Bilingualism: Language and Cognition

cambridge.org/bil

L2-L1 noncognate masked translation priming as a task-specific phenomenon

Mark J. McPhedran[1] and Stephen J. Lupker[2]

[1]Department of Psychology, Grenfell Campus, Memorial University of Newfoundland and [2]Department of Psychology, University of Western Ontario

Research Article

Cite this article: McPhedran MJ, Lupker SJ (2021). L2-L1 noncognate masked translation priming as a task-specific phenomenon. *Bilingualism: Language and Cognition* **24**, 507–523. https://doi.org/10.1017/S1366728920000504

Received: 18 March 2019
Revised: 7 July 2020
Accepted: 17 July 2020
First published online: 14 December 2020

Keywords:
masked translation priming; bilingualism; lexical decision; semantic categorization

Address for correspondence:
Mark McPhedran, Department of Psychology, Grenfell Campus, Memorial University of Newfoundland, 20 University Drive, Corner Brook, Newfoundland and Labrador, Canada, A2H 5G5,
E-mail: mmcphedran@grenfell.mun.ca.

Abstract

The masked translation priming effect was examined in Chinese–English bilinguals using lexical decision and semantic categorization tasks in an effort to understand why the two tasks seem to produce different patterns of results. A machine-learning approach was used to assess the participant-based factors that contribute to the sizes of translation priming effects in these tasks. As expected, the participant-based factors that predicted translation priming effects did vary across tasks. Priming effects in lexical decision were associated with higher self-rated listening, reading, and writing abilities in English. Priming effects in semantic categorization were associated with more frequent use of English in daily life, spoken English proficiency, and self-rated listening proficiency in English. These results are discussed within the framework of Multilink, the logic of which is then expanded in an attempt to account for these task differences.

Introduction

Questions of how the bilingual mental lexicon is organized are frequently addressed using data from the masked translation priming paradigm. In this paradigm, a prime is briefly presented (e.g., for 50 ms) in one language, and is sandwiched between a forward mask (e.g., #####) and a target, which is either a translation equivalent of the prime (e.g., 国王-KING), or is unrelated to the prime (咸肉-KING). The participant must then make a decision based on the target, typically either a word-nonword decision or a semantic categorization decision. If the two language representations are interconnected within lexical and semantic memory, translation equivalent primes should preactivate lexical and semantic information about the target, making decisions to the target faster than when such information is not preactivated.

One of the most well-replicated findings using this paradigm is the translation priming asymmetry in the lexical decision task (LDT). When the prime is presented in the participant's first language (L1) and the target is presented in the participant's second language (L2), significant facilitation is typically observed (e.g., de Groot & Nas, 1991). When the prime is presented in L2 and the target in L1, however, the translation priming effect is less reliable, with numerous studies finding null effects when the prime and target are noncognates (e.g., Chen, Zhou, Gao & Dunlap, 2014; Finkbeiner, Forster, Nicol & Nakamura, 2004; Gollan, Forster & Frost, 1997; Grainger & Frenck-Mestre, 1998; Jiang, 1999; Jiang & Forster, 2001; however, see Nakayama, Ida & Lupker, 2016). However, subsequent research has shown that this asymmetry is somewhat task-specific, as the semantic categorization task (SCT; e.g., Finkbeiner et al., 2004) typically shows significant L2-L1 translation priming effects (however, see Xia & Andrews, 2015, Experiment 2A, for an instance where the asymmetry arises even in an SCT). Further, the L2-L1 translation priming effect can be sensitive to the L2 proficiency of participants, in that L2-L1 translation priming effects can be produced in LDTs when the participants are highly proficient bilinguals (e.g., Nakayama et al., 2016; however, see Dimitropoulou, Duñabeitia & Carreiras, 2011).

Any model that would propose to provide an accurate description of bilingual language processing needs to account for these findings, and the debate over the theoretical mechanism responsible for producing the asymmetry remains unresolved. The model most frequently cited when discussing masked translation priming research is the bilingual interactive activation plus model (BIA+; Dijkstra & van Heuven, 2002). This model operates on the idea that language representations have different resting-level activations (RLAs). Because bilinguals are typically more proficient in their L1 than their L2, L1 representations are assumed to have higher RLAs than L2 representations. Consequently, L1 representations require less time to become activated than L2 representations, making L1 words more effective masked primes than L2 words. Critically, the BIA+ model assumes that the RLAs of L2 representations increase as a function of the frequency of L2 use, and, hence, the L2 proficiency of the bilingual. These assumptions provide a plausible explanation of why priming in the L2-L1 direction is often null, but can emerge for highly proficient bilinguals in the LDT.

Recently, Dijkstra et al. (2019) have developed Multilink, a localist-connectionist model of the bilingual language processing system, which combines characteristics of the BIA+ model with characteristics of prior models, such as the revised hierarchical model (RHM; Kroll & Stewart, 1994; Kroll & Tokowicz, 2001) and WEAVER++ (Roelofs, 2008), to provide an implemented account of the processes involved in bilingual word recognition and production. Multilink has several prominent features that allow it to simulate data from a number of bilingualism experiments. Multilink uses a layered network architecture, which contains bidirectionally connected units representing the orthography, phonology, semantics, and language membership of words. During the process of visual word recognition, a written input activates lexical-orthographic representations, which, in turn, activate semantic and phonological representations, as well as a language node, which denotes the language membership of the input. Because Multilink is an interactive model with bidirectional connections between different layers of units, Multilink assumes that the activation of units in one layer can influence and be influenced by the activation of units in the other layers. Within this architecture, words from both languages are represented in an integrated lexicon and, hence, words in both the target and non-target languages automatically become activated. As with the BIA+ model, Multilink also assumes that representations have different RLAs, which are determined by factors such as the frequency of word usage in each language. Multilink accounts for differences in RLAs for unbalanced bilinguals, for example, by assuming that the word usage frequency in the nondominant language would only be a fraction of the word usage frequency in a bilingual's dominant language. Therefore, for unbalanced Dutch–English bilinguals, an English word with an objective frequency of 100 occurrences per million words (opm) may have the same RLA as a Dutch word with a frequency of 25 opm.

Finally, consistent with the BIA+ model (Dijkstra & van Heuven, 2002), Multilink (Dijkstra et al., 2019) contains a task/decision subsystem that specifies the appropriate response, which depends on task context and the stimulus. Unlike the BIA+, however, Multilink has been designed to produce computational simulations of data, and has been used to simulate data from tasks such as word translation, lexical decision, and naming with Dutch–English bilinguals. Multilink assumes that task-specific effects arise because the representations that decision-making is based on vary between tasks. The component of the system responsible for producing such differences is the task/decision subsystem, which selects particular representations to be used for output, sets parameters, and specifies the responses depending on the nature of the task and the stimulus list. Depending on the task context, the system may base responses on language membership of the targets, or the degree of orthographic, phonological, or semantic activation. For example, in a visual LDT, a word decision is assumed to occur when lexical-orthographic representations within the model reach a critical threshold. In naming tasks, responses are assumed to occur when phonological representations of the target language surpass a critical threshold. It is plausible that this same task/decision system can also account for semantic categorization by assuming that responses in an SCT are made when the relevant components of the semantic representations of the target word surpass a critical threshold. Through these assumptions, Multilink models bilingual word recognition and production processes using a common underlying network of orthographic, phonological, semantic and language codes, while also accounting for differences that arise in language processing due to the use of different tasks.

The present research

The primary focus of the present research was investigating why the translation priming effect pattern differs as a function of the task, with an eye toward determining whether these task differences can be understood within the framework of Multilink (Dijkstra et al., 2019). Based on the assumptions of Multilink, one plausible explanation for why different tasks produce different priming patterns with the same participants would be that the underlying word processing system emphasizes different sources of information in lexical decision and semantic categorization, sources that play out differently in many individuals.

Multilink (Dijkstra et al., 2019) could account for the effects of translation primes by assuming that the presentation of the prime influences the activity of orthographic, phonological, and semantic representations for both the prime and its translation equivalent. The effectiveness of a prime, according to the Multilink account, would be determined by whether the prime can sufficiently increase the activation of these representations in its translation equivalent. If the prime can sufficiently activate the representations of the translation equivalent, the amount of time needed for those representations to surpass their critical activation threshold when the translation equivalent is presented as a target would be reduced. In this model, the influence of the prime can be different for different target representations, with the amount of priming being a function of how strongly the prime has activated the representation relevant to the task. In an LDT, for example, the lexical-orthographic representations would be used by the task/decision system to make word/nonword decisions, and the effectiveness of the prime in this task would be mainly determined by how the prime influences the accumulation of activation in this layer.

Due to the interactive nature of Multilink (Dijkstra et al., 2019), a second source of influence that the prime may have over responses stems from how the prime excites the representations of other layers of the model, which can influence the activity of the critical layer via the bidirectional connections that connect the semantic, phonological, and orthographic representations. For these reasons, Multilink can account for the differences one obtains with cognate versus noncognate translation primes. Because cognates overlap orthographically (i.e., when the languages use the same script), semantically, and phonologically, a cognate translation prime produces a significant increase in the activation of target representations, and allows activation to surpass the critical threshold faster once the target is presented. For noncognate translation primes, however, these primes only overlap semantically with the target. While this semantic overlap between the prime and target may be quite important in SCTs, such primes have a weaker effect on the activity of lexical-orthographic representations. As a result, the prime does not provide much assistance in allowing lexical-orthographic representations to surpass the critical threshold needed to make a lexical decision.

It should be noted that this explanation of cognate priming does not account for the possibility of lateral inhibition occurring during priming, as Dijkstra et al. (2019) did not run simulations using the lateral inhibition parameter settings. Other evidence suggests that cognate translation priming effects co-occur with a small orthographic inhibition effect when languages that have the same script are used (e.g., Dijkstra, Hilberink-Schulpen & van Heuven, 2010; Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010). Thus, there may be other differences between cognate and noncognate translation priming. However, this particular issue is beyond the scope of the present research.

The influence of the prime should also be affected by the proficiency of the bilingual. As discussed previously, Multilink (Dijkstra et al., 2019) models proficiency effects by assuming that proficiency in a language impacts the RLA of the language's word representations with the crucial factor being how frequently those representations are activated by the bilingual. Becoming more proficient in an L2 requires that the bilingual's L2 representations be accessed more frequently. As a result, word representations for more proficient bilinguals are assumed to have higher RLAs, and require a shorter period of time for their activity to surpass the critical threshold needed for a decision. Thus, L2 primes are more effective at influencing the activation of the target and reducing the time needed for activation to reach a critical threshold.

Because the source of information prioritized by the word processing system varies by task, factors affecting these respective domains may exert differing levels of influence on translation priming in each task. This idea leads to a number of general expectations. First, if decisions are based on the activity of lexical-orthographic representations surpassing a critical threshold, translation priming in LDTs will be more impacted by how the prime influences the activity of the orthographic representations of the target than translation priming in SCTs. This account may explain why noncognate L2-L1 translation priming is more difficult to obtain in an LDT than in an SCT (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; however, see Xia & Andrews, 2015). It may also explain why the absence of L2-L1 translation priming in LDTs is primarily seen when the bilingual's L1 and L2 have different scripts. That is, in contexts with no prime-target orthographic overlap, the utility of the prime in LDTs can be limited because it does not strongly affect the orthographic representations needed for processing the target. Essentially, in LDTs, it is only bilinguals with strong L2 orthographic knowledge (i.e., who are skilled readers and writers in the priming language) that would be expected to show priming.

