The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2*

IRINA ELGORT
Victoria University of Wellington
ANNA E. PIASECKI
University of the West of England

Deliberate vocabulary learning is common in the L2, however, questions remain about most efficient and effective forms of this learning approach. Bilingual models of L2 word learning and processing can be used to make predictions about outcomes of learning new vocabulary from bilingual (L2–L1) flashcards, and these predictions can be tested experimentally. In the present study, 41 late adult German–English bilinguals learned 48 English pseudowords using bilingual flashcards. Quality of component lexical representations established for the studied items was probed using form priming and semantic priming. The results show that, although all participants were able to establish robust orthographic representations of the studied items, only bilinguals with large L2 vocabularies established high-quality lexical semantic representations. With neither the Revised Hierarchical Model (RHM) nor the Sense Model able to fully account for these findings, an alternative explanation based on a distributed semantic features view of word learning is proposed. Learning implications of the findings are discussed.

Introduction

Vocabulary knowledge is fundamental to one's ability to understand and be understood. Late bilinguals, who start learning English after acquiring their L1, need to build their L2 vocabularies up to 8000–9000 word families in order to read authentic literature and periodicals with understanding (Nation, 2006). The L2 lexical knowledge can be acquired incidentally through reading, but it has been shown that this learning pathway is slow and uncertain (Cobb & Horst, 2004). Another common way of building up L2 vocabularies is deliberate learning, and learning from flashcards is one of the top independent word learning techniques reported by L2 learners (Oxford & Crookall, 1990). Elgort (2011) found that learning L2 (English) vocabulary using within-L2 flashcards may result in the establishment of robust lexical representations in the memory. After about four hours of learning spread across one week, intermediate and advanced proficiency bilinguals in Elgort's study developed high-quality formal-lexical and lexical semantic representations of deliberately studied L2 items that could be retrieved fluently in lexical decision.

High-quality lexical representations are necessary for rapid and automatic word recognition and are indicative of good readers (Andrews & Hersch, 2010), enabling them to easily and reliably access the word's meaning from its form. Lexical representations that are high-quality (or robust) are fully specified and consistently draw on the nexus of orthographic, phonological and semantic information (Perfetti & Hart, 2001). In Elgort's study, newly-learned pseudowords were used as form primes and semantic primes in three experimental tasks, and quality of their representations was estimated based on their ability to generate predicted form and semantic priming effects, previously observed with known words, in L1 and L2. While Elgort (2011) used within-L2 flashcards, it is equally (if not more) common for language learners to use bilingual flashcards, with L1 translation equivalents or definitions. Therefore, the question addressed in the present study is whether bilingual cards are also effective in establishing robust L2 word knowledge.

In the present study late German–English bilinguals used L2–L1 flashcards to learn 48 L2 (English) vocabulary items over one week. Their knowledge of the newly-learned items was tested using two primed speeded L2 lexical decision tasks that foregrounded either the form or the meaning of these items. An obvious difference between the bilingual and monolingual word learning

Address for correspondence:
Irina Elgort, Victoria University of Wellington, P.O. Box 600, Wellington, 6140, New Zealand
irina.elgort@vuw.ac.nz

modes is that, in the former mode, explicit connections are made between L2 forms and L1 translation equivalents or definitions, while in the latter mode, the practised connections are within the L2. Thus, while effective formal-lexical representations may be established for the target L2 items in both learning modes, high-quality L2 lexical semantic representations may be more problematic when learning from bilingual flashcards, resulting in a less robust overall quality of lexical knowledge (Perfetti & Hart, 2001).

Deliberate bilingual word learning has been examined in previous studies but, to our knowledge, most of these studies used tests that required the use of L1, such as translation priming task (Altarriba & Mathis, 1997; Duyck & Brysbaert, 2004, Experiment 3; Lotto & De Groot, 1998; Witzel & Forster, 2012, Experiment 2). One problem with testing L2 word knowledge in mixed-language tasks is that they provide an additional, external trigger for activating the already dominant L1 lexicon. This conjecture is supported in a recent event-related brain potentials (ERP) study by Guo, Misra, Tam and Kroll (2012), who report that, although L1 translation equivalents were activated by bilinguals when processing L2 words for meaning in a translation verification task, the time-course of this activation indicates that "access to the L1 translation equivalent FOLLOWS the retrieval of the meaning of an L2 word" (p. 17, our emphasis). These results show that meaning can be accessed directly from L2 form and that L1 activation is task dependent.

Results reported in bilingual vocabulary learning studies are mixed, with only some studies showing early semantic involvement in L2 learning. Altarriba and Mathis (1997), for example, found that direct links between L2 lexical and conceptual knowledge were established even for novice English–Spanish bilinguals. After learning L2 (Spanish) items using L1 (English) translation equivalents, English monolinguals (referred to as NOVICE BILINGUALS) were tested using a translation recognition task, in which Spanish words were paired with three types of stimuli: the correct English translation (e.g., *hilo–thread*), an orthographically-similar word different from the correct translation by one letter (e.g., *hilo–threat*) and an unrelated English word (e.g., *hilo–prison*). Participants were required to respond with the "yes" key if the English target word was the correct translation of the Spanish prime, and with the "no" key if it was not. The results showed a significant effect of translation condition on response latencies, with responses being slower in the orthographically related condition than in the unrelated condition. The same group of novice bilinguals was tested again using semantically-related English foils, e.g., Spanish word *hilo* was paired with the correct English translation, *thread*, or a semantically related English word, *needle*. The overall effect of translation condition was again significant, with the semantically related condition

producing significantly slower response latencies compared to the unrelated condition. On the basis of these findings, Altarriba and Mathis (1997, p. 554) concluded that "words in the second language are represented both lexically and conceptually very early in the process of acquisition", even for novice bilinguals. The results obtained by Altarriba and Mathis suggest that meaning information can be coded early in the process of bilingual L2 vocabulary learning, if the conceptual properties of the words are emphasized in the course of their study.

Early encoding of semantic information was also observed in Experiment 3 of Duyck and Brysbaert (2004), in which Dutch speakers learned Estonian (L2) pseudoword labels for numbers between one and fifteen. In this experiment, large semantic effects of number magnitude (i.e., a more rapid activation of the magnitude information for small than for large numbers) was observed in a number-word translation (naming) task, immediately following the learning phase. Moreover, this effect was reliable for both forward (L1 → L2) and backward (L2 → L1) translation, even though L2 number labels were learned just prior to the experiment. These findings suggest that newly-learned L2 number-word forms are "mapped onto existing abstract (magnitude-related) semantic information very early in the L2 acquisition process" (Duyck & Brysbaert, 2004, p. 899).

However, other deliberate bilingual word learning studies failed to demonstrate robust lexical semantic learning. In Witzel and Forster's (2012) study, for example, English monolingual participants learned the meanings of Basque words using English translations. After the training phase, these novice bilinguals completed L1 episodic recognition and lexical decision tasks, in both of which English translation equivalents from the training phase were primed by the newly-learned L2 (Basque) words. Although a robust priming effect was observed when the participants made old–new judgments on the English translation equivalents of the studied L2 words, no priming occurred when they made lexical decisions on the same words. Witzel and Forster took this result to support an earlier conjecture (Jiang & Forster, 2001) that L2 words may be represented in a different (episodic-like) memory system compared to lexically represented L1 words. Setting aside this conjecture, the absence of reliable semantic priming in lexical decision appears to suggest that the L2 words, learned bilingually, failed to establish lexical semantic representations. This finding is counter to the results reported by Altarriba and Mathis (1997) and Duyck and Brysbaert (2004). Therefore we believe further investigation of bilingual L2 word learning is warranted.

