The elusive impact of L2 immersion on translation priming

Adel Chaouch-Orozco[1] [ID], Jorge González Alonso[2,3] [ID], Jon Andoni Duñabeitia[2,3] [ID] and Jason Rothman[3,2]*

[1]The Hong Kong Polytechnic University, Hong Kong; [2]Universidad Nebrija, Madrid, Spain; [3]UiT The Arctic of Norway, Tromsø, Norway
*Corresponding author. Email: jason.rothman@uit.no

Abstract

A growing consensus sees the bilingual lexicon as an integrated, nonselective system. However, the way bilingual experience shapes the architecture and functioning of the lexicon is not well understood. This study investigates bilingual lexical-semantic representation and processing employing written translation priming. We focus on the role of active exposure to and use of the second language (L2)—primarily operationalized as immersion. We tested 200 highly proficient Spanish–English bilinguals in two groups differing in their societal language (immersed vs. nonimmersed) and amount of L2 use. L2 proficiency was controlled across participants, allowing us to disentangle its effects from those of L2 use. Overall, however, the immersion's impact on our data was minimal. This suggests a ceiling effect for the influence of active L2 use on bilingual lexical functioning when L2 development is maximal. The present data provide relevant insights into the nature of the bilingual lexicon, informing developmental models.

Introduction

Most current accounts of bilingual lexical organization assume that the bi-/multilingual lexicon is integrated (words from both languages are stored together; e.g., Meade et al., 2017; van Heuven et al., 1998). Moreover, it has been shown that first (L1) and second language (L2) words are activated in parallel during lexical access, even when the context calls for a fully unilingual mode (e.g., Thierry & Wu, 2007). Nonselective access has been observed in numerous studies examining comprehension and/or production of isolated words (e.g., Kroll et al., 2006). The most compelling evidence of nonselectivity comes from coactivation in sentence comprehension studies, where the sentential context could be expected to constrain the search space to only one of the languages (e.g., Duyck et al., 2007; van Assche et al., 2012).

An important question following from these findings is how exactly words from different languages are connected at each level of representation (sublexical, lexical,

semantic), and whether the nature and strength of these connections change dynamically as a function of relative experience with each language (e.g., Kroll & Stewart, 1994). One of the most direct ways to explore cross-language connections in the bilingual lexicon is to use priming techniques, with prime and target words belonging to different languages. By manipulating the factors postulated to constrain and regulate these connections and measuring how this affects priming, we can test theories about the architecture of the lexicon. In word recognition research, this logic has most often been embodied in lexical decision tasks with translation priming (e.g., Wen & van Heuven, 2017, for review).

In a primed lexical decision task, the subject is presented with a prime word followed by a string of letters upon which they make a lexical decision (i.e., a yes/no answer to the implicit question "is this a real word?"). In the critical condition, prime and target are related at some level of interest—for example, semantics, morphology, orthography—while in the control condition they are unrelated. In a *translation* priming paradigm, related primes and targets are translation equivalents (e.g., *flecha-ARROW*, in a Spanish–English experiment), and cross-language unrelated pairs constitute the control condition (e.g., *camisa*, Spanish for "shirt"-*ARROW*). Priming effects manifest as significantly different mean response times (RTs) and/or error rates between the two conditions, typically with shorter latencies and/or greater accuracy in the related condition. Under interactive theories of lexical-semantic processing (e.g., Collins & Loftus, 1975), priming effects are interpreted as a given amount of (pre)activation spreading from a related prime to a target within the lexical network, facilitating its processing and speeding up retrieval.

The priming literature on bilingual lexical access has used both cognate and noncognate words (e.g., Duñabeitia et al., 2010; see Sánchez-Casas & García-Albea, 2005, for an early review). Cognate words are etymologically related pairs in different languages that retain some similarity in both form and meaning. Noncognate translation equivalents refer to semantically related pairs with no overlap at the form level (e.g., English *dog* and Spanish *perro*). Because orthography and phonology are not shared in these pairs, priming effects between noncognate translation equivalents have been used to gauge the availability of links between such words at the lexeme and conceptual levels. A recurrent finding in these experiments is a priming asymmetry (Wen & van Heuven, 2017). While priming effects in the L1 (prime) to L2 (target) translation direction tend to be robust and have been replicated in numerous studies, L2-L1 effects are rare and almost invariably smaller. Although the effect has been mostly studied with masked priming paradigms (e.g., Schoonbaert et al., 2009; Wen & van Heuven, 2017), larger L1-L2 priming has been consistently found with unmasked primes as well (e.g., Chen & Ng, 1989; Jin, 1990; Keatley & Gelder, 1992; Keatly et al., 1994; Kiran & Lebel, 2007; Smith et al., 2019). This asymmetry is consistently observed with late bilinguals who are more proficient or dominant in one of their languages (generally the L1). High L2 proficiency (e.g., Nakayama et al., 2016; cf. Dimitropoulou et al., 2011) and early onset of bilingualism (e.g., Duñabeitia et al., 2010; Wang, 2013) have been reported to attenuate or eliminate the asymmetry. While these effects have been interpreted from various models of the bilingual lexicon (see the following text), all must account for why exactly factors relating to bilingual experience (e.g., language use, L2 proficiency) should strongly moderate the effect. A further complication is that relative language exposure/use and L2 proficiency seem to be correlated, making it difficult to tease apart their effects. The goal of the present study is to contribute useful data in this respect, by examining late bilinguals matched for L2 proficiency but differing in their amount of active L2 use.

To the previously mentioned end, we tested 200 highly proficient Spanish–English sequential bilinguals split into two groups equally, differing only in their societal language (L2 immersed vs. nonimmersed). Crucially, immersion proxied not only exposure to the L2 but also active use of that language, as confirmed by the significantly different mean scores of both groups in a linguistic background questionnaire (see the following text). We created a set of 314 noncognate translation pairs. In comparison to previous studies of this type, this generated a large number of observations, to which we applied a conservative analysis. In doing so, we answer recent calls for sufficiently powered studies in bilingualism (see Brysbaert, 2019, 2021; Brysbaert & Stevens, 2018). Following from current theories as described previously, we predicted that immersion—proxying the amount of active L2 exposure/use—would modulate priming effects, especially in the L2-L1 direction. As a result, participants with L2 English as their societal language were expected to show a larger advantage in related trials (i.e., those with L2 translation primes) over control trials as compared to the non-immersed group.

Models of bilingual lexical-semantic representation

Although several comprehensive theories have been advanced (see Kroll & Ma, 2017, for review), the two most prominent models of bilingual lexical-semantic processing are arguably the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994; Kroll et al., 2010) and Multilink (Dijkstra et al., 2019), which aims to integrate the tenets of the RHM and the Bilingual Interactive Activation model(s) (BIA/BIA+; Dijkstra & van Heuven, 2002).