In contrast, in SCTs, because categorization decisions are based on the activity of semantic representations, factors relating to semantic coding may play a larger role in translation priming. One factor that appears to capture the semantic development of a bilingual's L2 is the degree of cultural immersion, as research has shown that cultural immersion has effects on the conceptual representations of bilinguals above and beyond their L2 proficiency. For example, Malt and Sloman (2003) studied the effects of cultural immersion on L2 conceptual development by having English L2 learners provide typicality ratings for objects in English. Malt and Sloman found that participants immersed in an English-dominant cultural environment had typicality ratings consistent with those provided by native English speakers. Cultural immersion was, in fact, a better predictor of native-like ratings than formal instruction, and it may be an important factor in producing priming in the SCT. Note also that, although cultural immersion has typically been approximated by the number of years the learner has been living in an L2-dominant environment, this measure may be a poor approximation of cultural immersion, because it ignores the fact that much of that time may have been spent among others from the individual's original culture. Therefore, asking about the extent to which L2 learners use their L2 in real-world social interactions was used in this research as a stronger proxy measure.

To investigate these types of ideas, the present research examined whether L2-L1 translation priming in the two tasks might be based on the use of different types of L2-based skills (e.g., reading, writing, speaking, listening), and behaviors (e.g., use of L2 as compared to L1 by participants in social interaction and daily functioning). Such behaviors and skills may influence the utility of the L2 primes in LDTs versus SCTs. In addition, due to the possibility that the impact of the prime is partly dependent on the corresponding L1-based skills and behaviors, the influence of those factors was also investigated.

To this end, an LDT was used in Experiment 1 and an SCT was used in Experiment 2. Different stimuli were used in the two tasks, as participants took part in both experiments in the same session, with the order of tasks being counterbalanced. The basic question is, are there a set of measures that predict priming effects across individuals in the two tasks and, if so, are those measures, as hypothesized, different in the two tasks?

Experiment 1

Method

Participants

One-hundred-and-three Chinese–English bilingual undergraduate students (76 female, 27 male) at the University of Western Ontario participated in the two experiments for course credit. Participants ranged in age from 18 to 34 years old ($M = 19.29$, $SD = 1.69$). Five participants were excluded from the analyses due to not filling out their Language Experience Questionnaires (LEQs) properly (4.85% of the total data), leaving a total of 98 participants. Of these 98 participants, 97 reported speaking Mandarin as their native language. All of these participants were born in Mainland China. One participant spoke Cantonese and reported being born in Hong Kong, but indicated being able to read simplified Chinese script. The time that the participants had been living in Canada ranged from 0–21 years ($M = 2.92$, $SD = 3.69$) at the time of testing, with the majority of participants having spent less than five years in Canada. All participants had normal or corrected-to-normal vision.

Stimuli

Experiment 1 involved a set of 100 word and 100 nonword Chinese targets and 200 English word primes (e.g., bush-衬套, game-游戏), 100 of which were the translation equivalents of the Chinese words and 100 of which were selected to be used as primes for the nonwords. All word and nonword targets were composed of two simplified Chinese characters. For the nonword targets, while each character was a word on its own, the combination of the two characters was not (e.g., 石虎, or "rock-tiger"). For each participant, half the word targets were primed by an English translation prime and half were primed by an unrelated prime. All translation equivalents were validated as being accurate by a member of the lab, who was a native Chinese speaker and had lived in China until coming to graduate school in Canada. The unrelated prime-target pair for the word targets were created by re-pairing the primes and targets being used to create the unrelated condition for that participant. Hence, the unrelated primes consisted of English words which were translation equivalents of different words in the unrelated condition (e.g., game-衬套, bush-游戏). This procedure necessitated the creation of two counterbalancing lists to ensure that each target would be primed by its translation equivalent in one list, and an unrelated word in the other list. Each participant received only one list. Words and nonwords were matched on stroke count. Mean ratings for stroke count and log frequency for all targets, as derived from the Chinese Lexicon Project (Tse et al., 2017), as well as prime CELEX frequency (Baayen, Piepenbrock & Gulikers, 1995) and length can be found in Table 1. The stimuli used in Experiment 1 can be found in Appendix A.

Apparatus

Stimuli were presented on an LG Flatron W2242TQ-BF LCD monitor with a refresh rate of approximately 60 Hz. Recording of response latencies and accuracies was done using DMDX software (Forster & Forster, 2003).

Procedure

Information about participant background was obtained, and then participants completed a questionnaire to assess self-reported level of proficiency in English and the contexts in which participants used and acquired English. As noted, all participants completed both the LDT (Experiment 1), and the SCT (Experiment 2 – the details of which will be presented subsequently). Half of the participants completed the LDT first, and half completed the SCT first. Depending on experimenter native language, verbal instructions were either given in English or in Chinese. Letters of information, consent, and questionnaires were in English. The instructions for each experiment were exclusively written in simplified Chinese script.

For Experiment 1, participants were instructed to decide whether each target was a Chinese word or nonword as quickly and as accurately as possible, pressing the right shift key for a word, or the left shift key for a nonword. Participants received 6 practice trials before beginning the experiment. The experiment itself consisted of a single block of 200 trials, with each trial beginning with a forward mask (############) for 500 ms, followed by the prime for 50 ms, then a backward mask (&&&&&&&&&&&&) for 150 ms, and finally the target to which they had to respond. The purpose of using a backward mask for 150 ms was to replicate the procedure used by Finkbeiner et al. (2004). All masks and primes were presented in 14-point Courier New font, while the Chinese targets were presented in 14-point DengXian font.

Measures: Background information questionnaire

This questionnaire was used to collect basic demographic information, including age, gender, whether the participant was born in Canada or came from abroad, IELTS score, as well as the number of years that the participant had been living in Canada.

Measures: Language experience questionnaire (LEQ)

This questionnaire was largely based on the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007), which is a self-assessment measure involving several variables. The LEAP-Q measures language exposure across several domains. First, participants indicate their native country, native language, and their second language, and then indicate at what age they moved to Canada if Canada was not their native country. Afterwards, participants indicate the order in which they learned their languages, and order the languages they know from their most proficient to their least proficient. Participants are then asked about their use of English and Chinese in different environments and social contexts. Participants give estimates for the percentage of time that they used English and Chinese at home, at school, and in other social settings, and then rate their language proficiency in four domains: speaking, understanding, reading, and writing, in both English and Chinese, on a 10-point scale, ranging from 1 (very little proficiency in the language) to 10 (highly proficient in the language). The questionnaire takes approximately 5 minutes to complete, and consists of 21 questions. The self-reported proficiency measures were found to correlate reasonably well with scores from the International English Language Testing System (IELTS), (for self-rated reading, $r(82) = .42$, $p < .0001$, for writing, $r(82) = .39$, $p = .0002$, for speaking, $r(82) = .45$, $p < .0001$, and for listening proficiency in English, $r(82) = .49$, $p < .0001$), providing preliminary evidence that these predictors had good construct validity. The mean values for the LEAP-Q for the participants in Experiments 1 and 2 can be found in Table 2.

Table 1. Means and Standard Deviations for Prime CELEX, Prime Length, Target Log-Transformed Google Frequency, and Target Stroke Count for Words, Experiments 1 & 2.

	Experiment			
	LDT		SCT	
Factor	*M*	*SD*	*M*	*SD*
Prime CELEX	36.30	121.98	30.74	68.51
Prime Length	5.81	1.41	5.76	2.07
Target Google Frequency	5.84	0.55	5.45	0.41
Target Stroke Count	22.57	6.90	22.33	7.09

Note: LDT = Lexical Decision Task; SCT = Semantic Categorization Task.

Measures: Expected priming impact (ExPrime)

One important goal of the present research was to design an index, which we refer to as the Expected Priming Impact (ExPrime) that will allow us to determine what factors affect the mean priming effects in each task and, therefore, indicate which factors affect access of lexical and semantic information associated with L2 primes and L1 targets. While standardized measures of L2 proficiency such as the TOEIC have been shown to predict L2-L1 priming in lexical decision (e.g., Nakayama et al., 2016), such a measure is highly broad, which makes it impossible to determine whether L2-L1 priming is affected mainly by certain specific domains of L2 competency (e.g., reading) or by general L2 proficiency. Our approach to addressing this issue was to derive a set of weights for the individual difference factors we measured using linear modeling, and then to use those weights to compute a composite measure that can be used to predict the priming effect size. This goal can be accomplished by using a standard multiple regression procedure. However, using standard multiple regression runs into the problem of overfitting the data, and not providing a reliable predictive measure that can generalize to new data. Further, the inclusion of large numbers of factors in an analysis, as we have done here, can also increase the risk of overfitting the data. The objective here was to derive, for each task, a set of factors, which we refer to as ExPrime, that predict L2-L1 translation priming effect(s) without running into these problems.

One method for doing so is to regularize the linear models by constraining the coefficient estimates of predictors to make them as small as possible, a process that discourages the model from fitting overly complex patterns in the data. Another method for addressing the issue of overfitting is by making sure that we have extracted only the most relevant factors for predicting L2-L1 priming. To accomplish both of these goals in the

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

Table 2. Mean Language Experience Questionnaire Responses and IELTS scores for the Participants in Experiments 1 & 2.

Factor	*M*	*SD*
PEH	9.10	13.69 (60)
PES	65.53	25.41 (90)
PEO	36.46	29.75 (100)
ER	7.17	2.14 (7)
EW	6.34	2.10 (7)
EL	7.30	2.28 (7)
ES	6.73	2.27 (7)
CR	9.25	1.71 (3)
CW	8.63	1.92 (5)
CL	9.46	1.51 (4)
CS	9.38	1.53 (3)
IELTS	6.02	2.14 (5)

Note: PEH = Percentage of time English is used at home; PES = Percentage of time English is used at school; PEO = Percentage of time English is used in other social settings; ER = Self-rated English reading proficiency; EW = Self-rated English writing proficiency; EL = Self-rated English listening proficiency; ES = Self-rated English speaking proficiency; CR = Self-rated Chinese reading proficiency; CW = Self-rated Chinese writing proficiency; CL = Self-rated Chinese listening proficiency; CS = Self-rated Chinese speaking proficiency; IELTS = International English Language Testing System.

computation of ExPrime, a series of three machine-learning models were used: a ridge regression model (Hoerl & Kennard, 1970; Tikhonov, 1963), a lasso regression model (Tibshirani, 1996), and an elastic net regression model (Zou & Hastie, 2005). Each of these models is a regularized version of the standard linear (multiple) regression model while offering the advantage of constraining the model weights to reduce overfitting, and being robust when dealing with the problem of multicollinearity (e.g., Duzan & Shariff, 2015; Muhammad, Maria & Muhammad, 2013; Oyeyemi, Ogunjobi & Folorunsho, 2015). The fitting was done for the priming data for each task separately.