Two models of L2 word processing are particularly relevant to framing our predictions regarding the bilingual learning mode: The Revised Hierarchical Model (RHM; Kroll & Stewart, 1994) and The Sense Model (Finkbeiner,
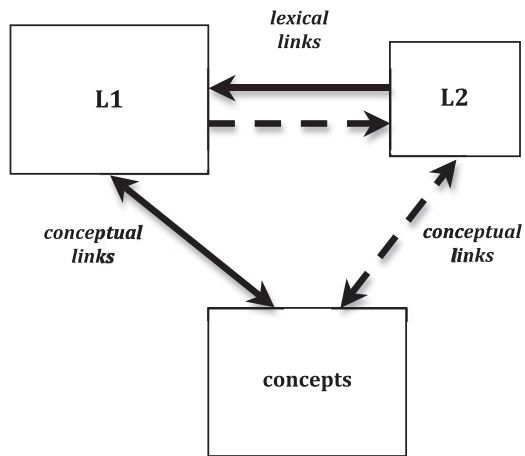
Figure 1. Revised Hierarchical Model (Kroll & Stewart, 1994).

Forster, Nicol & Nakumura, 2004). According to Kroll, Van Hell, Tokowicz and Green (2010), the central issue to which the RHM was addressed is "the way in which new lexical forms are mapped to meaning and the consequences of language learning history for lexical processing" (p. 373). One of the key assumptions of the RHM (Figure 1) is that of an asymmetry in the strength of links between word forms and the conceptual system, when unbalanced bilinguals process L1 and L2 words; i.e., conceptual links being stronger in the native language than in ancillary languages. Another key assumption of the model is that of strong lexical (form-level) connections in the L2 → L1 direction, due to a common practice of learning L2 words by associating them with L1 translation equivalents (precisely the kind of learning investigated in this study).

A number of cross-language semantic priming studies have been cited to support the predictions of the RHM. These studies demonstrate that masked cross-language (translation/semantic) priming in lexical decision occurs consistently when L2 targets are primed with L1 primes, but is considerably weaker or does not occur when L1 targets are primed with L2 primes (Altarriba, 1992; Fox, 1996; Jiang, 1999; Jiang & Forster, 2001; Keatley, Spinks & de Gelder, 1994; Schoonbaert, Duyck, Brysbaert & Hartsuiker, 2009; Schwanenflugel & Rey, 1986), an outcome that has become known as a TRANSLATION PRIMING ASYMMETRY EFFECT. According to the RHM, the bilingual word learning mode should facilitate LEXICAL (form-level) links between L2 and L1 translation equivalents and indirect (L1-mediated) connections between L2 formal-lexical representations and conceptual representations. Only very weak L2 lexical semantic representations are predicted to be established for L2 words learned in this manner (if at all), and access to these representations is predicted to be indirect and effortful (at least in the early stages of learning).

Thus, when used as semantic primes in a speeded within-L2 lexical decision task, vocabulary items learned using bilingual flashcards should not generate a robust priming effect because their L2 lexical semantic representations are likely to be too weak to facilitate the retrieval of semantically related targets.

Similar to the RHM, the Sense Model predicts the translation priming asymmetry effect but, rather than explaining it as a result of L1-mediation, Finkbeiner et al. (2004, p. 3) propose that "priming between semantically related words depends on the *proportion* of shared senses". Following the Distributed Feature Model of meaning processing (e.g., De Groot, 1992; Van Hell & De Groot, 1998), the Sense Model adopts a feature-based view of meaning representations, with bundles of features grouped into semantic senses. For bilinguals who learned L2 after acquiring L1 (and often through L1), it is conjectured that L2 representations are subsets of the semantic senses of their L1 translation equivalents (Wang & Forster, 2010, p. 337). This relative poverty of L2 lexical semantics underpins the translation priming asymmetry effect (Figure 2), which occurs because the proportion of an L2 target's senses activated by a sense-rich L1 prime is hypothesized to be larger compared to that activated by an L2 prime for an L1 target. The Sense Model is not limited to bilingual semantic priming, predicting asymmetries between semantically related words within the same language (see Finkbeiner et al., 2004, Experiment 4). Importantly, the Sense Model does not unequivocally predict attenuation of within-L2 semantic priming in lexical decision with bilingually learned words. If the proportion of an L2 target's senses primed by a recently-learned L2 prime is high enough, according to the Sense Model, reliable semantic priming should be observed.

The study design

The aim of this study is to evaluate whether the bilingual mode of learning is effective in establishing high-quality lexical representations of studied L2 words. Among key characteristics of high-quality lexical representations, Perfetti and Hart (2001) emphasize the reliability (consistency) of synchronous retrieval of orthographic, phonological and semantic information. In this study we focused on two out of the three types of representations – the formal-lexical (orthographic) and lexical semantic representations. We replicated the training regime and experimental design of an earlier L2 deliberate word learning study with monolingual flashcards (Elgort, 2011) because it allows us to compare the quality of L2 lexical representations established under bilingual and monolingual learning conditions; and because the within-L2 testing paradigm used in Elgort's study probes L2 word processing that is similar to real language use (e.g., word
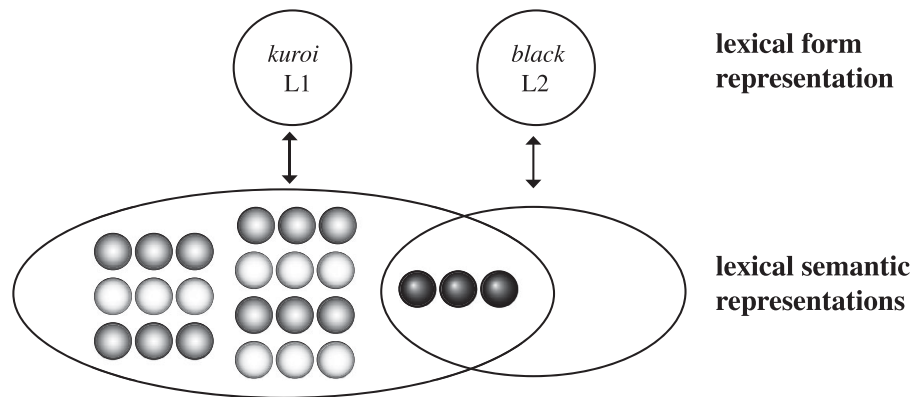
Figure 2. A schematic representation of two translation equivalents (Japanese–English) in the Sense Model (Finkbeiner et al., 2004, p. 9).

processing in L2 reading), moving away from translation priming or verification experiments.

Study participants were instructed to learn English (L2) pseudowords (critical items) from bilingual flashcards, following a recommended learning schedule. Each L2 pseudoword was printed on one side of a card and its short L1 (German) definition and a translation equivalent on the other. This card design ensured that learners could not see both the form and definition at the same time during practice (forcing a retrieval effort), and both languages were used during learning. Following the learning phase, two primed lexical decision tasks were used to examine participants' ability to access the form and meaning of the studied pseudowords online: form priming assessed the establishment of formal-lexical representations, while semantic priming assessed the establishment of lexical semantic representations for the critical items. To further verify whether the priming effects observed with critical items were aligned with those generated by real L2 word primes (for the same group of bilinguals), the experiments included trials with real L2 word primes that were likely to have been familiar to the bilinguals, under the same experimental conditions. The analyses also investigated whether the quality of lexical representations established as a result of deliberate bilingual word learning was a function of participants' L2 lexical proficiency.

Methodology

Participants

Study participants were 41 adult German–English bilinguals (15 male) studying or working in New Zealand, who responded to a call for participation. All participants were late bilinguals, having started to learn the L2 (English) as young adolescents at the average age of 10 ($SD = 2.3$). Their average age was 28.9 ($SD = 8.4$, *Median* = 25).