The RHM is a developmental model proposing qualitative differences in the way L1 and L2 words are represented and connected. L1 words have direct and robust links to the conceptual features that make up their meanings. For L2 words, however, lexical-semantic connections are weaker, at least at low proficiency. The RHM proposes that, over development, the bilingual lexicon bridges L2 lexeme-concept connections through L1 lexemes, which have more robust access to the conceptual level. As proficiency increases, L2 lexeme-concept links become stronger, granting direct, independent access from L2 words to concepts, and vice versa, without L1 mediation. Although the RHM was originally proposed to account for asymmetries in word production (see Brysbaert & Duyck, 2010, and Kroll et al., 2010, for discussion), it has been widely discussed in the word recognition literature, including translation priming. The RHM can predict the priming asymmetry only if we assume that translation priming is largely a semantic effect (e.g., Schoonbaert et al., 2009; Xia & Andrews, 2015) and if there are asymmetric connections between the L2 lexical forms and the semantic store, with the route from meaning to form being stronger. Then, recognition of the L1 word would activate the shared conceptual node, which in turn would preactivate the L2 word. Bilinguals with lower L2 proficiency have L2 weaker lexical-semantic connections. In translation priming experiments, this means that L2 primes cannot (sufficiently) stimulate conceptual features shared with the L1 target, which results in a weak or no observable preactivation advantage with respect to an unrelated control word. The contrary is expected in the L1-L2 direction. L1 primes activate shared conceptual nodes, and these in turn preactivate their L2 target counterparts. Priming effects in both directions should gradually become more symmetrical with increased L2 proficiency or L2 use, which are expected to reinforce L2 lexical-semantic connections.

Multilink, developed within a localist-connectionist framework, is a comprehensive computational model of word recognition and production. For Multilink, most of the differences between L1 and L2 word processing can be accounted for by an intrinsic property of lexical representations, independent of their language membership: their resting level activation (RLA). RLA is conceptualized as a word's baseline activation, from which task-related activation can push the lexical item over a given selection threshold. The model assumes that RLA is not static over time, and largely depends on subjective word frequency, defined as the speaker-specific frequency of each word (i.e., how many times a particular individual has encountered a particular word). Subjective frequency is, of course, not directly observable, but it may be proxied by different measurable factors, such as active language exposure/use or corpus word frequency. Like the RHM, Multilink predicts larger L2-L1 priming with more L2 experience. In this case, experience is assumed to increase the RLA for L2 words, speeding up prime recognition and increasing opportunities to preactivate the L1 target.

In sum, current models of bilingual lexical processing assume that L2 development brings about better connectivity and faster access for L2 words. This should result in less asymmetrical translation priming patterns between forward (L1-L2) and backward (L2-L1) translation directions. Whether patterns of L2 task performance can be faithfully captured, in this domain, through variables such as L2 proficiency or L2 active exposure/use is an open question.

L2 proficiency and L2 use in translation priming studies

Attempts have been made to assess the influence of L2 proficiency on translation priming effects. Dimitropoulou et al. (2011) tested Greek–English bilinguals in three groups with varying L2 proficiency (i.e., low, intermediate, high). Priming patterns in the three groups did not significantly differ, leading to the conclusion that L2 proficiency was not a deterministic factor explaining the asymmetry. In contrast, in a series of experiments, Nakayama and colleagues reported a major modulation of L2-L1 priming by L2 proficiency. In Nakayama et al. (2016), significant L2-L1 priming effects were obtained in two experiments with highly proficient Japanese–English bilinguals (TOEIC mean score [out of 990]: 872 and 917). It is worth noting that in one of the experiments in Nakayama et al. (2016), the same stimuli as in Experiment 2B of Nakayama et al. (2013) were used. Notably, these stimuli had previously failed to show significant L2-L1 priming at lower proficiency (TOEIC mean score: 740). A third experiment in Nakayama et al. (2016) with lower-proficiency subjects (mean TOEIC: 710) replicated the results in Nakayama et al. (2013, Experiment 2B). Taken together, this suggests that differences in L2 proficiency are a good candidate for explaining the misalignment of results in backward translation priming. That is, the 2016 findings overall seem to indicate that there is a lower bound of relatively high L2 proficiency required for L2-L1 priming.

Few studies have directly examined the role of language experience in noncognate translation priming with late bilinguals. In Experiment 1, Wang (2013) tested English–Chinese bilinguals who were more dominant in their L1, living in a bilingual society like Singapore. In Experiment 2, participants were more balanced bilinguals. Wang reports a priming asymmetry only in Experiment 1, suggesting an effect of dominance on priming effects. However, the cohort of participants in Experiment 2 was highly heterogeneous. For instance, 10 out of 20 subjects were early bilinguals, previously

shown to yield symmetric priming patterns (e.g., Duñabeitia et al., 2010), which might partly explain their results. Perhaps the most compelling evidence for the role of L2 use comes from Zhao et al. (2011). They tested four groups of Chinese–English bilinguals. In two of them, participants were highly proficient in their L2 but differed in whether their societal language was also English (L2-immersed vs. nonimmersed). The results showed that the size of the L2-L1 priming effect increased as a function of the amount of L2 experience. In particular, significant L2-L1 priming was only observed in the immersed, high-proficiency group (but not in a nonimmersed group with similar proficiency). However, the small number of observations potentially compromises their results—they tested 16 participants in the immersed group and employed 32 translation equivalents.

Taken together, these studies paint a mixed picture of the role of experiential factors in bilingual lexical processing and representation. Whereas some studies have suggested a fundamental role of L2 proficiency in the presence or absence of L2-L1 priming, others have failed to replicate these effects. This is true across the full range of L2 proficiency, even and most importantly for our purposes, at high levels of L2 proficiency where one would expect (cumulative) experience to be the most observable, if not testable. As per active L2 use, some findings seem to point toward a relevant involvement of this factor in modulating translation priming, but more research is needed to understand the magnitude of this role and disentangle it from those of L2 proficiency and language dominance.

It should come as no surprise that investigating such intertwined constructs results in a muddled picture. Indeed, the close relationship between proficiency and use is problematic for the study of the bilingual lexicon. Yet, it is hard to conceive why formal knowledge of a language—as L2 proficiency purportedly reflects—would be deterministic in lexical processing if this predictor were not intimately related to other aspects of bilingual experience. L2 proficiency, as an experimental construct, may be masking the contribution of other relevant factors, obscuring our understanding of the processes taking place within the lexical-semantic network. For instance, a bilingual with higher L2 proficiency will almost invariably have more frequent or intensive use of the L2 than someone with lower proficiency. L2 proficiency is a compound construct, necessarily including not only knowledge of a language but also *experience* with that language. Thus, proficiency perhaps introduces a confound in the equation, complicating the ability to accurately estimate the impact of language experience on bilingual lexical processing. In this sense, and despite the attention L2 proficiency has traditionally received, this construct may not be the best approximation to the contribution of language experience in the development of the bilingual lexicon. In contrast, the amount of active, meaningful experiences with the L2 may be more deterministic for dynamic changes in how (L1 and) L2 words are represented and processed.