The computation of ExPrime involved using responses on the LEAP-Q as predictors, and the mean priming effect for each participant as the dependent variable. All fitting was done using only the priming effects from positive trials (i.e., words in Experiment 1, exemplars in Experiment 2), although an examination of the relationship between ExPrime and priming effects for the nonexemplars will be presented following Experiment 2. A full description of the process of fitting these models is described in Appendix C.

Results

Data trimming

Data trimming was done in three steps for both the LDT and SCT. First, any items or participants with an accuracy below 50% were excluded from analyses. One item (0.50% of the total usable data), and three participants (3.05% of the total usable data) were excluded for this reason. Next, overall performance for participants and items was screened for multivariate outliers in speed-accuracy space using a Mahalanobis distance statistic and a p-value cut-off of .001 (Mahalanobis, 1936). Doing so eliminated three participants (3.05% of the usable data), and three items (1.41% of the usable data). This method was used to minimize the risk of the results being driven by specific items or participants. Finally, after this screening, trials with latencies that were faster than 200 ms and slower than 2000 ms, or deviated by more than three standard deviations from the participant's mean in that condition were removed (2.67% of the total data), and errors were removed (5.79% of the total data), leaving approximately 84% of the total usable data.

ExPrime coefficients

The coefficients for Experiment 1 ExPrime ($M = 189.78$, $SD = 27.84$) scores can be found in Table 3. The largest positive ExPrime predictors were self-rated listening and reading abilities in English, and self-rated speaking and listening proficiency in Chinese. Negatively associated predictors included self-reported reading and writing abilities in Chinese.

Prime x ExPrime

The raw response times and errors were submitted to a generalized linear mixed effects model using the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in R (R Core Team, 2017), with subjects ($SD = 47.68$) and items ($SD = 23.37$) treated as random effects (Baayen, Davidson & Bates, 2008), and prime type, ExPrime score, prime and target frequency, and previous trial RT were treated as fixed factors. Additional analyses were subsequently conducted to verify whether the individual components of ExPrime interacted with prime relatedness, and assess the effects of translation uniqueness on translation priming effects using data from the Wen and van Heuven (2017a) English-Chinese translation norms.[1] These supplementary analyses are found in Appendix D. The RT data were analyzed using an Inverse Gaussian distribution, while error data were analyzed using a binomial distribution. Due to issues with model convergence, no random slopes were included in these analyses. A bound optimization by quadratic approximation (BOBYQA; Powell, 2009) optimizer was used to ensure model convergence. In selecting models, the Bayes information criterion (BIC) and the Akaike information criterion (AIC; Akaike, 1973) were used to select the model most consistent with the data, and, if possible, minimized information loss. Finally, p-values for effects were computed using a Kenward-Roger approximation of degrees of freedom (Kenward & Roger, 1997; Luke, 2017) using the car package in R.

For RTs, the BIC and AIC both favored the following model: RT ~ Prime × ExPrime + Previous RT + Prime Frequency + Target Frequency + (1 ‖ Participant) + (1 ‖ Item). This model (AIC = 110336, BIC = 110414) was favored over the model which excluded all interaction terms (AIC = 110345, BIC = 110423), and the model which included all interaction terms (AIC = 110337, BIC = 110443). In these analyses, the main effects of ExPrime, $\beta = 13.62$, $SE = 2.78$, $t(8583) = 4.90$, $p < .0001$, target frequency, $\beta = -97.18$, $SE = 3.20$, $t(8583) = -30.39$, $p < .0001$, and previous trial RT were significant, $\beta = 23.89$, $SE = 1.89$, $t(8583) = 12.63$, $p < .0001$. Overall, response latencies were shorter for participants in Tertile 3 of the ExPrime scores ($M = 637$ ms) than they were in Tertiles 2 ($M = 674$ ms) or Tertile 1 ($M = 658$ ms), were significantly shorter for high-frequency targets ($M = 638$ ms) than they were for low-frequency targets ($M = 674$ ms), and were significantly shorter when the previous trial had a shorter latency ($M = 611$ ms) than when the previous trial had a longer latency ($M = 701$ ms). Most importantly, the two-way

[1]We would like to thank an anonymous reviewer of the paper for raising the issue of translation uniqueness effects. These analyses were performed as a result of that reviewer's comments.

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

Table 3. ExPrime Coefficients for Experiment 1.

Predictive Factor	ExPrime Coefficient Values
Chinese Spoken Proficiency	15.87
English Listening Proficiency	10.76
English Reading Proficiency	2.43
Chinese Listening Proficiency	2.17
English Writing Proficiency	0.46
Chinese Reading Proficiency	−1.69
Chinese Writing Proficiency	−7.89

interaction between prime type and ExPrime was significant, $\beta = -10.35$, $SE = 2.83$, $t(8583) = -3.66$, $p = .0003$.[2]

The nature of the two-way interaction is shown in Figure 1. Participants who reported higher listening and writing proficiency in English, as well as higher speaking and listening proficiency, but lower reading and writing proficiency in Chinese, produced larger priming effects. When dividing participants into tertiles by ExPrime score, the priming effect was larger in the top tertile (21 ms) than it was in Tertile 2 (-4 ms) or Tertile 3 (0 ms). A follow-up analysis was conducted on the ExPrime tertiles to evaluate the priming effects in the three tertiles. These analyses were computed using the phia package in R (De Rosario-Martinez, 2015), which produces a χ^2 value. The 21 ms priming effect in Tertile 1 was significant, $\chi^2(1) = 9.18$, $p = .0024$, while the 4 ms inhibition effect was nonsignificant in Tertile 2, $\chi^2(1) = 1.21$, $p = .27$, and the 0 ms priming effect in Tertile 3 was nonsignificant, $\chi^2 < 1$, $p > .70$. Further, an analysis of interaction contrasts showed that the two-way interaction was significant when comparing Tertiles 1 and 2, $\chi^2(1) = 9.69$, $p = .0019$, and Tertiles 1 and 3, $\chi^2(1) = 6.42$, $p = .011$, but the two-way interaction between Tertiles 2 and 3 was nonsignificant, $\chi^2 < 1$, $p > .45$. In the error analysis, none of the main effects or interactions were significant, all zs < 1.

Discussion

The purpose of Experiment 1 was to assess factors that contribute to L2-L1 masked translation priming in an LDT. Results from Experiment 1 have shown that facilitative L2-L1 translation priming effects were larger in the L2-L1 direction for individuals who reported weaker reading and writing abilities in their L1. Such a finding is similar to results reported in studies examining priming in the L1-L2 direction by Nakayama, Sears, Hino and Lupker (2013), who found that L1-L2 priming effects were larger when participants had lower as compared to higher proficiency in their L2. Additionally, larger translation priming effects were associated with higher self-rated L2 listening, reading, and writing abilities.

The pattern of findings observed in Experiment 1 fit reasonably well within the framework of Multilink (Dijkstra et al., 2019). Multilink assumes that decisions are made when the activation of representations surpasses a critical threshold, and also assumes that different tasks emphasize different sources of information in driving decisions. In an LDT, Multilink assumes that 'word' decisions are made when the activity of lexical-orthographic representations surpass a threshold. Factors that influence the activity of lexical-orthographic representations, such as the participant's orthographic knowledge in the prime and target languages may thus have an important role in determining whether translation priming is found in an LDT. For participants who are less skilled or experienced with orthographic processing in the target language, tasks that emphasize accessing orthographic lexical representations to make decisions would be more burdensome for them. From the perspective of Multilink, the resting-level activation of the target would generally be lower for such participants, and there should be greater opportunity for the prime to influence the processing of the target, resulting in a priming effect.

To produce L2-L1 priming, it is also critical that participants be reasonably familiarized and skilled with their L2 orthographic system, as the knowledge and familiarity of word forms in L2, reflected in measures of the participants' reading and writing abilities, would allow those primes to be effective. That is, without this necessary knowledge of L2 word forms, the prime's ability to activate lexical-orthographic representations enough that the activity will surpass its critical threshold during prime processing is limited, reducing the effectiveness of the prime.

Before evaluating the predictors for the SCT, there is one issue that should be commented on. Specifically, the RTs in Experiment 1 seemed to be relatively long for an LDT experiment. There would appear to be a couple of possible reasons for the longer latencies. The most obvious, and one that is consistent with the prior discussion, is simply that these bilinguals were not particularly proficient in their L1. Although this may have been the case, the majority of participants had been living in Canada (after moving from China) for less than three years, and it would seem unlikely that this sample of participants would have differed substantially in their L1 proficiency from participants in previous studies. A more feasible possibility is that there was a speed-accuracy trade-off in Experiment 1. Error rates to word targets in Experiment 1 were generally low (2.22%), whereas other LDT studies have reported error rates for words between 5–6% (e.g., Finkbeiner et al., 2004, Experiments 4a and 4b). It is possible, then, that at least part of the reason for the longer latencies here is that our participants were being a bit more careful and, hence, sacrificing a bit of speed in order to perform the task more accurately. However, there are several reasons to believe that the impact of any speed-accuracy trade-off on our analysis of the relationship between ExPrime and Priming would be minimal. First, while the error rates for word data were a bit lower than in some experiments reported in the literature, these error rates are within the range that has been reported in prior research (e.g., Finkbeiner et al., 2004, Experiment 2; Grainger & Frenck-Mestre, 1998). Hence, it does not appear that any speed/accuracy trade-off was outside of the normal range of what is found in these types of experiments. Second, the data cleaning procedure used in Experiment 1 screened participants and items for speed and accuracy on a multivariate basis, and any such participants and items that were extreme speed-accuracy outliers would have been removed prior to any analyses.

Experiment 2

As was noted previously, priming effects in the SCT tend to be more reliable than those in the LDT. Further, as hypothesized, the SCT may be affected by other factors than the LDT:

[2]When comparing participants' ExPrime scores to the priming effect that participants produced, the Pearson's correlation between ExPrime and the participant's priming effect revealed a significant positive relationship, $r(90) = .29$, $p = .006$, indicating that, as the main analysis indicates, larger ExPrime scores tended to be associated with larger priming effect sizes.

