Bilingualism: Language and Cognition 19 (5), 2016, 1001–1022 © Cambridge University Press 2015 doi:10.1017/S1366728915000462

Cross-script L2-L1 noncognate translation priming in lexical decision depends on L2 proficiency: Evidence from Japanese–English bilinguals*

MARIKO NAKAYAMA
Faculty of Arts, Letters and Sciences, Waseda University
KEISUKE IDA
Graduate School of Arts, Letters and Sciences, Waseda University
STEPHEN J. LUPKER
Department of Psychology, University of Western Ontario

(Received: October 8, 2014; final revision received: June 26, 2015; accepted: June 27, 2015; first published online 31 July 2015)

Previous research with unbalanced, different-script bilinguals has typically produced null L2-L1 noncognate masked translation priming effects in lexical decision tasks (LDT). Two novel models of the bilingual mental lexicon have emerged to account for these null results: the episodic L2 hypothesis and the Sense model. In contrast, the BIA+ model predicts significant priming whenever bilinguals are sufficiently proficient in L2. Using Japanese–English bilinguals, the role of L2 proficiency in L2-L1 noncognate translation priming in an LDT was examined. In Experiments 1 and 2, significant priming effects were observed for highly proficient bilinguals. In contrast, in Experiment 3, less-proficient bilinguals produced a null priming effect. This pattern demonstrates that L2-L1 priming effects do arise in an LDT and those effects are modulated by L2 proficiency, consistent with the BIA+ model's expectations. The pattern can be also explained by the episodic L2 hypothesis, provided that certain modifications are made to its assumptions.

Keywords: Masked priming, L2-L1 noncognate translation priming, Japanese–English bilinguals, Lexical decision task, L2 proficiency

In bilingual visual word recognition research, a critical question being asked by many researchers is how translation equivalents – words that have the same meaning in a bilingual's two languages – are represented in the bilingual lexicon. Much of the previous research on this topic has used the masked translation priming paradigm with a lexical decision task (LDT). In this paradigm, following a forward mask (e.g., ######), a prime word is briefly presented (e.g., for 40–60 ms). Immediately after the presentation of the prime, a target in the other language, either a translation equivalent of the prime or an unrelated word, is presented for a lexical decision (is it a word or not?). In this situation, the primes are highly unlikely to be identified by the participants and, therefore, any observed priming effects should reflect relatively 'automatic' lexical/semantic processing (see Kinoshita & Lupker, 2003, for a review of the masked priming technique).

The present research concerns masked priming effects for NONCOGNATE translation equivalents. Noncognates are words having the same meanings in the two languages but which are neither orthographically nor phonologically similar (e.g., *woman-mujer* or *woman-女性*). Further, as in most bilingual research, the bilinguals examined here were unbalanced bilinguals, individuals who are clearly dominant in their first language (L1). One typical result in situations of this sort is that lexical decision latencies to L2 targets (e.g., STORY) are significantly faster when they are primed by L1 noncognate translation equivalents (e.g., 物語, *story*) than by unrelated L1 words (e.g., 森林, *forest*) (e.g., Chen, Zhou, Gao & Dunlap, 2014; Gollan, Forster & Frost, 1997; Jiang, 1999; Jiang & Forster, 2001; Kim & Davis, 2003; Nakayama, Sears, Hino & Lupker, 2013; Voga & Grainger, 2007). This L1-L2 priming effect is presumed to be due to the L1 translation primes pre-activating the conceptual features they share with their L2 targets, resulting in faster and more accurate recognition of those L2 targets in comparison to when unrelated L1 words prime the targets. Significant priming effects in the L1-L2 direction have been observed not only for very proficient bilinguals (e.g., Dimitropoulou, Duñabeitia & Carreiras, 2011a; Nakayama et al., 2013) but also for less-proficient bilinguals (e.g., Chen et al., 2014; Dimitropoulou, Duñabeitia & Carreiras, 2011b; Nakayama et al., 2013), suggesting that L1-L2 conceptual links are established relatively early in the development of the bilingual lexicon. In addition, L1-L2 translation priming effects in LDTs are sensitive to bilinguals' L2

*This research was supported by a grant from the Japan Society for the Promotion of Science (JSPS) to Mariko Nakayama.

Address for correspondence:
Mariko Nakayama, Faculty of Letters, Arts and Sciences, Waseda University, 1-24-1 Toyama Shinjuku-ku Tokyo, Japan
mariko_nakayama@aoni.waseda.jp

processing fluency (Nakayama, Sears, Hino & Lupker, 2012; 2013). Priming effects are much larger for less-proficient than more-proficient bilinguals and for low-frequency than high-frequency targets, suggesting that this conceptually-based facilitation is greater when L2 target processing is more difficult.

When the direction of prime-target pairs is reversed, however, (i.e., L2-L1 noncognate translation priming) a somewhat different pattern emerges. Although a few studies have found significant L2-L1 priming (Dimitropoulou et al., 2011a; Duyck & Warlop, 2009; Jiang, 1999, Experiment 1; Schoonbaert, Duyck, Brysbaert & Hartsuiker, 2009), most studies have not (Chen et al., 2014; Dimitropoulou et al., 2011b; Finkbeiner, Forster, Nicol & Nakamura, 2004; Gollan et al., 1997; Grainger & Frenck-Mestre, 1998; Jiang, 1999, Experiments 2–5; Jiang & Forster, 2001; Nakayama et al., 2013; Sabourin, Brien & Burkholder, 2014; Wang, 2013, Experiment 1; Witzel & Forster, 2012; Xia & Andrews, 2015; Davis, Sánchez-Casas, García-Albea, Guash, Molero & Ferré, 2010), suggesting that L2 noncognate primes (e.g., story) typically do not facilitate lexical decision performance to L1 targets (e.g., 物語, *story*).

This conclusion does need to be qualified slightly, however. Simultaneous bilinguals and early bilinguals normally do show significant L2-L1 priming effects (Sabourin et al., 2014; Wang, 2013) that are often as large as L1-L2 priming effects (Basnight-Brown & Altarriba, 2007; Duñabeitia, Perea & Carreiras, 2010), suggesting the importance of native-level language proficiency and/or early acquisition of the L2 language in the emergence of an L2-L1 priming effect. It is only for unbalanced bilinguals that significant L2-L1 priming has been more the exception rather than the rule. In particular, with respect to these previous investigations of L2-L1 noncognate translation priming, one aspect of the results that stands out is that the effect is almost always null when the bilinguals are late L2 learners and the two languages involve completely different scripts (e.g., Hebrew–English, Chinese–English and Japanese–English bilinguals). Table 1 shows a selected review of previous studies examining L2-L1 priming in the LDT using this type of bilingual.

Models of the underlying mechanism of L2-L1 noncognate translation priming (or the lack thereof)

In order to explain the nature of L2-L1 noncognate translation priming for unbalanced (especially, different-script) bilinguals, two theoretical models have been proposed: the Sense model (Finkbeiner et al., 2004; Wang & Forster, 2010) and the episodic L2 hypothesis (Jiang & Forster, 2001; Witzel & Forster, 2012). The Sense model (Finkbeiner et al., 2004) proposes that L2-L1 priming only emerges in an LDT when L2 primes activate a sufficient proportion of semantic senses associated with L1 targets. For unbalanced bilinguals, L2 translation primes are inevitably semantically sparser than L1 targets. Therefore, no L2-L1 priming would be expected because L2 primes cannot activate a sufficient proportion of L1 targets' semantic senses to have a sufficient impact on processing.

Further, as Figure 1 indicates, the Sense model would predict that increased knowledge of L2 words' conceptual senses, once a bilingual becomes more proficient in their L2, will typically not lead to L2-L1 priming. The reason is that many of the senses associated with L2 words' meanings will likely be language specific. Therefore even when proficient bilinguals have learnt additional senses associated with an L2 word, those senses will not be shared with its L1 translation equivalent. This situation would be especially likely for bi-cultural bilinguals, including Japanese–English bilinguals and Chinese–English bilinguals. Significant effects in an LDT would only emerge if a bilingual were to become fully balanced *and* their translation equivalents had very similar sets of senses.

The Sense model also proposes that the proportional difference in cross-language semantic senses, that is, the reason why there is no L2-L1 priming in an LDT, should not be an issue when a semantic categorization task is used (e.g., Is it a color name?). In this situation, the conceptual senses relevant to making a "Yes" decision are those specified by the category name which typically are those senses common to the translation equivalents (black and 黒い/kuroi/). Therefore, in this task there is no proportional gap in the relevant senses between L2 and L1 words. Consistent with this reasoning, significant L2-L1 priming effects typically are observed in a semantic categorization task (e.g., Finkbeiner et al., 2004; Wang & Forster, 2010; Xia & Andrews, 2015).

The episodic L2 hypothesis (Jiang & Forster, 2001; Witzel & Forster, 2012) proposes that L2-L1 noncognate translation priming in an LDT is observed only when both translation equivalents are represented in lexical memory. According to this hypothesis, L2 words are stored in episodic memory whereas L1 words are stored in lexical memory unless bilinguals learn their two languages simultaneously from a very early age. For unbalanced bilinguals, who almost always learn their L2 later in life (e.g., 10–12 years of age), L2-L1 noncognate translation priming will not be observed in an LDT because episodically represented L2 translation primes do not facilitate lexically represented L1 targets. Thus, the episodic L2 hypothesis also predicts that there will be no L2-L1 priming for unbalanced bilinguals when the task is lexical decision.

As is true of the Sense model, the episodic L2 hypothesis does not predict that there is no task that would show L2-L1 priming. The episodic L2 hypothesis

Table 1. *Previous L2-L1 noncognate masked priming experiments testing* ***unbalanced different-script bilinguals*** *in an LDT with a prime duration between 40–67 ms. (n = number of items per cell, N = number of participants, PD = prime duration, BL = blank field, BW = backward mask, PE = priming effect in ms)*

	Language	L2 Prime		L1 Target				L2 Level			Priming Procedures				
	L2 (prime)- L1(target)	Frequency	Length (range)	Frequency	Length (range)	Item n	pps N	Proficiency		AoA	PD	BL	BW	SOA	PE
Chen et al. (2014)															
Exp.1	English-Chinese	493	4.6 (3-8)	215	2 (2-2)	14	44	Subj	2.4/6.0^{r}	11	50	-	150	200	4
Wang (2013)															
Exp.1	Chinese-English	115	1.5 (1-2)	45	4.9 (3-11)	40	20	Subj	5.9/7.0^{r}	5	50	-	-	50	0
Gollan et al. (1997)															
Exp.3	English-Hebrew	17	U.A	U.A (similar to primes)	U.A	16	40	-	*U.A*	U.A (early)	50	-	-	50	9
Exp.4	Hebrew-English	U.A (similar to targets)	U.A	17	U.A	16	30	-	*U.A (less proficient)*	U.A (early)	50	-	-	50	-4
Nakayama et al. (2013)															
Exp.2B	English-Japanese	51	4.4 (3-5)	8	2 (2-2)	30	32	Obj	TOEIC 740/990	11.4	50	-	-	50	-1
Jiang (1999)															
Exp.1	English-Chinese	191	6.5 (3-10)	157	2 (2-2)	16	52	Obj	TOEFL >550/677	U.A (-12)	50	-	-	50	13*
Exp.2	English-Chinese	39	5.7 (3-10)	34	2 (2-2)	16	44	Obj	TOEFL >550/677	U.A (-12)	50	-	-	50	3