Individual participant's lexical proficiency in L2 was assessed in terms of the quantity (vocabulary size) and quality (fluency of retrieval) of word knowledge. Their vocabulary size (in word families) was evaluated using a monolingual (English) vocabulary size test (VST; Nation, 2006). Their average VST was estimated to be 9444 word families ($SD = 1689$, *Min* = 5100, *Max* = 13800), indicating intermediate to high L2 proficiency. The fluency of lexical retrieval was estimated using a correlation between bilinguals' response latencies and their coefficient of variation (CV; Segalowitz & Segalowitz, 1993) in a lexical decision task conducted prior to the study (critical items were not used in this task). The CV is calculated as a ratio of an individual's standard deviation of response time (RT) over this person's mean RT, and is interpreted as an indicator of the relative deployment of controlled and automatic processes in behavioral tasks (Phillips, Segalowitz, O'Brien & Yamasakia, 2004; Segalowitz, 2000). 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 indicates heavy dependence on effortful processing and is observed for less skilled language users (Harrington, 2006; Segalowitz & Segalowitz, 1993, p. 381). In the present study, there was a significant positive correlation between participants' CV and RT ($r_s = .758$, $p < .01$), confirming their high L2 proficiency.

Critical items

Thirty-two out of 48 critical items in this study were seven- and eight-letter L2 (English) pseudowords (16 of each kind) from Elgort's (2011). In addition, 16 new six-letter pseudowords were used in this study instead of the nine-letter items from Elgort's study, as it was argued that nine-letter pseudowords were at the limit of the bilinguals' visual acuity (New, Ferrand, Pallier & Brysbaert, 2006),

potentially preventing participants from fully processing the pseudoword primes in the form priming task (Elgort, 2011). The critical items in the present study were all pronounceable English pseudowords, constructed by changing one letter in a real English word – the base word (e.g., *advern* from *adverb*) – that was unrelated in meaning to the base word (*advern* meaning *Säge* "saw"). The base words were used in the form priming experiment as related word targets. The base words were low frequency (CELEX occurrences per million (opm): *Mean* = 5.1, *SD* = 4.5) and had few orthographic neighbors (*Coltheart's N*: *Mean* = .65, *SD* = .93). The meanings assigned to the critical items were in one of the two categories – building (n = 24) or medical (n = 24) terms. Within each category, the meanings were thematically (rather than semantically) clustered, in order to avoid "proactive interference" from semantically similar items (Goggin & Wickens, 1971; Tinkham, 1997) and to facilitate learning, as thematic clustering allows learners to group new words to fit their existing schemata (Brewer & Nakamura, 1984; Mezynski, 1983) (e.g., *obsolate* = *chirurgisch entfernen* "surgically remove"; *regrain* = *Blutgerinnsel* "blood clot"; *aportle* = *Injektionsspritze* "syringe").

Learning materials and procedure

A set of bilingual flashcards was created with the critical L2 items printed on one side of each card and a short definition in German and a German translation equivalent (where possible) on the other. Translation equivalents were provided in parentheses after the definition, e.g., *circhit – Sterile Abdeckung, die auf eine Wunde gelegt wird, um jene vor Infektion oder weiterer Schädigung zu schützen. (Verband)* "sterile covering that is put on a wound to protect it from infection or further damage. (dressing)".[1] If German translation equivalents were cognates of the English words, the meanings of which were signified by critical items, no translation equivalent was included (e.g., for *surmit* meaning "bulldozer" or "tractor", translation equivalents – *Bulldozer* or *Traktor* – were not used). In these cases and for the pseudowords that did not have an L1 (German) translation equivalent, a close L1 semantic relative was embedded in the definition, e.g., for the pseudoword *discrent*, the L1 word *Bodenbelag* "surface/floor covering" was included in the definition. German definitions were constructed to emphasize the semantic features (senses) of the related word targets used in the semantic priming task, e.g., *discrent – Ein dicker, glatter Bodenbelag oder Anstrich* "a thick smooth floor covering or coating" was paired with a related target word *flooring*.

---

[1] Note that English translations in square brackets are supplied purely for the benefit of the reader; in the study, all definitions were presented to participants in German only.

Study participants attended an introductory learning session with one of the researchers during which they completed a computer-based learning procedure and the lexical proficiency tests (VST and lexical decision task) described above. Participants were seated in front of a computer and instructed to learn 48 novel English vocabulary items. Each item was first presented in the middle of the screen with no other text, accompanied by an audio recording of the item. On the following screen each item was presented again, in a bilingual (English–German) dictionary-style entry:

> **dragment** /ˈdræɡmənt/
> **Substantiv (zählbar); Plural: dragments**
> **Bedeutung:** Einrichtung zum Auf- und Abbefördern von Menschen oder schweren Gütern. (Hebezug)

Participants were instructed to memorize the meanings of the novel vocabulary, and were tested offline after each 24 items, using paper flashcards. This task was used to train participants in using flashcards and verify their understanding of the meanings conveyed in the definitions. The mean meaning retrieval score in the building category was 6.3 (*SD* = 3.6) and in the medical category, 8.0 (*SD* = 4.4) correct responses.

At the end of the face-to-face learning session, participants were given a set of flashcards to take home. They were instructed to practice passive (form-to-meaning) and active (meaning-to-form) retrieval of the critical items for one week, following a recommended spaced repetition schedule. Participants were also instructed to keep a learning log where they had to record the date and duration of each learning session, and the number of correct responses. They were instructed to aim to know all 48 items in both directions (form → meaning and meaning → form) by the end of the week. The learning procedure and materials were the same as in the monolingual flashcard study, reported in Elgort (2011), with the only difference that all instructions were in German. The learning logs were collected at the start of the testing session. On average, the participants completed 7.5 learning sessions (*SD* = 1.9, *Median* = 7) that took 140 minutes (*SD* = 45 minutes). Elgort (2011) reported that her participants completed on average 5.8 learning sessions, studying critical items on average for 243 minutes, overall. This means that it took participants in the L2-only study almost twice as long to learn the new vocabulary items, compared to the bilingual study mode, with L2–L1 flashcards.

The testing session

Participants returned for the second face-to-face (testing) session on day eight. They first completed a form priming and a semantic priming lexical decision task containing critical items. After this, they completed a

pen-and-paper task to estimate their explicit written productive knowledge of the critical items. Participants were given a list of L1 definitions (in a pseudorandom order) and were asked to write down the studied vocabulary items corresponding to these definitions. The average explicit knowledge score in this task was 45 items ($SD = 5.4$, $Median = 47$), confirming that the bilingual participants had created explicit bilingual form–meaning links for the studied pseudowords. The same score of 45 items ($SD = 4.2$, $Median = 47$) was recorded in the monolingual flashcard study (Elgort, 2011), indicating that bilinguals in both studies could, on average, explicitly retrieve 94% of critical word forms from their meaning. The main question of this study, however, is about the quality of lexical representations established for the newly-learned pseudowords in the bilingual flashcards learning mode. This question was addressed in the two priming experiments described below.

Experiment 1: Form priming

The form priming experiment in this study was conducted to test whether the bilingual learning method allowed participants to establish native-like formal-lexical representations. At the core of this experiment is the PRIME LEXICALITY EFFECT (PLE) first reported in Forster and Vereš (1998) in an L1 lexical decision task with unmasked primes, and later shown to be present under masked priming conditions (Davis & Lupker, 2006), and for recently-learned L1 words (Qiao & Forster, 2012). The PLE arises due to differences in outcomes of form priming with word and non-word primes: (i) when a word target (e.g., *FUNCTION*) is preceded by an orthographic neighbor that is not a real word (e.g., *bunction*), the target is recognized faster than in the control conditions (when it is preceded by an unrelated word); and (ii) when a word target is preceded by an orthographic neighbor prime that is a real word (e.g., *junction*) this positive priming is attenuated. Positive form priming with nonword primes results from the letter-level facilitation (due to the orthographic similarity between the prime and the target); however, with related word primes, this facilitation is attenuated as a result of the word-level competition between the lexical representations of the prime and the target (Davis & Lupker, 2006; for alternative explanations of the PLE, see Forster & Vereš, 1998; Qiao & Forster, 2012). Thus, it is expected that, once robust lexical representations of new vocabulary items are established, no significant priming should occur when these items are used as form-primes in lexical decision.