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

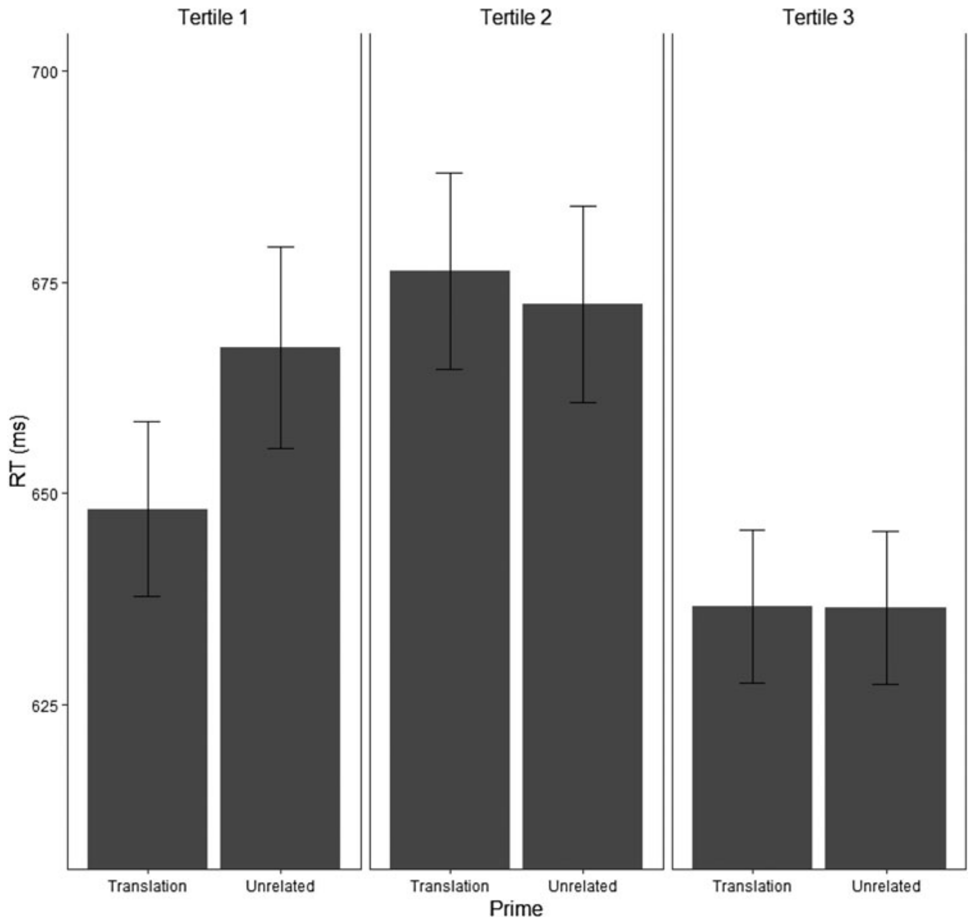

Fig. 1. Response times as a function of prime and ExPrime tertile, Experiment 1. Error bars represent 95% confidence intervals.

potentially, factors more associated with semantic processing in L2, such as the amount of time that participants use their L2 in their daily lives, across different social environments. That is, it is possible that participants who use their L2 at home, at school, and in other social contexts more frequently would have more opportunities to acquire a richer base of semantic knowledge associated with their L2. The SCT, while still requiring L2 orthographic knowledge, should not place as much emphasis on that knowledge as it does on the development of L2 semantic knowledge, as obtained through the use of the language in naturalistic social settings. Further, in the SCT, priming might be predicted to be less dependent on the strong activation of L2 word representations, with the more important issue being how rapidly the prime activates information about the category membership of the target. These ideas were examined in Experiment 2.

Method

Participants

Participants were the 103 participants who had also participated in Experiment 1.

Stimuli

Experiment 2 consisted of 200 trials across five blocks of 40 trials, with 20 exemplars and 20 nonexemplars of a selected category in each block. Five categories were used for the exemplars and nonexemplars: mammals, insects, body parts, vegetables/fruits, and clothing/accessories. Consistent with prior studies (e.g., Finkbeiner et al., 2004), each word appeared twice in the experiment, appearing as an exemplar in one block, and as a nonexemplar in another block. Nonexemplars in each block were taken from the four other categories, with five nonexemplars being taken from each category. Half of the exemplars and nonexemplars in each block were preceded by a translation prime, while the other half was preceded by an unrelated prime. For both exemplars and nonexemplars, unrelated primes were from a different semantic category than the target, however, not from the exemplar category for the trial block. The order of blocks was randomized across participants in a way that allowed each category block to occur approximately equally often in the first, second, third, fourth and fifth positions. None of the stimuli in Experiment 1 were used in Experiment 2. The stimuli used in Experiment 2 are shown in Appendix B.

Measures

The measures were identical to the measures used in Experiment 1. The only difference between the experiments was that the ExPrime measure was now based on the priming data from Experiment 2.

Procedure

Experiment 2 was completed in the same session as Experiment 1. Participants were instructed to indicate whether each target was a member of a target category specified at the beginning of the block or not, as quickly and as accurately as possible, by pressing either the right shift key for exemplars or the left shift key for nonexemplars. Participants initially received eight practice trials before beginning the experiment, in which the target category was weapons. After the practice trials, a new set of instructions was presented, allowing the participants to take a break, and informing them what the target category was going to be for the next block. The order of trials within each block was randomized. Participants completed five blocks of 40 stimuli and were always given a break with a new set of instructions about the new target category after the block. Upon completing both the SCT and LDT, the participants were then debriefed and dismissed.

Results

Data trimming

In the first phase of screening, two participants (2.02% of usable data) and two items (1% of usable data), were excluded due to having error rates that exceeded 50%. An additional participant (1.01% of usable data) and five items (2.42% of usable data) were excluded due to being extreme multivariate outliers in speed-accuracy space. Finally, errors (4.39% of usable data), and latencies faster than 200 ms or slower than 2000 ms, or that deviated from the participant's mean in that condition by more than three standard deviations (2.21% of usable data) were removed. In total, approximately 87% of the usable data was retained.

ExPrime coefficients

The coefficients for the model derived for Experiment 2's ExPrime scores ($M = 46.09$, $SD = 36.08$) can be found in Table 4. The largest positive predictors of priming effects in the SCT included self-reported L2 speaking and listening proficiency, as well as the percentage of time that participants used English in the school environment and in other social contexts. Negatively associated with priming effects were self-reported reading, writing, and speaking proficiency in Chinese.

Table 4. ExPrime Coefficients for Experiment 2.

Predictive Factor	ExPrime Coefficient Values
English Speaking Proficiency	13.44
English Use at School (%)	3.84
English Listening Proficiency	2.24
Chinese Listening Proficiency	1.67
English Use Outside of School & Home (%)	0.50
Chinese Reading Proficiency	−3.85
Chinese Speaking Proficiency	−3.95
Chinese Writing Proficiency	−3.95

Prime x ExPrime analysis

Due to issues with convergence, random slopes were excluded from these analyses, while retaining subjects ($SD = 52.02$ for exemplars, $SD = 51.89$ for nonexemplars) and items ($SD = 38.35$ for exemplars, $SD = 29.95$ for nonexemplars) as random effects. The models also failed to converge using an inverse Gaussian distribution, so the data were analyzed using a Gamma distribution with a BOBYQA optimizer (Powell, 2009). As with Experiment 1, additional supplementary analyses were conducted, and can be found in Appendix D. The relationship between prime type and ExPrime is shown in Figure 2 for the exemplars. For the exemplars, the model favored by the AIC and BIC was as follows: RT ~ Prime × ExPrime + Previous RT + Prime:Previous RT + (1 || Participant) + (1 || Item). This model (AIC = 109660, BIC = 109724) was favored over the model which included the main effects of prime and target frequency and excluded all interactions (AIC = 109667, BIC = 109730), and the model which included all two-way interactions between prime and ExPrime, prime and target frequency, and previous trial RT (AIC = 109663, BIC = 109755).

More specifically, in the RT analysis, there was a significant effect of prime type, $\beta = -14.87$, $SE = 2.84$, $t(8326) = -5.23$, $p < .0001$. Targets preceded by a translation prime ($M = 690$ ms) produced faster latencies than targets preceded by a control prime ($M = 704$ ms), replicating the translation priming effect found in prior research (e.g., Grainger & Frenck-Mestre, 1998). A significant effect of previous trial RT was also observed, $\beta = 39.28$, $SE = 3.38$, $t(8326) = 11.64$, $p < .0001$. Targets that were preceded by a fast trial ($M = 638$ ms) produced significantly shorter latencies than targets that were preceded by a slower trial ($M = 756$ ms). Most importantly, there was a significant two-way interaction between ExPrime and prime type, $\beta = -9.47$, $SE = 2.93$, $t(8326) = -3.23$, $p = .0012$, and prime and previous trial RT, $\beta = -10.49$, $SE = 3.79$, $t(8326) = -2.77$, $p = .006$. Participants who reported using English more frequently at school and in other social contexts, participants who reported being more proficient at speaking English, and participants who had relatively weaker reading, writing, and speaking proficiency in Chinese produced larger priming effects. An additional factor that predicted larger priming effects was listening proficiency in both English and Chinese. The priming effect was largest in the top tertile of ExPrime scores (30 ms), followed by Tertile 2 (14 ms), and Tertile 3 (-4 ms). Additionally, the priming effect was larger on trials that were preceded by a slower trial (18 ms) than it was on trials that were preceded by a faster trial (9 ms).[3]

The simple main effects analysis (paralleling that used in Experiment 1) revealed that the translation priming effect was significant in Tertile 1, $\chi^2(1) = 114.01$, $p < .0001$, and Tertile 2, $\chi^2(1) = 15.10$, $p = .0001$, but not in Tertile 3, $\chi^2 < 1$. Interaction contrasts revealed that the two-way interaction was significant when comparing the priming effect in Tertiles 1 and 2, $\chi^2(1) = 25.31$, $p < .0001$, Tertiles 1 and 3, $\chi^2(1) = 48.40$, $p < .0001$, and Tertiles 2 and 3, $\chi^2(1) = 8.82$, $p = .003$. In analyzing the effects of previous trial RT, simple main effects analyses revealed that the translation priming effect approached significance when the trial was preceded by a fast trial, $\chi^2(1) = 3.24$, $p = .07$, and was significant when the trial was preceded by a slow trial, $\chi^2(1) = 15.04$, $p = .0001$. However, the difference in priming effects between trials preceded by a slow trial and a fast trial was only marginally significant, $\chi^2(1) = 2.77$, $p = .10$.

As noted previously, an attempt was made to use the ExPrime values to predict priming effects in the nonexemplar data. In the nonexemplar data, the AIC and BIC favored the following model: RT ~ Prime + ExPrime + Previous RT + Prime Frequency + Target Frequency + (1 || Participant) + (1 || Item). This model

[3]Once again, the Pearson's correlation between ExPrime scores and the participants' priming effects revealed a significant positive relationship, $r(93) = .26$, $p = .011$, indicating that, as the main analysis indicates, larger ExPrime scores tended to be associated with larger priming effect sizes.