In another relevant study, Chaouch-Orozco et al. (2021) attempted to disentangle these effects by studying the interaction between L2 proficiency and use in a translation priming experiment. They tested Spanish–English bilinguals with English as their societal language (i.e., L2-immersed) and varying degrees of L2 proficiency and L2 use, both of which were operationalized as continuous variables. Participants' L2 proficiency ranged from upper-intermediate to advanced. The L2 use scores—obtained from a linguistic background questionnaire—ranged from those reflecting equal use of both languages to greater L1 use. Chaouch-Orozco et al. (2021) reported L2-L1 priming effects that were modulated by L2 use only, while L2 proficiency did not affect the priming effects significantly. Thus, the authors concluded that L2 use was a better predictor of L2 lexical processing than standard measures of L2 grammar knowledge.

At 60 participants and 50 word pairs, however, the dataset in Chaouch-Orozco et al. (2021) may have been underpowered to explore complex interactions of the type that the study focused on. Furthermore, the range of relative L1-L2 use was biased toward the L1 side of the scale, despite these being immersed L2 speakers.

The present study improves upon Zhao et al. (2011) and Chaouch-Orozco et al. (2021) with a much larger sample and a more extensive set of word pairs, which guarantee sufficient statistical power to investigate the interactions of interest. In addition, here we introduce a more systematic exploration of L1/L2 use—operationalized through immersion—while factoring out potential effects of L2 proficiency by controlling this factor across participants. The goal is to offer a robust dataset that sheds light on the role of language use in bilingual lexical-semantic processing as reflected by translation priming effects.

Method

Participants

Two hundred Spanish–English sequential bilinguals (see Table 1 for participant characteristics) took part in two translation priming lexical decision tasks (LDT) under overt priming conditions, one experiment per priming direction. Participants were recruited from two different populations. Half of them were L1-immersed, living in Spain; the other half were L2-immersed, living in the United Kingdom. L2 proficiency was controlled across participants to isolate the effect of L2 use and was assessed with the LexTALE test (Lemhöfer & Broersma, 2012), a validated measure of L2 vocabulary and knowledge. A minimum score of 80/100 was required to participate in the study. This threshold was based on Lemhöfer & Broersma's report of LexTALE correlating with the Oxford Quick Placement Test (OQPT; Oxford University Press, University of Cambridge, and Association of Language Testers in Europe, 2001). In particular, 80% correct responses in the OQPT, which corresponds to a CEFR (Common European Framework of Reference for Languages; Council of Europe, 2001) C1 level, corresponded to a LexTALE score of 80.5% in the authors' analyses (Lemhöfer & Broersma, 2012, p. 335). A two-sample t-test showed that the groups differed significantly in their LexTALE score ($t = -44.79$, $p < .001$; see Table 1 for averages), despite small numerical differences in mean and standard deviation. However, further exploration with a parsimonious mixed-effects model showed that the factor, treated continuously across the whole population, did not significantly modulate overall RTs nor priming effects. Moreover, we further inspected a potential effect of proficiency by subsetting the groups to have nonsignificant differences in LexTALE scores between them. We achieved this by removing eight participants in each group ($t = 0.47$, $p = 0.64$). We then ran a new model with this subset, which yielded remarkably similar outcomes to our final model reported in the text that follows. Therefore, the analysis continued as planned.

Table 1. Participant characteristics

	Age (years)	LexTALE	LSBQ	UK length of residence (years)
Spain	26 (4.5; 19–39)	89.7 (5.6; 80–100)	4.6 (3.1; −2.3–11.4)	–
UK	32 (4.9; 22–40)	88.1 (5.0; 80–100)	14.6 (2.9; 6.1–21.6)	6 (3.7; 1–21)

Note: Mean values (standard deviation; range). "LSBQ" column shows composite L2 use score across contexts (home, social, etc.).

Language use information was collected through the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018), which provides a fine-grained, context-dependent, and dynamic measure of relative L1/L2 use. Mean values differed significantly between these groups ($p < .05$), with the immersed group reporting more L2 use. Therefore, the intuition about immersion proxying not only exposure but also active use of the L2 was supported and deemed the critical manipulation of immersion adequate for our empirical purposes. The LexTALE and LSBQ scores were not correlated ($r = –0.11, p < .001$). All participants reported having started to learn English in primary school and never before age six. Only four participants in the Spain-based group reported previous immersion experience, but not within the 12 months before the experiment.

Task order was as follows: first direction of the translation priming LDT – LSBQ – second direction of the LDT. Order of LDT priming direction (L1-L2 or L2-L1) was counterbalanced across participants. Participants were recruited online and compensated with £20 (or the equivalent in euros) for their participation.

Materials

A total of 314 noncognate translation equivalent pairs were used in each translation direction (see Appendix A for the stimuli list and Table 2 for stimuli characteristics). Targets were extracted from a continuum of frequencies and concreteness. Given that the stimuli consisted of translation pairs, we opted for using only the English words' values to avoid employing different norms. Thus, each English word within each pair was given a concreteness value extracted from Brysbaert et al.'s (2014) norms. English word frequencies were obtained from the SUBTLEX$_{UK}$ corpus (van Heuven et al., 2014), whereas Spanish frequencies were extracted from SUBTLEX$_{ESP}$ (Cuetos et al., 2011). Mean values between languages did not differ significantly. Words in both languages were also matched for length and orthographic neighborhood.

To generate "no" trials necessary for lexical decision, 314 pseudowords were created for both translation directions with the Wuggy software (Keuleers & Brysbaert, 2010). These pseudowords matched their word counterparts on length of subsyllabic segments, letter length, transition frequencies, and two out of three segments. The pseudowords were paired with 314 different words that served as their primes. Four lists were created (two for each target language). For each language, one list had half the target words preceded by their translation equivalents and the other half by control primes, whereas the other list inverted these conditions for the same targets. Control primes were created by scrambling the related primes in the other list. We ensured that control pairs remained orthographically and semantically unrelated. The words in each list were matched for frequency, word length, and orthographic neighborhood. Each list began with 16 practice items.

Table 2. Stimuli characteristics

	Spanish	English
Frequency	4.3 (0.7; 2.5–6.1)	4.5 (0.6; 2.6–6.3)
Concreteness		4.0 (1.01; 1.19–5.0)
Length	5.5 (1.4; 3–8)	5.5 (1.4; 3–8)

Note: Mean values (standard deviation and ranges). Concreteness values for Spanish words are assumed to approximate that of their English translations.

To ensure that participants knew the English stimuli, they completed a picture-word matching task with the concrete stimuli, where they were presented with pictures depicting objects accompanied by two words in English: the correct picture name and a distractor. The lowest individual accuracy score was 89%. Only five words received responses with an accuracy lower than 80% overall. These were removed from the dataset. Knowledge of the abstract word pairs, which have much lower imageability, was evaluated through a translation recognition task. Five participants showed an accuracy below 85% and were removed from the dataset. Thirty-nine (abstract) words showed an accuracy below 80% and were removed from the dataset. In all cases, these tasks were conducted after the LDTs.

