiency group (HPG/LPG). Similar to the central-parietal N400 analysis, this analysis showed a main effect of meaning relatedness ($F_{(1,22)}$=17.51, p<.001, η_p^2=.44), with the mean N400 amplitude reliably less negative on related (M=−.85 μV, SE=.35) than on unrelated (M=−1.67 μV, SE=.30) trials. There was no interaction of relatedness and proficiency (F<1), indicating that the effect occurred for both proficiency groups (Figure 8). Among the conditions of the sentence final word, the N400 amplitude was least negative for probes preceded by the higher-frequency words (M=−.84, SE=.39) and most negative for probes preceded by critical words in the incongruous (M=−1.55, SE=.28) and new (M= −1.53, SE=.39) conditions (Figure 9); however the overall condition effect did not reach the conventional level of reliability ($F_{(3,66)}$=2.26, p=.089, η_p^2=.09). Finally, there was a strong trend ($F_{(1,22)}$=3.97, p=.06, η_p^2=.15) for the frontal N400 amplitudes to be less negative for the HPG (M=−.64 μV, SE=.48) compared to the LPG (M=−1.88 μV, SE=.40).

ERP analysis of the late congruence effect—Because visual inspection of the ERP recordings on the probe word (Figure 4 and 5) pointed to a sustained effect of meaning congruence extended through 700 ms for both proficiency groups, we carried out a repeated measures ANOVAs on the mean amplitudes in the 500-700 ms time window at central-parietal (Cz,C4,Pz,P4) and frontal (Fz, F4) electrode clusters separately. In the central-parietal analysis there was a reliable simple effect of meaning congruence between the critical word and the probe ($F_{(1,22)}$=26.77, p<.001, η_p^2=.55), qualified by the following two-way interactions: (1) relatedness and sentence-final word condition ($F_{(3.66)}$=3.16, p<.05, η_p^2=.13); (2) relatedness and anteriority; ($F_{(3.22)}$=6.13, p<.05, η_p^2=.22) (3) relatedness and laterality ($F_{(1,22)}$=4.51, p<.05, η_p^2=.17) (but no interaction with proficiency). The late congruence effect on the probe was most prominent at Cz, and this effect was the greatest when the newly-learned word was encountered in new congruous contexts and the smallest when it was encountered in incongruous contexts (Figure 10).

The analysis of the late ERPs at the two frontal regions showed a reliable simple effect of meaning congruence ($F_{(1,22)}$=17.63, p<.001, η_p^2=.45) that was qualified by a three-way interaction among relatedness, sentence-final word condition and proficiency that did not reach the conventional level of reliability ($F_{(3,66)}$=2.48, p=.07, η_p^2=.10) (Figure 11). For the HPG, the effect of relatedness was larger when the critical word was encountered in new and incongruous contexts and smaller when it was encountered in the old (repeated) context and for higher-frequency words. For the LPG, the effect was the largest for the higher-frequency word, followed by that in the old and new context conditions. When the critical word was embedded in incongruous contexts, no late frontal congruence effect on the probe was observed for the LPG (Figure 11 and 4a).

Higher frequency word processing by proficiency—Additional analyses were performed on the ERPs recorded on trials with the higher-frequency words that were used as a benchmark of meaning processing in L2. To test proficiency differences in responses to these higher frequency words, we carried out separate ANOVAs on ERP amplitudes at the central–parietal regions and the frontal regions. The central-parietal N400 analysis showed a reliable two-way interaction between relatedness and proficiency $F_{(1,22)}=4.41$, $p<.05$, $\eta_p^2=.17$ (Figure 12), reflecting a canonical N400 effect for the HPG, but not LPG. In contrast, the frontal N400 effect was reliable ($F_{(1,22)}=5.22$, $p<.05$, $\eta_p^2=.19$) with no interaction with proficiency. Furthermore, no interaction between relatedness and proficiency was observed in the late processing window (500-700 ma after probe onset), with the mean amplitudes being reliably less negative on the related than unrelated trials, for both proficiency groups ($F_{(1,22)}=6.20$, $p<.05$, $\eta_p^2=.22$).

Summary of findings

The behavioural and ERP results show that some incidental contextual L2 word learning occurred for all participants, but that there was a difference in the quality of knowledge established for the two proficiency groups. The behavioural semantic relatedness effect (Figure 2) and the N400 effect at frontal locations (Figure 4) were reliable for both groups, as was the late meaning congruence effect measured by the ERPs recordings in the 500-700 ms time window on the probe word. However, the semantic relatedness effect was reliably larger for the HPG than for the LPG (Figure 2), due to trials with critical words presented in new, congruous contexts (Figure 2a) – the condition that facilitates lexical semantic processing of the newly-learned word. Proficiency also modulated the central-parietal N400 effect, as indicated by a reliable interaction between meaning relatedness and proficiency. The N400 amplitude was significantly reduced on related compared to unrelated trials for higher-proficiency learners, but not for low-proficiency learners (Figure 6). These findings are in line with the prediction that more proficient bilinguals would be more successful than less proficient bilingual in establishing initial lexical semantic representations for contextually-learned L2 words [v].

The participants' performance in the semantic relatedness judgement task was also affected by the type of contexts in which the newly-learned word was presented. Although the HPG participants responded fastest when they encountered critical words in the same context as in the initial learning session ("old" condition), these responses reflected mostly episodic processing, as indicated by the attenuation of the semantic relatedness effect. The N400 analysis in the mid-parietal region showed that negativity was reduced to a larger extent when the newly-learned word was encountered in a new congruous sentence for both

[v]Following a reviewer's suggestion, we also conducted additional N400 analyses excluding incorrect trials, in order to check whether a different pattern of results would emerge when bilinguals were able to correctly judge the two stimuli as semantically related or unrelated. We argued that, if the N400 effect is observed at central and parietal regions for both proficiency groups in this analysis, our hypothesis about the quality of semantic learning for lower proficiency bilinguals would need to be revisited. After removing incorrect trials, all subjects were still suitable for the N400 analysis, but it was no longer possible to compare the context conditions, as not enough trials were left per condition. We therefore removed trials with higher frequency sentence-final word (to avoid conflating result for the newly- learned and previously known items) and conducted the N400 analysis with relatedness (related vs. unrelated) as a within-participant variable and proficiency (HPG vs. LPG) as a between-participant variable. The results of this analysis closely mimicked those of the N400 analysis on all trials. The repeated-measures ANOVA on the amplitudes at central and parietal regions (Cz, C4, Pz, P4) showed a reliable simple effect of relatedness $F_{(1,21)}=22.35$, $p<0.001$, $\eta_p^2=.52$ and a reliable interaction between relatedness and proficiency $F_{(1,21)}=10.50$, $p<0.01$, $\eta_p^2=.33$; the N400 semantic congruence effect was observed for the HPG, but not LPG. At frontal locations (Fz, F4), the N400 effect on the meaning probe was statistically reliable $F_{(1,21)}=5.84$, $p<0.05$, $\eta_p^2=.22$, but no interaction with proficiency was observed. These results corroborate our conclusion that quality of incidental contextual L2 word learning varies with L2 lexical proficiency.

proficiency groups (Figure 7). The new context advantage was maintained in the late ERP analysis for both proficiency groups. As predicted, new supportive contexts (“new” condition) stimulated the L2 readers’ access to the initially established semantic features of the contextually-learned words (and may have added further features) resulting in a robust semantic relatedness effect, for the HPG, and a larger meaning congruence effect on the probe, for both groups.

