# Individual differences in bilingual word recognition: The role of experiential factors and word frequency in cross-language lexical priming

Adel Chaouch-Orozco[1,*], Jorge González Alonso[2], Jason Rothman[2,3]

[1]University of Reading, United Kingdom, [2]UiT The Arctic University of Norway, Norway, and [3]Universidad Nebrija, Spain

*Corresponding author. Email: a.chaouchorozco@pgr.reading.ac.uk

In studies of bilingual word recognition with masked priming, L1 primes activate their L2 translation equivalents in lexical decision tasks, but effects in the opposite direction are weaker (Wen & Van Heuven, 2017). This study seeks to clarify the relative weight of stimulus-level (frequency) and individual-level (L2 proficiency, L2 exposure/use) factors in the emergence of asymmetrical priming effects. We offer the first dataset where L2 proficiency and L1/L2 exposure/use are simultaneously investigated as continuous variables, along with word frequency. While we replicate the asymmetry in priming effects, our data provide useful insights into the factors driving L2-L1 priming. These fall almost exclusively under the category of stimulus-level factors, with L2 exposure/use being the only experiential variable to show considerable influence, although complex interactions involving L2 proficiency and word frequency are also present. We discuss the implications of these results for models of bilingual lexical processing and for the appropriate measurement of experiential factors in this type of research.

**Keywords:** bilingual word recognition, individual differences, word frequency, L2 proficiency, bilingual lexicon, bilingual lexical access, priming asymmetry, lexical priming

***Introduction***

The literature on multilingual lexical organization has been dominated by two different but interconnected debates. The first focused on whether the languages of a bi-/multilingual are subserved by the same or different neural networks in the brain (i.e., *physically* stored together). The second debate asked whether the lexical items of different languages are *functionally* independent of each other, or rather lexical selection is open to competition among potential candidates from several languages, irrespective of what the response-relevant language is in a given context. Researchers speak of *language selective* vs. *language non-selective* lexical access, in reference to the disputed claim that, in a language selective model, only words from the response-relevant language may be considered for selection (e.g., Costa, Miozzo & Caramazza, 1999; Gerard & Scarborough, 1989; Macnamara & Kushnir, 1971; see Costa, 2005, for an extensive account).

In recent years, growing evidence has led to a moderate consensus around a view of the multilingual lexicon organised as a unitary system, where access occurs in a non-selective manner. That is, words from all languages are simultaneously active, to some degree, in comprehension and production (e.g., De Groot, Delmaar & Lupker, 2000; Dijkstra, Grainger & Van Heuven, 1999; Kroll & Stewart, 1994; Van Heuven, Schriefers, Dijkstra & Hagoort, 2008). Assuming this type of system, the focus must be placed on how exactly words from different languages are connected and interact, and what the nature of that relationship is—i.e., at which level of representation it is established (e.g., Brysbaert & Duyck, 2010; Brysbaert, Verreyt & Duyck, 2010; Dijkstra & Rekké, 2010; Kroll, Van Hell, Tokowicz & Green, 2010).

Most of the evidence supporting the non-selective view of the bi-/multilingual lexicon comes from studies where the degree of form and meaning similarity within the stimuli has been manipulated. The speed of access to cognate words, translation equivalents with a form and meaning overlap, has been shown to be faster than that to non-cognate words, even in

monolingual tasks (see Caramazza & Brones, 1979, for the first report on the effect; Van Hell & Dijkstra, 2002, for cognate effects on L1 lexical access, and Dijkstra et al., 1999, for L2; see also Lemhöfer, Dijkstra, & Michel, 2004, for cumulative effects in multilingualism). Words with similar orthography and/or phonology but with different meanings across languages, *interlingual homographs*, have also been exhaustively explored during the last decades. However, whether they yield facilitatory or inhibitory effects seems to be less clear, as this is dependent on factors such as the task employed or the stimulus list composition (e.g., Dijkstra et al., 1999; Brenders, van Hell & Dijkstra, 2011). The interaction of mental representations in the multilingual lexicon is not restricted to meaning and orthographic/phonological form. Cross-language activation has also been shown in priming studies exploring bilingual processing of compounds (Ko, Wang & Kim, 2011; Wang, 2010) and derivation (e.g., Duñabeitia, Dimitropoulou, Morris & Diependaele, 2013). What this body of research suggests is that words from different languages are activated and available for selection during production and comprehension, even in situations where only one of the languages is required.

While cognates and interlingual homographs are obvious candidates for shared or intimately related lexical representations across languages, it is likely that these are not the only points of contact within the multilingual lexicon. In that sense, non-cognate translation equivalents (e.g., English *arrow* and Spanish *flecha*), have been a major focus in research on bi-/multilingual lexical access for the past two decades. Because sublexical features (e.g., orthography, phonology) are not shared in these pairs, we may reasonably assume them to be connected, at least, through their largely overlapping conceptual semantics. The existence of priming effects between them suggests that translation equivalents, cognate or not, activate shared semantic representations (Xia & Andrews, 2015:295), and, therefore, have the potential to activate each other.