Procedure

All experiments were created and presented online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Given the limits that online presentation poses to the experimenter's role on controlling participants' performance, data quality control, and exclusion criteria were implemented to ensure participants' constant attention during the experimental tasks. First, there was a time limit (95 minutes—on average, a session took 60–70 minutes to complete) to finish each session. Attention checks were implemented, and their presentation was pseudorandomized (i.e., within blocks of 20 trials) so participants could not know when they would appear. Participants had to press "B" on the keyboard within 2 seconds from the instructions' onset. Participants failing to pass less than 95% of these checks were excluded from the study. We also examined their responses to ensure they were not blatantly random. Failing to meet these criteria resulted in exclusion from the study. Twenty-five participants out of 225 failed to meet these criteria, resulting in the final 200 participants whose data were analyzed.

Each trial began with a fixation cross on the centre of the screen (500 ms), followed by the prime in lowercase letters (200 ms) and the target in upper case letters, which remained on the screen until the subject provided a response. Right-handed participants had to press "0" on the keyboard to indicate YES, and "1" for NO. This order was inverted for left-handed participants. They were asked to respond as fast and as accurately as possible. Each task (priming direction) was further divided into 15 blocks of approximately 40 trials. Participants were given the chance to rest between these 40-trial blocks. They were asked to avoid any distractions during the session and to ensure their vision was corrected if needed. No participant completed the sessions at night, and they were encouraged not to participate when they felt tired. In sum, we paid special attention to simulating, to the extent possible, lab testing conditions.

Data analysis

Data and analysis code can be found in the first author's OSF repository (https://osf.io/yx6hw/). Besides the five participants excluded due to low accuracy on the translation recognition task, two more participants were removed for the same reason after inspecting the LDT data. The analysis continued with the remaining 193 participants (96 in the Spain group, 97 in the UK group) and 270 word pairs. Incorrect responses and pseudoword trials, as well as RTs below 200 ms (4 in total) and above 5,000 ms (80 in total), were removed (see Baayen and Milin, 2010). We transformed the latencies to obtain inverse Gaussian, log-normal, and BoxCox distributions. After visual

inspection (Q-Q plots) and Shapiro–Wilk tests, the inverse Gaussian distribution was selected to perform the analysis, as it provided a better correction of the skewness (inverse Gaussian: $p = .42$; BoxCox: $p = .33$; log-normal: $p = .08$). Sum contrasts were employed for categorical variables, and all continuous independent variables were scaled, centred, and converted to z units.

Error rates and response times were analyzed employing (generalized) linear mixed-effects models (Baayen et al., 2008) in *R* (version 3.6.1; R Core Team, 2021) with the *lme4* package (Bates et al., 2015). We followed Scandola and Tidoni (2021) for an optimal trade-off between maximal random structure specification, convergence, and computational power in random-effects specification and model selection. Scandola and Tidoni show that computational times are linked with convergence and overfitting issues. Consequently, in cases of high model complexity—as with our models—and relatively low computational power (standard lab equipment), they recommend employing Complex Random Intercepts (CRI). In a full-CRI model, (complex) random slopes (with many interactions) are replaced by different random intercepts for each grouping factor. The method minimizes Type-I error risk. For each analysis, we fitted a maximal model. If the model did not converge, we removed the CRI that explained the least variance and tried again until a maximal model converged. Further criticism was applied to this convergent model, including checking assumptions (e.g., normality of residuals' distribution, homoscedasticity) and removing observations with absolute standardized residuals above 2.5 *SD* (Baayen & Milin, 2010). Thus, we employed a maximal model approach, as suggested by Barr et al. (2013; but see also Brauer & Curtin, 2018; Scandola & Tidoni, 2021), because (i) it offered an optimal trade-off between Type-I and II errors (Scandola & Tidoni, 2021:13), and (ii) given our large number of observations, a more parsimonious method (Matuschek et al., 2017) did not seem as necessary.

We included main effects and interactions of interest as fixed effects in the analyses for both accuracy and RTs (Brauer & Curtin, 2018). The grouping factors were language (i.e., translation direction), prime type (related vs. control), group (immersed vs. nonimmersed), and their interactions; that is, the factors that varied within subjects, primes, and targets (Brauer & Curtin, 2018).[1] Thus, a full-CRI structure was specified with random intercepts for subjects, primes, and targets for each grouping factor.

Results

Response times

Table 3 summarizes RTs and error rates in all conditions. Appendix B provides the summary of the final model. Full specification and outcomes of other models can be found in the first author's OSF repository. In the main analysis of response latencies, the final model revealed a significant effect of language ($\beta = -0.05$, $t = -3.43$, $p < .001$),

---

[1]Note that, in cross-language priming studies, stimuli (primes and targets) are randomly sampled from two populations (i.e., languages). Therefore, taking advantage of the mixed-effects models' capabilities, random intercepts for primes and targets can better model the variance arising from random population sampling in each language. Notably, the suitability of this approach was confirmed by comparing the goodness of fit between three parsimonious models. Model 1 included random intercepts for primes and targets; model 2 included random intercepts for items (prime and target pairs); model 3 included random intercepts for targets and a random slope for targets within primes. An ANOVA test showed that model 1 offered the best fit to the data ($p < .001$).

Table 3. Mean response times (RTs, in milliseconds; standard errors), error rates (%), and priming effects (in milliseconds)

	Spain group				
	Related		Control		
	RT	Error rate	RT	Error rate	Priming
L1 to L2	643 (1.7)	1.5	731 (2.4)	3.3	88*
L2 to L1	625 (1.5)	0.9	696 (1.9)	2.1	71*
	UK group				
	Related		Control		
	RT	Error rate	RT	Error rate	Priming
L1 to L2	667 (1.7)	1.2	750 (2.3)	2.8	83*
L2 to L1	648 (1.6)	0.8	712 (2.0)	1.8	64*

*p < .05.

indicating that responses to Spanish targets were faster. The main effect of prime type was also significant ($\beta = -0.15$, $t = -26.93$, $p < .001$), revealing overall faster RTs to related trials. However, a significant interaction between language and prime type ($\beta = 0.03$, $t = 4.19$, $p < .001$) showed that priming effects were larger in the L1-L2 direction. The interaction between group and prime type was significant ($\beta = 0.02$, $t = 2.48$, $p = .014$), indicating that priming effects were larger for the nonimmersed participants in both translation directions. Finally, the three-way interaction between language, prime type, and group was nonsignificant ($\beta = 0.002$, $t = 0.28$, $p = .78$), suggesting no differential role of immersion between translation directions.

To follow up on this null effect of immersion as a modulator of the priming asymmetry, we conducted further analyses. First, given that our stimuli consisted of words from all frequencies and from the whole concreteness spectrum, we controlled for the effect of both factors by running two separate analyses with interactions with the factors of interest as well as frequency and concreteness specified in the models. The potential effects of prime and target frequency were analysed in separate models to avoid multicollinearity issues (because the frequencies of translation equivalents tend to be correlated). Results in all these models revealed the same effects as in the main model. That is, there were significant effects of language and prime type, as well as significant interactions between these two factors. Further, in all the models, the two-way interaction between group and prime type was significant (all $ps < .02$), and the three-way interaction between group, prime type, and language was nonsignificant. Finally, complex four- and five-way interactions involving group and frequency or concreteness were observed, although none of them substantially changed the findings of the main analysis with respect to immersion.

