

Research Article

Cite this article: Jia, R. and Jiang, N. (2025). Examining strength of L2 form-meaning connection: A study of intralingual L2 semantic priming. *Bilingualism: Language and Cognition* 1–11. <https://doi.org/10.1017/S1366728925100667>

Received: 25 February 2025

Revised: 13 July 2025

Accepted: 22 September 2025

Keywords:

the strength of L2 form-meaning connection; L2 semantic priming; task effects; lexical decision task; semantic categorization task; advanced L2 learners

Corresponding author:

Ruirui Jia;

Email: rjia@umd.edu

 This research article was awarded Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

Examining strength of L2 form-meaning connection: A study of intralingual L2 semantic priming

Ruirui Jia  and Nan Jiang

Second Language Acquisition, University of Maryland, College Park, MD, USA

Abstract

The study investigated the strength of L2 form-meaning connections among advanced L2 speakers. Two unmasked intralingual L2 semantic priming experiments were conducted, with lexical decision and semantic categorization tasks. Thirty-eight native English speakers and 40 advanced Chinese learners of English were tested in each task. The stimuli involved L2 word targets that were preceded by either a related L2 prime or an unrelated one. Previous research has used the lexical decision task in this investigation, and the semantic task was also used in the present study to boost the involvement of conceptual connections in L2 processing. Consistent with previous findings, native English speakers showed a reliable priming effect in both tasks, but English L2 speakers showed no priming effect in either task. No task effect was found in either group. The findings provided further evidence for a weaker L2 form-meaning connection among advanced L2 speakers.

Highlights

- Native English speakers demonstrated a significant semantic priming effect in both tasks.
- L2 semantic priming effect was not significant for L2 learners in either task.
- Task effects were not found for both native English speakers and L2 learners.
- L2 form-meaning connection is weak and nonnative-like for advanced L2 learners.

1. Introduction

A long-standing issue in bilingual processing research is the connection between L2 words and concepts or L2 lexical form-meaning connections. It consists of several specific questions. The first is whether a bilingual's two languages are linked to a shared or two separate semantic or conceptual systems. The second is whether L2 words are directly linked to conceptual representations or indirectly through L1 translations. A third issue is the strength of connections between L2 words and concepts if such direct connections exist. Earlier research seemed to have helped create a consensus regarding the first two questions. For example, researchers have compared response time between two tasks: L2 picture naming and L1–L2 translation. If the two languages are linked through shared conceptual representations, performing the two tasks would involve similar activation pathways, that is, picture-concept–L2 word in picture naming and L1 word–concept–L2 word in L1–L2 translation. As a result, bilinguals should complete the two tasks in a similar amount of time. However, if L2 words are only connected with L1 translations rather than via shared concepts, bilinguals should perform the translation task much faster than the L2 picture naming task. Several studies showed similar response latencies in the two tasks (e.g., Chen & Leung, 1989; Kroll & Curley, 1988; Potter et al., 1984), and these findings have been considered as evidence for a direct connection between L2 words and concepts. Similarly, comparable semantic priming effect size from interlingual and intralingual prime-target pairs (as reported by Caramazza & Brunes, 1980; Frenck & Pynte, 1987; Schwanenflugel & Rey, 1986) were taken as evidence for the view that “bilinguals have a single semantic representation that subserves two distinct sets of lexical entries for their two languages” (Caramazza & Brunes, 1980, p. 81). Thus, it is not surprising that current models of bilingual lexical representation such as the Revised Hierarchical Model (RHM, Kroll & Stewart, 1994), the distributed conceptual feature model (DCFM, de Groot, 1992, 1993), and the BIA+ model (Dijkstra & van Heuven, 2002) all recognized shared conceptual representations by the two languages and a direct L2–concept connection.

1.1. The strength of L2-concept connections: Conflicting findings

The third issue, that of the strength of connections between L2 word form and meaning, remains controversial. Many researchers endorse the view that L2 word–concept connections are weaker

in strength than L1 word-concept connections, and acknowledge, at the same time, that this connection can become stronger with increased L2 proficiency (e.g., Kroll & Tokowicz, 2005). Other researchers emphasize a robust and direct conceptual connection with L2 words, often based on the findings that a semantic effect can be observed in L2 processing that is similar in magnitude to that in L1 processing (e.g., Duyck & de Houwer, 2008).

Empirical evidence is mixed. Some studies reported comparable semantic processing effects in L1 and L2. For example, a similar imageability effect (de Groot & Poot, 1997) and a similar concreteness effect (de Groot et al., 1994) were found in forward (L1 to L2) and backward (L2 to L1) translations. De Groot et al. (2002) also reported that semantic properties (imageability, context availability, and definition accuracy) affected the Dutch-English bilinguals' lexical decision latencies in both L1 and L2. Duyck and Brysbaert (2004, 2008) examined the topic in terms of the number magnitude effect, for example, individuals responding to smaller numbers (e.g., 2 or 3) faster than larger numbers (e.g., 8 or 9), and they found that bilinguals showed a similar number magnitude effect in both translation directions. Duyck and de Houwer (2008) adopted a semantic Simon task to explore this issue. They presented animal names (e.g., fox) and occupation names (e.g., driver) to Dutch-English bilinguals in upper and lower cases and asked them to say "animal" to uppercase targets and "occupation" to lowercase targets. They manipulated the word type (animal or occupation names) and letter case (upper or lower) such that they may be congruent (e.g., FOX presented in uppercase) or incongruent (e.g., fox presented in lowercase). This setup allowed them to determine if there was a congruency effect (or a semantic Simon effect in their term) in processing L1 and L2. They found a comparable congruency effect of 28 ms and 23 ms for L1 and L2 items, respectively. The finding that newly learned words could produce reliable semantic priming effects, as reported by Elgort (2011) and Elgort and Piassecki (2014), is also consistent with immediate and robust semantic connections for L2 words.

Where other paradigms were adopted, many studies have shown a weaker semantic effect associated with an L2. Researchers have compared the emotional effects in L1 and L2 processing, for example. The emotional effect refers to an elevated response to emotional words (e.g., *murdur*, *war*) compared to neutral words (e.g., *sleep*, *jar*), as assessed in terms of recall accuracy, valence or arousal rating scores, response latencies, electrophysiological responses, level of skin conductance, or pupil size in visual word recognition. There is increasing evidence showing a reduced emotional effect in L2 processing. In one of the earliest studies, Anooshian and Hertel (1994) asked English-Spanish bilinguals to first rate the difficulty and emotionality of a set of emotional and neutral words, which was then followed by a surprise recall task. The participants showed a better recall rate for emotional words than for neutral words in their L1, but not in L2. Winskel (2013) tested Thai-English in a color decision task with emotional and neutral words displayed in different colors. The bilinguals showed a delay in responding to emotional words as compared to neutral words in L1, but they showed no such effect in L2. Eilola et al. (2007) and Eilola and Havelka (2010) demonstrated that while participants showed an elevated skin conductance response to L1 negative and taboo words, they did not show this pattern in L2. Similar results were also reported by Iacozza et al. (2017), where the emotional effect was measured in terms of emotionality rating scores and pupillary responses, that is, the size of the pupils as recorded by an eye tracker. Spanish-English bilinguals showed different pupillary response patterns while performing the task in

their L1 and L2. The change in pupil size between negative and neutral words was larger in L1 than in L2. Assuming that the emotional effect is semantic in nature, that is, arising from the meaning of the stimulus words, these findings suggested a weaker L2-concept connection.