Table 1. *(Cotinued)*

	Language	L2 Prime		L1 Target		Item	pps	L2 Level		Priming Procedures				
	L2 (prime)-L1(target)	Frequency	Length (range)	Frequency	Length (range)	n	N	Proficiency	AoA	PD	BL	BW	SOA	PE
Exp.3	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	16	Obj TOEFL >550/677	U.A (-12)	50	50	-	100	4
Exp.4	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	18	Obj TOEFL >550/677	U.A (-12)	50	50	150	250	7
Exp.5	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	18	Obj TOEFL >550/677	U.A (-12)	50	50	150	250	-2
Jiang & Forster (2001)														
Exp.1	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	24	Obj TOEFL >550/677	U.A (-12)	50	50	150	250	8
Exp.2	English-Chinese	191	6.5 (3-10)	157	2 (2-2)	16Old 16New	16	Obj TOEFL >550/677	U.A (-12)	50	50	150	250	9 3
Exp.3	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	18	Obj TOEFL >550/677	U.A (-12)	50	-	-	50	4
Exp.3	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	18	Obj TOEFL >550/677	U.A (-12)	50	50	150	250	9
Witzel & Forster (2012)														
Exp.1A	English-Chinese	178	6.7 (3-10)	171	2 (2-2)	16	32	Obj TOEFL >550/677	U.A (-12)	50	50	150	250	1

Table 1. *(Cotinued)*

	Language	L2 Prime		L1 Target		Item	pps	L2 Level			Priming Procedures				
	L2 (prime)- L1(target)	Frequency	Length (range)	Frequency	Length (range)	n	N	Proficiency		AoA	PD	BL	BW	SOA	PE
Xia & Andrews (2015)															
Exp.1B	English- Chinese	16	5.6 (3-10)	U.A	2 (2-2)	16	34	Subj Obj	4.9/7.0 IELTS > 6.0 /9.0	10	50	-	150	200	12
Exp.2B	English- Chinese	58	5.5 (3-10)	U.A	2 (2-2)	16	30	Subj Obj	5.2/7.0 IELTS > 6.0/9.0	U.A (-10)	50	-	150	200	14

Note 1. PE, $* = (p < .05)$
Note 2. Item n refers to the number of items in each cell.
Note 3. U.A = unavailable
Note 4. For the columns Prime Frequency and Target Frequency, means reported in log frequency were back converted to approximate raw frequency counts (per million) for ease of comparison across different experiments.
Note 5. For the column L2 proficiency, "Subj" refers to subjective measures and "Obj" refers to objective measures used to assess L2 proficiency. The small letters r next to subjective proficiency ratings refer to reading ability.

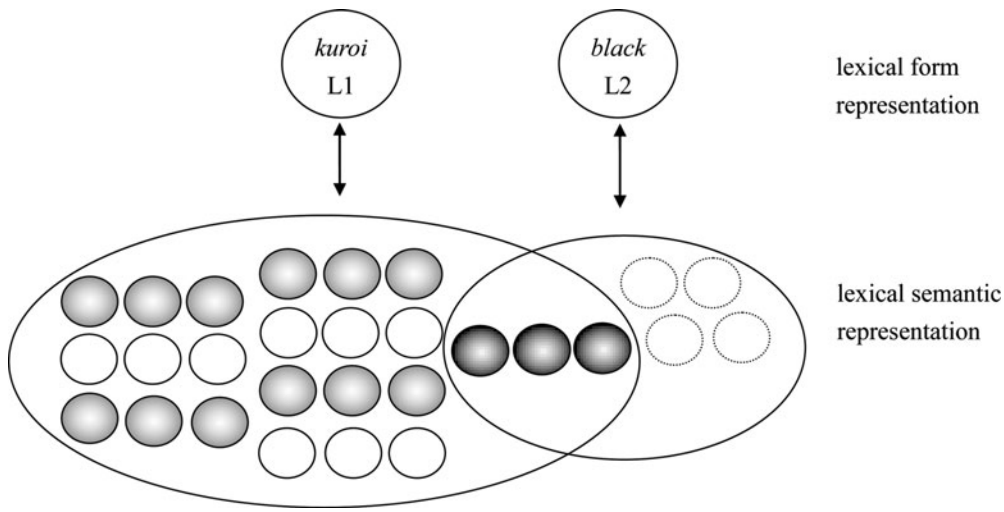

Figure 1. Schematic representation of the Japanese–English translation equivalents, 黒い (*kuroi*) and black, according to the Sense model (Finkbeiner et al., 2004). The semantic senses determining translation equivalence are depicted in a dark gray color and are shared by the two words. The L1 language specific senses are depicted in light gray or white. The Sense model assumes that the two translation equivalents are semantically associated in terms of their shared senses only. The dotted circles were added by the present authors in order to depict the L2 language specific senses of the English word “black” that have been learned by bilinguals with increased L2 proficiency. According to the Sense model, the semantic activation created by processing an L2 translation equivalent cannot activate a sufficient proportions of the semantic senses of its L1 translation equivalent (dark gray plus light gray and white circles), resulting in null L2-L1 priming in an LDT. Newly learned L2 specific senses (dotted circles, added by the authors) will not be of any help in activating L1 specific senses. Thus, increased L2 proficiency should not affect the pattern of a null L2-L1 priming effect.

proposes that significant L2-L1 priming would be observed when the task used is sensitive to episodic priming, in particular, in an episodic recognition task (ERT). In this task, participants first study (i.e., memorize) a list of L1 words, establishing episodic traces of those words. Then later in a test phase, participants are asked to determine whether a target item was in the studied list or not. Studied L1 items, that is, items that are otherwise stored in lexical memory but were recently episodically activated during the study phase, are typically recognized faster when they are primed by their L2 translation primes than by unrelated L2 words (Jiang & Forster, 2001; Witzel & Forster, 2012). The priming is presumed to be due to the fact that L2 words are also episodically represented, allowing L2 translation primes to facilitate processing of the studied L1 targets. That is, L2 primes help activate episodic records of L1 words that were stored during the study phase, leading to significant L2-L1 priming in an ERT. The secondary prediction, that words that were not studied do not show L2-L1 priming, has also been supported.

While the Sense model and the episodic L2 hypothesis seem to predict that L2-L1 noncognate priming will be always null for unbalanced bilinguals in an LDT, according to the BIA+ model (Dijkstra & van Heuven, 2002, see also the BIA-d model, Grainger, Midgley & Holcomb, 2010), whether or not L2-L1 noncognate translation priming will be observed mainly depends on the L2 proficiency of the bilinguals (e.g., Thomas & van Heuven, 2005, p. 208). The BIA+ model explains the difficulty in observing L2-L1 noncognate translation priming as being due to L2 words having low resting activation levels in general. It is difficult for masked L2 primes to facilitate L1 targets simply because those primes are unable to produce much conceptual activation. The BIA+ model assumes that L2 words will be activated more efficiently as L2 proficiency levels increase (as resting activation levels of L2 words would also increase), even for unbalanced bilinguals. Essentially, according to this model, L2-L1 priming effects will be more likely to emerge when highly proficient bilinguals are investigated. Further, these effects should also be more likely to emerge when using relatively high frequency L2 primes because the resting activation levels of such words should be higher than those of low frequency words. Therefore, the chances are higher that masked L2 primes will be activated to a level sufficient to produce conceptually-based priming.

The present research

The present experiments were an examination of L2-L1 noncognate translation priming effects with Japanese–English bilinguals using relatively high-frequency L2 primes. The experiments were conducted with two goals in mind. The first goal was to test the core expectation of the Sense model and the episodic L2 hypothesis, that

L2-L1 noncognate translation priming will not be observed in an LDT for unbalanced different-script bilinguals. Japanese–English bilinguals were deemed to be well suited to address this question because a driving force in the models' development was the null L2-L1 noncognate translation priming that has been repeatedly observed with Chinese–English and Japanese–English bilinguals.

Our second goal was simply to provide an empirical examination of the effect of L2 proficiency on L2-L1 noncognate translation priming in an LDT. Although L2 proficiency is a critical factor in the BIA+ model and proficiency has often been discussed as being an important factor in much of the previous literature (Grainger et al., 2010; Dimitropoulou et al., 2011a; 2011b; Hoshino et al., 2010; Perea, Duñabeitia & Carreiras, 2008), empirical data evaluating its impact are somewhat scarce. There appears to be only one direct investigation of this issue with unbalanced bilinguals; Dimitropoulou et al. (2011a). Those authors tested three groups of unbalanced Greek–English bilinguals who had different L2 English proficiency levels according to both subjective measures and objective measures. Atypically, there were L2-L1 priming effects for all three groups; however, L2 proficiency did not modulate the size of the priming effects (e.g., all were 11–14 ms effects), a result that was "unexpected" even to the authors themselves (p. 15). These results are, of course, not consistent either with any of the models discussed above or with much of the discussion of this issue in the literature.

In the present experiments, L2-L1 noncognate priming effects in an LDT were directly compared for highly-proficient bilinguals versus less-proficient bilinguals using the same set of stimuli, as was done in Dimitropoulou et al. (2011a). In our experiments, we used TOEIC scores as an objective measure of L2 proficiency. The objective index of L2 proficiency used by Dimitropoulou et al. was based on a categorical classification scheme (i.e., the Cambridge ESOL with FCE, CAE, and CPE categories; each category indicating respectively higher proficiency levels, with a certificate for each category being awarded for passing a specific test administered for that category). While there is no doubt that objective measures provide a much more concrete index of L2 proficiency than subjective self-ratings, one issue with respect to the Cambridge ESOL is that the test system allows proficiency overlap across its proficiency categories: for instance, a bilingual who barely passes a high-proficient category test can actually be less proficient than a bilingual who easily passes a low-proficient category test but never takes the proficiency test for the next higher category. In the TOEIC test, all test takers take a uniform test and a standardized score is calculated for each test taker (possible score range: 10–990). The TOEIC test WAS designed specifically to differentiate L2 proficiency levels of bilinguals and thus should capture the effects of L2 proficiency in a more sensitive fashion. The fact that the TOEIC test provides continuous proficiency scores is also advantageous as L2 proficiency can be treated as a continuous variable as well as a categorical (i.e., high vs. low) variable.

In sum, the purposes of the present experiments were to test whether L2-L1 priming in the LDT will be truly null for unbalanced but highly proficient different-script bilinguals and whether, in general, L2 proficiency plays a role in the pattern of the priming effects. With regard to the outcomes of the experiments, both the Sense model and episodic L2 hypothesis predict null L2-L1 priming effects even for highly proficient bilinguals. The BIA+ model, on the other hand, suggests that very proficient bilinguals would show L2-L1 priming, whereas less proficient bilinguals would show much weaker, or likely, null, priming effects.