However, to further inspect these interactions, we conducted separate analyses with subsets of the data. First, we looked at concreteness. The new models with subsets containing only concrete or abstract words, revealed the same pattern of results (i.e., a significant interaction between prime type and group; $ps < .001$). Then, we ran four new models with subsets splitting the data by prime and target frequency. The results showed that, with low-frequency stimuli, the small interaction between prime type and group disappeared. With high-frequency stimuli, however, the interactions between prime type and group were significant in the two models ($ps < .001$). Therefore, this

result suggests that the significant interaction between prime type and condition is mainly driven by the high-frequency stimuli.

Moreover, although our main analysis focused on the effect of immersion, individual variation in language experience could also impact the priming effects. To investigate this possibility, we conducted independent analyses on the participants of each group, replacing the categorical variable group with the continuous LSBQ score. The results of these analyses showed that the LSBQ score did not modulate the priming patterns. Finally, we wondered whether our main finding would be replicated if the LSBQ score was employed instead of the group variable in a model with all the participants' data. Notably, the results emerging from this new model mimicked those of the main model. We observed a significant interaction between prime type and LSBQ score ($\beta = 0.01$, $t = 2.26$, $p = .025$), indicating that priming was larger for the participants who reported using more the L1 (i.e., the nonimmersed ones).

Accuracy analysis

Accuracy was dummy-coded as 1 (correct) or 0 (incorrect). Generalized linear mixed-effects models with a binomial family were fit to the error data. Significant effects of language ($\beta = 0.36$, $z = 2.68$, $p < .001$), prime type ($\beta = 0.80$, $z = 9.61$, $p < .001$), and group ($\beta = 0.26$, $z = 2.01$, $p < .05$) were observed. This indicated that participants were more accurate when responding to Spanish targets, as well as in related trials. In addition, participants in the UK group were overall more accurate. The interaction between language and prime type was not significant, suggesting no priming asymmetry across tasks for accuracy. Note that accuracy analyses tend to be less sensitive to these experimental manipulations and, as usual in the relevant literature, were not central to the current study.

Discussion

We have presented data from a study investigating the effect of active L2 use on bilingual lexical representation and processing, employing two lexical decision tasks with noncognate translation priming. We tested highly proficient L1 Spanish-L2 English late bilinguals in two groups that differed in their societal language: L2-immersed versus nonimmersed. We ensured that the immersion factor accurately proxied for differences in L2 use between the groups by measuring this more precisely through a detailed questionnaire (LSBQ). A significant difference in LSBQ score between the groups suggests that the categorical split is justifiable in our sample. Furthermore, we controlled L2 proficiency across groups, which allowed us to isolate the potential effects of immersion/L2 use.

In line with much of the literature (see Wen & van Heuven, 2017) and despite our participants' high proficiency, we observed an asymmetry in priming effects between translation directions, with L2-L1 priming being significantly less pronounced. This finding aligns with the significant L2-L1 priming effects with high-proficiency Japanese–English bilinguals reported by Nakayama and colleagues (2016).

More important for our study is the impact of immersion. According to the models of bilingual lexical representation and processing we presented in the preceding text, bilingual experience-related factors such as immersion (and what it proxies, i.e., active exposure to and use of the L2) should play a prominent role in the representation and

functioning of the lexicon. In the context of translation priming, increased exposure to and active use of the L2 should lead to larger priming effects in the L2-L1 direction.

The Revised Hierarchical Model states that the links between L2 lexical representations and their meanings are relatively weak at low proficiency, in contrast with the fully developed L1 lexical-semantic connections. These architectural dissimilarities can explain the priming asymmetry, as long as translation priming is assumed to take place through coactivation of shared conceptual features between translation equivalents (and not through lexeme-to-lexeme links). At higher proficiencies, or with increased L2 use, stronger L2 lexical-semantic connections would ensure more direct access to the conceptual store for L2 translation primes. This would result in enhanced semantic activation between translation pairs and larger L2-L1 priming effects. In Multilink, resting level activation (RLA) is assumed to be sensitive to changes in the amount of use of the language(s), ultimately a proxy for how often a given lexical item may be retrieved. Hence, the more often a second language is used, the higher RLA should be for L2 words, which would translate into faster processing. With higher RLA, L2 related primes should be recognized faster and prove more effective in preactivating the L1 target, leading to larger L2-L1 priming effects.

Our results clearly challenge these hypotheses. First, although immersion (or more L2 use) did have a significant effect on translation priming effects, this did not result in the expected priming patterns. L2-L1 priming was in fact larger for those participants with less L2 use, contradicting our main hypothesis. Therefore, we cannot conclude that more active use of the L2 led our participants to benefit further from the presence of L2 primes. Analyses controlling for word frequency and concreteness further confirmed these results. Moreover, this modulation of priming effects by immersion was true of both translation directions, which prevents a straightforward interpretation by the RHM or Multilink. For both models, relative L1/L2 use should have a bearing on how fast prime and target words are retrieved, and this should in turn influence priming effects. In this sense, while larger L1-L2 priming for the Spain group could be explained by their comparatively higher L1 use, the same account fails to predict larger L2-L1 priming for the same group.

Crucially, other aspects of our data can offer some insights on the nature of this effect. A relevant difference between the groups is that the UK-based participants were overall more accurate. While there are several ways to interpret this, one may argue that the Spain group was thus slightly less confident in their responses (especially because they were not faster overall, which may have suggested a speed-accuracy trade-off). This makes longer trials more susceptible of showing a difference between the groups, as priming effects in standard (unmasked) priming paradigms are known to occur early or late in the RT distribution, potentially underlain by different mechanisms (e.g., Balota et al., 2008). A look at priming effects for each group across quantiles of the RT distribution suggests that this account might be on the right track. In Figure 1, we can see how participants in the Spain group obtained comparatively larger priming effects toward the highest quantiles, that is, in longer trials. Further left in the distribution, the two groups show priming effects that are similar in magnitude.[2] Furthermore, motivated by our result in the main analysis, we inspected the relation between word

---

[2]Note that Figure 1 reflects overall priming effects irrespective of language direction, in line with the significant two-way interaction observed in the RT analysis (i.e., prime type by group). We further visually inspected this interaction in each translation direction. The effect was comparable in both directions and followed the pattern showed when priming effects were plotted altogether (Figure 1), confirming the adequacy of plotting the overall priming effects in Figure 1.
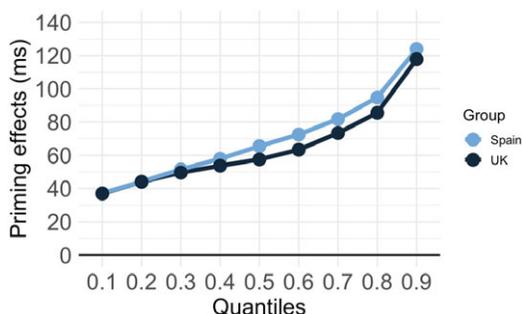
Figure 1. Plot of overall priming effects across quantiles for the two groups. Each point represents a quantile. Note that nine quantiles, 0.1 to 0.9, were employed for smoother curves.