The same pattern can be seen in false memory studies where L1 and L2 are compared. These studies often employed the Deese-Roediger-McDermott (DRM) paradigm. A set of words related to a theme is displayed to participants for memory in the study phase, for example, *beer*, *drunk*, *liquor*, *wine*, *glass*, and *alcohol*, as related to the "alcohol" theme. In the test phase, words are displayed for a recognition task in which participants have to decide if a word is one of those shown in the study phase. The test items included both previously displayed "old" words that require a positive response, and "new" words that were not displayed previously and thus require a negative response. Among the latter were words related or not related to the themes involved in the study phase. Individuals often produce an incorrect positive response to a theme-related word, for example, *whiskey*, than to a control word, which is referred to as the false memory effect. Several studies have shown that this effect was much weaker in L2 than in L1. For example, the false recognition rate was 29% and 15% in the dominant and nondominant languages among the English-Spanish bilinguals tested by Arndt and Beato (2017, Experiment 1). They also showed that more proficient L2 speakers produced a higher false recognition rate (22%) than less proficient bilinguals (13%). This L1-L2 difference has also been shown in several other studies (Anastasi et al., 2005; Cabeza & Lennartson, 2005; Howe et al., 2008; Miyaji-Kawasaki et al., 2004; Sahlin et al., 2005).

1.2. Semantic priming studies involving L2 primes: Further inconsistency

A frequently used research paradigm for studying word-concept connections, also the one employed in the present study, is the priming paradigm. Strength of conceptual connections for L1 and L2 words can be assessed by comparing the semantic priming effect size involving L1 and L2 primes. This can involve both intralingual (L1-L1, L2-L2) and interlingual (L1-L2, L2-L1) prime-target pairs. Semantic priming arising from semantically or associatively related prime-target pairs, such as *dog-cat* or *dog-mao* (for Chinese-English bilinguals), is more informative than translation pairs, as translation priming can potentially arise from direct lexical links.

Evidence from semantic priming studies is also quite inconsistent, though. Many early bilingual priming studies showed a relatively consistent pattern: interlingual semantic priming effects were stronger in the L1-L2 direction than in the L2-L1 direction. For example, Schwanenflugel and Rey (1986) showed a priming effect of 135 ms and 47 ms in the two priming directions in their Experiment 1, and that of 63 ms and 12 ms in Experiment 2, where the stimulus onset asynchrony was reduced from 300 ms to 100 ms. The priming effects in the two directions were 91 ms and 15 ms in Jin (1990), 95 ms and 10 ms and 45 ms and 5 ms in the two experiments reported by Tzelgov and Eben-Ezra (1992), and 38 ms and -5 ms in Keatley et al. (1994, Experiment 1, SOA 250 ms). This pattern has been replicated in more recent studies, for example, 53 ms and -3 ms in Smith et al. (2019). The priming effect in the L2-L1 direction was absent in multiple studies. These results suggest that L2-concept connections are weaker than L1-concept connections. Other studies, however, have reported comparable priming effects or an opposite pattern, for example, 20 ms and 33 ms in Keatley and de Gelder (1992, Experiment

1, high proportion of related items), 18 ms and 26 ms in Grainger and Beauvillain (1988, Experiment 2, long SOA), and 170 ms and 250 ms in Chen and Ng (1989, Experiment 2).

All these studies adopted an unmasked priming paradigm in which the prime was visible to a participant, which may give rise to guessing or anticipation once a participant noticed a semantic relationship between the prime and the target. Subsequent masked priming studies overcame this problem by presenting the prime subliminally, for example, for 50 ms, and by sandwiching it between a visual mask and the target. Two studies showed a semantic priming effect in both directions (Perea et al., 2008; Schoonbaert et al., 2009). Basnight-Brown and Altarriba (2007) also tested bilinguals in both priming direction, but reported priming only from the dominant language. Two additional masked priming studies tested bilinguals only in the L1–L2 direction. Williams (1994) showed a semantic priming effect only among items of strong semantic relation, but de Groot and Nas (1991) reported semantic priming from L1 to L2 only when the prime was visible; there was no masked semantic priming from L1 to L2.

In contrast to a relatively large number of interlingual semantic priming studies, only a small number of studies have examined intralingual semantic priming in L2. Frenck and Pynte (1987) tested English–French bilinguals with English or French targets (e.g., *sparrow* or *ormoineau*) displayed alone or preceded by a category name in the same language (e.g., *bird* or *oiseau*). They found an intralingual semantic priming effect of 81 ms and 21 ms in L1 and L2, respectively, among their skilled bilinguals. Grainger and Beauvillain (1988) also reported significant semantic priming in L1 and L2. de Groot and Nas (1991) included L2–L2 semantic priming involving noncognates in their Experiment 3, and a significant masked priming effect of 35 ms was found. Smith et al. (2019) tested unbalanced Hebrew–English bilinguals on semantically related and unrelated prime–target pairs in both L1 and L2 in a lexical decision task (LDT). The prime duration was 150 ms followed by 50-ms blank screen. They found a significant semantic effect of 41 ms in L2, but the 14 ms priming effect in L1 was not significant. The latter finding was inconsistent with those of previous studies in which reliable semantic priming was found in L1 (e.g., 28 ms of priming in Perea et al., 2008).

1.3. The present study

L2-concept connections represent a fundamental issue in bilingual processing research. In light of the inconsistent findings, the present study was intended as a further effort to examine this issue. Specifically, we adopted the unmasked priming paradigm to examine L2–L2 semantic priming among unbalanced but advanced bilinguals in two tasks: lexical decision and semantic categorization. These methodological decisions were made on the basis of several considerations that we explain below.

First, we examined L2–L2 semantic priming for three reasons. First, whether L2 primes were able to show a reliable semantic priming effect will shed light on the strength of L2-concept connections. A reliable L2–L2 semantic priming effect would provide evidence for a strong L2-concept connection. Previous research has already documented such priming effects, but methodological characteristics of these studies often prevent them from generalizing the findings. For example, Perea et al. (2008) reported L2–L2 semantic priming, but they tested balanced bilinguals. Smith et al. (2019) also reported this effect, but they adopted a high relatedness ratio (1:1) of their test items, which could have led to strategic effects. Second, compared to a large number of interlingual

semantic priming studies, there have been a small number of studies that examined intralingual semantic priming in L2. Finally, intralingual L2–L2 priming was preferred over L2–L1 interlingual priming not only to avoid language switching complications but also for a better chance of observing a reliable priming effect. As Keatly et al. (1994) argued, within-language association is generally stronger than cross-language association as words in the same language share the same semantic network, which would result in more efficient transfer of activation to related words stored in the same interrelated network.

We chose to test unbalanced L2 speakers because this is the population (in contrast to balanced bilinguals) that has been studied in most bilingual processing studies. Current models of bilingual representation, such as the revised hierarchical model of Kroll and Stewart (1994) and BIA+ model of Dijkstra and van Heuven (2002), are based on the findings from this population. We focused on advanced L2 speakers because substantial lexicosemantic development has occurred in these individuals, and their performance (in contrast to lower-proficiency L2 speakers) would be more informative about whether strong L2-concept connections can be developed in L2 learning.