Experiment 1

The present experiments employed standard masked priming procedures in which targets were presented immediately after the presentation of the primes. The prime duration was 60 ms, which was slightly longer than the 50 ms prime duration typically used in previous masked priming studies investigating L2-L1 priming (Dimitropoulou et al., 2011a; 2011b; Jiang, 1999, Experiments 1 and 2; Gollan et al., 1997). A 60 ms prime duration was chosen so that the English primes would be activated slightly more strongly than in previous studies, however, this prime duration is still well within the range of durations used in the masked priming literature (i.e., 40–67 ms). Experiment 1 involved high-proficient bilinguals and examined the straightforward question of whether one can observe L2-L1 noncognate translation priming with unbalanced Japanese–English bilinguals who, nonetheless, read English very well.

Method

Participants

Thirty-six proficient Japanese–English bilinguals from Waseda University (Tokyo, Japan) participated in Experiment 1. On average, they began to learn English at the age of 10.2 years, and had studied English for 11.2 years. The minimum TOEIC requirement to participate in Experiment 1 was 800 (possible test score range: 10–990). The mean TOEIC score of the bilinguals was 872 (range = 805–990). This mean score falls within the top 4 percent of the test score distribution according to the developer of the TOEIC, the Educational Testing Service (ETS).

Stimuli

The critical stimuli were 60 noncognate translation equivalents. The translation equivalents were dominant translations of each other, as verified by two proficient Japanese–English bilinguals including the first author. The Japanese targets were all two-character Kanji words (e.g., 屋根, /ya.ne/, *roof*) and had a mean written frequency of 68 per million (the NTT data base, Amano & Kondo, 2003). Each target was primed either by its English translation equivalent (e.g., roof) or by an unrelated English word (e.g., baby). English translation primes had a mean written frequency of 122 per million (Kučera & Francis, 1967) and were on average 4.7 letters long (range: 4–6) with 5.8 neighbors. Unrelated English primes were phonologically and conceptually unrelated to their targets and were matched to the English translation primes on their mean length ($M = 4.7$, range: 4–6), neighborhood size ($N = 5.8$), and word frequency ($M = 118$), all *ts* < 1. The prime-target pairs used in Experiment 1 are shown in Appendix A.[1] There were two counterbalancing lists for the word targets, so that targets primed by English translation equivalents in List A were primed by unrelated English primes in List B and vice versa. Participants were randomly assigned to one of the two presentation lists, with 18 participants in each list.

Sixty Japanese nonwords were created for the "No" trials. The nonwords were two-character Kanji strings which were created by combining two Kanji characters in such a way that that particular combination does not constitute a Japanese word (e.g., 世牧). Nonwords were primed by phonologically unrelated English words. The English primes preceding Japanese nonword targets were matched with those preceding Japanese word targets on their mean word frequency ($M = 116$), length ($M = 4.7$, range: 4–6), and neighborhood size ($N = 5.7$). Because the nonwords had no meanings, there was no counterbalancing of prime type (translation vs. unrelated). Therefore, there was only one presentation list for the nonwords.

Apparatus and procedure

Participants were tested individually. The experiment was programmed using the DMDX software package (Forster & Forster, 2003). Stimuli were presented on a 21-inch CRT display driven by a desktop computer. Each trial began with the presentation of a forward mask (######) for 500 ms followed by the 60-ms presentation of an English prime (in lower case letters). Immediately following the prime, a Japanese target was presented, which remained on the display until the participant responded. Because targets were often shorter than their primes, each target was flanked by arrows (>>>> and <<<<) so that a prime was completely backward masked by the target display. The existence of the primes was not mentioned. Participants were instructed to make a decision as to whether each target is a real Japanese word or not as quickly and accurately as possible and to respond by pressing the "Yes" or "No" button on a response box in front of them. Participants completed 16 practice trials to familiarize themselves with the task prior to the collection of data. None of the items in the practice trials was used in the experimental trials.

Results and Discussion

Correct response latencies longer than 1500 ms were replaced by this cut-off value (0.1% of the data). Response latencies and error rates were analyzed with a repeated measures ANOVA with the critical factor being Prime Type (translation vs. unrelated). List (A vs. B) was also included as a factor in the analysis in order to remove variance associated with that factor. Both subject and item analyses were conducted. The mean response latencies and error rates from subject analyses are shown in Table 2.

In the latency analysis, there was significant L2-L1 noncognate translation priming, $F_s(1, 34) = 7.33$, $p = .012$, $MSE = 263.0$, $\eta_p^2 = .17$; $F_i(1, 58) = 6.73$, $p = .014$; $MSE = 579.9$, $\eta_p^2 = .10$. L2 (English) translation primes significantly facilitated L1 (Japanese) target identification relative to unrelated L2 primes ($M = 508$ ms vs. 518 ms). The error analyses did not show significant priming ($M =$ 3.5% and 3.9%, for translation and unrelated conditions, respectively), both Fs < 1.

The results of Experiment 1 clearly indicate that an L2-L1 noncognate translation priming effect can be observed in an LDT for high-proficiency unbalanced different-script bilinguals. The significant 10 ms priming effect parallels the results of previous studies that also used the standard masked priming paradigm and relatively high-frequency stimuli, including Dimitropoulou et al.'s (2011a) studies with three-groups of Greek–English bilinguals (11–14 ms), and Jiang's (1999) Experiment 1 with Chinese–English bilinguals (13 ms).

Experiment 2

Although the size of priming effect observed in Experiment 1 was reasonably similar to the significant priming effects previously observed using similar procedures, the effect was somewhat small. Given that many prior studies did not find a significant L2-L1 noncognate translation priming effect for similar types of bilinguals (see Table 1), we thought it would be important to replicate the significant effect with a different set of stimuli and a different group of bilinguals.

The stimuli used in Experiment 2 were the same noncognate translation equivalents used by Nakayama

[1] The mean word frequencies for the related versus unrelated English primes did not significantly differ ($M = 115$ vs. 141, $t = 0.83$, $p > .40$) when a more up-to-date index of English word frequency, SUBTLEX (Brysbaert & New, 2009) was used. Such was also the case for the primes in Experiment 2 ($M = 67$ vs. 57, $t = 0.53$, $p > .52$).

Table 2. *Experiment 1: Mean lexical decision latencies (in milliseconds) and error rates for L1-Japanese targets primed by L2-English translation equivalents and by unrelated words.*

	Prime Type		
	Translation	Unrelated	Priming Effect
Example	roof-屋根	baby-屋根	
	508 (3.5%)	518 (3.9%)	+10 (+0.4%)

Note. The mean latency and error rate for Japanese nonword targets were 594 ms and 6.1%, respectively.

et al. (2013, Experiment 2B). As indicated above, less-proficient bilinguals (their mean TOEIC score was 740) did not show any sign of L2-L1 translation priming with this set of stimuli (i.e., a −1 ms effect) despite the fact that those translation pairs were selected in a way that appeared to maximize the chances of observing a significant effect (see the Method section below). By way of contrast, the Japanese–English bilinguals who showed significant priming in the present Experiment 1 were much more proficient in English (mean TOEIC score of 872) than Nakayama et al.'s bilinguals. If English proficiency matters, in line with the assumptions of the BIA+ model, Nakayama et al.'s stimulus set should produce a significant L2-L1 priming effect when proficient bilinguals are tested. In order to maximize the likelihood of producing a priming effect, the minimum TOEIC score required to take part in Experiment 2 was higher (>850) than the minimum used in Experiment 1 (>800).

Method

Participants

Thirty-four very proficient Japanese–English bilinguals from Waseda University participated in Experiment 2. None had participated in Experiment 1. This group began learning English at, on average, 9.2 years of age and had studied English for, on average, 11.8 years. Their mean TOEIC score was 917 (range = 850–990). This mean score falls within the top 2% of the test score distribution.

Stimuli

The critical stimuli consisted of 60 noncognate translation equivalent prime-target pairs. In this stimulus set, the Japanese target words were lower in frequency than their English primes which should increase the chance of observing significant L2-L1 priming. The rationale for this stimulus selection was that by presenting Japanese targets that have relatively low resting activation levels, L1 target recognition would be more likely to be facilitated by any conceptual activation available from high-frequency L2 (English) translation primes.

The Japanese targets were two-character Kanji words with a mean frequency of 8 per million (Amano & Kondo, 2003). The English translation primes (e.g., map-地図) had a mean frequency of 51 per million (Kučera & Francis, 1967) and were on average 4.4 letters long (range: 3–5) and had 6.2 neighbors. Unrelated English primes (e.g., lid-地図) had an identical average length (and range), and equivalent mean word frequencies (M = 50) and neighborhood sizes (N = 6.0) as the translation primes, all $ts < 1$. Unrelated primes were neither phonologically nor conceptually similar to their targets. There were two counterbalancing presentation lists for the word targets. Prime-target pairs used in Experiment 2 are shown in Appendix B.

The 60 Japanese nonword targets were two-character Kanji stimuli that are not words in Japanese (e.g., 日房). Nonword targets were primed by English words (e.g., goal) that were matched in length (M = 4.2, range = 3–5), word frequency (M = 50.0) and neighborhood size (N = 6.7) with the English primes preceding the word targets. For the nonword targets, there was only one presentation list.

Apparatus and procedure

The apparatus and procedure were the same as in Experiment 1 except that the prime duration was set to 50 ms to be consistent with the duration in Nakayama et al. (2013, Experiment 2B) against which the present results will be compared. This prime duration has also been used in many previous studies (e.g., Dimitropoulou et al., 2011a; 2011b; Gollan et al., 1997; Jiang, 1999, Experiments 1 and 2).

Results and Discussion

The mean response latencies and error rates from the subject analyses are shown in Table 3. As in Experiment 1, correct response latencies longer than 1500 ms were replaced by that value (0.5% of the data). The analyses were identical to those in Experiment 1.

Table 3. *Experiment 2: Mean lexical decision latencies (in milliseconds) and error rates for L1-Japanese targets primed by L2-English translation equivalents and by unrelated words for very high-proficient Japanese–English bilinguals.*

	Prime Type		
	Translation	Unrelated	Priming Effect
Example	map-地図	lid-地図	
	511 (5.7%)	533 (8.1%)	+22 (+2.4%)

Note. The mean latency and error rate for Japanese nonword targets were 583 ms and 4.3%, respectively.

Replicating the results of Experiment 1, but sharply different from the results of Nakayama et al. (2013, Experiment 2B), the L2-L1 noncognate translation priming effect was significant, $F_s(1, 32) = 11.08$, $p = .002$, $MSE = 747.1$, $\eta_p^2 = .26$; $F_i(1, 58) = 19.22$, $p < .001$, $MSE = 1038.3$, $\eta_p^2 = .25$. L1(Japanese) targets primed by L2 (English) translation primes were responded to significantly faster (511 ms) than when those same targets were primed by unrelated English primes (533 ms), a 22 ms priming effect. There also was a priming effect in the error data; Japanese targets primed by translation primes produced a lower error rate (5.7%) than targets primed by unrelated primes (8.1%), $F_s(1, 32) = 4.96$, $p = .033$, $MSE = 20.6$, $\eta_p^2 = .13$; $F_i(1, 58) = 4.88$, $p = .031$, $MSE = 36.9$, $\eta_p^2 = .08$.