frequency and the immersion effect. Visual examination suggested that immersion seemed not to impact priming effects at lower quantiles irrespective of frequency. When differences appeared at higher quantiles, they seemed to be driven by high-frequency words. To further inspect this effect, we run omnibus ANOVAs on the mean RTs for each participant in each quantile with prime type, group, frequency, and quantile and with each subset. The analysis showed that the larger priming effect in higher quantiles for the Spain group was not significant in any subset. That is, these analyses did not show that frequency significantly modulated the interaction between prime type, group, and quantile. However, more direct research would be needed to determine these patterns, as these priming effects were relatively small and distribu-tional analyses typically need larger numbers of observations to detect significant effects (Balota et al., 2008).

In the literature on semantic priming with monolingual speakers, priming effects at higher quantiles have been associated to processes where the prime-target relationship is checked in memory before providing a lexical decision (e.g., McKoon & Ratcliff, 1998; Thomas et al., 2012). That is, the more evidence on the prime-target translation relationship participants accumulate over time, the greater the (priming) benefit. In this light, both groups seem to have comparable priming effects of the early, more automatic type (a headstart effect through preactivation; Forster et al., 2003), but differ in the amount of priming caused by a prime-target compound cue. Note that inde-pendently from which model's tenets are on the right track, both the RHM's and Multilink's mechanisms are based on spreading activation, which, crucially, would take place early in the trial. Therefore, we can safely conclude, that, at least with regard to the implications for the models discussed, our two groups elicited similar priming effects in the two translation directions and behave similarly. While this is a more fine-grained description of our results, the ultimate reasons behind these different patterns of priming effects across groups remain unclear.

Overall, our results do not support an effect of immersion in the size of L2-L1 priming effects. This finding contrasts with the more deterministic role of this factor reported by Chaouch-Orozco et al. (2021) and Zhao et al. (2011). In both cases, L2 use clearly modulated L2-L1 priming effects. Discrepancies between the present study and Zhao et al.'s results are particularly intriguing. They observed that L2-L1 priming was significant in a group of highly proficient L2-immersed bilinguals, but not in a similarly proficient nonimmersed group, which is essentially at odds with our current results.

Differences in power (100 subjects judging 314 items vs. 16 subjects judging 32 items) might explain at least some of these divergencies.

One possible explanation for the minimal impact of L2 use on L2-L1 priming here is that very high-proficiency marks some upper boundary for the effects of L2 use, so that its effect becomes negligible past some threshold of proficiency. A weaker version of this hypothesis would be that experience-driven changes in the lexicon do not occur at the same pace throughout development but slow down (i.e., require more experience to maintain the same rate) toward the higher end of the proficiency spectrum. This account reconciles two aspects of our data: the absence of a strong immersion/L2 use effect in our high-proficiency sample and the fact that we observe a difference in the magnitude of priming effects between translation directions (i.e., a priming asymmetry), which suggests that RLA/semantic connectivity for L2 words has not reached its maximum. Further research is needed to specifically test this hypothesis.

Moving forward, future efforts could target highly proficient immersed bilinguals with a broader range of immersion time, heritage speakers with different code-switching profiles, professional interpreters, or passive bilinguals, among others. Each bilingual profile offers unique opportunities to disentangle the roles of different experiential factors and, with it, obtain an ever-so-slightly clearer picture of the bilingual lexicon's architecture.

Data Availability Statement. The experiment in this article earned Open Materials and Open Data badges for transparent practices. The materials and data are available at https://osf.io/yx6hw/.

References

Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*, 250–263.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407.

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

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

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*, 495–523.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

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

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*, 389.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*, 16.

Brysbaert, M. (2021). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, *24*, 813–818.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and Cognition*, *13*, 359–371.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*, 9.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Chaouch-Orozco, A., Alonso, J. G., & Rothman, J. (2021). Individual differences in bilingual word recognition: The role of experiential factors and word frequency in cross-language lexical priming. *Applied Psycholinguistics*, *42*, 447–474.

Chen, H. C., & Ng, M. L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory & Cognition*, *17*, 454–462.

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

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2011). SUBTLEX-ESP: frecuencias de las palabras espanolas basadas en los subtitulos de las peliculas. *Psicológica*, *32*, 133–144.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197.

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

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Two words, one meaning: Evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology*, *2*, 188.

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

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 663.

Forster, K. I., Mohan, K., Hector, J., Kinoshita, S., & Lupker, S. J. (2003). The mechanics of masked priming. In *Masked priming: The state of the art* (pp. 3–37). Psychology Press.

Jin, Y. S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills*, *70*, 1139–1154.

Keatley, C., & Gelder, B. D. (1992). The bilingual primed lexical decision task: Cross-language priming disappears with speeded responses. *European Journal of Cognitive Psychology*, *4*, 273–292.

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

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633.

Kiran, S., & Lebel, K. R. (2007). Crosslinguistic semantic and translation priming in normal bilingual individuals and bilingual aphasia. *Clinical Linguistics & Phonetics*, *21*, 277–303.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*, 119.

Kroll, J. F., & Ma, F. (2017). The bilingual lexicon. In E. M. Fernández & H. S. Cairns (Eds.), *The handbook of psycholinguistics* (pp. 294–319). Wiley.

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

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*, 373.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, *49*, 25–42.

Meade, G., Midgley, K. J., Dijkstra, T., & Holcomb, P. J. (2017). Cross-language neighborhood effects in learners indicative of an integrated lexicon. *Journal of Cognitive Neuroscience*, *30*, 70–85.

Nakayama, M., Ida, K., & Lupker, S. J. (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*, 1001–1022.

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

Oxford University Press, University of Cambridge, & Association of Language Testers in Europe. (2001). *Quick placement test: Paper and pen test*. Oxford University Press.

R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/

Sánchez-Casas, R., & García-Albea, J. E. (2005). The representation of cognate and noncognate words in bilingual memory. In J. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 226–250). Oxford University Press.

Scandola, M., & Tidoni, E. (2021, February 8). The development of a standard procedure for the optimal reliability-feasibility trade-off in multilevel linear models analyses in psychology and neuroscience. PsyArXiv preprint. https://doi.org/10.31234/osf.io/kfhgv

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

Smith, Y., Walters, J., & Prior, A. (2019). Target accessibility contributes to asymmetric priming in translation and cross-language semantic priming in unbalanced bilinguals. *Bilingualism: Language and Cognition*, *22*, 157–176.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*, 12530–12535.

Thomas, M. A., Neely, J. H., & O'Connor, P. (2012). When word identification gets tough, retrospective semantic processing comes to the rescue. *Journal of Memory and Language*, *66*, 623–643.