Materials and experimental design

In this experiment 48 English word targets (base words) were paired with the following three kinds of primes (i)

the 48 bilingually studied critical items (pseudowords) one letter different from the targets, (ii) 48 nonword primes one letter different from the targets and (iii) 48 unrelated (control) word primes (Appendix A). The nonword primes were also created by changing one letter in the base words (e.g., the nonword *engrive* and the critical item *entrave* were both created from a base word *ENGRAVE*), but the nonwords had not been seen by the participants prior to the task and did not have meanings, while the critical pseudowords did (*to entrave* – "to administer a drug"). For the purpose of creating a lexical decision task, 48 English word-like nonword targets (e.g., *SPRANKLE*) were included in the experiment; they were paired with orthographically related words (e.g., *sprinkle*), orthographically related nonwords (e.g., *sprandle*), and unrelated words (e.g., *goldfish*). With word and nonword targets preceded by orthographically related or unrelated word and nonword primes, and with a low proportion of trials (12.5%) with pseudoword primes per stimulus list in the critical set, the development of response strategies based on primes' lexicality or their orthographic relationship with the targets was unfeasible. The experiment also included 32 unrelated filler trials (with 16 word and 16 nonword targets) to equalize the number of orthographically related and unrelated trials. Additionally, a block of 39 word and 39 nonword targets (eight-letter long) from Forster and Vereš (1998) was also used in the experiment (in the same three experimental conditions), in order to compare the PLE for real English words and the newly-learned critical items, for the same group of bilinguals. The mean frequency (in CELEX) of word targets in the Forster and Vereš (FV) set was 17 opm ($SD = 26$), the mean number of orthographic neighbors (Coltheart's N) was 1.4 ($SD = 0.8$). In addition, 26 unrelated filler trials (13 with word and 13 with nonword targets) were added to equalize the number of related and unrelated trials in this set. The three experimental conditions – (i) related nonword primes (e.g., *engrive–ENGRAVE*; *deadlime–DEADLINE*), (ii) related pseudoword (or word) primes (e.g., *entrave–ENGRAVE*; *headline–DEADLINE*), and (iii) unrelated primes (e.g., *flaming–ENGRAVE*; *monarchy–DEADLINE*) – were counter-balanced across three experimental lists. Each list contained 232 trials – 128 in the critical set and 104 in the FV set. Targets appeared in each counter-balanced list only once, presented in one of the three conditions, and all targets were presented in all three conditions across the three lists. Participants also completed 22 practice trials to get used to the task.

Procedure

All experimental procedures were programmed in E-Prime (Psychological Software Tools, Inc., Pittsburgh,

PA) and presented on a DELL PC (Intel® Core™2 Duo CPU), with a DELL LCD monitor (screen area: 1280 by 1024 pixels; refresh rate: 60 Hz). The following sequence was used in each trial: a string of hash marks (#) → prime → target, all presented in the middle of the computer screen for 490 ms. On the basis of an earlier pilot, the duration of 490 ms was considered to be sufficiently long for the participants to fully retrieve recently-learned vocabulary, and sufficiently short to create time pressure and minimize the use of strategies. The target was replaced by a blank screen displayed until response (or a 3000 ms cut-off). Primes were presented in lowercase letters and targets in uppercase letters, to minimize the graphical shape overlap (Humphreys, Evett & Quinlan, 1990). Participants were instructed to make fast and accurate L2 lexical decisions only to the uppercase stimuli (targets) by pressing the YES or NO button on the response box connected to the computer.

Analysis and results

An initial data inspection led to the exclusion of one participant due to a high error rate (40%). In addition, two participants were excluded from the analysis because they received less than 66% on the pen-and-paper test that measured participants' explicit knowledge of the critical items. Subsequent inspection of the initial dataset resulted in the following exclusions: responses that were faster than 200 ms, and responses to three targets with an error rate higher than 50% (*MAGARINE*, *PHEASANT*, *PRETENSE*). Incorrect responses were excluded from the analysis of response time (RT) data. Inspection of the distribution of RTs revealed a marked non-normality. The RT data were inverse transformed to attenuate the observed non-normality. The dataset was split into two sets that were analyzed separately: (i) the CRITICAL SET that included vocabulary items studies using bilingual flashcards, and (ii) the FV SET that served as a benchmark of the PLE with real L2 words, for the same bilinguals.

In both experiments (form and semantic priming), linear mixed-effects modeling was used in the data analysis. All analyses included participants and items as crossed random effects. A minimally adequate statistical model was fitted to the RT data, using a stepwise variable selection and the likelihood ratio test for model comparisons (Baayen, Davidson & Bates, 2008). The resulting statistical model contained only variables that reached significance as predictors (i.e., their regression weights were significantly different from zero), improved the model fit, or were involved in interactions; all other predictors were excluded from further analysis. Next, the constructed regression model was subjected to model criticism. Potentially harmful outliers (i.e., data points with standardized residuals exceeding 2.5 standard

deviations) were removed and the model was refitted. The measure of statistical significance of the fixed effects in each model was based on Markov Chain Monte Carlo (MCMC) sampling (10,000 iterations; Baayen et al., 2008). For ease of reading, plots of results below are based on back-transformed estimates from the *lmer* models, i.e., with RTs expressed in milliseconds.

The initial statistical model fitted to the data included the experimental condition as the primary-interest predictor. The secondary-interest item variables in the model were ITEM Frequency, Length, and the number of English and German items in the targets' immediate neighborhood. The secondary-interest PARTICIPANT variables included in the initial model were participants' lexical proficiency (VST and CV values), their age and the age when they started learning English (AoA). Finally, two longitudinal predictors were also included, namely Trial Number and participants' RT on a preceding trial (PrevRT; Baayen & Milin, 2010). Predictors used in the final models for the critical and FV datasets (Table 1 and 2, respectively) are listed in summaries of the coefficients of fixed effects of these models. In the final model fitted to the critical dataset, the standard deviation for the by-item random intercept was 0.07, that for the by-participant random intercept was 0.20, and that of the residual error was 0.24. In the final model fitted to the FV set, the standard deviation for the by-item random intercept was 0.07, that for the by-participant random intercept was 0.20, and that of the residual error was 0.21.

For the critical dataset, the results revealed a clear PLE (Figure 3 – note, all plots are based on model predictions). A significant (46 ms; $t = -5.14$, $p < .001$) facilitation was observed when word targets were preceded by orthographically related nonword primes (e.g., *engrive–ENGRAVE*) compared to the control condition (e.g., *flaming–ENGRAVE*), but when these targets were preceded by orthographically related studied pseudoword primes (e.g., *entrave–ENGRAVE*, *entrave* meaning "to administer a drug"), the small 11 ms facilitation was not reliable ($t = -1.13$, $p = .266$). This pattern of results was similar to the PLE observed in the FV set, but the latter effect with real L2 word primes was even more pronounced (Figure 4): a significant 70 ms facilitation occurred with orthographically related nonword primes ($t = -9.69$, $p < .001$), and a non-significant inhibition of 4 ms was observed with related word primes ($t = 0.47$, $p = .620$).

Neither lexical proficiency measure interacted with priming, suggesting that both less and more proficient bilinguals developed formal-lexical representations of the newly-learned L2 items, robust enough to attenuate form priming. Numerical differences in the results for the two datasets will be addressed in the "General discussion and conclusions" section below.