Furthermore, the unmasked priming paradigm was chosen over the masked priming paradigm because semantic priming under the masked condition was often quite small even among L1 speakers, for example, 7 ms, 17 ms, and 11 ms with prime duration of 33 ms, 50 ms, and 67 ms, respectively, in Perea and Gotor (1997), 11 ms in Tan and Yap (2016), and 9 ms and 20 ms with the prime displayed for 40 ms and 60 ms, respectively, in Kiefer et al. (2023). Given slower processing associated with L2 words, using the masked priming paradigm risks missing the priming effect due to increased processing demand. We were aware of the risk of the participants developing strategic effects once they became aware of the semantic relationship between an unmasked prime and a target. This risk was dealt with by keeping a low proportion of related items, as was done in previous studies (e.g., Frenck & Pynte, 1987; Grainger & Beauvillain, 1988).

Finally, the comparison of semantic priming in the two tasks of lexical decision and semantic categorization was inspired by the task effect in translation priming. Several studies have demonstrated that masked L2 words often failed to prime their L1 translation in an LDT. However, when the task was switched to a semantic categorization task, reliable L2–L1 translation priming effects can be observed. This task effect has been demonstrated where the masked priming paradigm was adopted (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Xia & Andrews, 2015). The same task effect was also reported where a two-phase priming paradigm was adopted. In the latter case, participants were asked to perform a lexical task, such as the LDT or a semantic task, such as animacy judgment in L2, in the study phase, which was then followed by a test phase in which the participants performed the same tasks but with words in the other language. An L2–L1 translation priming effect was observed only when a semantic task was performed (e.g., Taylor & Francis, 2017; Zeelenberg & Pecher, 2003). The same task effect has also been observed in semantic priming in L1 (e.g., de Wit & Kinoshita, 2015a). The task effect suggests that word-concept connections are more likely to be involved or boosted in a semantic task than in a lexical task (Gollan & Kroll, 2001; Zeelenberg & Pecher, 2003). Thus, the semantic task offers a better opportunity for semantic priming to materialize.

Thus, the present study examined the strength of L2-concept connections via L2–L2 semantic priming. Two favorable conditions

were created for this priming effect to surface if strong L2-concept connections existed: the display of visible L2 primes and the adoption of a semantic task in comparison to a lexical task. A group of native speakers (NS) of English was included for comparison. We expected NS to show a reliable semantic priming effect in both tasks, thus replicating previous findings and confirming the adequacy of the design. They may produce a stronger semantic priming effect in the semantic task than in the lexical task. If strong L2-concept connections are present, we would expect advanced L2 speakers to show a semantic priming effect, maybe smaller in magnitude in comparison to NS, and at least in the semantic task. If L2 speakers did not show a priming effect in both tasks, the findings would suggest a weak word-concept connection in L2 speakers.

2. Experiment 1: L2 semantic priming in LDT

2.1. Method

2.1.1. Participants

The study tested 38 native English speakers and 40 L2 learners. Native English participants were recruited from the SONA system and were all undergraduate students at a mid-Atlantic university in the United States, with an average age of 19.28 ($SD = 1.16$). All the L2 learners were recruited through snowball sampling and were native Chinese speakers, learning English as a foreign or second language. 80% of them were current university students (10% undergraduate junior and senior students; 70% graduate students) in the United States, majoring in various disciplines (37.5% in language/linguistic-related fields, 37.5% in social sciences, and 25% in STEM), and the rest 20% were recent graduates who were now working full-time in the United States. All of them used English on a daily basis with an average age of 26.23 years old ($SD = 3.98$).

The language learning and use questionnaire, adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007), indicates that the average age of acquisition for the L2 learners was 8.12 ($SD = 2.69$). All of them have been living in an English-speaking country for an average of 3.98 years ($SD = 3.32$). 75% of the participants reported that they were generally more comfortable using Chinese, and the rest reported being equally comfortable using both English and Chinese, indicating that most of them were L1 dominated. Moreover, the participants scored on average 75.38% ($SD = 8.87$; Min = 58.75%; Max = 92.5%) in the LexTALE test, which confirmed their high English proficiency, ranging from level B2 to level C2 (Lemhöfer & Broersma, 2012). Their self-reported English proficiency on a 10-point Likert scale in reading ($M = 7.98$, $SD = 1.40$), listening ($M = 7.75$, $SD = 1.26$), speaking ($M = 7.1$, $SD = 1.52$), and writing ($M = 6.93$, $SD = 1.62$) also demonstrated their high English proficiency. All participants had normal or corrected-to-normal vision without any language or learning problems. They were given a \$10 Amazon gift card for their participation.

2.1.2. Materials

Thirty-six semantically related English word pairs were selected from the USF Free Association Norms (Nelson et al., 1998). The selected stimuli were all concrete nouns containing three to seven letters with forward association strength greater than 0.38 and word frequency larger than 10 times per million. These critical items were chosen based on the following criteria: 1) the forward association strength of words should be as large as possible to ensure a high

relatedness of the word pairs; 2) the frequency of each prime and target word should be as high as possible to ensure that participants were familiar with the stimuli; 3) the target words should be easily categorized into either the “man-made” or “natural” category to be used in both tasks, so we selected concrete nouns instead of other parts of speech; 4) the length of words should be kept minimum to facilitate recognition. By applying those criteria to material selection, 42 pairs of English words were initially selected. Five highly proficient Chinese learners of English who shared a similar language learning background with the participants were instructed to complete an English-Chinese translation task for all selected pairs. They also rated their familiarity with each word on a scale from 1 (least familiar) to 5 (most familiar). Six word pairs were removed from the list due to either low translation accuracy or low familiarity ratings. Eventually, there were a total of 36 semantically related pairs selected for the present study. The mean familiarity ratings for the primes and targets were 4.97 and 4.99, respectively, suggesting that Chinese learners of English were very familiar with these word pairs. The average word association strength of the final pairs was 0.55, which means that, on average, 55% of the native English speakers tested in Nelson et al. (1998) were able to produce the targets when presented with the primes.

Two counterbalanced lists were created. The target words that were primed by semantically related words in List 1 were primed by semantically unrelated words in List 2. The related primes had a mean frequency of 48.86 per million and were on average 5.00 letters long with 5.67 mean concreteness. The unrelated primes were matched to the related primes on word length, concreteness, and frequency. See Table 1 for the lexical properties of the items. To avoid strategic processes, the proportion of the related items was reduced to 20% (McNamara, 2005) by adding another 54 unrelated filler pairs to make up a total of 90 word pairs. Both filler primes and filler targets were matched with the related primes and targets on word length, concreteness, and frequency, respectively. Moreover, to ensure that the same material can be used in the following semantic categorization task, all the 90 word pairs had an equal number of word targets that can be categorized as either “natural” ($N = 45$) or “man-made” ($N = 45$). To construct a lexical decision task, 90 orthographically legal and pronounceable English nonwords were selected from the ARC non-word database (Rastle et al., 2002) by matching the word length to that of the critical word targets. Moreover, another 90 English words that match the preceding English primes on word length, concreteness, and frequency were selected to prime the nonwords. The same filler pairs and nonword pairs were incorporated into both lists. To help participants familiarize the task, six practice items were added at the beginning of each list. Those items were different from the items

Table 1. Lexical properties of the primes and targets

	Length	Concreteness	Frequency
Related primes	5.00	5.67	48.86
Unrelated primes	5.00	5.74	46.11
Filler primes	4.96	5.76	47.35
Nonword primes	4.89	5.59	48.82
Critical targets	4.42	5.82	117.63
Filler targets	4.79	5.69	111.81
Nonword targets	4.69	/	/

used in the main experiment, and responses to these items were not included in the data analysis. Thus, each list contained a total of 186 trials, including six practice pairs, 36 critical word pairs, 54 unrelated filler pairs, and 90 nonword pairs. The trials (except for the practice pairs) were pseudorandomized to ensure that participants did not provide the same response for more than three consecutive items. Additionally, participants were randomly assigned to one of the two lists and they only saw each target word once throughout the experiment.