Probe processing was most effortful when the critical word was encountered in the incongruous condition, as shown by 1) increased frontal N400 negativity and 2) slower semantic judgments by the HPG. For the LPG, semantic judgments on the incongruous (and new) conditions were affected by the probes’ length in letters, also pointing to more effortful processing [vi]. Although frontal N400 effects should be interpreted with caution, our results are consistent with the need to draw upon control networks in retrieving the meaning of partially- or weakly-known words (Frishkoff et al., 2009; Mestres-Messe et al., 2007) and are aligned with previous findings that lexical semantic processing is less automatic at lower proficiencies (Ardal et al., 1990).

General discussion and conclusions

To examine whether contextual learning of word meanings is modulated by L2 lexical proficiency, participants in this study made meaning relatedness judgment involving semantic probes and newly-learned words in the sentence-final position. The results show that both episodic and lexical semantic knowledge can be involved in retrieving recently-learned L2 words, depending on learners’ L2 lexical proficiency and on the context in which a word is presented. The processing of critical words in contexts identical to those in the learning session was based primarily on the fit of the word to the episodic (sentence) memory rather than semantic memory. When these words appeared in new congruous contexts, more proficient bilinguals were able to use the previously established lexical semantic representations and an additional contextual exposure on day two, resulting in reliably faster responses on related than on unrelated trials. For less proficient bilinguals, however, both old and new contexts resulted in a similarly small semantic relatedness effect, suggesting that they were less effective in contextual word learning. We've argued that ineffective contextual learning at lower L2 lexical proficiencies arises out of difficulty in integrating the meanings of new words into insufficiently-developed L2 lexical semantic networks. Having confirmed that working memory capacity of participants in the two proficiency groups and their sentence comprehension in the learning phase were comparable, we conclude that the differences in semantic learning can be attributed squarely to their lexical proficiency.

Our interpretation of these differences is further informed by the morphology and topography of the N400 effect for the two groups. The canonical central-parietal N400

[vi]The fact that probe word length was not a reliable main predictor of RT in the analysis of semantic decisions for either proficiency group (Appendix C, Table 2a and 2b) indicates that, overall, the probes were processes fluently by all participants in this study. It was only when the newly-learned words were presented in unfamiliar contexts that LPG showed the effect of probe length (Figure 2c), i.e., the difficulty in making semantic decisions was likely due to the extra effort involved in processing critical words in new and incongruous contexts, rather than to the processing of the probe words per se.

reduction on semantically related trials (with both newly-learned and familiar L2 words) was observed only for the higher-proficiency learners. This suggests that, at higher proficiencies, encountering unfamiliar L2 words in supportive contexts quickly leads to some encoding of context-related semantic features, facilitating lexico-semantic learning. This conclusion is supported by the following results: 1) more proficient participants showed clear benefits of encountering recently-learned words in new supportive contexts for sustained contextual learning; 2) although misleading (incongruous) contexts interfered with semantic processing of the critical words (no behavioural relatedness effect was observed), there was no interaction between type of context and the central-parietal N400 effect, for the higher proficiency bilinguals. The latter finding suggests that even under adverse conditions on day two, the more proficient bilinguals were able to access semantic knowledge of the new words (gained in the learning session) fluently, as indicated by the N400 effect on the meaning probe [vii]. At lower proficiencies, no fluent processing of the meaning was observed, either for the newly-learned or for the familiar L2 words (as shown by the absence of the N400 effect at central-parietal regions in the 300-500 ms time window). Therefore, we conclude that contextual learning trajectories and quality of lexical semantic knowledge are different for higher and lower proficiency bilinguals.

The frontal N400 effect observed for both groups suggests an additional mechanism at play in making relatedness judgments – most likely, an effortful process requiring a higher degree of executive control (Abutalebi, 2008). More effortful processing evidenced by the frontal meaning congruence effect also has been reported in L1 contextual word learning studies (Frishkoff et al., 2010; Mestres-Messe et al., 2007). In L2 studies with known words, similar topographic differences for the N400 appear to be proficiency-related (Ardal et al., 1990; Kotz & Elston-Güttler, 2004; Kutas & Kluender, 1994; Midgley et al., 2009; Newman et al., 2012; Phillips et al., 2004). In our study similar proficiency-based differences in the topography and latency of the N400 effect were found for the familiar higher-frequency L2 words. Importantly, to our knowledge, this is the first study showing that lexical proficiency-related differences emerge for L2 learners even when the learning conditions are calibrated for meaning acquisition (when high-constraining lexically-simplified L2 learning contexts are used).

Although the precise causal component of this proficiency effect remains an open question, its general form may lie in differences in the quality of individual lexical networks. Since learning the meaning of a new word requires its integration into existing L2 lexical semantic networks of the learner (e.g., by virtue of shared semantic features or bundles of features, i.e., overlapping neural networks), the quality of the networks (e.g., the precision of form knowledge and the mapping of forms to meanings) affects learning outcomes. Since, at lower proficiency, lexical semantic representations of L2 words may not be fully specified and may have fewer and weaker connections, opportunities for incremental semantic learning from reading are reduced, even in supportive contexts. In a sparse L2 vocabulary (LPG), a new word establishes few lexical semantic connections (making context rehearsal

[vii] A separate N400 analysis with trials where the critical word was presented in the incongruous context at central-parietal locations (Cz,C4,Pz,P4) showed a reliable interaction between relatedness and proficiency $F_{(1,22)}$=8.00, p=0.01, η_p^2=.17, such that the N400 meaning congruence effect was observed only for the HPG (see also Figure 4a).

preferable over semantic rehearsal, Rodríguez-Fornells et al., 2009); and its retrieval therefore may rely more on the episodic (sentence) memory. When the existing L2 lexical semantic knowledge system is robust, initial lexical semantic representations are established quickly when a new word is encountered in informative contexts, facilitating further contextual learning. When the L2 lexical semantic system is underspecified (e.g., in its early developmental stages), initial contextual encounters fail to extract salient lexical semantic features of new words, and further contextual learning is less effective.

A more general implication of these results concerns the limits of incidental learning conditions for acquiring new word meanings. Learning words incidentally from language contexts is undoubtedly the most common and thus most important way of building new word knowledge. However, some additional, more deliberate learning opportunities may be needed when learners have low proficiency in the language.

Acknowledgements

This research was supported in part by National Institutes of Health Grant 115260 to the University of Pittsburgh (Charles Perfetti, PI) and National Science Foundation grant SBE-0836012 to the Pittsburgh Science of Learning Center.

We thank anonymous reviewers for their comments on previous versions of this manuscript.

Appendices

Appendix A

Participants' characteristics

	LPG	HPG	All
No.	14	10	24*
Age	23.1 (SD = 5.8)	26.1 (SD = 5.3)	24.4 (SD = 5.7)
Sex	female = 9	female = 8	female = 17
Age of L2 acquisition	11.6 (SD = 2.6)	9.0 (SD = 2.6)	10.5 (SD = 2.8)

*Only participants whose data were included in the analysis (24 out of the original 26 participants) are included in this table.

Appendix B. Materials

Appendix B. Table 1

Materials A cycle of the learning session (below critical words are italicized for clarity, but they were not highlighted in any way in the learning session).

No.	Set	Sentence
1	1	The *evanescence* of Sam's passions became apparent when they gave way to jealousy.
2	1	The *spate* of water sent crashing through the valley after the storm caused much damage to their camp-site.
3	1	The philosopher's idea of the perfect society was dismissed as being too *laputan* for today's world.
4	1	At the soldier's funeral, the president gave an unreserved *encomium* appropriate for the national hero.
5	1	Pregnant women are stereotypically thought of as being quick to get *petulant* with their husbands.