As noted, the stimuli used in Experiment 2 were the same stimuli used in Nakayama et al.'s (2013) Experiment 2B. We have now demonstrated, using the exact same set of stimuli, different patterns of L1-L2 noncognate translation priming for very proficient bilinguals (our Experiment 2) versus less proficient bilinguals (Nakayama et al., 2013, Experiment 2B). It does appear, therefore, that L2 proficiency does modulate the L2-L1 noncognate translation priming effect in an LDT. In order to specifically test the effect of L2 proficiency on L2-L1 priming, we conducted a regression analysis on the data from both the present Experiment 2 and Nakayama et al.'s Experiment 2B treating TOEIC scores as a continuous variable.[2] In this analyses, the priming effect for each bilingual was the criterion variable and his/her TOEIC score was a predictor variable. The List factor was dummy-coded and entered into the model in order to remove variance associated with it. Higher TOEIC scores were significantly associated with larger L2-L1 priming effects, partial $r(63) = .37$, $t = 3.15$, $p < .01$. This result further suggests that L2 proficiency plays a significant role in producing L2-L1 noncognate translation priming effects.[3]

Experiment 3

The results of Experiments 1 and 2 showed that L2-L1 noncognate translation priming is a real phenomenon for high-proficient different-script bilinguals. In contrast, using the same stimuli as used in Experiment 2, less-proficient bilinguals had not shown significant priming previously (Nakayama et al., 2013, Experiment 2B). In addition, the regression analyses revealed that L2 proficiency was a significant predictor of the size of L2-L1 priming effects. Our results therefore strongly indicate that L2 proficiency is an important determinant of L2-L1 noncognate translation priming in an LDT. Experiment 3 was conducted to confirm this conclusion. In this experiment, the same stimuli used in Experiment 1 were presented to somewhat less-proficient Japanese–English bilinguals. The expectation was that those bilinguals would not show significant L2-L1 priming.

Method

Participants

Thirty-four Japanese–English bilinguals from Waseda University participated in Experiment 3. This group began to learn English at, on average, 11.3 years of age and had studied English for, on average, 9.5 years. Their mean TOEIC score was 710 (range = 625–795). This mean score falls within the top 25% of the test score distribution.

[2] In this analysis, an upper cut-off of 1200 ms was applied to the present data in order to be consistent with the cut-off used in Nakayama et al.'s (2013) Experiment 2B. The use of this cut-off value did not change the size of the priming effect in the present Experiment 2 (22 ms).

[3] It would be expected that prime frequency would also be related to the size of the priming effect. However, because almost all of the primes used in this stimulus set were higher frequency words, their frequency range was quite small. Therefore, it would be unlikely that one would find a significant relationship between prime frequency and the size of the priming effect when using these primes.

Table 4. *Experiment 3: Mean lexical decision latencies (in milliseconds) and error rates for L1-Japanese targets primed by L2-English translation equivalents and by unrelated words for less-proficient Japanese–English bilinguals.*

	Prime Type		
	Translation	Unrelated	Priming Effect
Example	roof-屋根	baby-屋根	
	506 (3.6%)	509 (2.8%)	+3 (-0.6%)

Note. The mean latency and error rate for Japanese nonword targets were 589 ms and 3.9%, respectively.

Stimuli

The same stimuli used in Experiment 1 were used.

Apparatus and procedure

These were identical to those used in Experiment 1.

Results and Discussion

One participant was removed and replaced by a new participant because the RT of that participant was exceptionally slow (more than 4.5 standard deviations above the overall mean). Correct response latencies longer than 1500 ms would have been replaced by that value; however, all data points fell within the acceptable range. Thus no data point was replaced. The data analyses were conducted identically to those in the previous experiments. The mean response latencies and error rates from the subject analyses are shown in Table 4.

As expected, no significant L2-L1 priming was observed for these bilinguals. L1 (Japanese) targets primed by L2 (English) noncognate translation primes were responded to 3 ms faster (506 ms) than when those same targets were primed by unrelated English primes (509 ms) with this effect being far from significant, both Fs < 1. There also was no priming effect in the error data, both $Fs < 1$. Because our *a priori* prediction for Experiment 3 was null effect, we conducted a Bayesian factor analysis testing the likelihood of the result being truly null (e.g., Masson, 2011; Wagenmakers, 2007). According to Masson (2011), P_{BIC} $(H_1|D) > .75$ is positive evidence supporting the null hypothesis, and P_{BIC} $(H_1|D) > .95$ is strong evidence supporting the null hypothesis. P_{BIC} $(H_1|D)$ for our response latency data was .99, supporting our prediction that, as a group, bilinguals at this level of proficiency do not produce L2-L1 noncognate translation priming in the LDT.

In order to directly examine the effect of L2 proficiency, we again conducted a regression analysis using the data from Experiments 1 (high-proficient bilinguals) and 3 (less-proficient bilinguals) by treating L2 proficiency as a continuous variable and, again, entering the List factor as a dummy-coded factor. The results replicated the pattern reported following Experiment 2. There was a significant positive association between the individual's priming effect and their TOEIC score, $r(67) = .29$, $t = 2.48$, $p < .05$.

The results of Experiment 3 might seem somewhat surprising. A null L2-L1 priming effect was observed in that experiment even though the bilinguals involved were actually quite proficient, achieving an average TOEIC score in the upper 25% of the bilingual distribution. Crucially, however, the contrasting results of significant L2-L1 priming for high-proficient bilinguals (Experiment 1, Experiment 2) and a null effect for less-proficient bilinguals (Experiment 3, Nakayama et al., 2013, Experiment 2B), along with the results of the two regression analyses do provide quite strong support for the idea that L2 proficiency plays a key role in determining whether there will be significant L2-L1 noncognate translation priming in an LDT and that the proficiency level has to be quite high for an effect to emerge.

General Discussion

Previous studies have reported somewhat mixed results in terms of whether L2-L1 noncognate masked translation priming in an LDT exists for unbalanced bilinguals. In particular, although some researchers have found L2-L1 priming, this effect has been particularly elusive, especially for different-script bilinguals. This typically observed null effect has led to the creation of two theoretical models: the Sense model and episodic L2 hypothesis. Both models present novel ideas about how L2 words are represented and how L2-L1 words are connected in bilingual memory: The episodic L2 hypothesis explains the null L2-L1 priming in an LDT as being due to the fact that L2 words' representations are episodic, rather than lexical, and the Sense model explains the null effect in an LDT as being due to semantically sparse L2 primes failing to sufficiently activate sufficient conceptual senses of their L1 translation equivalents. One goal of the present research was to empirically examine whether L2-L1 priming in the LDT truly is nonexistent for unbalanced different-script bilinguals as is predicted by those two models. The other goal of the present research was to examine the effect of L2 proficiency in L2-L1 noncognate translation priming empirically, given that the BIA+ model and much of the previous literature assume that factor to be critical, although it has only been directly empirically examined once (Dimitropoulou et al., 2011a).

With respect to the first question of whether L2-L1 noncogante translation priming is truly nonexistent for unbalanced different-script bilinguals, the present experiments indicate that such is not the case. Masked

L2-L1 noncognate translation priming does arise in an LDT for proficient bilinguals as shown in Experiments 1 and 2. Such results, therefore, question the theoretical premises of the Sense model and the episodic L2 hypothesis, as those models have no obvious means of explaining L2-L1 priming in an LDT with unbalanced different-script bilinguals. We will return to a discussion of these models subsequently.

With respect to the second question of whether L2 proficiency modulates the L2-L1 noncognate translation priming effect, the answer is clearly yes. Sets of stimuli that failed to produce significant priming effects for less proficient bilinguals produced significant effects for proficient bilinguals (Experiment 1 vs. 3 and Experiment 2 vs. Nakayama et al.'s, 2013, Experiment 2B). Plus, in both contrasts, higher TOEIC scores were significantly associated with larger priming effects. Thus, consistent with the BIA+ model, bilinguals' L2 proficiency is indeed a crucial factor in observing L2-L1 noncognate translation priming in an LDT.

The effects of L2 proficiency vs. age of L2 acquisition (L2 AoA) on L2-L1 priming

As noted, Dimitropoulou et al.'s (2011a) experiments were the first and only ones that directly tested the impact of L2 proficiency on L2-L1 noncognate translation priming in an LDT. Those authors found significant L2-L1 priming in all of their groups, with their priming effect sizes being virtually identical (11–14 ms effects). If higher L2 proficiency results in more efficient access to conceptual representations by L2 primes, allowing the emergence of significant L2-L1 translation priming, bilinguals with higher L2 proficiency should have produced larger priming effects than less proficient ones. Their results therefore are NOT consistent with our results or with the BIA+ model's prediction.

One possible reason why Dimitropoulou et al. (2011a) did not find an effect of L2 proficiency was that, although their three groups of bilinguals did vary to some degree in measured L2 proficiency, the absolute between-group differences in proficiency in L2 (English) were not particularly large. Several aspects of Dimitropoulou et al.'s results provide evidence for that claim. For one, when these bilinguals made lexical decisions to L2 targets, overall response latencies were not statistically different between their most and least proficient groups (671 ms vs. 699 ms). Second, more critically, the three groups of bilinguals also produced virtually identical priming effects in the L1-L2 direction (i.e., 28–31 ms effects). Recall that Nakayama et al. (2013, Experiment 1) found that L1-L2 noncognate priming was statistically larger for less- than more-proficient bilinguals and also for low- than high-frequency targets (see also Nakayama et al., 2012; 2013 for the same pattern of effects obtained with cognate translation priming pairs). According to Nakayama et al.'s results, less fluent processing of L2 words should lead to larger L1-L2 priming. Thus, the equivalent L1-L2 priming effects observed by Dimitropoulou et al. indicate that their bilingual groups were essentially equally proficient in processing L2 words, that is, those words that were presented as primes when examining priming in the L2-L1 direction. As such, it is actually not surprising that Dimitropoulou et al. found no effect of L2 proficiency on L2-L1 priming across their three participant groups.[4]

Another potential issue that is worth discussing in the context of Dimitropoulou et al.'s (2011a) results is the effect of L2 AoA, the age at which bilinguals began to acquire their second language. In Dimitropoulou et al.'s experiments, the three groups of Greek–English bilinguals were matched in their L2 AoA, with their mean first exposure to English writing being between 8.6 and 8.8 years of age. On the other hand, the L2 AoAs of our Japanese–English bilinguals were not matched. The mean L2 AoA of the very proficient bilinguals in the present Experiment 2 and the less proficient bilinguals in Nakayama et al.'s (2013) Experiment 2B were 9.2 and 11.4, respectively, with this difference being statistically significant ($p < .01$). The L2 AoAs of the bilinguals tested in Experiments 1 and 3 were numerically, but not statistically, different ($M = 10.2$ vs. 11.3, $p = .11$). In both of these contrasts, however, only the group of bilinguals who started learning English (slightly) earlier showed significant L2-L1 priming. This analysis raises the possibility that the discrepancy between the results of Dimitropoulou et al. and ours may have been due to bilinguals' L2 AoAs. Therefore, we examined the effect of L2 AoA apart from the effect of L2 proficiency by conducting hierarchical regression analyses.