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

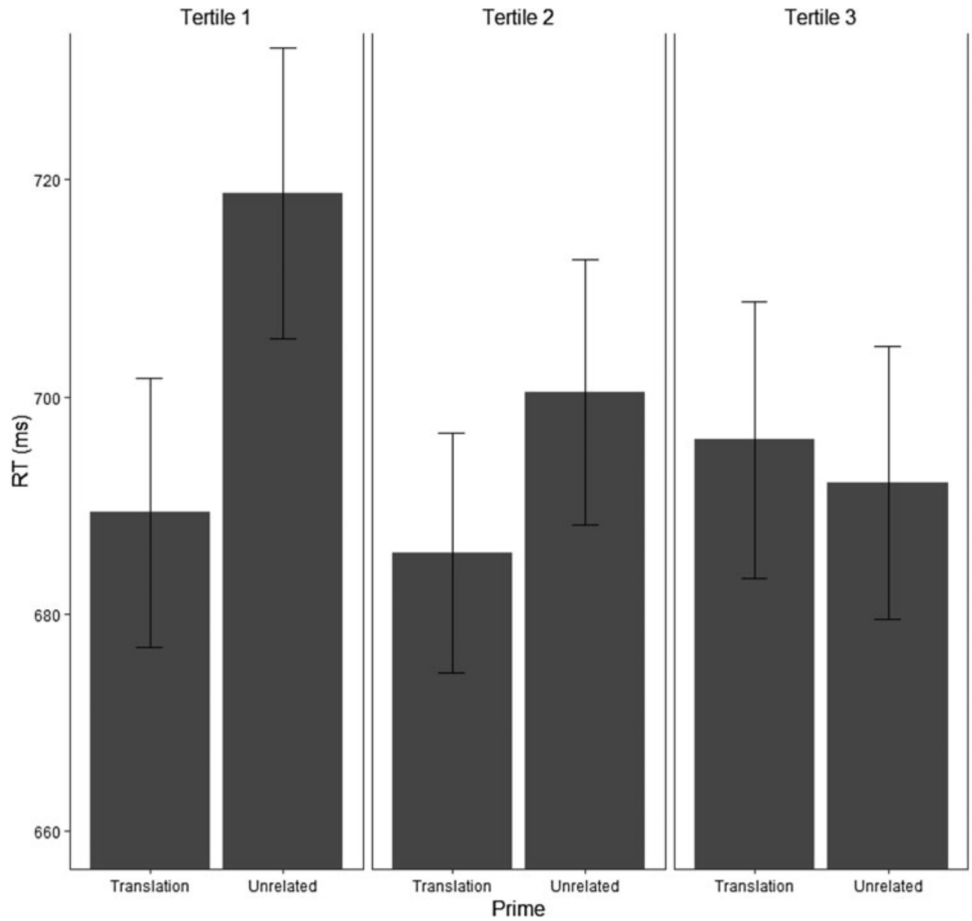

Fig. 2. Response times as a function of prime and s ExPrime tertile, Experiment 2 exemplars. Error bars represent 95% confidence intervals.

(AIC = 113372, BIC = 113436) over the model that included the interactions (AIC = 113385, BIC = 113632). This model revealed significant effects of ExPrime, $\beta = 19.46$, $SE = 3.19$, $t(8696) = 6.10$, $p < .0001$, and previous trial RT, $\beta = 31.20$, $SE = 2.37$, $t(8696) = 13.19$, $p < .0001$. A simple main effects comparison found that latencies were longer in Tertile 1 (i.e., those individuals who showed the largest priming effects in the exemplar data, $M = 715$ ms) than they were in Tertile 2 ($M = 685$ ms), $\chi^2(1) = 83.88$, $p < .0001$, and Tertile 3 ($M = 679$ ms), $\chi^2(1) = 29.46$, $p < .0001$. The difference between Tertiles 2 and 3 was nonsignificant, $\chi^2 < 1$, $p > .75$. Latencies were also shorter for nonexemplar trials that were preceded by a fast trial ($M = 637$ ms) than for slow trials ($M = 748$ ms). Critically, the priming effect was not significant, nor was the interaction between prime and ExPrime, *ts* < 1. Those effects were, instead, limited to the exemplar data (see Figure 3). In the error analysis, neither the effect of prime type, nor the two-way interaction between prime type and ExPrime were significant for exemplars, all *ts* < 1.16, *ps* > .25, or nonexemplars, all *zs* < 1.52, *ps* > .12.

Discussion

Experiment 2 produced three notable findings. First, replicating prior research on L2-L1 translation priming in semantic categorization (e.g., Finkbeiner et al., 2004; Grainger & Frenck-Mestre, 1998; Wang & Forster, 2010; Xia & Andrews, 2015), Experiment 2 produced a significant L2-L1 translation priming effect despite the same participants producing a null priming effect in an LDT in Experiment 1. Second, a priming effect was only found for words that were exemplars of the target category. Finally, and most relevant to the present discussion, the expectation was that there would be evidence of a dissociation between the L2 skills and behaviors (and, potentially, the L1 skills and behaviors) that would predict priming effects in an LDT and an SCT. The results of Experiment 2 confirmed this expectation, demonstrating that translation priming in semantic categorization was associated with participants' spoken proficiency of English, and with the amount of time the participant used English in their daily lives. These results stand in contrast to the results of Experiment 1, where translation priming effects were related to participants' reading, writing, and listening proficiency in English.

Note also, however, that there was a parallel between Experiments 1 and 2 in that L1 reading and writing proficiency was a significant negative predictor of priming in both experiments. The implication is simply that individuals less proficient in L1 are better able to have their L1 processing primed by a translation prime regardless of the L1 processing goal. A more complete overview of how these results could be accounted for is discussed below.

General discussion

The present research was an attempt to examine L2-L1 masked translation priming effects to evaluate whether the skills and linguistic behaviors most predictive of priming in lexical decision and semantic categorization were different and, if so, to determine

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

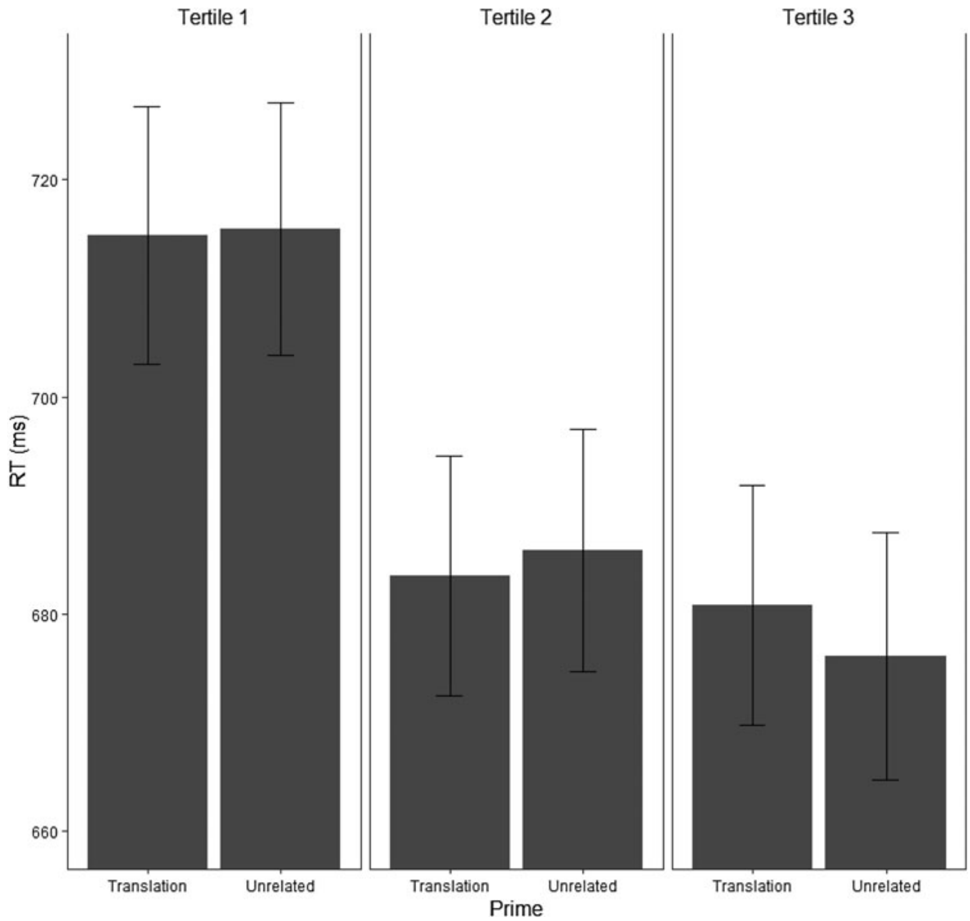

Fig. 3. Response times as a function of prime and ExPrime tertile, Experiment 2 nonexemplars. Error bars represent 95% confidence intervals.

whether the relevant specific skills and behaviors might be consistent with various representational assumptions made by current models of bilingual word processing – in particular, Multilink (Dijkstra et al., 2019). This research produced the following results. First, Experiments 1 and 2 replicated results from prior studies, producing a null overall priming effect in lexical decision and a significant priming effect in semantic categorization (e.g., Grainger & Frenck-Mestre, 1998). Second, both tasks showed an interaction between prime and language proficiency, as measured by the ExPrime score computed based on the associated task. That is, it was possible to create ExPrime scores based on various language skills which differed between tasks such that participants with higher ExPrime scores in the relevant task produced larger priming effects than participants with lower ExPrime scores. Finding that each set of ExPrime scores interacted with priming in the associated task, then, provides good evidence that the priming effects are sensitive to certain, and somewhat different, dimensions of L2 proficiency.

These results contribute to the evidence that L2-L1 priming in the LDT is related to competency of participants in their L2 (e.g., Nakayama et al., 2016), as well as, for the first time, providing a demonstration that priming in the SCT is as well. The more important fact, however, is our demonstration that, while domain-general L2 proficiency may play a role in driving translation priming, masked translation priming effects in LDTs and SCTs are partially driven by different components of L2 knowledge. Specifically, these results show that participants who are skilled readers and writers in their L2, and who have good comprehension of their spoken L2, tend to produce larger translation priming effects in LDTs, while participants who are more fluent speakers and listeners, and who actively use their L2 more frequently, tend to produce larger priming effects in SCTs.

Such findings appear to be accommodated by Multilink (Dijkstra et al., 2019), which proposes a common, underlying language processing system that, depending on the task that needs to be performed, can use different criteria to make decisions during processing. Multilink assumes that the general processing that takes place in lexical decisions and semantic categorizations is essentially the same. Both tasks must involve the activation of orthographic and semantic codes during processing. Where the tasks differ is in the code that is used by the task/decision subsystem to make a decision. In lexical decision, the task/decision subsystem is assumed to base word/nonword decisions on whether activity within the lexical-orthographic representations of the model surpass a critical threshold, while, in semantic categorization, exemplar/nonexemplar decisions are assumed to be based on whether the activity of semantic representations (or, at least, certain aspects of those semantic representations) surpass a threshold. Such an explanation is capable of accounting for why factors such as L2 reading and writing proficiency, which may reflect the activation of representations within the lexical-orthographic level of the word identification subsystem, would be particularly important predictors of L2-L1 translation priming in LDTs. For bilinguals who are highly skilled and

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

familiarized with the orthographic system of their L2, the resting-level activity of representations within their lexical-orthographic system would be higher due to the greater frequency of use and exposure to the L2 orthography. When a translation equivalent is used as a prime in such circumstances, the prime is more efficient at preactivating L1 representations, bringing the activity in that system to critical threshold faster.