No.	Set	Sentence
6	2	Theatre shows come and go, making actors realize the *evanescence* of their craft.
7	2	The words flowed from my lips like a river in *spate*.
8	2	The young scientist's ideas were too *laputan* to put into practice in applied research.
9	2	An *encomium* to the city of Florence declared its superiority over every other Italian city.
10	2	The older man's seemingly *petulant* manner was just a cover for a kind heart.
11	3	The World Trade Center memorial fountains capture the idea of permanence versus *evanescence*.
12	3	He finally snapped and started shouting out insults and curses in a violent *spate*.
13	3	Her suggestions were always too difficult to implement and came across as *laputan*.
14	3	At his mother's 80th birthday party her dedication to her family and community was portrayed in a heartfelt *encomium*.
15	3	If the little boy didn't have some sleep during the day he was likely to become *petulant*.
Answer:		**Comprehension Questions:**
TRUE		The city of Florence was declared superior to every other Italian city.
TRUE		The camp-site located in the valley was damaged by the violent storm.

Appendix B. Table 2

Presentation conditions of the critical words, *evanescence* and *cogent,* and the corresponding higher-frequency words, *imbalance* and *persuasive,* across all versions of the semantic relatedness judgment experiment

Old context	New context	Incongruous context	Higher-freq. word	Semantic probe
The World Trade Center memorial fountains capture the idea of permanence versus evanescence.	The feelings of joy and sorrow eventually go away; we've all experienced their evanescence.	She pulled up in front of the house taking up all the space in the driveway without evanescence.	Tears were an emotional problem, not a chemical imbalance.	INSTABILITY
No one knew what the verdict would be because the arguments from both sides were so cogent.	She could always convince me to do what she wanted because her arguments were cogent.	What scares scientists most is that a new virus hybrid will be both deadly and cogent.	He originally felt that the war was justified for a million reasons, none of which now seemed particularly persuasive.	CONVINCING

Appendix C. The semantic relatedness judgment task: behavioural analyses

Appendix C. Table 1

The semantic relatedness judgment task: behavioural analyses Coefficients of the fixed effects in the regression model for the response latencies in the semantic relatedness judgment task, 95% Highest Posterior Density (HPD) intervals, and *p*-values based on 10,000 Markov Chain Monte Carlo samples of the posterior distributions of the parameters. Intercept levels: *New sentence context, Unrelated condition, HPG.*

	Coef. β	*t* value	MCMC mean	HPD95 lower	HPD95 upper	*p*
(Intercept)	6.11	52.83	6.11	5.89	6.32	1.0E-04 ***

	Coef. β	t value	MCMC mean	HPD95 lower	HPD95 upper	p
Type=hf	−.12	−1.78	−.12	−.25	.01	.07 ·
Type=inc	.03	.49	.03	−.10	.16	.63
Type=old	−.13	−1.91	−.13	−.27	2.9E-03	.05 *
Rel=yes	−.29	−4.19	−.30	−.43	−.16	1.0E-04 ***
Prof=LPG	.02	.15	.02	−.23	.27	.87
PrbLength	.02	3.44	.02	.01	.03	1.0E-03 ***
Type=hf:Rel=yes	.22	2.28	.22	.03	.40	.02 *
Type=inc:Rel=yes	.34	3.13	.34	.13	.56	2.0E-03 **
Type=old:Rel=yes	.33	3.30	.34	.14	.53	6.0E-04 ***
Type=hf:Prof=LPG	.14	1.50	.14	−.04	.32	.17
Type=inc:Prof=LPG	−.16	−1.64	−.15	−.35	.03	.11
Type=old:Prof=LPG	.07	.67	.07	−.13	.25	.48
Rel=yes:Prof=LPG	.21	2.17	.22	.02	.41	.03 *
Type=hf:Rel=yes:Prof=LPG	−.23	−1.69	−.23	−.49	.03	.08 ·
Type=inc:Rel=yes:Prof=LPG	−.18	−1.22	−.18	−.48	.10	.22
Type=old:Rel=yes:Prof=LPG	−.28	−2.00	−.28	−.55	−1.4E-02	.04 *

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

Appendix C. Table 2a

Results summary for the fixed effects in the regression model for the response latencies in the semantic relatedness judgment task for the HPG. Intercept levels: *New sentence context, Unrelated condition.*

	Coef. β	t value	MCMC mean	HPD95 lower	HPD95 upper	p
(Intercept)	6.30	69.20	6.29	6.11	6.48	1.0E-04 ***
Type=hf	−.12	−1.92	−.12	−.24	.01	.06 .
Type=inc	.04	.60	.04	−.09	.17	.55
Type=old	−.14	−2.06	−.14	−.27	−.01	.04 *
Rel=yes	−.31	−4.55	−.31	−.45	−.18	1.0E-04 ***
Type=hf:Rel=yes	.24	2.55	.24	.06	.42	.01 *
Type=inc:Rel=yes	.33	3.11	.32	.13	.54	1.2E-03 **
Type=old:Rel=yes	.34	3.53	.34	.16	.53	2.0E-04 ***

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

Appendix C. Table 2b

Results summary for the fixed effects in the regression model for the response latencies in the semantic relatedness judgment task for the LPG. Intercept levels: *New sentence context, Unrelated condition.*

	Coef. β	t value	MCMC mean	HPD95 lower	HPD95 upper	p
(Intercept)	6.13	23.54	6.14	5.64	6.65	1.0E-04 ***
Type=hf	.05	.15	.05	−.61	.73	.89
Type=inc	.30	.86	.30	−.34	1.01	.38

	Coef. *β*	*t* value	MCMC mean	HPD95 lower	HPD95 upper	*p*
Type=old	−.37	−1.01	−.37	−1.09	.33	.31
Rel=yes	−.68	−2.14	−.68	−1.29	−.05	.03 *
PrbLength	.02	.74	.02	−.03	.07	.45
Type=hf:Rel=yes	.50	1.15	.51	−.35	1.32	.23
Type=inc:Rel=yes	−.07	−.15	−.06	−1.02	.86	.89
Type=old:Rel=yes	1.04	2.32	1.03	.21	1.94	.02 *
Type=hf:PrbLength	−3.3E-03	−.09	.00	−.08	.07	.93
Type=inc:PrbLength	−.05	−1.20	−.05	−.12	.03	.22
Type=old:PrbLength	.03	.85	.03	−.05	.11	.40
Rel=yes:PrbLength	.07	1.95	.07	.00	.13	.05 *
Type=hf:Rel=yes:PrbLength	−.06	−1.19	−.06	−.15	.04	.22
Type=inc:Rel=yes:PrbLength	.03	.48	.03	−.08	.13	.64
Type=old:Rel=yes:PrbLength	−.11	−2.26	−.11	−.21	−.01	.02 *

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

References

Abutalebi J. Neural aspects of second language representation and language control. Acta Psychologica. 2008; 128:466–478. [PubMed: 18479667]

Anderson JE, Holcomb PJ. Auditory and visual semantic priming using different stimulus onset asynchronies—An event-related brain potential study. Psychophysiology. 1995; 32:177–190. [PubMed: 7630983]

Ardal S, Donald MW, Meuter R, Muldrew S, Luce M. Brain responses to semantic incongruity in bilinguals. Brain and Language. 1990; 39:187–205. [PubMed: 2224493]

Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.

Baayen, RH.; Piepenbrock, R.; Gulikers, L. The CELEX lexical database. Linguistic Data Consortium, University of Pennsylvania; Philadelphia: 1995.

Balass, M. Word Experience, Familiarity, and Meaning Acquisition. University of Pittsburgh; 2011. Unpublished doctorate dissertation.

Balass M, Nelson JR, Perfetti CA. Word learning: An ERP investigation of word experience effects on recognition and word processing. Contemporary Educational Psychology. 2010; 35:126–140. [PubMed: 22399833]

Balota DA, Boland JE, Shields L. Priming in pronunciation: Beyond pattern recognition and onset latency. Journal of Memory and Language. 1989; 28:14–36.