We first analyzed the response latency priming effect data from the present Experiment 2 and Nakayama et al.'s (2013) Experiment 2B, since the difference between bilinguals' L2 AoAs was larger for those two groups ($M = 9.2$ vs. 11.4). As expected, L2 AoA was significantly correlated with TOEIC scores for these individuals; the earlier bilinguals started learning English, the higher their TOEIC score, $r(64) = -.41$, $p < .01$. The analyses however, revealed that L2 AoA, when controlling for TOEIC scores, did not predict priming effect sizes, partial $r(62) = -.03$, $t < 1$. In fact, L2 AoA was not significantly related to the observed priming effect sizes even when L2 AoA was entered into the regression model

[4] Our L2 proficiency manipulation in the present paper was stronger than in our previous experiments reporting a significant interaction between L2 proficiency and L1-L2 translation priming. For instance, in Nakayama et al. (2013, Experiment 1), the difference in the mean TOEIC scores between high and low proficient bilinguals was 142 points. In the present paper, the differences were 162 points (between Experiment 1 and 3) and 177 points (between Experiment 2 and Nakayama et al., 2013, Experiment 2B).

before the TOEIC scores were entered, partial $r(63) = -.18$ $t = -1.42$, $p > .15$. On the other hand, TOEIC scores, when controlling for L2 AoA, were a significant predictor of priming effect sizes, partial $r(62) = .33$, $t = 2.76$, $p < .01$. The same pattern was observed in the analysis of the data from Experiments 1 and 3. In this data set, the correlation between L2 AoA and TOEIC scores was modest but significant, $r(68) = -.24$, $p < .05$. Again, there was no relationship between L2 AoA and priming effect sizes when controlling for L2 proficiency, partial $r(66) = .03$, $t < 1$ and L2 AoA, when entered before TOEIC scores were entered into the model, had no predictive value, partial $r(67) = -.05$, $t < 1$. In contrast, the relationship between TOEIC scores and priming effect sizes when controlling for L2 AoA was statistically significant, partial $r(66) = .29$, $t = 2.44$, $p < .05$.

The results of these regression analyses provide good support for the idea that L2 proficiency is a more important factor than L2 AoA in determining whether or not there will be an L2-L1 noncognate translation priming effect in an LDT, at least for those individuals whose mean L2 AoAs are in the 9–12 years of age range. The results do not imply, of course, that L2 AoA plays no role in the emergence of an L2-L1 priming effect in an LDT. In fact, Sabourin, Leclerc, Burkholder and Brien (2014) have observed effects of L2 AoA using a much wider range of L2 AoAs, 3 years–19 years, than the range provided by the present participants (note that the L2 AoAs in Sabourin et al.'s study refer to the age of immersion into the L2 environment, rather than the age of first contact with the L2 language).

Why has no L2-L1 noncognate translation priming been observed in an LDT with proficient Chinese–English bilinguals?

As noted, the literature contains repeated observations of null L2-L1 noncognate priming effects with unbalanced, different-script bilinguals (e.g., Chinese–English and Japanese–English bilinguals), results that gave birth to both the Sense model and episodic L2 hypothesis. Those results, of course, differ noticeably from the results reported here. In discussing this apparent empirical contradiction, we will focus on previous studies that used an objective measure of L2 proficiency (e.g., TOEFL, IELTS) as those types of measures provide more reliable indices of proficiency levels than subjective measures of L2 proficiency. In all of these studies, the minimum English requirement to take part was a TOEFL score of 550 (Paper-based, possible score range: 300–677) or equivalent (IELTS 6.0 on a 9-point scale). Thus, all participants would have had decent English proficiency. Of 12 such experiments (see Table 1), only one showed L2-L1 noncognate translation priming in an LDT (Jiang, 1999, Experiment 1).[5] Certainly, the null L2-L1 noncognate translation priming reported in these experiments serves as strong empirical support for the absence of an effect even for proficient different-script bilinguals, which was one of the main motivations for the development of the two new models. Why, then did we find significant L2-L1 priming with proficient different-script (i.e., Japanese–English) bilinguals while almost all of the previous experiments did not?

Notably, all of the previous studies employing objective measures of L2 proficiency involved Chinese–English bilinguals (e.g., Jiang, 1999; Jiang & Forster, 2001; Witzel & Forster, 2012; Xia & Andrews, 2015). One obvious possibility for the difference between those results and our results is that the bilinguals' L1 languages were different (Chinese vs. Japanese). However, it seems unlikely that Chinese–Japanese language differences would provide a reasonable explanation of the different patterns. In particular, the two types of languages/bilinguals are actually quite similar in a number of ways, for example, their L1s and L2s do not have any orthographic similarity and also their L1 noncognate translation equivalents are written in logographic form. Therefore, although it may be premature to exclude this possibility entirely, there seems to be no strong reason to suspect that bilinguals' L1 language should affect the presence/absence of L2-L1 noncognate translation priming when Chinese–English and Japanese–English bilinguals are being compared.

A second possibility is that our proficient Japanese–English bilinguals were simply more proficient in their L2 than the Chinese–English bilinguals who showed no priming. The ETS, the developer of both TOEIC and TOEFL, provides an estimate of score equivalency between the two tests (Educational Testing Service, Canada, 2003). According to the equivalency table, the minimum TOEIC score requirements for the present Experiments 1 and 2 (800 and 850) correspond to TOEFL scores of 569 and 588, respectively. These scores are slightly higher than the minimum score requirement for

[5] In this section, we purposely did not discuss an LDT experiment reported by Finkbeiner et al. (2004, Experiment 2), even though that experiment tested proficient Japanese–English bilinguals. There appears to have been a critical methodological problem in that particular experiment. Specifically, all of their nonword targets were two-character Kanji items (e.g., 幹鈴). However, only 35% of their word targets were two-character Kanji items (e.g., 雑誌, *magazine*), whereas 55% of their word targets were either one character Kanji items (e.g., 黒, *black*) or three character Kanji items (e.g., 心理学, *psychology*). The remaining 10% of their word targets were Kana words (e.g., カリウム,カブトムシ,アリ, *potassium, beetle, ant,* respectively). Therefore, the number of Kanji characters or the script type of the target could easily have been used as a cue to the correct (word) decision. Participants could always respond "Yes" if a target was not a two-character Kanji item. Such an experimental set up is far from ideal for testing for the absence/presence of L2-L1 translation priming in an LDT.

the previous studies testing Chinese–English bilinguals (a TOEFL of 550 or equivalent). The score equivalency table also indicates that a TOEFL score of 549 is roughly equivalent to a TOEIC score of 750. Recall that the mean TOEIC score of the less-proficient Japanese–English bilinguals in Experiment 3 was 710 and in Nakayama et al. (2013, Experiment 2B) it was 740. Both of those groups showed no L2-L1 priming. These facts suggest that at least some of the Chinese–English bilinguals participating in the previous experiments were not as proficient as our proficient bilinguals and, possibly, were, therefore, not proficient enough to produce L2-L1 noncognate translation priming effects in an LDT. Unfortunately, however, it is impossible to do more than speculate on this issue because neither the Chinese–English bilinguals' mean TOEFL scores nor their score ranges were reported in the previous studies (only the minimum score used was reported).[6]

A third possibility is related to the characteristics of the stimuli used. Notably, many of the previous experiments with Chinese–English bilinguals used the same set of stimuli to test L2-L1 noncognate translation priming. With the exception of Xia & Andrews's (2015) experiments, the same set of the stimuli (Jiang, 1999, Experiment 3), or a subset of the stimuli taken from Jiang (Experiment 1), was used in seven of the other experiments: (Jiang, 1999, Experiments 4 and 5; Jiang & Forster, 2001, Experiments 1, 2 and two sub-experiments of Experiment 3; Witzel & Forster, 2012, Experiment 1A). One major difference between Jiang's stimuli and our stimuli is that their L2 (English) primes were significantly longer ($M = 6.5$, range: 3–10) than our English primes ($M = 4.7$, range: 4–6 and $M = 4.4$, range: 3–5 in Experiments 1 and 3), both ps < .001. In fact, the vast majority of Jiang's primes were longer than 6 letters (80%), with the half of those primes being 7–10 letters long whereas almost all of our English primes were shorter than 5 letters, (i.e., 87% and 100% in Experiments 1 and 2). Xia and Andrews's Experiments 1B and 2B involved two different sets of stimuli, but both sets included many long English primes ($M = 5.6$, range: 3–10 and $M = 5.5$, range = 3–8), with about half of their primes having 6 or more letters. As such, the ranges of English primes' letter lengths hardly overlapped between most of the stimuli used in the Chinese–English experiments and our stimuli.

Would the primes' length affect the pattern of priming effects? It would seem reasonable that L2-L1 priming would be difficult to observe when long L2 primes are used, especially for bilinguals whose first language involves a different script (e.g., Chinese–English or Japanese–English bilinguals). The rationale for this suggestion is that different-script bilinguals would benefit much less from their well-developed, more automatic L1 visual word processing routines when processing L2 words, in contrast to bilinguals whose languages share a script (e.g., French–English bilinguals, and to some degree, Greek–English bilinguals). Therefore, different-script bilinguals would be somewhat inefficient at encoding lower-level properties of L2 words (e.g., feature and letter level properties). Long L2 primes, especially when they are masked and presented very briefly, would be most likely to suffer in this situation, slowing their conceptual processing as well. Consequently, L2-L1 noncognate translation priming in an LDT, which would require conceptual level processing, would be somewhat difficult to observe when different-script bilinguals are presented with relatively long L2 primes.

Lastly, we should also note that, as Dimitropoulou et al. (2011a) pointed out, previous studies reporting null L2-L1 noncognate translation priming in an LDT, including ones involving Chinese–English bilinguals, may have had low statistical power, either due to the use of small numbers of participants or the use of small numbers of items per condition, or both (see Table 1). Given that the expected effect size of L2-L1 noncognate translation priming in an LDT is not large (ranging from 10 to 26 ms with the actual size depending on the proficiency of the bilinguals in question), it is certainly possible that some of the previous null effects were at least partly due to low statistical power (Type II errors).

Why have L2-L1 priming effects been observed in other tasks?

In the preceding section, we discussed possible factors that could have resulted in the null L2-L1 priming effects observed in previous LDT experiments with Chinese–English bilinguals. We should, however, note that, for the same groups of Chinese–English bilinguals, the very same long English primes that did not produce L2-L1 priming in an LDT, nevertheless did produce L2-L2 (i.e., repetition) priming in that same task (Jiang, 1999, Experiments 3 and 5) although parallel data patterns are not found in all situations (Xia & Andrews, 2015, Experiment 1B and 2B).[7]

[6] Japanese uses two syllabic scripts, Katakana and Hiragana, in addition to logographic Kanji whereas Chinese only uses logographs. As Ken Forster has suggested in his role as reviewer, in general, it may be easier for Japanese readers to cope with briefly presented English characters because of their familiarity with multiple scripts. Therefore, differences in TOEIC scores between Japanese and Chinese readers may actually slightly underestimate the proficiency difference between the two populations of readers in terms of their processing of English primes in a masked priming environment.

[7] In Xia and Andrews's (2015) Experiment 2B, significant L2-L2 repetition priming was observed but only when the word/nonword distinction was fairly obvious (50% of the nonwords were orthographically illegal letter strings). In their Experiment 1B,

In addition, previous experiments also reported that the same, relatively long, English primes that produced no L2-L1 priming in a LDT produced significant priming when the task was episodic recognition (Jiang & Forster, 2001, Experiments 1 and 3; Witzel & Forster, 2012, Experiment 1A) or semantic categorization (Xia & Andrews, 2015). Results such as these suggest that the lack of L2-L1 priming effects in an LDT is not simply because bilinguals were unable to process masked L2 primes effectively.