2.1.3. Procedure

The experiment was conducted online. The participants first read and signed the consent form before moving on to the main experiment. The experiment was delivered via PsychoPy (2022.2.4). Both the primes and targets were displayed in lowercase to facilitate recognition by Chinese learners of English (see Jiang, 2021). In the lexical decision task, participants were instructed to decide whether the letter strings shown on the screen were real English words or not. They were directed to press the “a” button on the keyboard if the letter strings were words and “l” if they were not. Participants were encouraged to create tags or take notes to help them remember which button represents which decision to avoid confusion during the experiment. They were also instructed to respond as quickly and accurately as possible. Their reaction times were recorded from the onset of the target until they provided a response. The experiment began with a fixation (+) appearing on the screen for 500 ms, followed by a prime lasting for 250 ms and by a target which remained on the screen until the participants responded or for a total of 3 seconds. After the main experiment, participants were instructed to complete the LexTALE test and fill out the revised LEAP-Q questionnaire. The sequence of the experiment was to avoid test effects, as the LexTALE and the main experiment share a similar task format.

2.1.4. Data analysis

Before the main analyses, the data were first cleaned and trimmed. Participants whose error rate was above 20% were excluded from the final dataset ($N = 2$). The RT analysis focused only on the 36 critical items. We first deleted all the incorrect responses. This led to the exclusion of 2.34% of the data in the native speakers and 1.61% in the L2 learners. RT outliers were also removed as they may not reflect a genuine word recognition process (Jiang, 2012). We excluded RTs longer than 2,500 ms and shorter than 300 ms and RTs that deviated from each participant’s mean RT by 2.5 standard deviations (Jiang, 2012). This procedure resulted in the deletion of another 3.89% of the data in the native speakers and 2.97% in the L2 learners. The whole data trimming procedure affected a total of 6.23% of the data for the native speakers and 4.58% for the L2 learners. After data trimming, the data were inverse transformed ($-1000/\text{RT}$) to improve distribution normality (as judged from the histograms, the inverse transformation functioned better than the log transformation).

Then, the trimmed and transformed RT data were analyzed through linear mixed effects (LME) models. All the analyses were conducted using R (Version 4.2.1: R Core Team, 2022) in the R Studio environment with the *lme4* (Version 1.1–33: Bates et al., 2015) and the *lmerTest* (Version 3.1–3: Kuznetsova et al., 2017) packages. The transformed RT data were treated as the outcome variable and were continuous in nature. Condition (Related vs. Unrelated), Group (Native vs. L2), Association Strength, Target Word Frequency, Concreteness, and Length were taken as fixed effects, with Association Strength, Target Word Frequency,

Concreteness, and Length being continuous and Condition and Group being categorical. All the continuous variables were grand-mean-centered, and the two categorical variables were dummy-coded with the related condition and native speakers as the reference groups. Item and Participant were treated as random intercepts, and both by-item and by-participant random slopes were considered in each analysis.

To fit the random effects, the maximal random effects structures recommended by Barr et al. (2013) were used, meaning that all the by-participant and by-item random intercepts and slopes were first included in the model. We examined whether deleting one random component would significantly decrease the model fit. If removing one random component led to a significant decrease in the model fit, the more complex model would be retained. Otherwise, a simpler model should be chosen. Models were estimated using the maximum likelihood technique. The χ^2 likelihood ratio test with its associated p -value was examined for model comparison in this stepwise procedure.

To evaluate the fixed effects, we adopted a confirmatory approach. This was achieved by treating the transformed RT data as the outcome variable, Condition, Group, and the interaction between Condition and Group as the fixed effects, with Association Strength, Target Word Frequency, Concreteness, and Length as covariates. The semantic priming effect for native speakers was indicated by the main effect of Condition when the reference group was set as native speakers. We relevelled the reference group to L2 learners to evaluate the L2 semantic priming effect for the L2 learners. We acknowledge the use of generalized mixed-effects models (GLMM) with inverse-Gaussian or gamma distributions to handle positively skewed RT data. However, we did not adopt this approach because such models often present convergence issues and are sensitive to complex random-effects structures, as in the present case. Complete mixed-effects model output and R code for all the analyses are available on Open Science Framework (OSF): <https://osf.io/d2ph4>

2.2. Results and discussion

Table 2 displays the descriptive statistics of the average RT and ER of the native and non-native English participants’ performance in the LDT. The ERs did not differ systematically across the related and unrelated conditions for both the native English speakers and L2 learners.

When we set native English speakers as the reference group, the best-fitting LME model with by participant random slopes for Condition and Strength and by-item random intercept demonstrates that there was a significant main effect of Condition $b = 0.074$, $SE = 0.022$, $p = .001$, indicating that native English

Table 2. Native English speakers’ and L2 learners’ mean RT (in ms) and ER (in percentage) in the LDT (standard deviation in parentheses)

	Native English speakers		L2 learners	
	RT	ER	RT	ER
Related	607.24 (168.48)	2.05 (14.17)	668.34 (149.15)	1.46 (12.01)
Unrelated	642.50 (208.08)	2.63 (16.02)	670.76 (150.84)	1.75 (13.14)
Priming	35.26*	0.58	2.42	0.29

Note: Asterisk means statistically significant at $p < .05$ level. Inferential statistics were computed on inverse-transformed RTs, and the raw means should be interpreted descriptively only.

speakers performed significantly slower in the unrelated condition than in the related condition, meaning that the semantic priming effect was significant for native speakers. However, after releveling the reference group to L2 learners, the main effect of Condition was not significant, even controlling for covariates: $b = 0.005$, $SE = 0.022$, $p = .817$, indicating that L2 learners did not respond significantly differently in the related and unrelated conditions. The result suggests a non-significant L2 semantic priming effect for L2 learners. The interaction between Condition and Group was significant: $b = -0.069$, $SE = 0.031$, $p = .027$, indicating that the priming effect differed significantly between native speakers and L2 learners. This suggests that the two groups engaged in qualitatively different processing in the LDT (i.e., the NSs showed an effect, whereas the NNSs did not).

In sum, Experiment 1 examined L2 intralingual semantic priming with both native English speakers and L2 learners in the LDT. The results demonstrate a significant semantic priming effect for native English speakers, but a null effect for L2 learners. The nonsignificant L2 semantic priming effect was consistent with the findings of many previous bilingual processing studies (e.g., Chen & Ng, 1989; Jin, 1990; Smith et al., 2019), indicating a weak connection between L2 words and their meanings.

3. Experiment 2: L2 semantic priming in SCT

As reviewed earlier, previous translation priming studies (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998) have consistently found a more robust presence of L2 translation priming effects in a semantic task than in a lexical decision task. Given that translation priming can be confounded by language switching and form-level activation across languages, intralingual L2 semantic priming may be a better paradigm to examine the strength of L2 form-meaning connections. However, whether a significant L2 semantic priming effect could be found in a semantic task among advanced L2 learners is unknown. Thus, Experiment 2 was conducted to see whether a robust L2 semantic priming effect could be observed in a semantic categorization task.