Balota DA, Cortese MJ, Sergent-Marshall SD, Spieler DH, Yap MJ. Visual word recognition of single-syllable words. Journal of Experimental Psychology: General. 2004; 133:283–316. [PubMed: 15149254]

Batterink L, Neville H. Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. Journal of Cognitive Neuroscience. 2011; 23:3181–3196. [PubMed: 21452941]

Beck, IL.; McKeown, MG.; Kucan, L. Bringing words to life: Robust vocabulary instruction. Guildford Press; New York, NY: 2002.

Bolger DJ, Balass M, Landen E, Perfetti CA. Contextual variation and definitions in learning the meaning of words. Discourse Processes. 2008; 45:122–159.

Borovsky A, Elman JL, Kutas. M. Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. Language Learning and Development. 2012; 8:278–302. [PubMed: 23125559]

Borovsky A, Kutas. M, Elman JL. Getting it right: Word learning across the hemispheres. Neuropsychologia. 2013; 51:825–837. [PubMed: 23416731]

Brown R. How shall a thing be called? Psychological Review. 1958; 65:14–21. [PubMed: 13505978]

Burgess C, Lund K. Modelling parsing constraints with high-dimensional context space. Language and Cognitive Processes. 1997; 12:177–210.

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

Cree G, McRae K, McNorgan C. An attractor model of lexical conceptual processing: Simulating semantic priming. Cognitive Science. 1999; 23:371–414.

Daneman M, Green I. Individual differences in comprehending and producing words in context. Journal of Memory and Language. 1986; 25:1–18.

Davis MH, Di Betta AM, Macdonald MJE, Gaskell MG. Learning and consolidation of novel spoken words. Journal of Cognitive Neuroscience. 2009; 21:803–820. [PubMed: 18578598]

Elgort I, Warren P. L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. Language Learning. 2014; 64:365–4141.

Ferree TC. Spherical splines and average referencing in scalp electroencephalography. Brain Topography. 2006; 19:43–52. [PubMed: 17019635]

Ferree TC, Luu P, Russell GS, Tucker DM. Scalp electrode impedance, infection risk, and EEG data quality. Journal of Clinical Neurophysiology. 2001; 112:536–544.

Ferrel Tekmen EA, Daloğlu A. An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. Foreign Language Annals. 2006; 39:220–243.

Finkbeiner MS, Forster K, Nicol J, Nakamura K. The role of polysemy in masked semantic and translation priming. Journal of Memory and Language. 2004; 51:1–22.

Frishkoff GA, Perfetti CA, Collins-Thompson K. Lexical quality in the brain: ERP evidence for robust word learning from context. Developmental Neuropsychology. 2010; 35:376–403. [PubMed: 20614356]

Frishkoff GA, Perfetti CA, Collins-Thompson K. Predicting robust vocabulary growth from measures of incremental learning. Scientific Studies of Reading. 2010; 15:71–91. [PubMed: 21442050]

Gernsbacher MA, Varner KR, Faust ME. Investigating differences in general comprehension skill. Journal of Experimental Psychology: Learning, Memory, & Cognition. 1990; 16:430–445.

Harrington M. The lexical decision task as a measure of L2 proficiency. EUROSLA Yearbook. 2006; 6:147–168.

Hinton GE, Shallice T. Lesioning an attractor network: Investigations of acquired dyslexia. Psychological Review. 1991; 98:74–95. [PubMed: 2006233]

Hintzman DL. “Schema abstraction” in a multiple-trace memory model. Psychological Review. 1986; 93:411–428.

Holcomb PJ. Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. Psychophysiology. 1993; 30:47–61. [PubMed: 8416062]

Horst M, Cobb T, Meara P. Beyond a clockwork orange: Acquiring second language vocabulary through reading. Reading in a Foreign Language. 1998; 11:207–223.

Jeon EH, Yamashita J. L2 reading comprehension and its correlates: A meta-analysis. Language Learning. 2014; 64:160–212.

Juffs A, J. Harrington M. Aspects of working memory in L2 learning. Language Teaching. 2011; 44:137–166.

Junghöfer M, Elbert T, Tucker DM, Braun C. The polar average reference effect: A bias in estimating the head surface integral in EEG recording. Clinical Neurophysiology. 1999; 110:1149–1155. [PubMed: 10402104]

Kiger JI, Glass A. The facilitation of lexical decisions by a prime occurring after the target. Memory and Cognition. 1983; 11:356–365. [PubMed: 6633254]

Koriat A. Semantic facilitation in lexical decision as a function of prime-target association. Memory and Cognition. 1981; 9:587–598. [PubMed: 7329239]

Kotz SA. Neurolinguistic evidence for bilingual language representation: A comparison of reaction times and event-related brain potentials. Bilingualism: Language and Cognition. 2001; 4:143–154.

Kotz SA, Elston-Güttler K. The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. Journal of Neurolinguistics. 2004; 17:215–235.

Kroll, JF.; Dijkstra, A. The bilingual lexicon.. In: Kaplan, R., editor. The Oxford handbook of applied linguistics. Oxford University Press; Oxford: 2002. p. 301-321.

Kroll JF, Michael E, Tokowicz N, Dufour R. The development of lexical fluency in a second language. Second Language Research. 2002; 18:137–171.

Kutas M, Federmeier KD. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences. 2000; 4:463–70. [PubMed: 11115760]

Kutas M, Federmeier KD. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). Annual Review of Psychology. 2011; 62:621–647.

Kutas M, Hillyard SA. Reading senseless sentences: brain potentials reflect semantic incongruity. Science. 1980; 207:203–205. [PubMed: 7350657]

Kutas M, Hillyard SA. An electrophysiological probe of incidental semantic association. Journal of Cognitive Neuroscience. 1989; 1:38–49. [PubMed: 23968409]

Kutas, M.; Kluender, R. What is who violating? A consideration of linguistic violations in light of event-related brain potentials.. In: Heinze, HJ.; Münte, TF.; Mangun, GR., editors. Cognitive electrophysiology: Basic and clinical applications. Birkhäuser; Boston: 1994. p. 183-210.

Kutas, M.; Van Petten, C. Psycholinguistics electrified: event-related brain potential investigations.. In: Gernsbacher, MA., editor. Handbook of Psycholinguistics. Academic; San Diego, CA: 1994. p. 83-143.

Landauer TK, Dumais S. A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review. 1997; 104:211–240.

Lindsay S, Gaskell MG. A complementary systems account of word learning in L1 and L2. Language Learning. 2010; 60S2:45–63.

Masson MEJ. A distributed memory model of semantic priming. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1995; 21:3–23.

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

McRae K, de Sa VR, Seidenberg MS. On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology: General. 1997; 126:99–130. [PubMed: 9163932]

Mestres-Missé A, Rodríguez-Fornells A, Münte TF. Watching the brain during meaning acquisition. Cerebral Cortex. 2007; 17:1858–1866. [PubMed: 17056648]

Midgley KJ, Holcomb PJ, Grainger J. Language effects in second language learners and proficient bilinguals investigated with event-related. Journal of Neurolinguistics. 2009; 22:281–300. [PubMed: 19430590]

Murphy, GL. The big book of concepts. MIT Press; Cambridge, MA: 2004.

Nadel L, Moscovitch M. Memory consolidation, retrograde amnesia and the hippocampal complex. Current Opinion in Neurobiology. 1997; 7:217–227. [PubMed: 9142752]

Nation ISP. How large a vocabulary is needed for reading and listening? The Canadian Modern Language Review. 2006; 63:59–82.