The point to be made here, however, is that significant priming from L2 primes in some experimental paradigms does not, necessarily, imply that those primes were processed to the conceptual level, which is, presumably, what is needed in order to observe significant L2-L1 priming in an LDT. For instance, significant L2-L2 repetition priming can be expected whenever a prime's lexical representations are sufficiently activated, even if its conceptual representations are not activated to any important degree (see Hoshino, Midgley, Holcomb & Grainger, 2010 for a similar argument, see Lin & Ryan, 2007, for neurological evidence supporting this view, and also see Forster, 2013 for recent evidence that repetition priming consists of two separate sources of facilitation, at the lexical form level and at the meaning level). In fact, given that L2-L2 priming involves a full repetition of the word, a priming effect can be generated from either sub-lexical phonological or orthographic similarity as well. This situation is different from L2-L1 noncognate priming where the source of priming must be predominantly conceptual. Therefore, although significant L2-L2 priming clearly shows that the bilinguals can process masked L2 primes to at least some degree, significant L2-L2 priming does not imply that bilinguals can automatically activate conceptual information from L2 primes to any noticeable degree.

What, then, can be inferred from the significant L2-L1 priming observed in episodic recognition and semantic categorization tasks? Given that the significant effects in these tasks vs. the null effects observed in the LDT are crucial to the theoretical underpinnings of the episodic L2 hypothesis and the Sense model respectively, in the following section, we will review these task-specific effects in detail.

Episodic recognition task (ERT)

As noted above, in an ERT, participants are first presented with a set of L1 words to memorize. Participants are then presented with both "old" (studied) and "new" (unstudied) L1 words that are primed by L2 translation equivalents or by unrelated L2 words, and asked to judge whether or not the L1 words had been presented in the study phase. Those previous experiments consistently showed that the "Yes" responses to "old" (studied) Chinese targets, but not to "new" (unstudied) targets, were faster when they were primed by English translation equivalents than by unrelated English words even when the same pairs did not produce priming in an LDT (Jiang & Forster, 2001; Witzel & Forster, 2012).

The significant L2-L1 priming observed for "old" items in the ERT (noted just above) indicates that those Chinese–English bilinguals were capable of processing masked L2 primes to some degree. As noted above, however, one needs to realize that significant L2-L1 priming in an ERT does not necessarily imply that the primes were processed to the conceptual level, since the ERT is not one of the paradigms that necessarily involves conceptual processing. One result supporting this claim is that the ERT also produces significant repetition priming for nonword items (e.g., Forster, 1985; Rajaram & Neely, 1992). Another important fact, more relevant to the present experiments, is that significant L2-L1 priming has been observed in an ERT in situations where conceptual activation from L2 primes would be either nonexistent or extremely weak. Specifically, in Witzel and Forster's (2012) Experiment 2, significant L2-L1 priming (for "old" items) was observed after one experimental session of a study phase in which participants were taught unfamiliar, Basque translation equivalents of L1 English words. In such a situation, it is very difficult to attribute the significant priming effect to conceptual activation of Basque primes. These results indicate that in an ERT significant priming is quite possible even if conceptual information is not activated by the primes.

The priming effect in the ERT coupled with the assumed lack of a priming effect in the LDT is, however, theoretically very important to the episodic L2 hypothesis (Jiang & Forster, 2001; Witzel & Forster, 2012) because that hypothesis holds that significant priming occurs in the ERT because bilinguals who learned their L2 late in life, even including highly-proficient ones, represent L2 words in their episodic memory. In their original proposal, Jiang and Forster (2001) suggested that significant L2-L1 priming occurs in an ERT because an episodically represented L2 word, when presented as a prime, automatically pre-activates an episodic record of its L1 translation equivalent if there was such a record (i.e., if that word had been studied). Pre-activation of such a record biases an "old" decision to the L1 target, producing a significant priming effect. The pre-activation of an episodic record of an L1 target, on the other hand, is irrelevant in an LDT and does not produce priming in that task, because a decision is made based on activity in the word's lexical representation.

In Witzel and Forster's (2012) more recent statement of the episodic L2 hypothesis the assumption is slightly different. In that account, significant L2-L1 priming occurs in an ERT because the task guides bilinguals to

involving only word-like nonwords, equivalently proficient Chinese–English bilinguals did not produce a L2-L2 repetition priming effect.

access episodic memory where L2 words are represented. As a result, conceptual information concerning L2 primes is successfully retrieved which then facilitates L1 target identification. In other words, Witzel and Forster's account posits that the priming is conceptually-based. In contrast, significant L2-L1 priming does not occur in an LDT because the task requires bilinguals to access lexical memory where no information about the L2 prime is available. (We should note that this latter account would force Witzel and Forster to claim that the above mentioned effects with Basque–English translation equivalents that they reported must have been conceptually-based.) Although the two accounts propose somewhat different mechanisms for the ERT specific effect, they, nevertheless, both assume that the task specific effects occur because L2 words are represented in episodic memory whereas L1 words are represented only in lexical memory unless they have been studied recently.

In principle, the episodic L2 hypothesis would predict significant L2-L1 priming for "old" L1 targets in any ERT. One problem for this hypothesis, however, is that that pattern is not always found. In the previous ERT experiments, significant L2-L1 priming did not occur when the task employed a standard three-field masking procedure, where a target word is presented immediately after the presentation of a prime word (Jiang & Forster, 2001; Witzel & Forster, 2012). That is, significant L2-L1 priming has, in fact, been found only when the masking procedure involved the presentation of a blank screen between prime and target words (i.e., 50 ms prime > 50 ms blank screen > 150 ms backward mask and then a target word, a SOA of 250 ms). Witzel and Forster (2012) argued that this priming procedure produces priming because it gives L2 primes sufficient processing time. However, using the same 250 ms SOA, no L2-L1 priming was observed when the blank screen was eliminated and a backward mask was instead presented for 200 ms (Finkbeiner, 2005). Further, significant L2-L1 priming was also observed when the 50 ms blank screen was kept but the backward mask was eliminated (a 100 ms SOA), (Witzel & Forster, 2012). These results indicate that it was the presence of blank screen, rather than the length of the SOA, that plays the key role in observing L2-L1 priming in an ERT.

The use of a blank screen in masked priming situations is known to produce a "ghosting effect" – a perceptual phenomenon where masked primes look salient and has been associated with conscious awareness of the primes (e.g., Finkbeiner, 2005). It is unlikely, however, that significant L2-L1 priming in an ERT occurred due to a conscious strategy induced by visible primes (e.g., overt translation of L2 primes) because the same masking procedure does not produce L2-L1 priming in an LDT (Jiang & Forster, 2001; Witzel & Forster, 2012). Therefore, the role of the blank screen would also seem to be task specific. One possibility as to why the insertion of a blank screen was needed to produce significant L2-L1 priming in an ERT could be that that specific priming procedure creates episodic traces of L2 primes. The episodic trace may somehow help to produce priming in an ERT, possibly by pre-activating pre–existing episodic connections between L2-L1 translation equivalents. At the same time, it may not be surprising that the episodically activated L2 primes do not produce priming in an LDT as the task requires a different type of activation (i.e., lexical activation). In fact, given that all of the significant L2-L1 priming observed in the previous ERT experiments employed the specific priming procedure described above, one could even argue that the significant effects emerged because episodic representations of L2 primes were being ACTIVATED, not because L2 words are episodically REPRESENTED.

Semantic categorization task (SCT)

In an SCT, participants are presented with a category name (e.g., furniture, color, etc.) and asked to judge whether a target word is a member of the category or not. Previous experiments (Finkbeiner et al., 2004; Wang & Forster, 2010; Xia & Andrews, 2015) found that "Yes" responses are significantly faster when L1 targets were primed by their L2 translation equivalents than by unrelated L2 words. Significant L2-L1 priming effects in an SCT do suggest that unbalanced different-script bilinguals can process masked L2 primes up to the conceptual level. Thus, if one assumes that the presence/absence of L2-L1 noncognate translation priming depends on how efficiently and strongly bilinguals process L2 masked primes, bilinguals who show significant L2-L1 priming effects in an SCT should also show similar effects in an LDT. Such, however, is not the case (e.g., Xia & Andrews, 2015).

As was discussed earlier, the Sense model assumes that the lack of L2-L1 priming in the LDT is due to semantically sparser L2 primes being unable to activate a high enough proportion of the semantic senses of their L1 translation equivalents. That is, although bilinguals can activate masked L2 primes up to the conceptual level, they cannot pre-activate enough semantic senses of their L1 translation equivalents to produce an impact in an LDT. Significant priming emerges in an SCT, because a category name filters out irrelevant semantic senses and restricts semantic senses to those that are core to the translation equivalence, eliminating the proportional gap (e.g., those depicted in dark gray in Figure 1).

One alternative explanation for this discrepancy, of course, is that, in general, an SCT is more sensitive at picking up conceptual facilitation than an LDT. The SCT indeed seems to do a better job at showing conceptually-based facilitation than the LDT, as the task does reliably show significant priming in situations where the LDT rarely does, namely, masked semantic/associative priming (e.g., de Wit & Kinoshita, 2014). However, some additional facts suggest that this alternative cannot be the whole story. In particular, significant L2-L1 priming effects in an SCT have been typically observed only in experimental settings where relatively small size categories are used (vegetables, body parts, colors, etc.) and trials are blocked by category (Finkbeiner et al., 2004; Wang & Forster, 2010; Xia & Andrews, 2015). While these types of results are consistent with Finkbeiner et al.'s notion of the category name restricting the number of senses relevant to the task, they are also consistent with the possibility that participants may consciously generate a set of candidates for possible target items. That is, because participants are responding to L1 targets, they may generate a set of names of category members (that would also be L1 words) with their generation producing the conceptual activation of many words in the category. As such, when L2 noncognate translation equivalents primes are presented, those L2 primes may be able to semantically facilitate L1 targets to a measurable degree because relevant conceptual information had already been partially activated. It would be somewhat difficult to distinguish between this type of account and Finkbeiner et al.'s restricted sense account. Either way, however, the fact that no L2-L1 priming was observed when the category was specified at the end of the trial (Wang & Forster, 2010) and the fact that L2 translation primes do not facilitate semantic categorization performance for non-exemplars (Wang & Forster, 2010; Xia & Andrews, 2015) do suggest that L2 primes provide little conceptual facilitation UNLESS the task provides strong contextual cues as to what the possible targets would be (e.g., a category has been activated ahead of time).

Note also that the Sense model's assumption that the task specific effects reflect an underlying proportional gap between L1 and L2 semantic senses has been recently challenged by Xia and Andrews (2015). As just discussed, the Sense model is based on the idea that when the task is changed to an SCT, the proportional gap is essentially eliminated or "the proportion of activation realized across the relevant sense is 1.0 in both L1-L2 and L2-L1 directions" (Finkbeiner et al., 2004, p. 9). Thus, priming effects will be equally large regardless of the prime-target language direction. Contrary to this prediction, Xia and Andrews found that unbalanced Chinese–English bilinguals in an SCT showed a priming effect that was significantly larger in the L1-L2 direction than in the L2-L1 direction (the priming effect was also observed for non-exemplars in the L1-L2 direction but not L2-L1 direction). These results suggest that priming in an SCT is dependent on how efficiently primes are processed rather than reflecting underlying differences in the semantic senses of L1-L2 translation equivalents.