3.1. Method

3.1.1. Participants

Another 38 native English speakers and 40 highly proficient L2 learners were recruited. The adoption of a between-group design was to make sure that no participant would see the same targets twice throughout the experiments. To ensure the equivalence of learner characteristics, Experiment 1 and Experiment 2 were conducted concurrently, and the participants were randomly assigned to either the LDT or the SCT. Two L2 participants who reported feeling more comfortable using English than Chinese were excluded. Thus, there were a total of 38 L2 learners. Table 3 displays the learner characteristics of the L2 learners recruited for the two

experiments. A set of independent samples t -tests or Wilcoxon rank sum tests (for nonnormally distributed data) demonstrate that there was no statistically reliable evidence for a difference between the two groups in terms of age: $W = 772.5$, $p = .904$, $r(\text{rank biserial}) = 0.02$, 95%CI $[-0.24, 0.27]$, age of acquisition: $t(76) = 1.30$, $p = .20$, Cohen's $d = 0.29$, 95%CI $[-0.15, 0.75]$, length of residence: $W = 610$, $p = .134$, $r(\text{rank biserial}) = -0.20$, 95%CI $[-0.43, 0.06]$, LexTALE score: $t(74) = -0.79$, $p = .430$, Cohen's $d = -0.181$, 95%CI $[-0.63, 0.27]$, self-reported reading: $W = 786.5$, $p = .788$, $r(\text{rank biserial}) = 0.03$, 95%CI $[-0.22, 0.28]$, listening: $W = 724$, $p = .714$, $r(\text{rank biserial}) = -0.05$, 95%CI $[-0.30, 0.21]$, writing: $W = 754.5$, $p = .959$, $r(\text{rank biserial}) = -0.007$, 95%CI $[-0.26, 0.25]$, and speaking proficiency: $W = 700$, $p = .542$, $r(\text{rank biserial}) = -0.08$, 95%CI $[-0.32, 0.18]$, revealing that the two groups possessed a comparable level of English proficiency and shared a similar language learning background.

3.1.2. Materials

The same 36 critical word pairs and 54 filler pairs used for Experiment 1 were adopted in Experiment 2. Similar to Experiment 1, the critical items were counterbalanced into two lists. The filler items were the same across the two lists. Thus, each presentation list includes 36 critical word pairs and 54 unrelated filler pairs displayed pseudo-randomly in each list along with six practical pairs at the beginning of each list.

3.1.3. Procedure

After the consent process, participants were asked to complete a semantic categorization task where they needed to decide whether the target word refers to something that is “man-made” or “natural.” If a target word refers to something made by humans (e.g., *table*), participants were instructed to press the “a” button on the keyboard. If a target word represents something that comes from nature (e.g., *ocean*), they were asked to press the “l” button. All other details remained the same as those outlined in Experiment 1.

3.1.4. Data analysis

The same data trimming and transformation procedure, statistics models, and R packages were used to analyze the RT data obtained from the SCT. One native English speaker was excluded due to a high error rate. The deletion of the incorrect responses led to 7.88% of the data loss in the native speakers and 7.53% in the L2 learners. The removal of outliers further affected 3.50% of the data in the native speakers and 3.40% in the L2 learners. The total number of data excluded due to the cleaning procedure accounted for 11.38% of the data in the native speakers and 10.93% in the L2 learners. The same statistical analyses conducted in Experiment 1 were also applied to Experiment 2. In addition, an extra analysis was conducted to explore task effects. This process involves combining the data collected from the LDT and the SCT. The transformed RT was the outcome variable, Condition, Group, Task (LDT vs. SCT; LDT as the reference group), and their interaction were treated as fixed

Table 3. Characteristics of the L2 participants (mean and standard deviation in parentheses)

	Age (year)	Age of acquisition (year)	Length of residency (year)	LexTALE (%)	Reading	Listening	Writing	Speaking
Exp.1: LDT (N = 40)	26.23 (3.98)	8.12 (2.69)	3.98 (3.32)	75.38 (8.87)	7.97 (1.40)	7.75 (1.26)	6.92 (1.62)	7.10 (1.52)
Exp.2: SCT (N = 38)	26.34 (4.07)	7.34 (2.63)	5.45 (4.16)	77.07 (9.93)	7.92 (1.15)	7.87 (1.32)	6.97 (1.15)	7.34 (1.34)

Table 4. Native English speakers' and L2 learners' mean RT (in ms) and ER (in percentage) in the SCT (standard deviation in parentheses)

	Native English speakers		L2 learners	
	RT	ER	RT	ER
Related	922.05 (320.55)	7.66 (26.61)	1096.96 (372.49)	7.16 (25.81)
Unrelated	957.43 (350.06)	8.11 (27.32)	1087.61 (366.92)	7.89 (26.99)
Priming	35.38*	0.45	−9.35	0.73

Note: Asterisk means statistically significant at $p < .05$ level. Inferential statistics were computed on inverse-transformed RTs, and the raw means should be interpreted descriptively only.

effects with appropriate by-item and by-participant random intercepts and slopes being entered into the initial model. The evaluation of both the random and fixed effects followed the same procedure as described in Experiment 1.

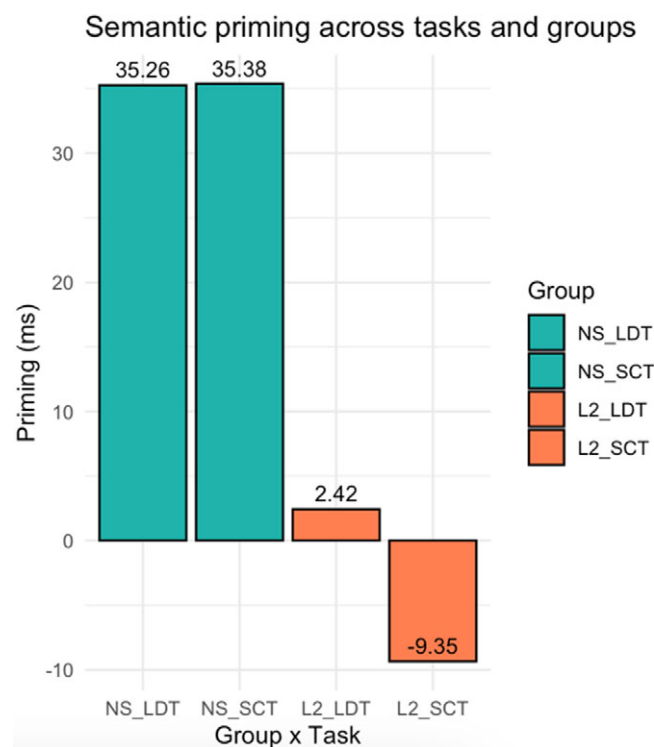
3.2. Results and discussion

Table 4 presents the descriptive statistics of the average RT and ER of the native and non-native English participants' performance in the SCT. The ERs of both the native and L2 speakers did not differ systematically across the related and unrelated conditions.

The best-fitting LME model with by-participant and by-item random intercepts demonstrates that there was a significant main effect of Condition: $b = 0.041$, $SE = 0.015$, $p = .008$, meaning that native speakers responded significantly slower in the unrelated condition than in the related condition. The result suggests that a significant semantic priming effect was found for native English speakers. However, after releveling the reference group to L2 learners, Condition was not a significant predictor of RT: $b = -0.002$, $SE = 0.015$, $p = .877$, indicating that L2 semantic priming effects were not significant for L2 learners even controlling for covariates. The interaction between Condition and Group was significant: $b = -0.043$, $SE = 0.021$, $p = .045$, indicating that the semantic priming effects were significantly different between native English speakers and L2 learners. This finding suggests that the two groups engaged in qualitatively different processing in the SCT.