Newman AJ, Tremblay A, Nichols ES, Neville HJ, Ullman MT. The influence of language proficiency on lexical semantic processing in native and late learners of English. Journal of Cognitive Neuroscience. 2012; 24:1205–1223. [PubMed: 21981676]

Nobre AC, McCarthy G. Language-related ERPs: scalp distributions and modulations by word type and semantic priming. Journal of Cognitive Neuroscience. 1994; 6:233–255. [PubMed: 23964974]

Ojima S, Nakata H, Kakigi R. An ERP study of second language learning after childhood: Effects of proficiency. Journal of Cognitive Neuroscience. 2005; 17:1212–1228. [PubMed: 16197679]

Perfetti CA, Wlotko EW, Hart LA. Word learning and individual differences in word learning reflected in event-related potentials. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2005; 31:1281–1292.

Phillips NA, Segalowitz N, O'Brien I, Yamasakia N. Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. Journal of Neurolinguistics. 2004; 17:237–262.

Plaut DC, Booth JR. Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. Psychological Review. 2000; 107:786–823. [PubMed: 11089407]

Pulido D. The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. Applied Linguistics. 2007; 28:66–86.

Pulido D, Hambrick DZ. The virtuous circle: Modeling individual differences in L2 reading and vocabulary development. Reading in a Foreign Language. 2008; 20:164–190.

R Development Core Team. R: A language and environment of statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2011. ISBN 3-900051-07-0, URL http://www.R-project.org

Reichle ED, Perfetti CA. Morphology in word identification: A word experience model that accounts for morpheme frequency effects. Scientific Studies of Reading. 2003; 7:219–237.

Rodríguez-Fornells A, Cunillera T, Mestres-Missé A, de Diego-Balaguer R. Neurophysiological mechanisms involved in language learning in adults. Philosophical Transactions of the Royal Society B: Biological Sciences. 2009; 364:3711–3735.

Segalowitz, NS. Automaticity and attentional skill in fluent performance.. In: Riggenbach, H., editor. Perspectives on fluency. University of Michigan Press; Ann Arbor, MI: 2000. p. 200-219.

Segalowitz NS, Segalowitz SJ. Skilled performance, practice, and the differentiation of speed-up from automatisation effects: Evidence from second language word recognition. Applied Psycholinguistics. 1993; 14:369–385.

Swaab TY, Baynes K, Knight RT. Separable effects of priming and imageability on word processing: An ERP study. Cognitive Brain Research. 2002; 15:99–103. [PubMed: 12433385]

Tokowicz N, Michael EB, Kroll JF. The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. Bilingualism: Language and Cognition. 2004; 7:255–272.

Tucker DM. Spatial sampling of head electrical fields: The geodesic sensor net. Electroencephalography and Clinical Neurophysiology. 1993; 87:154–163. [PubMed: 7691542]

Turner ML, Engle RW. Is working memory capacity task dependent? Journal of Memory and Language. 1989; 28:127–154.

van Berkum JJ, Hagoort P, Brown CM. Semantic integration in sentences and discourse: evidence from the N400. Journal of Cognitive Neuroscience. 1999; 11:657–671. [PubMed: 10601747]

Wang X, Forster K. Masked translation priming with semantic categorization: Testing the Sense Model. Bilingualism: Language and Cognition. 2010; 13:327–340.

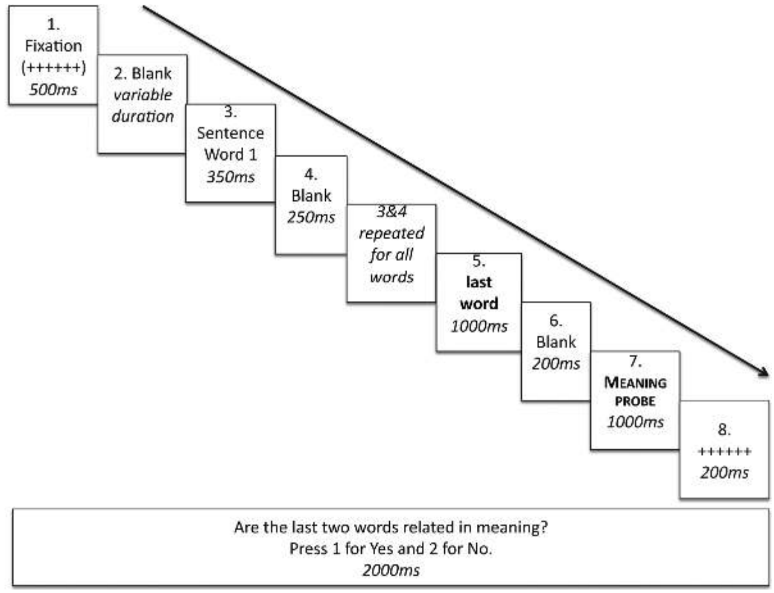


Figure 1.

Semantic relatedness judgement experimental procedure.

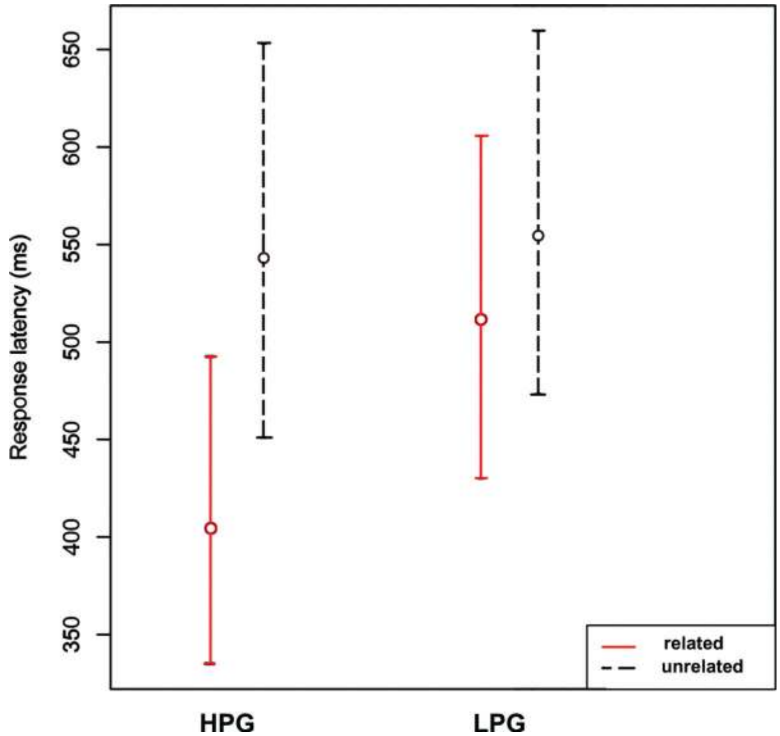


Figure 2.

A partial effects plot showing mean *RTs* (ms) and *mcmc* confidence intervals on related and unrelated trials by proficiency group in the semantic relatedness judgment task.

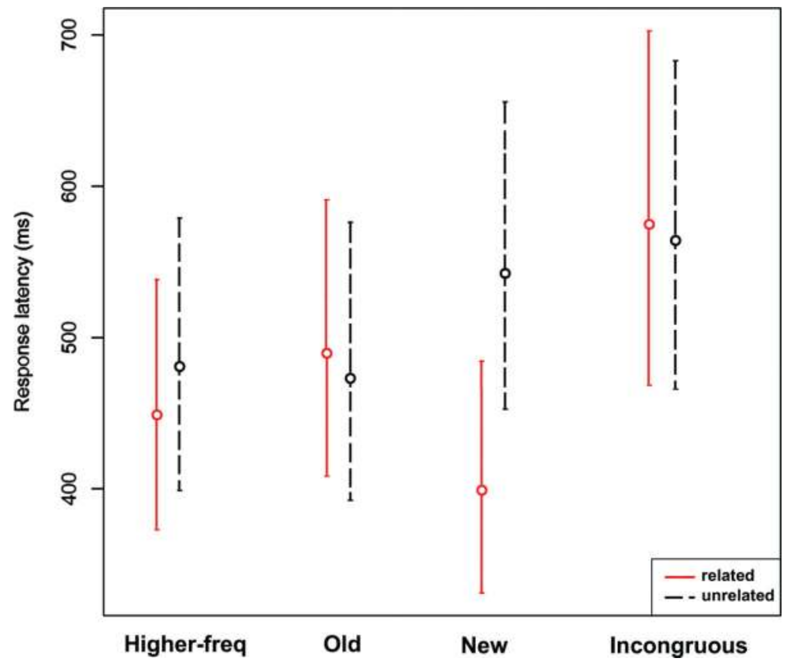


Figure 2a.