Table 1. *Coefficients of the fixed effects in the regression model for the response latencies in Experiment 1 (form priming – the critical dataset), estimated t-values, 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 level for Cond = Control. Trial, PrevRT and CV were centered to avoid collinearity. E_N_TGT stands for Number of English items in the targets' immediate neighborhood.*

	Estimate	*t*-value	MCMC mean	HPD95 lower	HPD95 upper	*p*
Intercept	−1.2432	−33.97	−1.2480	−1.3071	−1.1888	.0001
Cond = Rel.nw.prime	−0.0782	−5.14	−0.0785	−0.1092	−0.0487	.0001
Cond = Rel.pw.prime	−0.0174	−1.13	−0.0171	−0.0487	0.0116	.2664
Trial	−0.0008	−1.53	−0.0008	−0.0017	0.0003	.1422
PrevRT	0.1267	3.69	0.1387	0.0837	0.1991	.0001
E_N_TGT	−0.0292	−1.98	−0.0293	−0.0577	−0.0010	.0450
CV	1.0026	2.89	0.9649	0.4434	1.5327	.0006

Table 2. *Results summary for the fixed effects in the regression model for the response latencies in Experiment 1 (form priming – the FV dataset). Intercept level for Cond = Control. PrevRT and CV were centered.*

	Estimate	*t*-value	MCMC mean	HPD95 lower	HPD95 upper	*p*
Intercept	−1.2933	−34.28	−1.2939	−1.3529	−1.2336	.0001
Cond = Rel.nw.prime	−0.1403	−9.69	−0.1400	−0.1705	−0.1119	.0001
Cond = Rel.wd.prime	0.0070	0.47	0.0073	−0.0231	0.0358	.6198
PrevRT	0.0315	1.78	0.0414	0.0044	0.0759	.0200
Frequency	−0.0021	−4.10	−0.0021	−0.0032	−0.0012	.0001
CV	1.1524	3.32	1.1409	0.6555	1.6639	.0002
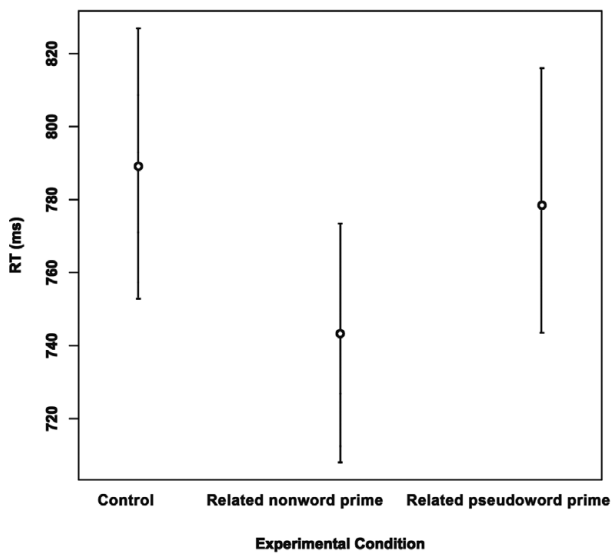
Figure 3. Form priming results for the critical dataset based on model predictions.

Figure 4. Form priming results for the FV (Forster & Vereš, 1998) dataset based on model predictions.

Experiment 2: Semantic priming

The second experiment was conducted with the same participants to evaluate the quality of lexical semantic representations established for the newly-learned items. The design of the experiment was the same as Experiment 3 in Elgort (2011). Since it has been shown

that semantically related primes facilitate lexical decisions on word targets (Collins & Loftus, 1975; McClelland, 1987; McRae & Boisve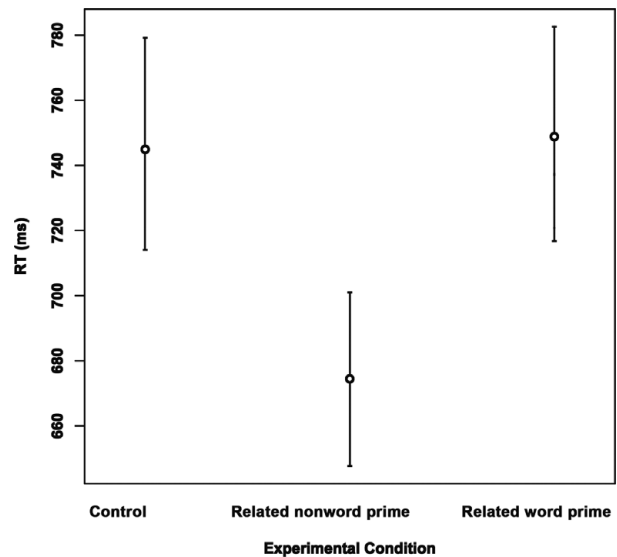rt, 1998), semantic priming with critical item primes can be used to test the quality of their lexical semantic representations. If robust lexical semantic representations of the critical pseudowords had been established, they would facilitate responses to related L2 word targets in lexical decision; however, if only weak (or no) semantic representations were established, there would be no effect (or an inhibition, as explained below).

Semantic priming with newly-learned words

In an earlier semantic priming study with newly-learned L1 words, Dagenbach, Carr and Barnhardt (1990) found a 64-ms facilitation effect on semantically related trials when the meanings of the studied word primes were both recognized and recalled, and a 64-ms inhibition when the meanings were recognized in a multiple-choice task, but not retrieved in a cued recall task. Dagenbach et al. (1990) argued that participants' inability to fully access the meaning of the primes was the cause of the observed inhibition. This is because, in the course of processing a partially-known prime with a weakly established semantic representation, all of its competing semantic neighbors need to be temporarily suppressed (inhibited), making them harder to recognize if they are used as related word targets immediately after the prime. Thus, if the experimental design requires participants to actively attend to partially-known semantic primes, the related targets may be processed slower than in the control condition, i.e., the semantic priming effect may be inhibitory.

Materials and experimental design

In the present semantic priming experiment, 48 critical pseudoword primes were matched with semantically related word targets in such a way that the prime and target shared the semantic senses (microfeatures) foregrounded in the L1 definitions of the pseudowords, used in the bilingual flashcards. For example, the pseudoword *advern*, meaning "*Eine mehrzweck Säge, die in verschiedenen Baubranchen benutzt wird*" "a multi-purpose saw, which is used in various construction industries" was used as a related prime with the target *handsaw*, sharing such semantic features as <is used as a tool>, <is found on building/construction sites>, <is used by builders>. In addition, these targets were also matched with semantically related real word primes, with which they shared the same or similar semantic senses (e.g., *hammer – handsaw*). This related word prime condition was used to verify that semantic priming could be achieved with the given levels of semantic-feature overlap, with the same group of bilinguals. In the control (unrelated)

condition, targets (e.g., *handsaw*) were paired with primes (e.g., *chickenpox*), with which they had no common semantic senses (Appendix B). In addition to the feature-overlap, functional (e.g., *surgeon–scalpel*) and thematic (e.g., *excavate–digger*, *inject–morphine*) relationships were used as the basis of priming (Ferretti, McRae & Hatherell, 2001; Moss, Ostrin, Tyler & Marslen-Wilson, 1995). Three counter-balanced presentation lists were constructed in such a way that each target appeared only once in each list, and was presented in all three conditions (with a RELATED PSEUDOWORD PRIME, RELATED WORD PRIME, or UNRELATED WORD PRIME) across the three lists. To construct the lexical decision task and reduce the proportion of related trials, each experimental list included 144 filler targets (96 nonword and 48 word) in an unrelated condition, in addition to the 48 critical targets.