It seems, therefore, that the previous pattern, that L2-L1 priming is observed in an SCT but not in an LDT, can be explained without adopting the original assumptions of the Sense model. That is, the discrepancy can be explained by simply assuming that those bilinguals are not able to process L2 primes effectively at the conceptual level unless the task provides a contextual cue (see Xia & Andrews, 2015, for similar argument). This proposal can be empirically examined in future studies by testing high-proficient unbalanced bilinguals. It is logical to assume that high-proficient bilinguals, who show significant L2-L1 priming in an LDT, are able to access conceptual information from L2 primes without any task cues. Such bilinguals should also show significant L2-L1 priming in an SCT. Critically, however, high-proficient bilinguals, as opposed to low-proficient bilinguals, will produce priming not only for exemplars but also for non-exemplars in an SCT. Such effects, if found, would increase the plausibility of our explanation for the task specific effects observed in the LDT and SCT.

Theoretical implications of L2-L1 noncognate translation priming in an LDT

As noted, different theoretical accounts have been proposed to explain the absence of L2-L1 noncognate translation priming in an LDT: the Sense model, the episodic L2 hypothesis and the BIA+ model. In this section, we return to a discussion of these accounts vis a vis the present results.

The present results seem to be best explained by the BIA+ model. In this model, significant facilitation is possible if bilinguals are proficient enough to process L2 primes rapidly and efficiently to the conceptual level. The model assumes that with increasing L2 proficiency (and higher word frequencies), representations for L2 words gain higher resting activation levels in the lexicon and consequently, less intense bottom-up input will be required to sufficiently activate words' representations at the conceptual level. Null L2-L1 priming is explained as being due to bilinguals' weaker access to conceptual representations of L2 primes when L2 proficiency is lower. The fact that significant L2-L1 priming was observed for high-proficient bilinguals (Experiments 1 and 2) but not for low-proficient bilinguals (Experiment 3 and Nakayama et al., 2013, Experiment 2B) is consistent with the BIA+ model's assumptions. Further, the results of our regression analyses showing that larger L2-L1 priming is associated with increasing L2 proficiency also lend support to the BIA+ model.

What should be noted, however, is that the above conclusion is based on the patterns of priming effects observed only in the LDT, and it remains to be seen if the BIA+ model will also be able to explain priming patterns when different tasks are used (e.g., an ERT and an SCT).

Although proponents of the BIA+ model have not yet addressed the nature of processing in either the ERT or the SCT, there doesn't seem to be any aspect of the model that would readily allow it to explain priming in either of these tasks in situations where there is no priming in the LDT (i.e., with less proficient bilinguals). One could argue that describing performance in the ERT is actually beyond the present scope of the BIA+ model because the model doesn't have an episodic store. Thus, it might not be too difficult to extend the model to account for priming for less proficient bilinguals in that task. Explaining priming in an SCT when there is none in an LDT, however, may be more difficult because processing in the two tasks would seem to involve quite similar sets of operations.

When considering the performance of the various models, another point that should be noted is that in the case of different-script bilinguals, increased L2 proficiency may not only result in higher resting activation levels of L2 words but may also result in more established L2 feature/letter representations and more proficiency in L2 feature/letter decoding. L2 proficiency would not seem to modulate processing for same-script bilinguals in these latter respects because their L2 feature/letter system is identical to their L1 feature/letter system, in comparison to different script bilinguals who must newly develop an L2 feature/letter system in order to be able to read L2 words. Thus, it seems natural to assume that L2 proficiency affects how different-script bilinguals process L2 words both at the lexical and sub-lexical levels. While skills at both levels would develop hand in hand during L2 learning, understanding how overall L2 proficiency is affected by lower-level (sub-lexical) processing proficiency will be required in order to gain a full picture of how different-script bilinguals learn to read in their L2.

Turning next to the Sense model (Finkbeiner et al., 2004), it predicts that L2-L1 priming will never be observed in an LDT for unbalanced bilinguals, because semantic senses associated with L2 words are inevitably much sparser than those associated with L1 words, and L2 translation primes cannot, therefore, sufficiently activate a high enough proportion of the semantic senses associated with their L1 targets to produce priming in an LDT. In addition, although increased L2 proficiency may lead to wider activation of various semantic senses associated with L2 words, doing so would not go very far in reducing the gap because many senses associated with each of the translation equivalents are language specific, and therefore are not shared by the two words, particularly when those words are learned at different times in a bilingual's life. Therefore, the Sense model cannot explain the L2-L1 noncognate translation priming observed in the present Experiments 1 and 2, especially given that the Sense model specifically maintains that Japanese and English noncognate translation equivalents do not share many senses, a claim which does appear to be true (see Nakayama et al., 2013, Experiment 1). Our results, along with the results of Xia and Andrew (2015), therefore, challenge the Sense model's core assumption that the lack of L2-L1 noncognate priming is due to L2 primes' inability to activate a sufficient PROPORTION of L1 targets' meaning senses.

Finally, the episodic L2 hypothesis (Jiang & Forster, 2001; Witzel & Forster, 2012) also faces trouble explaining the significant L2-L1 priming observed in present experiments. This model assumes that unless bilinguals learn both their L1 and L2 very early in life, L2 words will be represented in episodic, rather than lexical, memory. Therefore, L2-L1 priming should not be observed in an LDT because the task is not sensitive to processes occurring within episodic memory.

We have discussed earlier that significant L2-L1 priming observed previously in an ERT may have been due to the specific masking procedure used in those experiments, and significant L2-L1 priming in that task may not necessarily imply that L2 words are stored in episodic memory. However, it is important to note also that there is additional empirical evidence suggesting that L2 words are stored in episodic memory. Specifically, Witzel and Forster (2012, Experiment 3), using English repetition prime-target pairs (e.g., party-PARTY), rather than L2-L1 translation pairs, reported that Chinese–English bilinguals showed significant priming both for "old" (studied) and "new" (unstudied) English targets in an ERT. In contrast, native English speakers showed significant priming only for "old" (studied) English targets. Critically, these experiments used a normal three-field masked priming paradigm (i.e., no blank screen was inserted in between) and, thus, the result cannot be due to potential spurious effects associated with the ghosting effect.

Witzel and Forster (2012) interpreted those results as implying that L2 words, because they are presumed to be episodically stored (i.e., English words for Chinese–English bilinguals), produce significant priming in an ERT regardless of whether they had been studied or not. On the other hand, lexically-stored items (i.e., English words for native English speakers) produce significant priming only when the items have been encoded episodically in a study phase. These results appear to be consistent with the view that L2 words are represented episodically, at least for less proficient bilinguals (those who do not show L2-L1 priming in an LDT).

Given this evidence supporting the importance of the episodic representations of L2 words, a reasonable question is whether the present data patterns could be reconciled with the episodic L2 hypothesis framework. The answer may well be yes. This hypothesis currently proposes that only simultaneous and fully balanced bilinguals store words of their two languages in lexical memory and, therefore, significant L2-L1 noncognate translation priming in an LDT would only be observed for perfectly balanced bilinguals (e.g., Duñabeitia et al., 2010; Wang, 2013, Experiment 2). However, what the significant L2-L1 priming in an LDT, as observed in the present experiments, may indicate is that unbalanced bilinguals do, eventually, develop representations for L2 words in lexical memory, whereas the lack of such effects for less proficient bilinguals is consistent with the idea that they have not yet done so. That is, as a bilingual gains in L2 proficiency, the representations of L2 words may shift from being episodic to lexical in nature. Although unbalanced bilinguals who start learning their L2 relatively later in life would initially store L2 information in episodic memory, their representations may eventually be assimilated into lexical memory as their L2 proficiency increases. If a further assumption is added such that the representational shift would occur gradually (e.g., on a word by word basis), then the hypothesis could also explain why the size of L2-L1 priming effects become larger as L2 proficiency increases.

Testing these ideas

The present experiments showed that L2-L1 priming in the LDT can be reliably observed for unbalanced bilinguals, and the effect is modulated by L2 proficiency. What needs to be clarified in future research is whether increasing L2 proficiency leads to a quantitative (as would be assumed by BIA+) or qualitative (as would be assumed by our modified episodic L2 hypothesis) change in how L2 words are represented. Unfortunately, the results of present experiments do not provide a clear answer this question, because the two accounts would predict the same pattern of results as a function of L2 proficiency. That is, larger L2-L1 priming for more proficient bilinguals can be perfectly explained by assuming that L2 words have higher resting activation levels (conceptual information becomes more easily retrieved) or by assuming that many of L2 words have been lexicalized.

An interesting way to find out whether increased L2 proficiency will lead to a quantitative or qualitative change in the representations of L2 words would be to compare the pattern of L2 repetition priming effects for very proficient bilinguals and less proficient bilinguals in an ERT. That is, repetition priming in an episodic recognition task may provide a potential opportunity to discover which of the two memory stores L2 words are represented in. If the representational shift (i.e., from episodic to lexical) truly occurs for unbalanced bilinguals with increasing L2 proficiency, then very proficient bilinguals, who show significant L2-L1 priming in an LDT, should produce significant repetition priming only for "old" L2 items in an ERT, because for them, those L2 words would be lexically stored. In contrast, less proficient bilinguals, who do not show significant L2-L1 noncognate translation priming in an LDT, should produce significant priming both for old and new items, because, for them, all L2 words would be episodically stored.

Another way to tease apart whether the emergence of L2-L1 priming in an LDT would be better explained by the BIA+ or by the episodic L2 hypothesis would be to test the effects of prime duration in an LDT, because the two accounts would seem to predict different patterns of priming effects. If the lack of a significant priming effect is due to lower resting activation of L2 words in general, then giving more processing time to L2 primes would help bilinguals activate their representations more strongly. If so, L2-L1 priming effects should increase as the prime durations are increased (using the range of prime durations where primes are not visible and, therefore, still prohibiting participants from consciously manipulating the prime words). That is, if the absence/presence of L2-L1 priming in an LDT reflects a quantitative difference in how L2 words are represented, then less proficient bilinguals, who do not produce priming with 50–60 ms prime durations, may produce priming if slightly longer prime durations were used. On the other hand, if L2 words are stored in episodic memory, the use of longer prime durations will not give rise to L2-L1 priming for those individuals in an LDT. That is, although L2 words may be activated more strongly, such activation should take place in episodic memory only. Increased activation levels of L2 words in episodic memory is irrelevant in producing significant priming in an LDT, because it is the activation of the prime's lexical representation that facilitates target identification in that task. Thus, if the absence/presence of L2-L1 priming in an LDT reflects qualitative changes to L2 word representations, then the use of longer prime durations will not help produce significant priming for less-proficient bilinguals, as they represent L2 words only in episodic memory.