A partial effects plot showing mean *RTs* (*ms*) and *mcmc* confidence intervals on related and unrelated trials by condition for the HPG in the semantic relatedness judgment task.

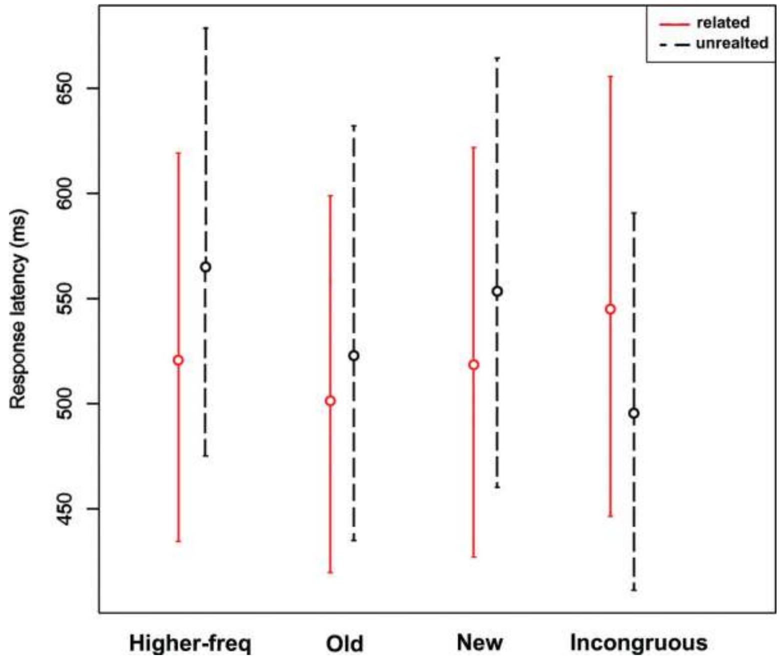


Figure 2b.

A partial effects plot showing mean *RTs* (*ms*) and mcmc confidence intervals on related and unrelated trials by condition for the LPG in the semantic relatedness judgment task.

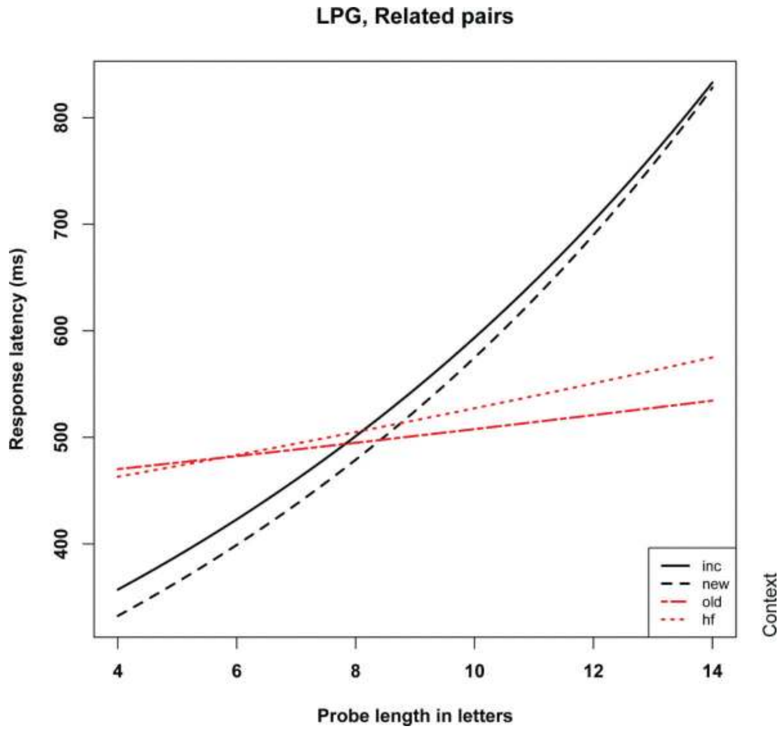


Figure 2c.

A partial effects plot showing the interaction between probe word length and experimental condition in the *RT* analysis for the LPG in the semantic relatedness judgment task.

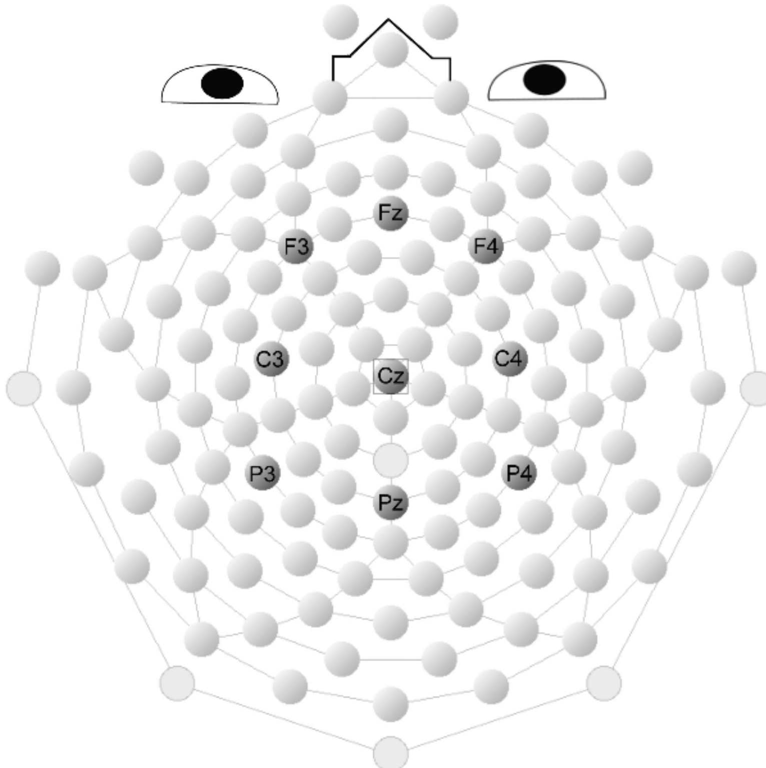


Figure 3.
International 10-20 system of electrode placement.

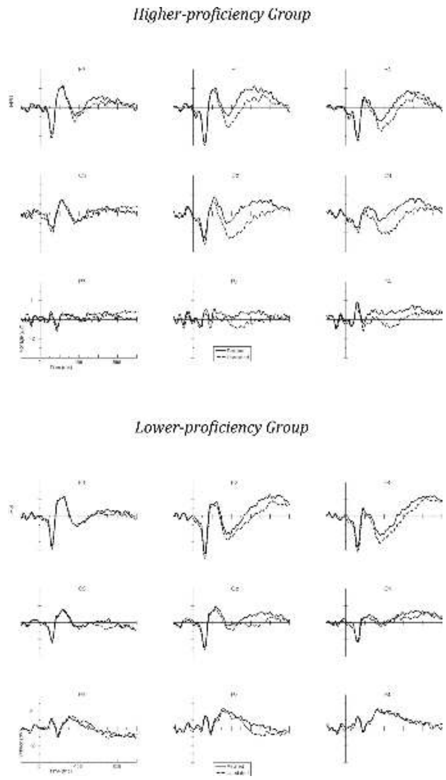


Figure 4.
N400 (μV) effects from nine sites (see Fig. 3 for electrode locations) for the HPG and LPG on the meaning probe, for related and unrelated pairs.