In semantic categorization, however, where decision-making is largely based on semantic information, factors that would be expected to affect the enrichment of L2 semantic representations should affect how the prime influences semantic-level activity during word processing. One such factor that would be expected to influence semantic development is cultural immersion, or the immersion of the L2 learner in an L2-dominant environment. In terms of the L2 cultural immersion, more frequent use of the L2 in social interactions likely reflects a greater degree of immersion in the L2-dominant culture. Such immersion may have a number of effects on L2 language processing. It may affect the frequency with which the L2 learner is exposed to particular words, or the frequency with which the L2 learner uses particular words, and may also directly affect the learner's understanding of such concepts in their L2. As prior research has shown that the influence of cultural immersion on L2 conceptual development is not simply measured by overall L2 proficiency (e.g., Malt & Sloman, 2003), it follows, therefore, that L2-L1 priming in Experiment 2 would, instead, likely be better predicted by the amount of time that L2 learners used their L2, as that factor is more likely to reflect the impact of cultural immersion on conceptual development.

Broadly speaking, while this account accommodates the present findings, there may be other factors that contribute to the differences observed between L2-L1 translation priming effects in SCTs and LDTs. One potential source for these differences may stem from the nature of the task setup and the composition of stimuli in each task. In semantic categorization, participants are informed what the target category will be (e.g., mammal, insect, body part, clothing, fruit/vegetable) prior to beginning the experiment. Because participants are only responding positively to an extremely limited set of possible words, participants may strategically generate some of the exemplars of the target category. In doing so, the translation equivalents of these exemplars would become activated due to the parallel nature of language activation. A similar interpretation has also been given by Xia and Andrews (2015), who argued that the instructions in an SCT provide a cue which results in semantic features that are relevant to the category becoming activated above baseline, facilitating retrieval and decision-making processes for category members. Prior to even the first trial, then, many of the exemplars that participants respond to in the task, as well as category-relevant semantic features, will already be activated above their baseline activation levels. To the extent that a participant exploits such a strategy, larger priming effects would be expected in an SCT. Evidence consistent with this idea can be seen in Experiment 2, in which priming effects were only found on exemplar trials. Such results may suggest that translation priming effects are affected by differences in how the task activates concepts, and aspects of the verification process involved in the task.

In contrast, in lexical decision, participants only know that targets will be either words or nonwords. Unlike the SCT, the instructions in an LDT do not provide any information about what kind of words the participant can expect to respond to. The number of possible targets that would require a "yes" response is far greater than in an SCT, and retrieving a set of words that one can anticipate encountering in the task is no longer an effective strategy. As such, the setup and instructions of an LDT do not provide any useful cues to preactivate the representations of a word target, and the priming effect may be smaller as a consequence.

Challenges

Before concluding, it may be useful to consider how the present research can be fruitfully extended. First, the precision and accuracy of the ExPrime measures are only as good as the factors that they were composed of. There may be factors that were not considered here that would help build a better representation of ExPrime in the two tasks (in the sense of better understanding the driving factors in translation priming in the tasks). For example, the role of receptive and productive vocabulary size in translation priming is currently not well understood, and was not measured here. Potentially, it could be a measure that might contribute to ExPrime in either task. Further, a recent meta-analysis conducted by Wen and van Heuven (2017b) has shown that one of the most important moderators of L2-L1 translation priming effect sizes in lexical decision is the number of items per cell. Specifically, a larger number of items per cell (e.g., 80 items per cell) has generally been associated with larger L2-L1 priming effects in that task (e.g., Luo et al., 2013; Schoonbaert, Holcomb, Grainger & Hartsuiker, 2011). For comparison, Experiment 1 in the present research used only 50 items per cell. It is unclear, of course, what it is about having large numbers of items per cell that might affect processing in this task. However, discovering why the number of items per cell has an impact, as well as discovering a more comprehensive set of factors that contribute to translation priming, should allow for a refinement of the crucial factors in the ExPrime measures.

Second, it should be noted that the computation of ExPrime in the present research largely was based on self-report measures. Therefore, the estimates that were used to make predictions about L2-L1 translation priming effects relied on the accuracy of self-assessments of L2 proficiency. It would be useful if future research could use more objective measures of L2 proficiency and vocabulary knowledge (e.g., LexTALE – Lemhofer & Broersma, 2012; the Nelson-Denny Reading Test – Brown, Fishco & Hanna, 1993; the individual components of IELTS, TOEIC, or TOEFL). Alternatively, it may also be useful to compare different types of measures, self-report or standardized, on how predictive they are of translation priming effects. Additionally, there were several measures that were not considered in the present experiments that may have been valuable to consider. These include the number of years that the participant had been speaking their L2, the age at which their formal L2 instruction began, and the age of acquisition of their L2. In future research, such measures will be considered to create a more thorough profile of participant L2 use and proficiency.

Finally, it will be useful to extend this research beyond the Chinese–English bilinguals examined here to involve different scripts, languages, and orthographies, such as Hebrew, Korean, or Japanese. Such orthographies are nonalphabetic and have tended to produce the interaction between translation priming and the two tasks used here when the L2 is an alphabetic language. The use of ExPrime in the present research thus represents only the first step towards developing a more sophisticated understanding of the factors that contribute to translation priming, and how those factors differ across different tasks.

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

Conclusions

The present experiments were an attempt to better understand the apparent task-specific nature of the masked translation priming effect that has been reported in prior studies (e.g., Finkbeiner et al., 2004; Gollan et al., 1997; Grainger & Frenck-Mestre, 1998; Jiang & Forster, 2001). More specifically, the present research investigated the question of how masked translation priming effects are affected by the interaction between these task-specific demands and participant L2 skills and usage patterns. To that end, a machine-learning algorithm was used in an effort to understand how L2 skills and usage patterns contribute to masked translation priming in two different tasks. The relevant analyses have shown evidence that the factors that contribute to the ability of translation primes to activate the relevant representations of their targets are reasonably task-specific. In lexical decision, priming effects were larger for participants who reported having better speaking, reading, and writing abilities in English, and relatively weaker reading and writing abilities in Chinese. In semantic categorization, participants who reported using English more frequently in daily living, participants who reported being more proficient at speaking English, and those who had relatively weaker reading, writing, and speaking proficiency in Chinese showed larger priming effects.

In general, there are two important implications of these experiments. First, these experiments highlight how a single, underlying word processing system can flexibly account for the different patterns of findings observed in the masked translation priming literature. These experiments support the notion that the performance of the underlying word processing system is fine-tuned by how task demands place greater emphasis on specific sources of information during word processing, and how certain participant variables influence the processing of this information. Ultimately, the present research suggests that the apparently inconsistent findings observed in the masked translation priming literature can be reconciled by such an account. Second, the present research represents an important step towards developing a large-scale, data-driven approach to understanding how bilingual memory processes influence the process of visual word recognition, and how those processes vary as a function of task demands. Future research will hopefully allow for the creation of more comprehensive data-driven tools, leading to a more sophisticated understanding of how second language acquisition affects the development of lexical and conceptual memory for words in both L1 and L2.

Acknowledgements. This research was based on Mark McPhedran's doctoral dissertation at the University of Western Ontario, under the supervision of Stephen J. Lupker. We thank Debra Jared, Albert Katz, Ken McRae, Yasaman Rafat and Alain Desrochers for their comments and suggestions, and Huilan Yang and Zian Chi for their assistance with designing the experiments and with the data collection.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In BN Petrov & F Csaki (Eds.), *2nd International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akadémiai Kiadó.

Baayen RH, Piepenbrock R and Gulikers L (1995) *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Baayen RH, Davidson DJ and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**(4), 390–412. DOI: 10.1016/j.jml.2007.12.005

Bates D, Maechler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**(1), 1–48. DOI: 10.18637/jss.v067.i01

Brown JA, Fishco VV and Hanna G (1993) Nelson-Denny reading test: Manual for scoring and interpretation, Forms G & H. Rolling Meadows, IL: Riverside Publishing.

Chen B, Zhou H, Gao Y and Dunlap S (2014) Cross-language translation priming asymmetry with Chinese–English bilinguals: A test of the sense model. *Journal of Psycholinguistic Research* **43**(3), 225–240. DOI: 10.1007/s10936-013-9249-3

de Groot AMB and Nas GLJ (1991) Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language* **30**(1), 90–123. DOI: 10.1016/0749-596X(91)90012-9

De Rosario-Martinez H (2015) *phia: Post-hoc interaction analysis*. R package version 0.2-1.

Dietterich TG (2000) Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1–15). Springer, Berlin, Heidelberg.

Dijkstra T, Hilberink-Schulpen B and van Heuven WJB (2010) Repetition and masked form priming within and between languages using word and nonword neighbors. *Bilingualism: Language and Cognition* **13**(3), 341–357. DOI: 10.1017/S1366728909990575

Dijkstra T, Miwa K, Brummelhuis B, Sappelli M and Baayen H (2010) How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language* **62**, 284–301.

Dijkstra T and van Heuven WJB (2002) The architecture of the bilingual word recognition system. From identification to decision. *Bilingualism: Language and Cognition* **5**(3), 175–197. DOI: 10.1017/S1366728902003012

Dijkstra T, Wahl A, Buytenhuijs F, Van Halem N, Al-Jibouri Z, De Korte M and Rekké S (2019) Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition* **22**(4), 657–679.

Dimitropoulou M, Duñabeitia JA and Carreiras M (2011) Masked translation priming effects with low proficient bilinguals. *Memory & Cognition* **39**, 260–275. DOI: 10.3758/s13421-010-0004-9

Duzan H and Shariff NSBM (2015) Ridge regression for solving the multicollinearity problem: Review of methods and models. *Journal of Applied Sciences* **15**, 392–404. DOI: 10.3923/jas.2015.392.404

Finkbeiner M, Forster K, Nicol J and Nakamura K (2004) The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language* **51**(1), 1–22. DOI: 10.1016/j.jml.2004.01.004

Forster KI and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* **35**(1), 116–124. DOI: 10.3758/BF03195503

Géron A (2017) *Hands-on machine learning with scikit-learn & TensorFlow: Concepts, tools, and techniques to build intelligent systems*. Sevastopol, CA: O'Reilly.

Gollan TH, Forster KI and 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**(5), 1122–1139. DOI: 10.1037/0278-7393.23.5.1122

Grainger J and Frenck-Mestre C (1998) Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes* **13**(6), 601–623. DOI: 10.1080/016909698386393

Hoerl AE and Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**(1), 55–67. DOI: 10.2307/1267351

Jiang N (1999) Testing processing explanations for the asymmetry in masked cross-language priming. *Bilingualism: Language and Cognition* **2**(1), 59–75.