In this experiment, the average target word frequency (in CELEX) was 7 opm ($SD = 8.1$) and length-in-letters was 7.5 ($SD = 1.4$); for the related word primes, the average frequency was 6.6 opm ($SD = 9.2$) and length-in-letters was 7.3 ($SD = 1.5$); for the unrelated word primes, the average frequency was 4.7 opm ($SD = 5.5$) and length-in-letters was 7.4 ($SD = 1.4$). The online LATENT SEMANTIC ANALYSIS (LSA) tool (http://lsa.colorado.edu) was used to obtain similarity values on related and unrelated trials. Latent Semantic Analysis represents meaning similarity statistically, using distributional characteristics of words in a large body of text (Landauer, Foltz & Laham, 1998). Similarity scores obtained using the LSA approach have been shown to closely match those of human similarity judgments. The semantic similarity values between primes and targets in the related word prime and unrelated word prime conditions (calculated using the LSA tool) in this experiment were significantly different from each other: $F_{(1,47)} = 90.2, p < .001$ ($\eta_p^2 = .657$), with the mean score of .32 ($SE = .028$) for the related condition and .03 ($SE = .008$) for the unrelated control condition. Although, to our knowledge, there are no direct interpretations of the LSA scores, the LSA website gives the following as example input: *cat/mouse* .34; *house/dog* .02, suggesting that the related condition similarity values in Experiment 2 were appropriate for the task.

Procedure

The experimental procedure was the same as in Elgort (2011, Experiment 3), which was based on McRae and Boisvert (1998, Experiment 1). A single-item continuous presentation (without explicit prime-target pairing) and low proportion of related trials (16.7%) were used to minimize the use of decision strategies (McNamara & Altarriba, 1988; Perea & Rosa, 2002; Shelton & Martin, 1992). Primes and targets were presented to the participants in lowercase, one stimulus at a time. The

Table 3. *Results summary for the fixed effects in the regression model for the response latencies in Experiment 2 (semantic priming). Intercept level for Cond = Control. PrimeRT, Trial, T.CELEX (Target frequency), VST and CV were centered.*

	Estimate	*t*-value	MCMC mean	HPD95 lower	HPD95 upper	*p*
Intercept	−1.1841	−33.20	−1.1851	−1.2416	−1.1223	.0001
Cond = Rel.pw.prime	0.0038	0.28	0.0045	−0.0230	0.0309	.7320
Cond = Rel.wd.prime	−0.0641	−4.79	−0.0642	−0.0904	−0.0372	.0001
PrimeRT	0.0928	3.43	0.0894	0.0410	0.1380	.0002
Trial	−0.0009	−3.18	−0.0009	−0.0014	−0.0004	.0012
T.CELEX	−0.0493	−4.22	−0.0493	−0.0701	−0.0281	.0001
VST	−3.1e−05	−1.72	0.0000	−0.0001	0.0000	.0422
CV	0.7650	2.55	0.7629	0.2983	1.2398	.0010
Cond = Rel.pw.pr:VST	−1.8e−05	−2.20	0.0000	0.0000	0.0000	.0230
Cond = Rel.wd.pr:VST	−7.2e−06	−0.88	0.0000	0.0000	0.0000	.3976

participants were instructed to make lexical decisions as quickly and accurately as possible to each stimulus that appeared in the middle of the screen after a 200-ms inter-trial interval, during which a blank screen was displayed.

Analysis and results

The same three participants were excluded from the analysis as in Experiment 1. An inspection of the initial dataset resulted in the following additional exclusions: responses that were faster than 200 ms and slower than 3000 ms, and responses to three targets with an error rate higher than 50% (*ailment*, *pliers*, *ulceration*), as well as their corresponding primes. Inspection of the distribution of RTs revealed a marked non-normality. The RT data were inverse transformed to attenuate the observed non-normality. Only correct responses were included in the RT data analysis.

The initial statistical model fitted to the data included the experimental condition as the primary-interest predictor. The secondary-interest item and participant variables in the model were the same as in Experiment 1. Two longitudinal predictors – Trial Number and participants' RT on a preceding trial were also included (PrimeRT). For the final set of predictors used in the model, see Table 3. In the final model fitted to the semantic priming dataset, the standard deviation for the by-item random intercept was 0.10, that for the by-participant random intercept was 0.17, and that of the residual error was 0.22.

The analysis revealed a reliable (37 ms; *t* = −4.79, *p* < .001) facilitation effect on related trials when targets were preceded by semantically related word primes, but not when they were preceded by related pseudoword primes (*t* = 0.28, *p* = .73), with the latter condition resulting in a numerical 5-ms inhibition.
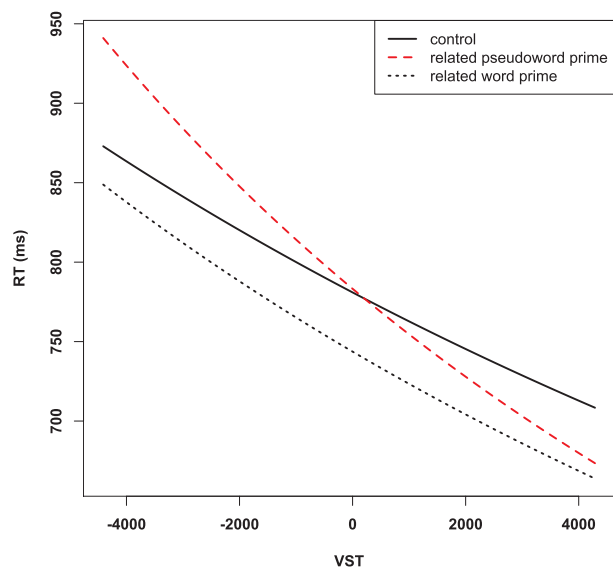

Figure 5. (Colour online) A partial effects plot showing the interaction between participants' L2 vocabulary size and semantic priming. Note that the VST predictor was centered to avoid collinearity. The VST was measured in word families (*Min* = 5100, *Max* = 13800).

The model fit was significantly improved as a result of a reliable (*p* < .05) interaction between priming in the pseudoword condition and participants' lexical proficiency (their estimated vocabulary size) (Figure 5). For the participants with larger L2 vocabularies, studied pseudoword primes facilitated responses to semantically related targets, compared to the control condition, with the facilitation effect size approaching that of semantic priming with real words. However, for the bilinguals with smaller L2 vocabularies, pseudoword priming had an inhibitory effect. Semantic priming with real word primes was unaffected by participants' L2 vocabulary size,

with semantically-related primes consistently facilitating lexical decisions on related trials, compared to control trials.

Summary of findings

The quality of lexical representations established for a set of 48 English pseudowords learned deliberately using bilingual flashcards was evaluated using two speeded lexical decision tasks – a form priming and semantic priming task. The prime lexicality effect (PLE) was used as a test of the formal-lexical representations, while the semantic priming effect was a test of lexical semantic representations of the newly-learned L2 items. The results of the form priming experiment are straightforward – a clear and robust PLE was observed in the critical set with studied pseudowords. This effect mimicked that observed with real L2 word primes in the FV set, for the same group of bilinguals. These results show that lexical representations established for the studied L2 pseudowords were robust enough to generate the PLE – a litmus test of lexicality in the L1 and L2.

The PLE was, however, numerically larger with real word primes than with the studied pseudowords. There are two primary reasons for this result. Firstly, previously known words are likely to be stronger competitors for orthographically related targets than newly-learned words and, therefore, they should be more effective at attenuating positive priming, arising as a result of the letter-level facilitation. This was indeed the case, with a 4-ms inhibition recorded in the word priming condition in the FV set and an 11-ms facilitation in the critical set, with the newly-learned items.[2] Secondly, word targets in the FV set were longer (eight-letter words) than in the critical set (averaging seven letters). Facilitation in form priming with nonword primes tends to increase as the stimulus length increases, because the proportion of an overlap between the prime and the target (in letters) is greater for longer words (Davis & Lupker, 2006). This was true in Experiment 1; the facilitation effect in the nonword priming condition was 24 ms greater in the FV set than in the critical set. A combination of the robust attenuation of priming in the word priming condition and a larger facilitation effect in the nonword priming condition

are the likely reasons for a larger PLE in the FV set. Nevertheless, the fact that a reliable PLE was observed in the critical set is a clear indication that high-quality L2 lexical-orthographic representations were established for the studied items.