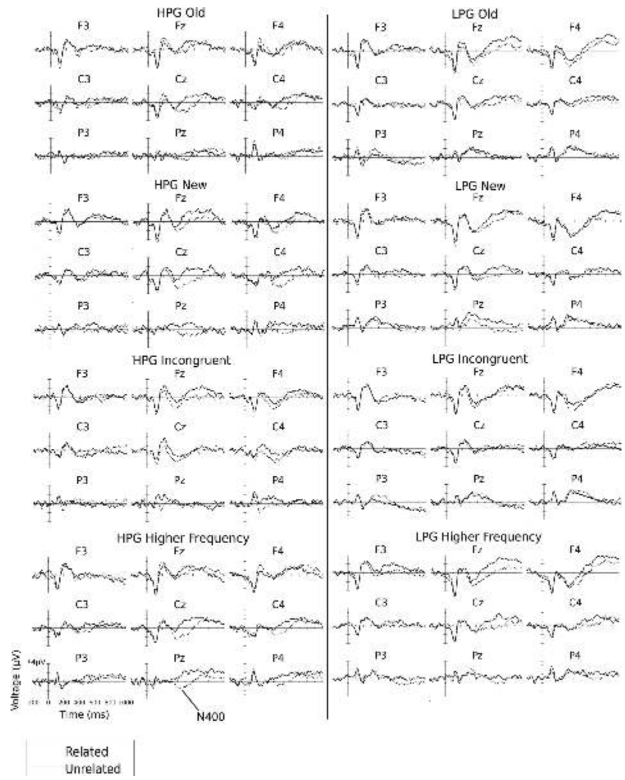


Figure 4a.

N400 (μV) effects from nine sites (see Fig. 3 for electrode locations) for the HPG and LPG on the meaning probe, by sentence-final word condition and relatedness.

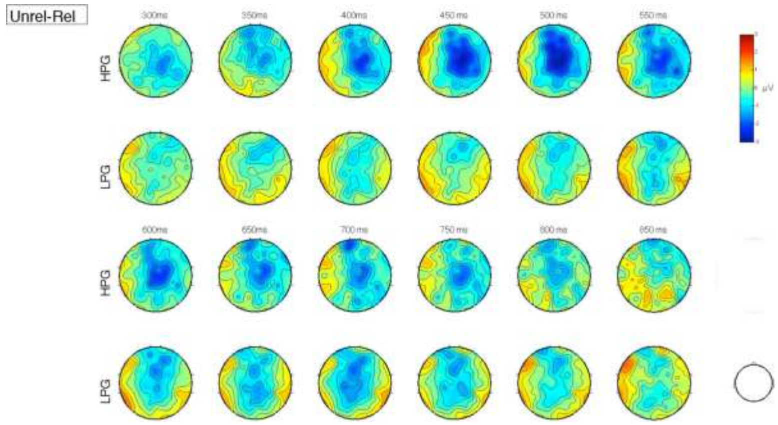
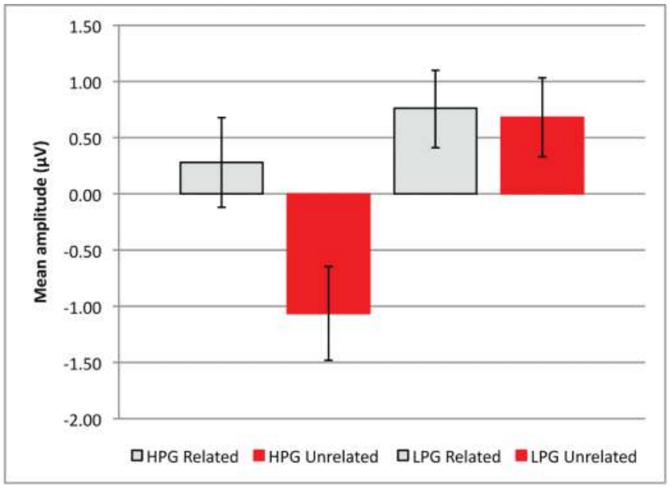
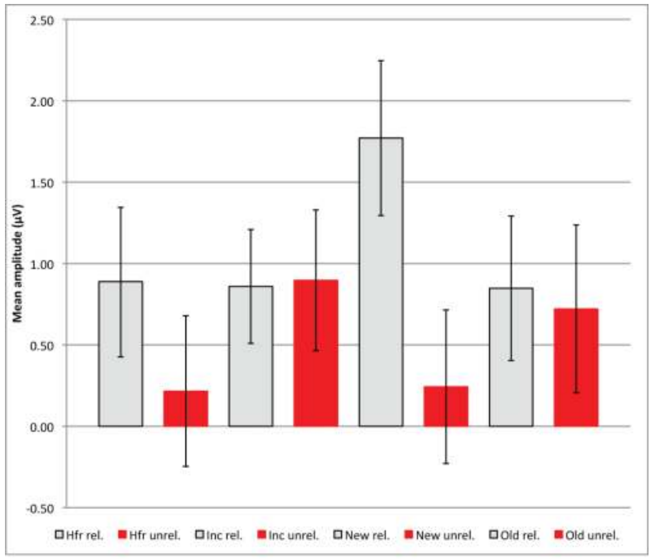


Figure 5.
Scalp voltage maps computed by subtracting the ERPs to meaning probes on related trials from the ERPs to meaning probes on unrelated trials.



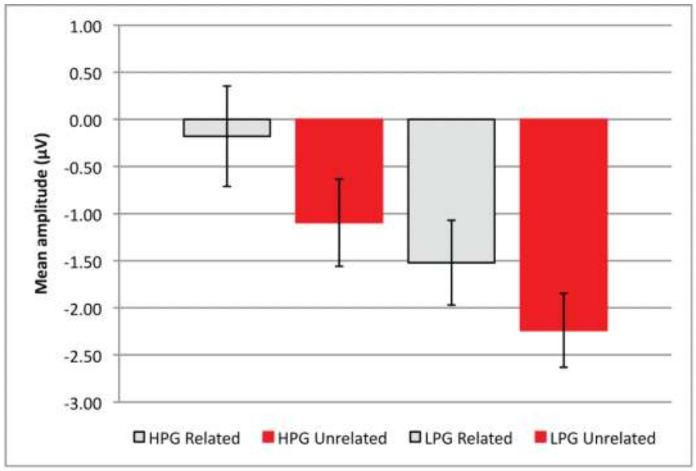
Proficiency	Congruence	Mean	Std. Error	95% Confidence Interval		N400 effect
				Lower Bound	Upper Bound	
HPG	Related	.28	.40	-.55	1.11	1.34
	Unrelated	-1.06	.42	-1.93	-.20	
LPG	Related	.76	.34	.06	1.46	.08
	Unrelated	.68	.35	-.05	1.41	

Figure 6.
Mean N400 amplitude (µV) averaged across central and parietal electrode clusters (Cz, C4, Pz, P4), by proficiency group and meaning congruence (relatedness).