Related to this second proposal, we should finally note that Wang and Forster (in press) recently reported that proficient Chinese–English bilinguals did not show L2-L1 priming in an LDT even when an 80 ms prime duration was used. These results appear to be consistent with the episodic L2 hypothesis. Wang and Forster's experiment did not involve a manipulation of the prime durations in their LDT, however. Future research manipulating the prime exposure factor should give us a more detailed idea as to how representations of L2 words change with increasing L2 proficiency. An additional point that should be made is that it is also important that future research test bilinguals having various language combinations in order to rule out the possibility that the task specific effects discussed here reflect something unique to Chinese/Japanese–English bilinguals' visual word recognition process.

Conclusions

The present experiments demonstrated that L2-L1 noncognate translation priming in an LDT is reliably observed for unbalanced bilinguals if their L2 proficiency is sufficiently high. Our results also suggest that L2 proficiency, rather than L2 AoA, plays the main role in determining the absence/presence of L2-L1 noncognate translation priming in an LDT for unbalanced bilinguals. The generally null results reported in previous studies appear to be due to differences in the L2 proficiency of the bilinguals participating, potentially along with differences in stimulus characteristics (e.g., word frequency and length), and low statistical power. Further research would be necessary to determine whether the emergence of L2-L1 noncognate translation priming truly reflects a quantitative change in underlying resting activation levels of L2 words as proposed by the BIA+ model or a qualitative change in representations of L2 words as proposed in our modification of the episodic L2 hypothesis.

Appendix A ***Noncognate Translation Primes, Unrelated Primes, and Targets (with Phonological transcriptions based on Tamaoka & Makioka, 2004) used in Experiments 1 and 3***

Translation	Unrelated	Target
hell	beat	地獄 /zigoku/
hero	task	英雄 /eijuR/
rate	stay	割合 /wariai/
role	wish	役割 /jakuwari/
hope	mass	希望 /kibou/
type	cost	種類 /sjurui/
name	week	名前 /namae/
angel	pause	天使 /teNsi/
habit	quest	習慣 /sjuRkaN/
cause	month	原因 /geNiN/
skill	count	技術 /gizjutu/
story	space	物語 /monogatari/
value	force	価値 /kacji/
danger	minute	危険 /kikeN/
effort	manner	努力/dorjoku/
fate	mood	運命 /uNmei/
rule	pair	規則 /kisoku/
loss	vote	損失 /soNsitu/
fear	rise	恐怖 /kjouhu/
past	west	過去 /kako/
plan	love	計画 /keikaku/
trip	flow	旅行 /rjokou/
devil	shock	悪魔 /akuma/
crime	trend	犯罪 /haNzai/
proof	anger	証拠 /sjouko/
magic	glory	手品 /tezina/
peace	issue	平和 /heiwa/
sense	thing	感覚 /kaNkaku/
memory	chance	記憶 /kioku/
result	second	結果 /keQka/
roof	baby	屋根 /jane/
bomb	coat	爆弾 /bakudaN/
cell	moon	細胞 /saibou/
park	sign	公園 /koueN/
meal	wood	食事 /sjokuzi/
farm	step	農場 /nouzjou/
ring	lake	指輪 /jubiwa/
slave	smoke	奴隷 /dorei/
bank	foot	銀行 /giNkou/
train	drink	電車 /deNsja/
blood	plane	血液 /ketueki/
child	table	子供 /kodomo/
woman	river	女性 /zjosei/
family	office	家族 /kazoku/
doctor	window	医師 /isi/
milk	nose	牛乳 /gjuRnjuR/
seat	bird	座席 /zaseki/
phone	bread	電話 /deNwa/
room	case	部屋 /heja/
test	ball	試験 /sikeN/
army	date	軍隊 /guNtai/
city	door	都市 /tosi/
sugar	piano	砂糖 /satou/
metal	chair	金属 /kiNzoku/
staff	horse	職員 /sjokuiN/
earth	floor	地球 /cjikjuR/
music	court	音楽 /oNgaku/
store	judge	商店 /sjouteN/
animal	dinner	動物 /doubutu/
magnet	poison	磁石 /zisjaku/

Appendix B ***Noncognate Translation Primes, Unrelated Primes, and Targets (with Phonological transcriptions) used in Experiment 2***

Translation	Unrelated	Target
map	lid	地図 /cjizu/
fat	guy	脂肪 /sibou/
coin	acid	硬貨 /kouka/
doll	tune	人形 /niNgjou/
boss	trap	上司 /zjousi/

Translation	Unrelated	Target
aunt	drug	叔母 /oba/
tail	pond	尻尾 /siQpo/
beef	root	牛肉 /gjuRniku/
tool	skin	道具 /dougu/
ring	wage	指輪 /jubiwa/
gold	snow	黄金 /ougoN/
roof	nose	屋根 /jane/
king	gain	王様 /ousama/
gray	hill	灰色 /haiiro/
race	film	人種 /ziNsju/
army	firm	軍隊 /guNtai/
ghost	troop	幽霊 /juRrei/
lover	slope	愛人 /aiziN/
clock	pupil	時計 /tokei/
smile	phone	笑顔 /egao/
slave	widow	奴隷 /dorei/
bride	smell	花嫁 /hanajome/
fruit	skill	果物 /kudamono/
queen	stick	女王 /zjoou/
uncle	taste	叔父 /ozi/
beach	score	海岸 /kaigaN/
sleep	motor	睡眠 /suimiN/
sweet	round	甘味 /amami/
fight	speed	喧嘩 /keNka/
green	teeth	緑色 /midoriiro/
hat	sum	帽子 /bousi/
gun	arm	拳銃 /keNzjuR/
pork	zinc	豚肉 /butaniku/
wool	tear	羊毛 /joumou/
nude	tile	裸体 /ratai/
soap	slim	石鹸 /seQkeN/
bath	sand	風呂 /huro/
bomb	myth	爆弾 /bakudaN/
seed	wire	種子 /sjusi/
milk	soul	牛乳 /gjuRnjuR/
hero	soil	英雄 /eijuR/
lady	lord	淑女 /sjukuzjo/
pain	rose	苦痛 /kutuR/
poet	clay	詩人 /siziN/
farm	wish	農場 /nouzjou/
dairy	tower	日記 /niQki/
angel	waist	天使 /teNsi/
flood	trick	洪水 /kouzui/
actor	print	俳優 /haijuR/
fiber	trail	繊維 /seNi/
lunch	stuff	昼食 /cjuRsjoku/
sugar	porch	砂糖 /satou/
magic	reply	魔法 /mahou/
limit	steel	限度 /geNdo/
honey	storm	蜂蜜 /hacjimitu/
metal	scale	金属 /kiNzoku/
chair	youth	椅子 /isu/
drink	train	飲物 /nomimono/
price	truth	値段 /nedaN/
blood	plant	血液 /ketueki/

References

Amano, N., & Kondo, T. (2003). *NTT database series: Lexical characteristics of Japanese language* (version 2). Tokyo: Sanseido. [CD-ROM]

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*, 953–965.

Brysbaert, M., & New, B. (2009). Moving beyond Kućera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Chen, B., Zhou, H., Gao, Y., & Dunlap, S. (2014). Cross-language translation priming asymmetry with Chinese–English bilinguals: A test of the sense model. *Journal of psycholinguistic research, 43*, 225–240.

Davis, C., Sánchez-Casas, R., García-Albea, J. E., Guash, M., Molero, M., & Ferré, P. (2010). Masked translation priming: Varying language experience and word type with Spanish-English bilinguals. *Bilingualism: Language and Cognition, 13*, 137–155.

de Wit, B., & Kinoshita, S. (2014). The masked semantic priming effect is task dependent: Reconsidering the automatic spreading activation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1733–1744.

Dijkstra, A., & 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.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011a). Two words, one meaning: Evidence of automatic co-Activation of translation equivalents. *Frontiers in Psychology, 288*, 1–20.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011b). Masked translation priming effects with low-proficient bilinguals. *Memory & Cognition, 39*, 260–275.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology, 57*, 98–107.

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language. New evidence from Dutch-French Bilinguals. *Experimental Psychology, 56*, 173–189.

Educational Testing Service. (2003). Comparing TOEIC and TOEFL scores. http://www.etscanada.ca/teachers/compare.php (retrieved February 25, 2014).

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.

Finkbeiner, M. (2005). Task-dependent L2–L1 translation priming: An investigation of the separate memory systems account. In J. Cohen, K. T. McAlister, K. Rolstad & J. MacSwan (eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp. 741–750). Somerville, MA: Cascadilla Press.

Forster, K. I. (1985). Lexical acquisition and the modular lexicon. *Language and Cognitive Processes, 1*, 87–108.

Forster, K. I. (2013). How many words can we read at once? More intervener effects in masked priming, *Journal of Memory and Language, 69*, 563–573.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116–124.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1122–1139.

Grainger, J., & Frenck-Mestre, C. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes, 13*, 601–623.

Grainger, J., Midgley, K.J., & Holcomb, P.J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickman (eds), *Language acquisition across linguistic and cognitive systems*, pp. 267–284. Philadelphia: John Benjamins.

Hoshino, N., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2010). An ERP investigation of masked cross-script translation priming. *Brain Research, 1344*, 159–172.

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.

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language, 49*, 484–499.

Kinoshita, S., & Lupker, S. J. (2003). *Masked priming: The state of the art*. New York: Psychology Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Lin, C-Y., & Ryan, L. (2007). Repetition priming without identification of the primes: Evidence for a component process view of priming. *NeuroImage, 38*, 589–603.

Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavioral Research Methods, 43*, 679–690.

Nakayama, M., Sears, C. R., Hino, Y., & Lupker, S. J. (2012). Cross-script phonological priming for Japanese–English bilinguals: Evidence for integrated phonological representations. *Language and Cognitive Processes, 27*, 1563–1583.

Nakayama, M., Sears, C.R., Hino, Y., & Lupker, S.J. (2013). Masked translation priming with Japanese–English bilinguals: Interactions between cognate status, target frequency, and L2 proficiency. *Journal of Cognitive Psychology, 25*, 949–981.

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.

Rajaram, S., & Neely, J. H. (1992). Dissociative masked repetition priming and word frequency effects in lexical decision and episodic recognition tasks. *Journal of Memory and Language, 31*, 152–182.

Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition*, *17*, 542–555.

Sabourin, L., Leclerc, J.-C., Burkholder, M., & Brien, C. (2014, May). Bilingual lexical organization: Is there a sensitive period? Presented at the Annual Meeting of the Canadian Linguistic Association, St. Catherine's, ON

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.

Tamaoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, & Computers, 36*, 531–547.

Thomas, M. S., & van Heuven, W. J. (2005). Computational models of bilingual comprehension. In J. F. Kroll & A.M.B. de Groot (eds), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 202–225. New York: Oxford University Press.

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition, 35*, 938–952.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*, 779–804.

Wang, X. (2013). Language dominance in translation priming: Evidence from balanced and unbalanced Chinese–English bilinguals. *The Quarterly Journal of Experimental Psychology, 66*, 727–743.

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

Wang, X., & Forster, K. I. (in press). Is translation priming asymmetry due to partial awareness of the prime? *Bilingualism: Language and Cognition*.

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.

Xia, V., & Andrews, S. (2015). Masked translation priming asymmetry in Chinese–English bilinguals: Making sense of the Sense Model. *The Quarterly Journal of Experimental Psychology, 68*, 294–325.