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology*, *3*, 174.

Van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*, 458–483.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190.

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

Wen, Y., & van Heuven, W. J. (2017). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Psychonomic Bulletin & Review*, *24*, 879–886.

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

Zhao, X., Li, P., Liu, Y., Fang, X., & Shu, H. (2011). Cross-language priming in Chinese-English bilinguals with different second language proficiency levels. In L. Carlson, C. Hölscher & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 801–806). Austin, TX: Cognitive Science Society.

Appendix

Appendix A. Complete list of stimuli

Table A1. Prime and target words and pseudowords

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)
ley	law	ler	baw
año	year	abe	croding
amor	love	anur	wixth
lío	mess	túo	pess
odio	hate	edia	hamp
daño	harm	gazo	hask
tía	aunt	lúa	aste
jefe	boss	joñe	bomp
lado	side	bano	sipe
frío	chill	flúo	chall
noche	night	gorre	nimes
gemelo	twin	necilo	knin
fiesta	party	deusta	manty
lástima	pity	víntima	moty
mezcla	mix	mechra	mox
tamaño	size	vanazo	rize
viaje	trip	guije	spip
enfado	anger	envate	arver
sueño	dream	ruebo	bleam
manojo	bunch	sacozo	budes
suciedad	dirt	suriadal	dort
prisa	haste	briga	haits
mitad	half	pital	harf
vida	life	nila	libe
lugar	place	rumar	plawn
brillo	glow	chirro	prash
rugido	roar	dusedo	rour
olor	smell	ecor	skell
trabajo	work	pradaño	wolt
ejército	army	exáncito	angy
ajedrez	chess	ajegrua	chend
multitud	crowd	ductitul	croif
fin	end	fen	ews
hecho	fact	horro	farp
ajetreo	hustle	ajebleo	huggle
mañana	morning	pafala	perning
sonrisa	smile	sarcisa	smale
choque	clash	chehue	spash
presión	strain	cremión	strail
huelga	strike	hesiga	strind
tregua	truce	chegué	trurn
nivel	level	nibal	tuvel
rebote	bounce	depose	bouths
soborno	bribe	soponlo	clibe
hito	feat	lico	fout
caída	plunge	meína	plurse
oración	prayer	osacuas	praire
sequía	drought	señida	prought
subida	raise	duveda	ralps
alcance	scope	armange	scode

(*Continued*)

Table A1. (*Continued*)

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)
capítulo	chapter	camónumo	shalter
verdad	truth	hordad	sluth
gripe	flu	fribe	spu
tarea	task	facia	tage
dueño	owner	huejo	uffer
invierno	winter	infiarmo	warter
juventud	youth	juvendid	yeath
ansia	craving	ancio	ac
zumbido	buzz	zórnido	burf
encanto	charm	endisto	chawl
diseño	design	simejo	denifs
aumento	surge	eisanto	fleed
resaca	hangover	demasa	foleover
hambre	hunger	fimbre	henser
broma	prank	frema	outsheek
revés	setback	repís	pedback
ráfaga	flurry	díjaga	plerry
elogio	praise	ecobia	pralps
perfil	profile	ponfil	progale
alivio	relief	acegio	reroof
guión	script	quión	scrimb
reunión	meeting	teulión	soating
tristeza	sadness	prosteña	sumness
tiempo	weather	tienzo	weinter
creencia	belief	pleangia	berieu
cambio	change	campia	chathe
engaño	deceit	esvazo	decoal
deleite	delight	deciose	detisse
ayuda	help	aelta	doduanty
vistazo	glimpse	tastaño	glitzed
invitado	guest	osminado	gurnt
locura	madness	nevura	macless
amigo	friend	arezo	pove
búsqueda	pursuit	térqueda	purgoot
altura	height	etrura	reight
regaño	scolding	devaho	scumping
sentido	sense	dindido	serbs
escasez	shortage	escalot	shuntage
rasgo	trait	resno	trarf
lesión	injury	reción	uncupy
vacío	void	tadúo	voir
anchura	width	enfrora	yeam
coartada	alibi	coínnado	atipo
placer	pleasure	grader	clealure
frialdad	coldness	crieldad	cortless
deceso	demise	demico	denite
clamor	outcry	dragor	eattry
ensayo	essay	encaua	empay
vuelo	flight	guero	flinge
bondad	goodness	rondal	gan
fantasma	ghost	vintalma	ghoms
puñado	handful	tufalo	handcal
farsa	hoax	garga	hoaf
ingesta	intake	insonta	intive

(*Continued*)

Table A1. (*Continued*)

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)
tumulto	mayhem	turuzno	magwem
tropiezo	misstep	trobuejo	misgrap
hipoteca	mortgage	bitoneta	modlfage
popurrí	medley	poduchá	mutley
ruido	noise	tiado	noits
nómina	payroll	jánina	paydods
retirada	retreat	decenada	replout
milagro	miracle	sicaclo	silatre
amenaza	threat	alacava	squeat
pandilla	gang	serdilla	gank
éxito	success	áhilo	suybess
susurro	whisper	pucucho	swismer
hambruna	famine	fimpluna	taline
belleza	beauty	tebreja	wemming
premio	award	prodio	abail
activo	asset	atnizo	ampet
primo	cousin	brico	coomin
derrota	defeat	rechola	defoul
trama	plot	brada	drot
esfuerzo	effort	esluenjo	effall
diosa	goddess	gaisa	gommess
pista	clue	minta	grue
cosecha	harvest	cocigra	hanrest
prueba	proof	prieva	preaf
raza	breed	laba	bried
entierro	burial	endiatro	felial
pereza	laziness	semepa	fowiness
jugada	gamble	nudana	gattle
chisme	gossip	chosbe	gostup
culpa	guilt	celba	goult
salud	health	talid	heanse
demencia	insanity	decensio	inmitaty
riesgo	jeopardy	reusno	junkainy
salto	jump	taldo	junt
medida	measure	sereda	moolure
retrato	portrait	degraso	porstaim
reino	realm	toino	reapt
asunto	affair	acosto	shidecut
alma	soul	amba	soal
sigilo	stealth	dipito	steanse
calor	warmth	macor	warque
consulta	query	conguara	whety
consejo	advice	conciño	adhace
negocio	business	nedomia	bereness
cielo	heaven	miero	hooden
brote	outbreak	grete	leep
molestia	nuisance	polintia	nuisudes
atajo	shortcut	acage	rurge
parecido	likeness	manalido	takeness
llegada	arrival	plecana	ancipal
cierre	closure	cuelle	closand
gente	people	garte	daople
muerte	death	miante	deeth
entrega	delivery	enchega	demitoly

(*Continued*)

Table A1. (*Continued*)