Jiang N and Forster KI (2001) Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language* **44**(1), 32–51. DOI: 10.1006/jmla.2000.2737

Kenward MG and Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**(3), 983–997. DOI: 10.2307/2533558

Kroll JF and Stewart E (1994) Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory

representations. *Journal of Memory and Language* **33**(2), 149–174. DOI: 10.1006/jmla.1994.1008
Kroll JF and Tokowicz N (2001) The development of conceptual representation for words in a second language. In JL Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 49–71). Malden, MA: Blackwell.
Lemhofer K and Broersma M (2012) Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research Methods* **44**(2), 325–343. DOI: 10.3758/s13428-011-0146-0
Luke SG (2017) Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* **49**, 1494–1502. DOI: 10.3758/s13428-016-0809-y
Luo X, Cheung H, Bel D, Li L, Chen L and Mo L (2013) The roles of semantic sense and form-meaning connection in translation priming. *The Psychological Record* **63**(1), 193–208. DOI: 10.11133/j.tpr.2013.63.1.015
Mahalanobis PC (1936) On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India* **2**(1), 49–55.
Malt BC and Sloman SA (2003) Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition* **6**(1), 47–67. DOI: 10.1017/S1366728903001020
Marian V, Blumenfeld HK and Kaushanskaya M (2007) The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* **50**(4), 940–967. DOI: 10.1044/1092-4388(2007/067)
Muhammad I, Maria J and Muhammad AR (2013) Comparison of shrinkable regression for remedy of multicollinearity problem. *Middle East Journal of Scientific Research* **14**(4), 570–579. DOI: 10.5829/idosi.mejsr.2013.14.4.488
Nakayama M, Ida K and Lupker SJ (2016) Cross-script L2-L1 noncognate translation priming in lexical decision depends on L2 proficiency: Evidence from Japanese–English bilinguals. *Bilingualism: Language and Cognition* **19**(5), 1001–1022. DOI: 10.1017/S1366728915000462
Nakayama M, Sears CR, Hino Y and Lupker SJ (2013) Masked translation priming with Japanese–English bilinguals: Interactions between cognate status, target frequency, and L2 proficiency. *Journal of Cognitive Psychology* **25**, 949–981. DOI: 10.1080/20445911.2013.839560
Oyeyemi GM, Ogunjobi EO and Folorunsho AI (2015) On performance of shrinkage methods: A Monte Carlo study. *International Journal of Statistics and Applications* **5**(2), 72–76. DOI: 10.5923/j.statistics.20150502.04
Powell MJD (2009) The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/0, University of Cambridge, Cambridge*, 26–46.
R Core Team (2017) R: A language and environment for statistical computing. Vienna, Austria. http://www.R-project.org/
Roelofs A (2008) Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **34**(2), 353–368. DOI: 10.1037/0278-7393.34.2.353
Schoonbaert S, Holcomb PJ, Grainger J and Hartsuiker RJ (2011) Testing asymmetries in noncognate translation priming: Evidence from RTs and ERPs. *Psychophysiology* **48**(1), 74–81. DOI: 10.1111/j.1469-8986.2010.01048.x
Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288.
Tikhonov AN (1963) Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics* **4**, 1035–1038.
Tse CS, Yap MJ, Chan YL, Sze WP, Shaoul C and Lin D (2017) The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods* **49**(4), 1503–1519. DOI: 10.3758/s13428-016-0810-5
Wang X and Forster KI (2010) Masked translation priming with semantic categorization: Testing the Sense Model. *Bilingualism and Cognition* **13**(3), 327–340. DOI: 10.1017/S1366728909990502
Wen Y and van Heuven WJB (2017a) Chinese translation norms for 1,429 English words. *Behavior Research Methods* **49**(3), 1006–1019. DOI: 10.3758/s13428-016-0761-x
Wen Y and van Heuven WJB (2017b) Non-cognate translation priming in masked priming lexical decisions: A meta-analysis. *Psychonomic Bulletin & Review* **24**(3), 879–886. DOI: 10.3758/s13423-016-1151-1
Xia V and Andrews S (2015) Masked translation priming asymmetry in Chinese–English bilinguals: Making sense of the Sense Model. *The Quarterly Journal of Experimental Psychology* **68**(2), 294–325. DOI: 10.1080/17470218.2014.944195
Zou H and Hastie T (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B* **67**, 301–320.

Appendix A Materials used in Experiment 1

Appendix A

Translation Prime	Control Prime	Target
Advice	heron	忠告
border	table	边境
chance	safety	机会
Dance	bicycle	舞蹈
government	Energy	政府
Land	soldier	土地
parrot	pocket	鹦鹉
Quail	rope	鹌鹑
secret	campaign	秘密
theory	legend	理论
Beach	bottle	海滩
candle	dive	蜡烛
college	lane	学院
energy	government	能源
Hotel	problem	旅馆
minute	captain	分钟
poetry	carpet	诗歌
Road	college	道路
Steam	vulture	蒸汽
Vote	window	投票
Beard	coffee	胡子
captain	minute	队长
comedy	reward	喜剧
Forest	bridge	森林
Idea	country	理念
mirror	lunch	镜子
Post	handsome	邮寄
Rope	quail	绳子
sunset	luck	夕阳
vulture	steam	秃鹰
Album	traffic	专辑
Bottle	sand	瓶子
Cliff	game	悬崖
discussion	clown	讨论
Guitar	career	吉他
legend	theory	传说
pencil	bacon	铅笔
record	metal	记录
Sink	camera	水槽

(*Continued*)

Appendix A (*Continued.*)

Translation Prime	Control Prime	Target
Ticket	piano	车票
author	penguin	作者
bridge	forest	桥梁
Clock	beer	时钟
Dive	candle	潜水
handsome	post	英俊
Luck	sunset	运气
penguin	author	企鹅
reptile	sponge	爬虫
Skate	swan	滑冰
Toilet	turkey	厕所
Beer	clock	啤酒
Car	instinct	汽车
computer	season	电脑
Friend	wall	朋友
instinct	car	直觉
morning	business	早上
problem	hotel	问题
Safety	chance	安全
Swan	skate	天鹅
Wall	friend	墙壁
bicycle	dance	单车
career	guitar	事业
country	idea	国家
Game	cliff	游戏
Juice	profit	果汁
Music	research	音乐
Profit	juice	利润
Salt	doctor	食盐
Table	border	桌子
whistle	customer	哨子
Bacon	pencil	咸肉
business	morning	商业
Clown	discussion	丑角
doctor	salt	医生
Health	kitchen	健康
Lunch	mirror	午餐
Piano	ticket	钢琴
research	music	研究
soldier	land	军人
traffic	album	交通
Bank	tape	银行
Bill	glass	法案

(*Continued*)

Appendix A (*Continued.*)

Translation Prime	Control Prime	Target
camera	sink	相机
coffee	beard	咖啡
Dollar	puppet	美元
Heron	advice	白鹭
Metal	record	金属
pocket	parrot	口袋
reward	comedy	奖励
sponge	reptile	海绵
turkey	toilet	火鸡
carpet	poetry	地毯
customer	whistle	顾客
Glass	bill	玻璃
kitchen	health	厨房
neighborhood	voice	邻里
puppet	dollar	木偶
season	computer	季节
Tape	bank	胶带
window	vote	窗户

Appendix B Materials used in Experiment 2

Appendix B

Translation Prime	Control Prime	Target
Bat	sweater	蝙蝠
camel	centipede	骆驼
Cow	watermelon	母牛
Fox	fly	狐狸
Goat	shoulder	山羊
hedgehog	moth	刺猬
hippopotamus	beetle	河马
kangaroo	lemon	袋鼠
lion	locust	狮子
monkey	throat	猴子
mouse	tooth	老鼠
orangutan	nose	猩猩
panda	eye	熊猫
rabbit	scarf	兔子
rhino	tie	犀牛
seal	coat	海豹
squirrel	wasp	松鼠
tiger	banana	老虎

(*Continued*)

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

Appendix B (*Continued.*)

Translation Prime	Control Prime	Target
whale	cicada	鲸鱼
zebra	plum	斑马
ant	chest	蚂蚁
bee	blouse	蜜蜂
beetle	hippopotamus	甲虫
butterfly	cherry	蝴蝶
caterpillar	lips	毛虫
centipede	camel	蜈蚣
cicada	whale	蝉鸣
cockroach	olive	蟑螂
cricket	onion	蟋蟀
dragonfly	belt	蜻蜓
earwig	eyeglasses	蠼螋
flea	necklace	跳蚤
fly	fox	苍蝇
grasshopper	gloves	蚱蜢
locust	lion	蝗虫
louse	mushroom	头虱
mantis	slippers	螳螂
mosquito	pear	蚊虫
moth	hedgehog	飞蛾
wasp	squirrel	黄蜂
apple	apron	苹果
banana	tiger	香蕉
beet	muscle	甜菜
celery	pyjamas	芹菜
cherry	butterfly	樱桃
corn	chin	玉米
cucumber	crown	黄瓜
grape	hat	葡萄
lemon	kangaroo	柠檬
lettuce	pancreas	生菜
mushroom	louse	冬菇
olive	cockroach	橄榄
onion	cricket	洋葱
orange	stomach	橙子
pear	mosquito	鸭梨
pineapple	liver	菠萝
plum	zebra	李子
strawberry	heart	草莓
tomato	arm	番茄

(*Continued*)

Appendix B (*Continued.*)

Translation Prime	Control Prime	Target
watermelon	cow	西瓜
apron	apple	围裙
belt	dragonfly	腰带
blouse	bee	衬衫
boots	thumb	靴子
bra	skin	胸罩
coat	seal	上衣
crown	cucumber	皇冠
eyeglasses	earwig	眼镜
gloves	grasshopper	手套
hat	grape	帽子
necklace	flea	项链
pyjamas	celery	睡衣
sandals	back	凉鞋
scarf	rabbit	围巾
shoes	ear	鞋子
skirt	skull	短裙
slippers	mantis	拖鞋
socks	finger	袜子
sweater	bat	毛衣
tie	rhino	领带
arm	tomato	胳膊
back	sandals	背部
chest	ant	胸部
chin	corn	下巴
ear	shoes	耳朵
eye	panda	眼睛
finger	socks	手指
heart	strawberry	心脏
lips	caterpillar	嘴唇
liver	pineapple	肝脏
muscle	beet	肌肉
nose	orangutan	鼻子
shoulder	goat	肩膀
skin	bra	皮肤
skull	skirt	头骨
stomach	orange	肠胃
throat	monkey	喉咙
thumb	boots	拇指
tongue	leopard	舌头
tooth	mouse	牙齿

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

Appendix C Computing ExPrime

To derive ExPrime, the predictors (i.e., English and Chinese reading, writing, speaking, listening proficiency, and estimated use of English in the home, school, and other environments) were first rescaled as standard scores, and the priming effects were mean centered. The priming and predictor data were then split into a training and testing set. The training set (approximately 80% of the dataset) was used to fit the models to the priming data and tune the hyperparameters of the models. Hyperparameters are parameters whose values are set before the learning process begins, rather than being derived through training. Tuning the hyperparameters of a model provides the benefit of minimizing the cost function, while ensuring that the model is not overfitting the data. The testing set was used to validate that the predictions of the model generalized to new data. Once each model was fit on the training data, its predictions were compared to actual priming effects and error rates, and both the mean squared error (*MSE*) and the root mean squared error (*RMSE*) were computed.