The results of the semantic priming experiment are more complex: although no semantic priming was recorded with pseudoword primes overall, a reliable interaction between priming and vocabulary knowledge revealed that positive semantic priming was likely to occur for the bilinguals with larger L2 vocabularies, while an inhibitory effect was more likely to occur for less lexically proficient bilinguals (Figure 5 above). Since robust facilitation was observed for all participants on trials with real L2 word primes, which had similar relationships with the targets as the studied pseudowords, we can be reasonably sure that the semantic similarity between primes and targets was sufficient to produce positive priming. We conjecture, therefore, that the bilinguals' L2 lexical proficiency must have affected the trajectory of L2 vocabulary learning from bilingual flashcards, slowing down the establishment L2 lexical semantic representations for bilinguals with smaller L2 vocabularies. Potential reasons and implications of this finding for L2 vocabulary learning are discussed below.

General discussion and conclusions

Let us first consider our findings in the light of the two models discussed in the introduction. The RHM predicts that only weak connections between L2 form and meaning are established at early stages of learning L2 words, if these words are learned through the L1. This is because the primary connections are created at the level of the form (L2 form → L1 form, e.g., *advern* → *Säge*; *circhit* → *Verband*), with ancillary, mediated connections between the form of an L2 word and the meaning of its L1 translation equivalent. This learning should, therefore, result in lower quality L2 lexical representations, lacking a robust lexical semantic component without which within-L2 automatic semantic priming is unlikely. However, our results show that, at least for more proficient bilinguals, the newly-learned primes facilitated the processing of semantically related L2 word targets. Thus the RHM-based prediction of the absence of the overall semantic priming in Experiment 2 was confirmed, but the facilitation for more proficient participants and the inhibition for less-proficient participants cannot be explained by this model. According to the RHM, participants should process a newly-learned L2 item by associating it with its L1 translation equivalent, activating the L1 lexical semantic representation. Since access to L1 lexical semantic knowledge is automatic and effortless, there is no need

---

[2] We are grateful to Marc Brysbaert for pointing out another potential contributor to the numerical 11-ms facilitation in the pseudoword priming condition – in the experiment the learned pseudowords were always followed by word targets, requiring a "yes" response. However, because the proportion of pseudoword primes in each experimental list was low and the nonword primes were created in the same way as the pseudowords, the likelihood of strategy-based facilitation was predicted to be low. Critically, strategy-based facilitation (if any) was attenuated, as a result of the word-level competition generated by the newly-learned pseudowords, further confirming the establishment of their formal-lexical representations.

to inhibit semantic neighbors, thus no inhibition effect is predicted.

In terms of the Sense Model, Finkbeiner et al. (2004, p. 8) argue that "the semantic sense(s) determining the translation equivalency" are shared between L1 and L2 lexical entries. We expect, therefore, that bilingual flashcard learning leads to a close match between the semantic senses of a studied L2 item and a subset of semantic senses of its L1 translation equivalent. In our study, these overlapping senses are likely to be those foregrounded in the definitions used in the flashcards. With the same senses also underpinning semantic similarity between the prime and the target in the within-L2 semantic priming task (Experiment 2), facilitation is predicted with the newly-learned pseudoword primes, provided a substantial proportion of the target's senses is shared with the prime. Since our findings show facilitation for participants with larger L2 vocabularies, it appears that the Sense Model predictions are partially confirmed. However, if quality of semantic learning is assumed to be the same across different L2 proficiencies, the Sense Model would predict MORE facilitation for less lexically proficient bilinguals and LESS facilitation for more proficient bilinguals. This is because, at higher proficiencies, known L2 words should have richer semantic senses, and a smaller proportion of these senses would be activated by the newly-learned prime (cf. the translation priming asymmetry effect and within-language priming asymmetry effect, Experiment 4 in Finkbeiner et al., 2004). Since this prediction is clearly in the opposite direction for the present findings, the assumption that more and less proficient bilinguals are equally effective and efficient at establishing robust lexical semantic representations of new L2 words is incorrect.

So how can we explain the different outcomes observed for more and less proficient learners in this study? It appears that neither model is able to account for the positive semantic priming for more proficient bilinguals and the inhibitory priming for less proficient bilinguals observed in the study. In search of an explanation for the dynamics of our findings, we turn to alternative connectionist models, particularly those that incorporate the semantic-features view of the semantic domain (Cree, McRae & McNorgan, 1999; Hinton & Shallice, 1991; McRae, de Sa & Seidenberg, 1997) because these models operate "at a level of detail that has a more transparent relationship to underlying [semantic] similarity structure" (Mirman & Magnuson, 2008, p. 66). Similar to the Sense Model, word meanings in the semantic-features models comprise multiple microfeatures that are reused (as building blocks) in creating semantic representations of multiple words. On the neurological level, these microfeatures are represented by consistent patterns of activation. Learning the meaning of a new word involves the strengthening of links between semantic features

(or bundles of features) that represent its meaning. By virtue of sharing semantic features with other words, learning a new word implies integrating its representation into existing lexical semantic memory networks of the learner. Presumably, L2 lexical semantic networks of less lexically proficient bilinguals are less developed, i.e., lexical semantic representations of L2 words may not be fully specified and may have fewer and weaker connections. This insufficient density and efficiency of the L2 lexical semantic networks at lower proficiencies leads to reduced chances of fast effective L2 lexical semantic learning. When learning in a bilingual mode, less proficient bilinguals may therefore rely more on their L1 to help them commit meanings associated with new L2 forms to memory (e.g., use L1 as a mnemonic device). However, fluent L2 word processing, such as that involved in within-L2 semantic priming, requires reliable synchronous retrieval of the L2 formal-lexical and lexical semantic representations.

This distributed microfeatures-based view of lexical semantics and learning is aligned with our results, i.e., only more lexically proficient bilinguals (but not less proficient bilinguals) established high-quality L2 lexical semantic representations that pre-activated overlapping semantic features of the targets in the semantic priming task, facilitating their recognition. The inhibitory priming tendency observed for less proficient participants is also in line with this view, because it predicts much weaker L2 lexical semantic representations for less proficient participants. Referring to Dagenbach et al. (1990), we argue that the inhibition effect is caused by the need for less proficient bilinguals to inhibit competing semantic neighbors (i.e., words with common semantic features) in order to recognize weakly-learned pseudowords in lexical decision (e.g., the word *bandage* is inhibited in the course of lexical decision on the pseudoword *circhit*).

Implications of these findings for approaches to L2 word learning need to be considered. Taken together with the results reported in Elgort (2011), it is clear that deliberate word learning from flashcards results in the establishment of high-quality lexical representations – a prerequisite of the L2 word recognition and processing in real language use. However, unlike the within-L2 learning mode that was equally effective for intermediate and advanced bilinguals (Elgort, 2011), bilingual (L2–L1) flashcards appear to produce better results for bilinguals with larger L2 vocabularies, than for less lexically proficient bilinguals. And yet, it is primarily less proficient language learners who tend to opt for the bilingual learning mode, because it is easier to process the meanings of new L2 words in the L1. This ease of processing, however, may be at the core of the problem with the bilingual learning mode. Recall that it took bilinguals in Elgort's (2011) study nearly twice as long

to learn the critical items in the within-L2 learning mode (using concise L2 definitions with controlled vocabulary), compared to the reported learning time in the present study. This shows that an ability to explicitly retrieve a word's meaning from its form, and vice versa, does not tell us much about the quality of lexical knowledge. Although the within-L2 flashcard learning mode is less efficient in achieving explicit knowledge of L2 words, it appears to be more effective, at least at lower proficiencies. This may be because it encourages learners to rehearse the retrieval of within-L2 form–meaning connections and facilitates integration of new word meanings into the L2 lexicon. Following these findings, we conclude that learners (particularly those at lower-intermediate lexical proficiencies) are likely to benefit more from using within-L2 (rather than bilingual) flashcards, because they facilitate the development of high-quality L2 lexical semantic representations that are needed in real language use, and that are at the core of robust lexical semantic networks in the L2.