Regarding task effects, the best-fitting model, which included by-participant random slopes for Condition and by-item random slopes for Tasks, failed to reveal a significant interaction between Condition and Task for native English speakers: $b = -0.035$, $SE = 0.028$, $p = .203$, meaning that the semantic priming effects were similar across the two tasks. Thus, no task effect was found for native speakers. A similar pattern was also observed for L2 learners. After releveling the reference group to L2 learners, the interaction between Condition and Task was not significant either: $b = -0.037$, $SE = 0.027$, $p = .891$, signaling the lack of a task effect for L2 learners. Figure 1 plots the semantic priming effects across tasks and groups.

In sum, Experiment 2 investigated the L2 intralingual semantic priming effect via a semantic categorization task. A significant semantic priming effect was only observed among native English speakers, but not among L2 learners. Additionally, no task effect was found for either group. The results were inconsistent with the findings of previous translation priming studies (e.g., Grainger & Frenck-Mestre, 1998; Finkbeiner et al., 2004; Xia & Andrew, 2015), where a significant L2 translation priming was more robustly found in a semantic task among unbalanced bilinguals.

**Figure 1.** The semantic priming effect across tasks and groups.

4. General discussion

To summarize the results of both experiments, native English speakers demonstrated a significant semantic priming effect in both the LDT and the SCT, and the priming magnitude was similar across tasks. This finding confirmed the adequacy of the experimental design: a reliable priming effect could be obtained with the experimental setup if strong word-concept links were present. The lack of the task effect among native speakers was not consistent with the expectation and the results of de Wit and Kinoshita (2015a). However, similar priming effect sizes have been reported in previous studies involving native speakers. For example, in an unmasked priming study with an SOA of 250 ms, Sanchez-Casa et al. (2006) reported a semantic priming effect of 29 ms and 27 ms in the lexical and semantic tasks, respectively, for their “very close” semantic pairs, and the priming effect was 14 ms and 17 ms in the two tasks for the “close” semantic pairs. No task effect was observed. Thus, the task effect among native speakers can be an evasive phenomenon and is dependent on the circumstances of the experimental setup.

In contrast to the reliable semantic priming effect among native speakers, advanced L2 learners failed to produce a significant L2 semantic priming effect in both tasks. The non-significant L2 semantic priming effect found in the LDT corroborated the results of many previous bilingual processing studies showing no semantic priming when L2 words served as primes (e.g., Chen & Ng, 1989; Jin, 1990; Keatley et al., 1994; Smith et al., 2019). The finding that semantic priming was hard to come by even when a semantic task was adopted provided stronger evidence for a weak word-concept connection among L2 speakers. It helped to rule out the possibility that a lack of semantic priming effects among L2 speakers was a result of adopting a lexical task.

These findings were inconsistent with those of several previous studies that demonstrated L2–L2 semantic priming effects. We

believe methodological differences may underlie the discrepancy. The first one was the linguistic profile of the participants. In the present study, we tested advanced unbalanced bilinguals who were L1 dominant. Even though they were living in an English-speaking country at the time of testing, they were classroom L2 learners with very limited access to natural input and interaction opportunities as their English emerged. In contrast, among the studies that produce L2–L2 semantic priming, Grainger and Beauvillain (1988) tested early English–French bilinguals who had “acquired both languages in infancy” (p. 268). They were a different population with a much earlier onset age and much higher L2 proficiency. The Dutch–English bilinguals tested in de Groot and Nas (1991) were unbalanced bilinguals, but as bilingual speakers of two typologically related languages who lived in an environment with easier access to L2, they were likely to be more proficient in L2 than the participants tested in our study. There were no data for noncognates in their L1 Dutch and L2 English for comparison, but the L1–L2 difference in RT on cognates (e.g., in their Experiment 2) was smaller than the difference shown by our participants. This proficiency difference could have contributed to the difference in the results.

A second methodological difference was the percentage of related items in the stimuli. In the present study, we adopted a low percentage of related items (20%) by adding unrelated filler items in order to minimize the chance for the participants to notice the prime–target relationship, thus producing a strategic effect. The unbalanced bilinguals in Smith et al. (2019)’s study showed a reliable L2–L2 semantic priming effect, but they did not use filler items to reduce the proportion of related items. As a result, 50% of their items were related items, compared to 20% in the present study, and 16% and 33% in Grainger and Beauvillain (1988) and French and Pynte (1987), respectively. Relatedness proportion is known to affect the semantic priming effect size (de Wit & Kinoshita, 2015b), with higher proportions associated with larger priming effects. Under these circumstances, at least some priming effect may be strategic in nature.

A third possible methodological cause was the test items used. The study reported by French and Pynte (1987) produced a reliable L2–L2 semantic priming effect among unbalanced bilinguals with a low related item ratio, so these two methodological features cannot explain the discrepancy. However, they used categories and exemplars (e.g., *bird-sparrow*) as prime–target pairs. Six categories (e.g., clothing, body, vegetable) were involved, each with six exemplars. In contrast, we selected our stimuli by relying on associative strength, resulting in a mixture of different types of prime–target pairs, such as members of the same category (e.g., *table-chair*), pairs with semantic feature overlap (e.g., *forest-tree*), and associative pairs with limited semantic feature overlap (e.g., *nest-bird*). A related difference was that the test items used in French and Pynte involved narrower categories compared to the much broader distinction of natural and man-made objects adopted in the present study. Previous research has demonstrated that a priming effect was more likely to occur when the stimuli involved smaller or narrower categories. For example, in a semantic judgment task, Forster (2004) reported a significant semantic priming effect (in the form of the congruency effect) with narrower categories such as Months (e.g., related/unrelated pairs being *january/machine-AUGUST*), but no such priming effect was observed in an earlier study involving a broader category (Forster et al., 2003). Whether or how these stimulus-related differences contributed to the different findings has yet to be determined.

Another point worth considering is why a task effect was observed in translation priming but not in L2–L2 semantic priming. Another way to frame this question is why L2–L1 translation priming, but not L2–L2 semantic priming, can be observed in a semantic task, even though L2 primes were involved in both cases and even when similar participants were involved (e.g., Finkbeiner et al., 2004; Xia & Andrew, 2015). This difference can be explained in terms of the type of connections involved. Assuming that the same conceptual representations are linked to a bilingual’s two languages, translation priming involves a single shared conceptual representation, thus with a relatively short path of connections: L2 word prime to shared concept to L1 translation target. In contrast, a longer activation pathway is involved in L2–L2 semantic priming: L2 word prime to concept to related concept to L2 word target. The difference between translation priming and semantic priming can also be seen in a distributed model of semantic representation. Translation pairs are likely to share more semantic features than semantically related pairs. Consistent with this analysis, the translation priming effect was often found to be larger than the semantic priming effect in the same study (e.g., Schoobaert et al., 2009; Smith et al., 2019). Thus, the semantic task may boost the involvement of L2 word–concept links to uncover L2–L1 translation priming due to its short activation pathway, leading to the task effect. However, it may not be enough to trigger L2–L2 semantic priming because of the longer activation pathway involved. It is worth noting in this context that Guasch et al. (2011) examined interlingual translation and semantic priming among highly proficient Catalan–Spanish and Spanish–Catalan bilinguals in a lexical and a semantic task, and the priming effect was similar between the two tasks, as well.