To ensure that the models were not overfitting the data, the models were further regularized by performing a randomized search to find the optimum combination of hyperparameters using a specified subset of the hyperparameter space for each model (Géron, 2017). This randomized search was then evaluated using a k-fold cross-validation method. In a k-fold cross-validation method, the sample of data is partitioned at random into k equal sized subsamples. One of these subsamples is then used as validation data, and the remaining subsamples are used as training data. This process is iterative, in that each subsample is used as validation data once. The cross-validation procedure used in the present research used 10 folds using this method. The fit of each iteration was evaluated using the negative mean squared error (*NMSE*). The randomized search was carried out for five thousand iterations per model. The set of hyperparameters which produced the best fit for each of the three models were selected. Finally, the newly derived models were validated on the testing set, and a set of coefficients was derived.

Once all three models were tuned, trained, and validated, a final k-fold cross-validation was performed on each model using the testing dataset, and the performance of each model for each iteration was scored using the *NMSE* of the predictions. The *RMSE*s were then derived from this final cross validation, and the mean and standard deviation of the *RMSE* for each model was then computed. Using the mean and standard deviations of the *RMSE* for each model allowed a comparison of how each model performed, as well as the coefficient weights for each model. The mean and standard deviations of the *RMSE* for each model for both experiments can be found below.

Mean and Standard Deviations for the Residual Mean Squared Errors for Each Model.

ExPrime was then computed using an ensemble method (e.g., Dietterich, 2000), in which the predictions of all three models were aggregated into a single, final prediction. Ensemble regressors can often perform better than any single regression model by capitalizing on the strengths of each model, and compensating for the weaknesses of each model. For the purpose of ExPrime, a simple averaging method was used, where the final coefficients used to compute ExPrime reflected the weighted average of the coefficients derived from the three models that were fit. The best-performing model coefficients were weighted three times as much as those of the other two models. Using these coefficients, ExPrime was computed by aggregating the weighted sum of the predictor values from the ensemble measure for each participant. Finally, for inclusion in the analyses, the ExPrime score was scaled using a z-score method.

	Model					
	Ridge		Lasso		Elastic Net	
Experiment	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	31.94	6.13	31.80	5.46	30.52	5.98
2	43.19	12.26	43.93	10.31	43.35	11.39

Appendix D Supplementary Analyses

Experiment 1 supplementary analyses

ExPrime component verification analyses

Analyses were performed to verify whether individual components of ExPrime significantly interacted with prime relatedness when all other ExPrime components were treated as covariates. The model structure was as follows: RT ~ Prime × ExPrime Component 1 + ExPrime Component 2 + ExPrime Component 3 + ExPrime Component 4 + ExPrime Component 5 + ExPrime Component 6 + (1 ‖ Participant) + (1 ‖ Item). This analysis revealed significant two-way interactions between prime and a) L2 reading ability, $\beta = -2.76$, $SE = 1.3$, $t(8584) = -2.12$, $p = .034$, b) L2 listening ability, $\beta = -5.62$, $SE = 2.45$, $t(8584) = -2.30$, $p = .022$, c) L1 reading ability, $\beta = -1.35$, $SE = 0.48$, $t(8584) = -2.85$, $p = .004$, d) L1 speaking ability, $\beta = -10.06$, $SE = 0.53$, $t(8584) = -18.90$, $p < .0001$, and e) L1 writing ability, $\beta = -2.30$, $SE = 0.95$, $t(8584) = -2.42$, $p = .02$, and as well as an interaction that approached significance between prime and L2 writing ability, $\beta = -2.62$, $SE = 1.43$, $t(8584) = -1.84$, $p = .066$.

Translation uniqueness effects

One factor that was not initially considered is whether the results of Experiment 1 were affected by the uniqueness of the translation pairs used. That is, when selecting the stimuli for Experiment 1, we did not distinguish between translation pairs in which the prime represented a) a unique translation of the target or b) one of a number of possible translations of the target. In an attempt to address this issue, subsequent analyses were conducted to examine whether translation uniqueness might have affected the results.

Data about translation uniqueness was collected from Wen and van Heuven's (2017a) English-Chinese translation norms. It was found that 68 of the words used in Experiment 1 appeared in these norms, of which 24 were unique translation pairs. Thirty of these words had multiple translation pairs with the translation used here being the most frequently-occurring Chinese translation equivalent. Only 14 of these translation pairs did not involve the most frequently-occurring Chinese translation equivalent. Of these 14 pairs, one pair (i.e., bacon-咸肉) involved a more frequent translation equivalent, but this translation equivalent was an English loanword (i.e., 培根, or péigēn in Pinyin). A uniqueness score was calculated by using the count data from Wen and van Heuven's norms. For example, a word such as *bottle* has two translation equivalents in Chinese (瓶子 and 瓶). However, the two-character translation equivalent had an observed count of 25, and the one-character translation equivalent had an observed count of only one. The uniqueness score for *bottle*-瓶子 was thus calculated as 96.15%. Overall, of the 68 pairs used in Experiment 1 that appeared in these norms, the average uniqueness score was 72% ($SD = 32.65$).

This uniqueness score was added as a covariate in follow-up analyses, using only the stimuli that appeared in Wen and van Heuven's (2017a) norms, these analyses were conducted using prime, ExPrime, previous trial RT, uniqueness, and prime and target frequency as fixed factors, and participant ($SD = 53.61$) and item ($SD = 20.86$) as random factors. The best fitting model was as follows: RT ~ Prime × ExPrime + Previous RT + Prime:Previous RT + Uniqueness + Prime Frequency + Target Frequency + (1 ‖ Participant) + (1 ‖ Item). This model (AIC = 76823, BIC = 76896) was favored over the model that excluded all interactions but retained prime frequency as a fixed effect (AIC = 76829, BIC = 76896), and the model which included the interaction term involving uniqueness (AIC = 76825, BIC = 76905).

This follow-up analysis once again revealed a significant two-way interaction between prime and ExPrime, $\beta = -8.35$, $SE = 2.85$, $t(5979) = -2.93$, $p = .0034$, and a two-way interaction between prime and previous trial RT, $\beta = -9.51$, $SE = 3.09$, $t(5979) = -3.08$, $p = .0021$. The priming effect was significant in Tertile 1 (19 ms), $\chi^2(1) = 32.08$, $p < .0001$, but was not significant in Tertile 2 (−3 ms), $\chi^2 < 1$, $p = .50$, or Tertile 3 (0 ms), $\chi^2 < 1$, $p = .81$. The priming effect was significantly larger in Tertile 1 than it was in Tertile 2, $\chi^2(1) = 21.11$, $p < .0001$, and Tertile 3, $\chi^2(1) = 18.32$, $p < .0001$. The priming effect was significant when the trial was preceded by a slow trial (18 ms), $\chi^2(1) = 6.19$, $p = .013$, but not when the trial was preceded by a fast trial (−6 ms), $\chi^2(1) < 1$, $p > .90$. The difference in the priming effect for trials preceded by fast and slow trials was significant, $\chi^2(1) = 8.02$, $p = .005$.

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press

Experiment 2 supplementary analyses

ExPrime component verification analyses

Analyses were conducted to verify whether the individual components of ExPrime significantly interacted with prime relatedness by assessing the two-way interaction between prime and an ExPrime component while treating all other components as covariates. The model structure was as follows: RT ~ Prime × ExPrime Component 1 + ExPrime Component 2 + ExPrime Component 3 + ExPrime Component 4 + ExPrime Component 5 + ExPrime Component 6 + (1 ‖ Participant) + (1 ‖ Item). This follow-up analysis revealed significant two-way interactions between prime and a) spoken English proficiency, $\beta = 5.67$, $SE = 1.74$, $t(8323) = 3.26$, $p = .001$, b) spoken English comprehension, $\beta = 15.95$, $SE = 1.65$, $t(8323) = 9.67$, $p < .0001$, c) use of English at school, $\beta = 9.41$, $SE = 1.94$, $t(8323) = 4.86$, $p < .0001$, d) Chinese writing proficiency, $\beta = 31.82$, $SE = 3.92$, $t(8323) = 8.12$, $p < .0001$, e) spoken Chinese proficiency, $\beta = 17.50$, $SE = 2.14$, $t(8323) = 8.18$, $p < .0001$, and f) spoken Chinese comprehension, $\beta = 19.43$, $SE = 3.06$, $t(8323) = 6.35$, $p < .0001$.

Translation uniqueness effects

It was found that 58 of the words used in Experiment 2 appeared in the Wen and Van Heuven (2017a) norms, of which 38 involved a unique translation pair. Fourteen of these words had multiple translation pairs, but the most frequently-occurring Chinese translation equivalent had been used in Experiment 2. Only six of these trials did not use the most frequently-occurring Chinese translation equivalent. The average uniqueness score for these words was 87.15% ($SD = 24.10$), indicating that the stimuli used in Experiment 2 largely reflected the most frequently used translation pairs based on the information in the Chinese–English translation norms.

A generalized linear mixed-effects model was constructed using prime, ExPrime, previous trial RT, uniqueness, and prime and target frequency as fixed factors, and participant ($SD = 57.40$) and item ($SD = 41.11$) as random factors. The model favored by the AIC and BIC criterion was as follows: RT ~ Prime × ExPrime + Previous RT + Prime:Previous RT + Prime Frequency + Target Frequency + (1 ‖ Participant) + (1 ‖ Item). This model (AIC = 65019, BIC = 65084) provided a better fit to the data than the model that excluded all interactions (AIC = 65027, BIC = 65092), the model which included uniqueness as a covariate (AIC = 65022, BIC = 65100), and the model which included the two-way interaction between prime and uniqueness (AIC = 65019, BIC = 65104). This follow-up analysis once again found significant two-way interactions between prime and ExPrime, $\beta = -14.14$, $SE = 3.60$, $t(4913) = -3.92$, $p < .0001$, and prime and previous trial RT, $\beta = -9.34$, $SE = 4.49$, $t(4913) = -2.08$, $p = .038$. The priming effect was significant in Tertile 1 (26 ms), $\chi^2(1) = 36.2$, $p < .0001$, and in Tertile 2 (16 ms), $\chi^2(1) = 9.78$, $p = .0018$, whereas an inhibitory effect that was observed in Tertile 3 (−19 ms) that was significant, $\chi^2(1) = 9.71$, $p = .0018$. Additionally, the difference in priming effects were significant between Tertiles 1 and 2, $\chi^2(1) = 3.84$, $p = .05$, Tertiles 1 and 3, $\chi^2(1) = 77.91$, $p < .0001$, and Tertiles 2 and 3, $\chi^2(1) = 30.08$, $p < .0001$. The priming effect was significant for trials preceded by a slow trial (16 ms), $\chi^2(1) = 8.52$, $p = .0035$, but not when the trial was preceded by a fast trial (1 ms), $\chi^2(1) < 1$, $p > .50$. The difference observed in the priming effects between trials preceded by a slow trial and a fast trial was also significant, $\chi^2(1) = 9.29$, $p = .0023$.

https://doi.org/10.1017/S1366728920000504 Published online by Cambridge University Press