Appendix A. Experiment 1: Stimuli used in critical trials.

Related peudoword prime	Related nonword prime	Unrelated word (Control)	Word target
abstair	abutain	gazette	ABSTAIN
advern	adverk	woolly	ADVERB
imputate	ampucate	flooring	AMPUTATE
animote	animage	transit	ANIMATE
antidoth	angidote	sublease	ANTIDOTE
aportle	epostle	camping	APOSTLE
bankrust	sankrupt	leafless	BANKRUPT
beacos	beacoy	comply	BEACON
bracenet	brafelet	stagnate	BRACELET
briening	briefins	sedative	BRIEFING
bockle	bickle	strung	BUCKLE
carnivat	carnital	shrewdly	CARNIVAL
circhit	cirnuit	outdoor	CIRCUIT
colonias	stectral	bruising	COLONIAL
custony	dustody	acclaim	CUSTODY
discrent	dimcreet	refinery	DISCREET
emback	embalk	funnel	EMBARK
entrave	engrive	flaming	ENGRAVE
erramic	euratic	harshly	ERRATIC
evotic	emotic	madman	EXOTIC
pluency	fluenty	aerosol	FLUENCY
dragment	flagment	adhesive	FRAGMENT
gatebay	gatsway	compass	GATEWAY
teometry	geobetry	abdicate	GEOMETRY
slobes	plobes	errand	GLOBES
imigate	umitate	callous	IMITATE
infecent	indewent	fixation	INDECENT
injent	invens	squire	INVENT
wockey	tockey	script	JOCKEY
jeking	jaking	grudge	JOKING
maxidise	maximide	ligament	MAXIMISE
mercusy	mewcury	profile	MERCURY
obsolate	oksolete	mythical	OBSOLETE
treacher	preather	swinging	PREACHER
proster	prospet	honesty	PROSPER
recibe	recile	mortar	RECIPE
regrain	refruin	dictate	REFRAIN
rebailer	retaiver	wardrobe	RETAILER
totate	cotate	muzzle	ROTATE
perial	derial	touchy	SERIAL
scother	smothey	romance	SMOTHER
surmit	sulmit	thrill	SUBMIT
tainor	taifor	bruise	TAILOR
telerant	toleract	immobile	TOLERANT
prolley	tropley	obscene	TROLLEY
utilisk	ukilise	awfully	UTILISE
banity	kanity	dozens	VANITY
wateny	wately	bomber	WATERY

Appendix B. Experiment 2: Stimuli used in critical trials.

Related word prime	Related pseudoword prime	Unrelated word prime	Word target	Category
clotting	regrain	gravel	artery	Medicine
dressing	circhit	wrench	bandage	Medicine
inhaler	perial	foundation	respirator	Medicine
fillings	wockey	stairway	dentist	Medicine
spectacles	slobes	dynamite	eyesight	Medicine
drowsy	evotic	sprinkler	fatigued	Medicine
dislocation	bracenet	lighting	fracture	Medicine
sinuses	beacos	linoleum	infection	Medicine
psychosis	custony	excavate	insanity	Medicine
inject	entrave	unblock	morphine	Medicine
midwife	treacher	meter	newborn	Medicine
soothing	jeking	trainee	painful	Medicine
abdomen	proster	chalet	pelvis	Medicine
dispense	imputate	threshold	pharmacist	Medicine
medic	telerant	artwork	physician	Medicine
anatomy	teometry	carpenter	physiology	Medicine
antibiotic	antidoth	calculate	pneumonia	Medicine
chickenpox	colonias	veranda	scarring	Medicine
saliva	mercusy	contractor	secretion	Medicine
allergy	wateny	decorate	sneezing	Medicine
scalpel	utilisk	tinted	surgeon	Medicine
diagnose	injent	handsaw	symptom	Medicine
remove	obsolate	cupboard	tumour	Medicine
syringe	aportle	wiring	vaccine	Medicine
calculate	bockle	allergy	angles	Building
trainee	tainor	spectacles	apprentice	Building
carpenter	bankrust	diagnose	bricklayer	Building
basin	carnivat	inhaler	bucket	Building
wiring	animote	pulse	cables	Building
plastering	maxidising	heartbeat	ceiling	Building
concrete	infecent	drowsy	cement	Building
chalet	gatebay	saliva	cottage	Building
dynamite	surmit	syringe	demolish	Building
excavate	recibe	dispense	digger	Building
threshold	briening	scalpel	doorway	Building
meter	rebailer	compress	electrician	Building
linoleum	discrent	psychosis	flooring	Building
gravel	erramic	antibiotic	footpath	Building
artwork	banity	kidney	fresco	Building
hammer	advern	chickenpox	handsaw	Building
stairway	abstair	anatomy	ladder	Building
crane	dragment	abdomen	lifting	Building
unblock	imigate	midwife	plumber	Building
veranda	emback	medic	porch	Building
heating	pluency	clotting	radiator	Building
wrench	scother	armpit	screws	Building
tinted	totated	inject	shaded	Building
plank	prolley	hiccup	timber	Building

References

Altarriba, J. (1992). The representation of translation equivalents in bilingual memory. In R. J. Harris (ed.), *Cognitive processing in bilinguals*, pp. 157–174. Amsterdam: North-Holland.

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36,* 550–568.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139,* 299–318.

Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language*, pp. 15–38. Amsterdam: John Benjamins.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428.

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

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

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

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

Dagenbach, D., Carr, T. H., & Barnhardt, T. M. (1990). Inhibitory semantic priming of lexical decisions due to failure to retrieve weakly activated codes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 328–340.

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 668–687.

De Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (eds.), *Orthography, phonology, morphology, and meaning*, pp. 389–412. Amsterdam: Elsevier.

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 889–906.

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

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44,* 516–547.

Finkbeiner, M., Forster, K. I., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language, 51,* 1–22.

Forster, K. I., & Vereš, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 498–514.

Fox, E. (1996). Cross-language priming from ignored words: Evidence for a common representational system in bilinguals. *Journal of Memory and Language, 35,* 353–370.

Goggin, J., & Wickens, D. D. (1971). Proactive interference and language change in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 453–458.

Guo, T., Misra, M., Tam, J. W., & Kroll, J. F. (2012). On the time course of accessing meaning in a second language: An electrophysiological investigation of translation recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1165–1186.

Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. In S. Foster-Cohen, M. M. Krajnovic & J. M. Djigunovic (eds.), *EUROSLA Yearbook*, pp. 147–168. Amsterdam: John Benjamins.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

Humphreys, G. W., Evett, L. J., & Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology, 22,* 517–560.

Jiang, N. (1999). Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition, 2,* 59–75.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language, 44,* 32–51.

Keatley, C. W., Spinks, J. A., & de Gelder, B. (1994). Asymmetrical cross-language priming effects. *Memory & Cognition, 22,* 70–84.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149–174.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition, 13,* 373–381.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259–284.

Lotto, L., & De Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48,* 31–69.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (ed.), *Attention and performance XII: The psychology of reading*, pp. 3–35. Hillsdale, NJ: Lawrence Erlbaum.

McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language, 27,* 545–559.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 558–572.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126,* 99–130.

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

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 65–79.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 863–883.

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

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin and Review, 13,* 45–52.

Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal, 7,* 9–30.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research, 66,* 180–194.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (ed.), *On the consequences of meaning selection*, pp. 67–86. Washington, DC: American Psychological Association.

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

Qiao, X., & Forster, K. I. (2012). Novel word lexicalization and the Prime Lexicality Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1064–1074.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition, 37,* 569–586.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25,* 605–618.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (ed.), *Perspectives on fluency*, pp. 200–219. Ann Arbor, MI: University of Michigan Press.

Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatisation effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14,* 369–385.

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

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

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1,* 193–211.

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

Witzel, N. O., & Forster, K. I. (2012). How L2 words are stored: The episodic L2 hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1608–1621.