Context	Relationship	Mean	Std. Error	95% Confidence Interval		N400 effect
				Lower Bound	Upper Bound	
Hfr	related	0.89	0.46	-0.06	1.83	0.67
	unrelated	0.22	0.46	-0.74	1.18	
Inc	related	0.86	0.35	0.13	1.58	-0.04
	unrelated	0.90	0.43	-2.0E-03	1.80	
New	related	1.77	0.48	0.78	2.76	1.53
	unrelated	0.24	0.47	-0.74	1.22	
Old	related	0.85	0.44	-0.07	1.77	0.13
	unrelated	0.72	0.52	-0.35	1.79	

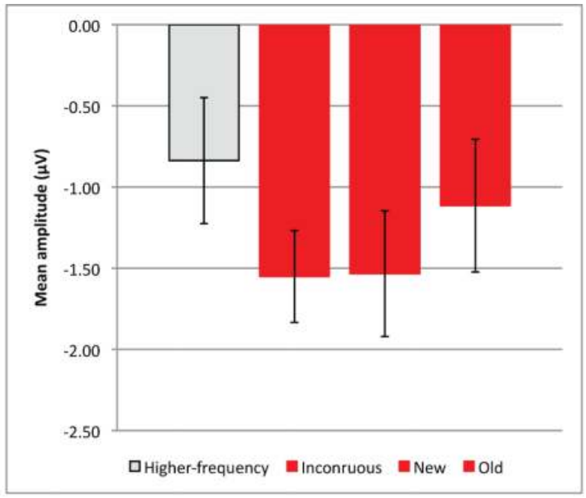
Figure 7.
Mean N400 amplitude (µV) at Pz, by sentence-final word condition and meaning congruence (relatedness).



Proficiency	Congruence	Mean	Std. Error	95% Confidence Interval		fN400 effect
				Lower Bound	Upper Bound	
HPG	Related	-.18	.53	-1.28	.93	0.92
	Unrelated	-1.10	.46	-2.06	-.14	
LPG	Related	-1.52	.45	-2.45	-.59	0.72
	Unrelated	-2.24	.39	-3.05	-1.43	

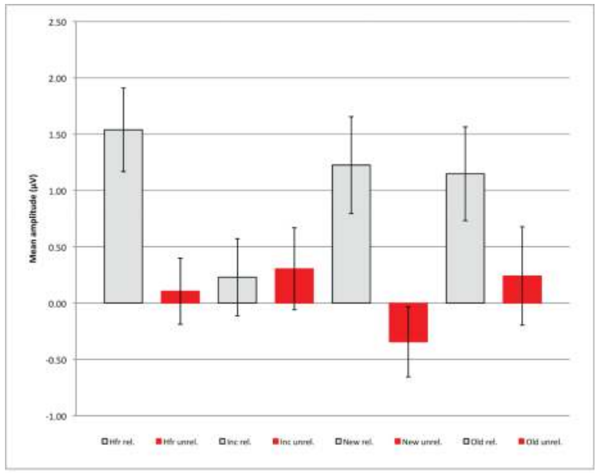
Figure 8.

Mean N400 amplitude (μV) averaged across frontal (Fz & F4) electrode clusters, by proficiency group and meaning congruence (relatedness).



Proficiency	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Higher-frequency	-.84	.39	-1.64	-.03
Incongruous	-1.55	.28	-2.14	-.96
New	-1.53	.39	-2.34	-.73
Old	-1.11	.41	-1.96	-.27

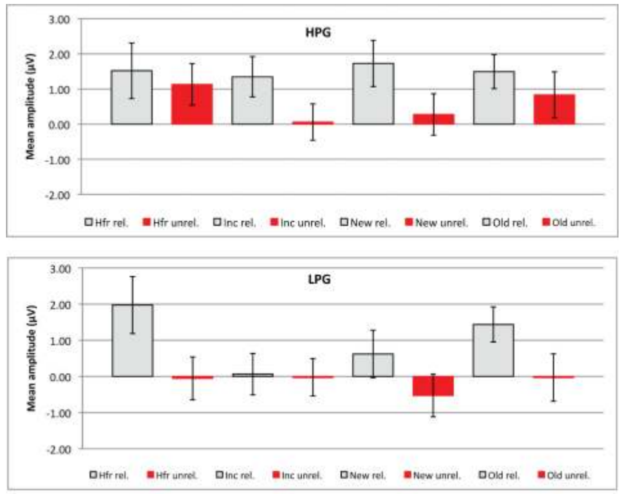
Figure 9.
Mean N400 amplitude (μV) averaged across frontal electrode clusters, by experimental condition.



Context	Relatedness	Mean	Std. Error	95% Confidence Interval		Late meaning congruence effect
				Lower Bound	Upper Bound	
Hfr	related	1.54	.37	.77	2.31	1.43
	unrelated	.11	.29	-.51	.72	
Inc	related	.23	.34	-.48	.94	-.08
	unrelated	.31	.36	-.45	1.06	
New	related	1.23	.43	.33	2.12	1.57
	unrelated	-.34	.31	-.99	.30	
Old	related	1.15	.42	.29	2.01	.91
	unrelated	.24	.44	-.67	1.15	

Figure 10.

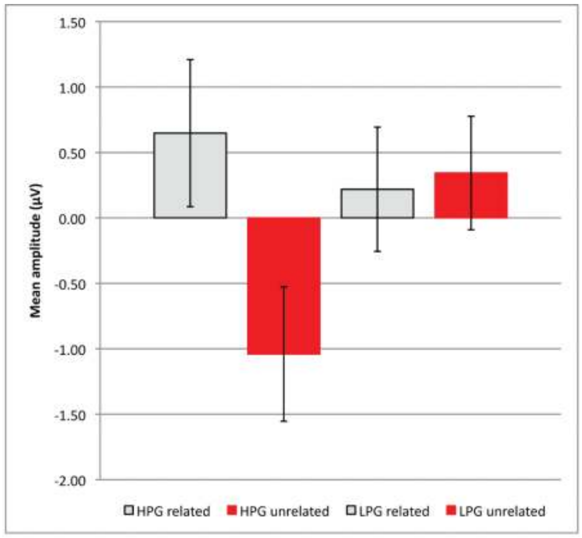
Mean late (500-700 ms after the probe word) ERP amplitude (µV) at central and parietal regions (Cz, C4, Pz, P4), by sentence-final word condition and meaning congruence (relatedness).



Proficiency	Context	Relatedness	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Late meaning congruence effect
HPG	Hfr	related	1.52	.79	-.12	3.15	.39
		unrelated	1.13	.59	-.10	2.35	
	Inc	related	1.35	.57	.16	2.54	1.29
		unrelated	.06	.52	-1.01	1.13	
	New	related	1.73	.66	.37	3.09	1.45
		unrelated	.28	.60	-.94	1.49	
	Old	related	1.49	.48	.49	2.50	.66
		unrelated	.83	.65	-.52	2.19	
LPG	Hfr	related	1.98	.67	.59	3.36	2.03
		unrelated	-.05	.50	-1.09	.98	
	Inc	related	.06	.48	-.94	1.07	.09
		unrelated	-.02	.44	-.93	.88	
	New	related	.62	.56	-.53	1.77	1.15
		unrelated	-.53	.50	-1.56	.50	
	Old	related	1.44	.41	.59	2.29	1.47
		unrelated	-.03	.55	-1.17	1.12	

Figure 11.

Mean late ERP amplitude (µV) at frontal regions (Fz, F4), by sentence-final word condition, meaning congruence (relatedness) and proficiency group.



Proficiency	Relatedness	Mean	Std. Error	95% Confidence Interval		N400 effect
				Lower Bound	Upper Bound	
HPG	related	.65	.56	-.52	1.81	1.69
	unrelated	-1.04	.51	-2.11	.03	
LPG	related	.22	.48	-.77	1.20	-.12
	unrelated	.34	.43	-.56	1.24	

Figure 12.

Mean amplitude (µV) at central-parietal regions (Cz, C4, Pz, P4), by meaning congruence (relatedness) and proficiency group, on trails with higher-frequency words.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Mean response latencies (ms) and percent accuracy (in parenthesis) in the semantic-judgment task, by the sentence-final word condition.

Condition	HPG	LPG
Higher-frequency word	531 (89%)	632 (77%)
Old context	531 (84%)	599 (69%)
New context	561 (88%)	620 (62%)
Incongruous context	635 (63%)	610 (55%)