To summarize, the lack of intralingual L2 semantic priming in both tasks corroborates existing findings in demonstrating a weaker word–concept link in L2 compared to L1 in these unbalanced L1-dominant bilinguals. These existing findings include a weaker interlingual semantic priming effect in the L2–L1 direction than in the L1–L2 direction (e.g., Jin, 1990; Schwanenflugel & Rey, 1986), a lower percentage of meaning-related responses in word association in L2 than in L1 (e.g., Jiang & Zhang, 2021), a weaker false memory effect among L2 speakers (e.g., Beato & Arndt, 2021; Howe et al., 2008), and a reduced emotional resonance in processing L2 (e.g., Jończyk et al., 2024; Toivo & Scheepers, 2019).

We recognize that some studies have shown reliable intralingual semantic priming effects in L2. Other studies have demonstrated comparable semantic effects between bilinguals’ two languages or among L1 and L2 speakers. We explained some discrepancies in terms of methodological differences. However, further research is needed to sort out the circumstances where semantic effects are comparable and different, and to explore L2–concept connections in general.

5. Conclusion

We want to conclude with a discussion of the limitations of the study and some ideas for future research. One potential limitation was related to the test items. We inadvertently include three critical items and two control items that may be considered to belong to different categories for the natural/man-made classification (i.e., juice – orange; garden – flower; net – fish; café – bird; cousin – job). This incongruity in category membership between the prime and the target may have an impact on a participant’s

response to the target. However, after deleting these five pairs, the post-hoc analysis still did not find a significant L2 semantic priming effect in both the LDT: $b = 0.002$, $SE = 0.022$, $df = 69.450$, $p = .941$ and the SCT: $b = 0.001$, $SE = 0.016$, $df = 1948$, $p = .940$, suggesting that the non-significance was not due to these items. This incongruity was also present in some filler items. They could have had some carryout effects, thus affecting the responses to some critical items, but we suspect the impact should be limited.

A second issue related to the test items was the number of items. According to Brysbaert and Stevens (2018), a minimum of 40 participants and 40 items is recommended to detect a median effect size. After data trimming, there were fewer than 36 critical items for each participant and fewer than 40 participants per task. This may compromise the effect size to some extent when using mixed-effects models. More items and participants are desirable to increase statistical power. Finally, with a focus on selecting prime-target pairs that were strong in associative relationships, we ended up including a mixture of items with different prime-target relationships, some being members of the same category and others being more associative in nature. In future research, it is desirable to include a more homogeneous set of items in this regard.

In addition, it is also worthwhile comparing participants of different L2 proficiencies, particularly very advanced L2 speakers, in this research. We classified our participants as advanced L2 speakers, but due to typological differences between their L1 and L2, they may be better classified as developing L2 speakers. Data from more advanced L2 speakers would be more informative for understanding this topic. Furthermore, given the null effect among L2 speakers, it is desirable to include a repetition priming manipulation to demonstrate that the L2 primes are processed by the L2 participants. Finally, one may also ask to what extent an associative relationship based on native speakers' performance also applies to nonnative speakers. For example, *bread-butter* may be a strong associate for English native speakers, but less so for Chinese ESL speakers because butter may not be part of their diet. It is desirable to select test items by norming them among prospective participants first, rather than based on L1 norming data.

Data availability statement. The data and material that support the findings of this study are openly available in OSF at: <https://osf.io/d2ph4>.

Competing interests. The authors declare none.

References

- Anastasi, J. S., Rhodes, M. G., Marquez, S., & Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory*, 13(8), 815–828. <https://doi.org/10.1080/09658210444000421>.
- Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition and Emotion*, 8(6), 503–514. <https://doi.org/10.1080/02699939408408956>.
- Arndt, J., & Beato, M. S. (2017). The role of language proficiency in producing false memories. *Journal of Memory and Language*, 95, 146–158. <https://doi.org/10.1016/j.jml.2017.03.004>.
- Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure in mixed-effects models: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, 35(5), 953–965. <https://doi.org/10.3758/bf03193468>.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Beato, M. S., & Arndt, J. (2021). The effect of language proficiency and associative strength on false memory. *Psychological Research*, 85(8), 3134–3151. <https://doi.org/10.1007/s00426-020-01449-3>.
- Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1). <https://doi.org/10.5334/joc.10>.
- Cabeza, R., & Lennartson, E. R. (2005). False memory across languages: Implicit associative response vs fuzzy trace views. *Memory*, 13(1), 1–5. <https://doi.org/10.1080/09658210344000161>.
- Caramazza, A., & Brunes, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 34(1), 77–81. <https://doi.org/10.1037/h0081016>.
- Chen, H. C., & Leung, Y. S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2), 316–325. <https://doi.org/10.1037/0278-7393.15.2.316>.
- Chen, H. C., & Ng, M. L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory & Cognition*, 17, 454–462. <https://doi.org/10.3758/BF03202618>.
- de Groot, A. M., Borgwaldt, S., Bos, M., & Van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, 47(1), 91–124. <https://doi.org/10.1006/jmla.2001.2840>.
- de Groot, A. M., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30(1), 90–123. [https://doi.org/10.1016/0749-596X\(91\)90012-9](https://doi.org/10.1016/0749-596X(91)90012-9).
- 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). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62805-8](https://doi.org/10.1016/S0166-4115(08)62805-8).
- de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27–51). John Benjamins Publishing Company.
- de Groot, A. M. B., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, 33(5), 600–629. <https://doi.org/10.1006/jmla.1994.1029>.
- de Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning*, 47(2), 215–264. <https://doi.org/10.1111/0023-8333.71997007>.
- de Wit, B., & Kinoshita, S. (2015a). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(4), 1062–1075. <https://doi.org/10.1037/xlm0000074>.
- de Wit, B., & Kinoshita, S. (2015b). An RT distribution analysis of relatedness proportion effects in lexical decision and semantic categorization reveals different mechanisms. *Memory & Cognition*, 43, 99–110. <https://doi.org/10.3758/s13421-014-0446-6>.
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197. <https://doi.org/10.1017/S1366728902003012>.
- 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 & Performance*, 30, 889–906. <https://doi.org/10.1037/0096-1523.30.5.889>.
- Duyck, W., & Brysbaert, M. (2008). Semantic access in number word translation: The role of crosslingual lexical similarity. *Experimental Psychology*, 55(2), 102–112. <https://doi.org/10.1027/1618-3169.55.2.102>.
- Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review*, 15(5), 961–966. <https://doi.org/10.3758/PBR.15.5.961>.
- Eilola, T. M., & Havelka, J. (2010). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism*, 15(3), 353–369. <https://doi.org/10.1177/1367006910379263>.
- Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion*, 21(5), 1064–1076. <https://doi.org/10.1080/02699930601054109>.

- Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61(2), 367–413. <https://doi.org/10.1111/j.1467-9922.2010.00613.x>.
- Elgort, I., & Piasecki, A. E. (2014). The effect of a bilingual learning mode on the establishment of lexical semantic representations in the L2. *Bilingualism: Language and Cognition*, 17(3), 572–588. <https://doi.org/10.1017/S1366728913000588>.
- Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, 51, 1–22. <https://doi.org/10.1016/j.jml.2004.01.004>.
- Forster, K. I. (2004). Category size effects revisited: Frequency and masked priming effects in semantic categorization. *Brain and Language*, 90(1–3), 276–286. [https://doi.org/10.1016/S0093-934X\(03\)00440-1](https://doi.org/10.1016/S0093-934X(03)00440-1).
- Forster, K. I., Mohan, K., & Hector, J. (2003). The mechanics of masked priming. In S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 3–37). Psychology Press.
- Frenck, C., & Pynte, J. (1987). Semantic representation and surface forms: A look at across-language priming in bilinguals. *Journal of Psycholinguistic Research*, 16(4), 383–396. <https://doi.org/10.1007/BF01069290>.
- Gollan, T., & Kroll, J. (2001). Bilingual lexical access. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 321–345). Psychology Press.
- Grainger, J., & Beauvillain, C. (1988). Associative priming in bilinguals: Some limits of interlingual facilitation effects. *Canadian Journal of Psychology*, 42(3), 261–273. <https://doi.org/10.1037/h0084193>.
- Grainger, J., & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes*, 13(6), 601–623. <https://doi.org/10.1080/016909698386393>.
- Guasch, M., Sánchez-Casas, R., Ferré, P., & García-Albea, J. E. (2011). Effects of the degree of meaning similarity on cross-language semantic priming in highly proficient bilinguals. *Journal of Cognitive Psychology*, 23(8), 942–961. <https://doi.org/10.1080/20445911.2011.589382>.
- Howe, M. L., Gagnon, N., & Thouas, L. (2008). Development of false memories in bilingual children and adults. *Journal of Memory and Language*, 58, 669–681. <https://doi.org/10.1016/j.jml.2007.09.001>.
- Iacozza, S., Costa, A., & Duñabeitia, J. A. (2017). What do your eyes reveal about your foreign language? Reading emotional sentences in a native and foreign language. *PLoS One*, 12(10), Article e0186027. <https://doi.org/10.1371/journal.pone.0186027>.
- Jiang, N. (2012). *Conducting reaction time research in second language studies*. Routledge.
- Jiang, N. (2021). The role of L1 in L2 word recognition: A case for case. *Journal of Second Language Studies*, 4, 1–18. <https://doi.org/10.1075/jsls.19039.jia>.
- Jiang, N., & Zhang, J. (2021). Form prominence in the L2 lexicon: Further evidence from word association. *Second Language Research*, 37(1), 69–90. <https://doi.org/10.1177/0267658319827053>.
- Jin, Y. S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual & Motor Skills*, 70, 1139–1154. <https://doi.org/10.2466/PMS.70.4.1139-1154>.
- Jończyk, R., Naranowicz, M., Bel-Bahar, T. S., Jankowiak, K., Korpala, P., Bromberek-Dyzman, K., & Thierry, G. (2024). Brain potentials reveal reduced sensitivity to negative content during second language production (pp. 1–12). *Bilingualism: Language and Cognition*. <https://doi.org/10.1017/S1366728924000075>.
- Keatley, C., & de Gelder, B. (1992). The bilingual primed lexical decision task: Cross-language priming disappears with speeded responses. *European Journal of Cognitive Psychology*, 4(4), 273–292. <https://doi.org/10.1080/095414492084061888>.
- Keatley, C. W., Spinks, J. A., & de Gelder, B. (1994). Asymmetrical cross-language priming effects. *Memory & Cognition*, 22, 70–84. <https://doi.org/10.3758/BF03202763>.
- Kiefer, M., Harpaintner, M., Rohr, M., & Wentura, D. (2023). Assessing subjective prime awareness on a trial-by-trial basis interferes with masked semantic priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(2), 269–283. <https://doi.org/10.1037/xlm0001228>.
- Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (pp. 389–395). John Wiley & Sons.
- 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. <https://doi.org/10.1006/jmla.1994.1008>.
- Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking Back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). Oxford University Press.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *The Journal of Speech, Language, and Hearing Research*, 50, 940c967. [https://doi.org/10.1044/1092-4388\(2007\)067](https://doi.org/10.1044/1092-4388(2007)067).
- McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press. <http://site.ebrary.com/id/10163350>.
- Miyaji-Kawasaki, Y., Inoue, T., & Yama, H. (2004). Cross-linguistic false recognition: How do Japanese-dominant bilinguals process two languages: Japanese and English? *Psychologia: An International Journal of Psychology in the Orient*, 46(4), 255–267. <https://doi.org/10.2117/psych.2003.255>.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language*, 58, 916–930. <https://doi.org/10.1016/j.jml.2008.01.003>.
- Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition*, 62(2), 223–240. [https://doi.org/10.1016/S0010-0277\(96\)00782-2](https://doi.org/10.1016/S0010-0277(96)00782-2).
- Potter, M. C., So, K. F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning & Verbal Behavior*, 23(1), 23–38. [https://doi.org/10.1016/S0022-5371\(84\)90489-4](https://doi.org/10.1016/S0022-5371(84)90489-4).
- R Core Team (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, 55A, 1339–1362.
- Sahlin, B. H., Harding, M. G., & Seamon, J. G. (2005). When do false memories cross language boundaries in English-Spanish bilinguals? *Memory & Cognition*, 33(8), 1414–1421. <https://doi.org/10.3758/BF03193374>.
- Sanchez-Casas, R., Ferré, P., García-Albea, J., & Guasch, M. (2006). The nature of semantic priming: Effects of the degree of semantic similarity between primes and targets in Spanish. *European Journal of Cognitive Psychology*, 18(2), 161–184. <https://doi.org/10.1080/09541440500183830>.
- 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(5), 569–586. <https://doi.org/10.3758/MC.37.5.569>.
- 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(5), 605–618. [https://doi.org/10.1016/0749-596X\(86\)90014-8](https://doi.org/10.1016/0749-596X(86)90014-8).
- Smith, Y., Walters, J., & Prior, A. (2019). Target accessibility contributes to asymmetric priming in translation and cross-language semantic priming in unbalanced bilinguals. *Bilingualism: Language and Cognition*, 22, 157–176. <https://doi.org/10.1017/S1366728917000645>.
- Tan, L. C., & Yap, M. J. (2016). Are individual differences in masked repetition and semantic priming reliable? *Visual Cognition*, 24(2), 182–200. <https://doi.org/10.1080/13506285.2016.1214201>.
- Taylor, R. S., & Francis, W. S. (2017). Between-language repetition priming in antonym generation: Evidence that translation-equivalent adjectives have

- shared conceptual representations across languages. *Memory*, **25**(3), 344–349. <https://doi.org/10.1080/09658211.2016.1173704>.
- Toivo, W., & Scheepers, C.** (2019). Pupillary responses to affective words in bilinguals' first versus second language. *PLoS One*, **14**(4), e0210450. <https://doi.org/10.1371/journal.pone.0210450>.
- Tzelgov, J., & Eben-Ezra, S.** (1992). Components of the between language semantic priming effect. *European Journal of Cognitive Psychology*, **4**(4), 253–272. <https://doi.org/10.1080/09541449208406187>.
- Williams, J. N.** (1994). The relationship between word meanings in the first and second language: Evidence for a common, but restricted, semantic code. *European Journal of Cognitive Psychology*, **6**(2), 195–220. <https://doi.org/10.1080/09541449408520143>.
- Winkel, H.** (2013). The emotional Stroop task and emotionality rating of negative and neutral words in late Thai–English bilinguals. *International Journal of Psychology*, **48**(6), 1090–1098. <https://doi.org/10.1080/00207594.2013.793800>.
- Xia, V., & Andrews, S.** (2015). Masked translation priming asymmetry in Chinese–English bilinguals: Making sense of the sense model. *Quarterly Journal of Experimental Psychology*, **68**(2), 294–325. <https://doi.org/10.1080/17470218.2014.944195>.
- Zeelenberg, R., & Pecher, D.** (2003). Evidence for long-term cross-language repetition priming in conceptual implicit memory tasks. *Journal of Memory and Language*, **49**, 80–94. [https://doi.org/10.1016/S0749-596X\(03\)00020-2](https://doi.org/10.1016/S0749-596X(03)00020-2).