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)
caza	hunt	mafa	hund
hoy	today	hol	logay
deber	duty	vejer	rury
fuente	source	hiente	shouch
ocio	leisure	osia	toesure
apoyo	support	adoez	aud
siglo	century	riplo	cerrugy
valor	courage	galir	cootage
marca	brand	canca	crand
ingreso	income	incrito	incert
olvido	oblivion	utrido	ospetion
lema	motto	loda	petto
cuenta	account	coista	acceine
cena	dinner	pona	panner
sed	thirst	sey	thisle
vista	view	hesta	yiew
boda	wedding	roga	geakness
mal	evil	ral	eryl
resto	leftover	sento	loleoper
mes	month	med	mopse
lucha	struggle	ligra	struttle
pena	sorrow	leta	dorrop
fallo	failure	bacho	forture
peso	weight	ceto	wought
carga	burden	marla	bermen
ira	wrath	ula	wrass
nana	lullaby	gara	funkavy
risa	laughter	sina	linchter
paso	step	laro	stup
moda	fashion	cona	tushion
fe	faith	ci	faire
baja	casualty	dapa	jush
capa	layer	mava	tager
nariz	nose	jaréz	nuse
oso	bear	eno	lear
sol	sun	sod	sep
abrigo	coat	allibo	roat
cisne	swan	cisde	drap
tacón	heel	tagón	hool
peine	comb	piole	cacs
puño	fist	pubo	fims
percha	hanger	pastra	hanker
rey	king	rez	kint
luna	moon	lura	moop
bruja	witch	bruba	datch
jabón	soap	jajón	sein
río	river	lúo	sover
clavo	nail	llavo	norn
reloj	watch	telop	wamps
leche	milk	reche	misk
búho	owl	lího	orm
nieve	snow	jaipe	whew
cadena	chain	cavena	chail
uva	grape	ubo	grame
charco	puddle	trarno	moddle

Table A1. (*Continued*)

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)
nido	nest	vimo	nugs
foca	seal	feda	sool
bolsillo	pocket	bolpillo	sicket
muro	wall	cino	wams
pájaro	bird	májaro	birk
cabra	goat	mabra	goot
cuerno	horn	cuergo	hoil
miel	honey	mial	honem
cuello	neck	ciello	nids
horno	oven	forgo	oken
cuchara	spoon	cucrara	sleen
dinero	money	dicero	soney
camarero	waiter	cararero	waider
cartera	wallet	cancera	gallet
guante	glove	guaste	glink
mofeta	skunk	poñeta	glone
traje	suit	chave	goam
mechero	lighter	petrero	latcher
bufanda	scarf	tufanda	scalf
oveja	sheep	oveza	shoon
caracol	snail	caracod	snain
ballena	whale	lachena	snite
flecha	arrow	plecha	andow
barba	beard	barla	bearn
esquina	corner	encaina	cark
payaso	clown	payuso	clode
tobillo	ankle	bodillo	arwhe
manzana	apple	marzana	apste
cajón	drawer	mabón	draxer
ojo	eye	ezo	eys
araña	spider	ecafa	flider
mujer	woman	muver	goman
imán	magnet	idín	madnet
pezón	nipple	mejón	napple
cuadro	painting	cultro	paunting
llave	key	claje	tox
cebolla	onion	cegolla	unoon
toro	bull	relo	bams
silla	chair	rilla	chasp
vestido	dress	tentido	cluss
ataúd	coffin	atail	coddin
puente	bridge	peinte	crorks
puerta	door	pierta	foor
langosta	lobster	vangosta	losster
cama	bed	cafo	ped
gatillo	trigger	narillo	prigger
calabaza	pumpkin	calabava	pullcin
lápiz	pencil	lópiz	puncil
trigo	wheat	chiso	sneat
paraguas	umbrella	paraguar	umblella
ala	wing	ado	wint
aguacate	avocado	abuacate	axocado
pepino	cucumber	sedeno	cucolser
erizo	hedgehog	elepo	hedgerig
avestruz	ostrich	avestriz	ostript

(*Continued*)

Table A1. (*Continued*)

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)
conejo	rabbit	corejo	rabbot
escudo	shield	espudo	shoard
ventana	window	tontana	windig
iglesia	church	inhesia	chorth
dedo	finger	hodo	fanger
bosque	forest	burque	hunest
hada	fairy	gapa	lamey
tijeras	scissors	higiras	plundors
ducha	shower	gurra	shamer
falda	skirt	dalda	skipe
ardilla	squirrel	arnilla	sprirrul
espejo	mirror	escezo	suppar
sombra	shadow	sostra	bradow
aguja	needle	aduza	neeble
hombro	shoulder	fomplo	choolder
pulgar	thumb	pulmar	thurl
armario	wardrobe	asparia	wardrolk
camarera	waitress	calarera	waseress
caja	box	paña	boc
mano	hand	mado	habs
pelo	hair	pono	hact
mesa	table	moma	lable
pulpo	octopus	pumzo	octovis
cereza	cherry	ceseja	shippy
morsa	walrus	mersa	wamnus
pato	duck	labo	dack
huevo	egg	duejo	erv
burro	donkey	lullo	monrey
queso	cheese	quido	sheese
goma	rubber	gedas	dubber
reina	queen	toina	snoup
maleta	suitcase	parita	moanpind
valla	fence	harra	ferks
pavo	turkey	mapo	furkey
lobo	wolf	lopo	woit
boca	mouth	reca	mouch
pollo	chicken	locho	pricken
rana	frog	laga	snaw
hilo	thread	vido	squead
vela	candle	dola	banble
cocina	kitchen	coreta	ketchen
bala	bullet	hara	bellet
cara	face	cajo	filt
loro	parrot	bero	pandot
codo	elbow	mogo	elpow
casa	house	mada	touse

Appendix B. Maximal and final main model for the RTs analysis

Maximal model

invRT ~ Language + Prime type + Group + Language : Prime type + Language : Group + Prime type : Group + Language : Prime type : Group + (1 | Participant) + (1 | Participant : Language) + (1 | Participant : Prime type) + (1 | Participant : Language : Prime type) + (1 | Target) + (1 | Target : Prime type) + (1 | Prime) + (1 | Prime : Prime type)

Final model after reduction of random structure and further criticism

invRT ~ Language + Prime type + Group + Language : Prime type + Language : Group + Prime type : Group + Language : Prime type : Group + (1 | Participant) + (1 | Participant : Language) + (1 | Participant : Prime type) + (1 | Participant : Language : Prime type) + (1 | Target) + (1 | Prime) + (1 | Prime : Prime type)

Appendix C

Table C1. Summary of final model for the analysis of RTs, including intercept and factors and their coefficients, standard errors, t-values, and p-values

	Coefficient	Std. Error	t-value	p-value
Intercept	−1.56	0.01	−123.56	<.001
Language	−0.05	0.02	−3.44	<.001
Prime Type	−0.15	0.01	−26.93	<.001
Group	0.05	0.02	2.08	.04
Language by Prime Type	0.03	0.01	4.19	<.001
Language by Group	−0.01	0.02	−0.41	.68
Prime Type by Group	0.02	0.01	2.48	.014
Language by Prime Type by Group	0.002	0.01	0.28	.78