The masked translation priming paradigm, which employs the same mechanisms of subliminal priming originally devised by Forster and Davis (1984), has become one of the most common experimental set-ups in bi-/multilingual lexical processing research. In the typical procedure, a forward mask (e.g., #####) is displayed for a short period of time (typically 500 ms) and replaced by a word in one of the multilingual's languages: the prime (e.g., *flecha*, Spanish for 'arrow'). This is usually followed by the target word (or a backward mask), which in critical trials is the prime's translation equivalent in another of the participant's languages (e.g., *arrow*). Response times in these trials are compared to those in control ones, where the prime and the target also belong to different languages but bear no resemblance in meaning or form. As in standard masked priming, two measures are taken to ensure that the prime is processed only subconsciously. The first is to reduce the perceptual saliency of its onset and outset by means of forward and backward masking (note that in standard procedures the target itself acts as a backward mask); the second is to reduce display time to only a few milliseconds, typically between 40 and 70 and never above 85 (Clahsen, Balkhair, Schutter & Cunnings, 2013), to avoid the risk of entering into the conscious processing time window—at about 100 ms prime duration, most subjects can report the primes.

The masked translation priming paradigm has most often been used in combination with lexical decision tasks (LDTs). In a (visual) LDT, participants are asked to indicate whether the letter string presented on screen is a word in the target language. For this reason, half of the target items in a standard LDT are nonwords. Studies employing masked translation priming in LDTs have consistently reported an asymmetry in the direction of the priming effects obtained with non-cognate translation equivalents. Priming effects are robust and widely attested with L1 primes and L2 targets (e.g., De Groot & Nas, 1991; Jiang, 1999; Xia & Andrews, 2015). However, in the opposite translation direction (L2 primes, L1 targets) priming effects are either absent (e.g., Gollan, Forster, & Frost, 1997; Grainger & Frenck-Mestre, 1998)

or significantly smaller than those produced by L1 primes on L2 targets (e.g., Basnight-Brown & Altarriba, 2007; Schoonbaert, Duyck, Brysbaert & Hartsuiker, 2009; see Wen & van Heuven, 2017, for a comprehensive review).

### *The asymmetry in models of bilingual lexical processing*

We briefly introduce here two models of bilingual lexical processing: the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994; Kroll, van Hell, Tokowicz, & Green, 2010), and Multilink (Dijkstra, Wahl, Buytenhuijs, Halem, Al-Jibouri, De Korte & Rekké, 2019a). The RHM and the Bilingual Interactive Activation + model (BIA+; Dijkstra & Van Heuven, 1998, 2002), the model from which Multilink has evolved, have been by far the most influential proposals to date. They focus predominantly on word production and translation (RHM) and word recognition (BIA+). Multilink essentially continues in the computational tradition of the BIA+, while incorporating insights from the RHM. Regardless of the type of data that initially motivated these models, the architectures they propose for the mental lexicon should hold both in production and comprehension (Brysbaert et al., 2010).

### *The Revised Hierarchical Model*

Like most current models of the multilingual lexicon, the Revised Hierarchical Model is a three-store proposal (see Paradis, 2004): words from different languages are represented separately but share access to conceptual representations. These relationships between words and conceptual features are established through links that vary in intensity. L1 words are strongly connected to the conceptual system, reflecting the fact that an L1 lexicon is completely developed by the time late L2 learners start acquiring the new language (Kroll & Tokowicz,

2005). Conversely, the lexico-semantic mapping is typically weak(er) for L2 words, especially in low-proficiency bilinguals, who rely on L1 words to access semantic information as L2 words are generally learned through their L1 translation equivalents. In other words, a strong lexical connection in the L2-L1 direction allows L2 words to access L1 lexical representations, which, in turn, activate the shared conceptual nodes, indirectly connecting the L2 words with the relevant semantic information.

Xia and Andrews (2015) discuss a way in which the RHM could account for the priming asymmetry. If we assume that priming between (non-cognate) translation equivalents obtains exclusively through semantic mediation (and, crucially, *not* via lexical links), the model would predict that an L1 word can prime the lexical representation of its L2 translation equivalent because it can easily activate the shared conceptual nodes; however, since L2 primes cannot reliably stimulate these shared conceptual representations (or, at least, not fast enough), they fail to produce priming in the L2-L1 direction. The RHM states that the connections between L2 lexical items and conceptual representations become stronger as a direct function of L2 proficiency, which would eventually allow L2 primes to activate shared concepts in a similar way to that of L1 primes. Studies showing cross-language priming effects with simultaneous or balanced bilinguals (e.g., Basnight-Brown & Altarriba, 2007; Duñabeitia, Dimitropoulou, Uribe-Etxebarria, Laka & Carreiras, 2010; Duñabeitia, Perea & Carreiras, 2010) could, in principle, support this prediction. However, recall that research with unbalanced bilinguals specifically testing the role of proficiency has reported mixed results (e.g., Dimitropoulou et al., 2011; cf. Nakayama et al., 2016). While the model's proponents have gradually abandoned the idea of L1-mediated access to conceptual representations for L2 words, they maintain that conceptual links are weaker in the L2, even at higher levels of proficiency (Kroll et al., 2010), and that this has proven to be more noticeable in the concept-to-word direction than the other way around—which would predict differences between comprehension and production data.

### *Multilink*

In spreading/interactive-activation accounts of language processing (see Collins & Loftus, 1975; McClelland & Rumelhart, 1981), the activation level of a node in the network—in this case, a lexical entry—has to rise from its Resting Level Activation (RLA) to a certain threshold for it to become active (Jiang, 2015)—and thus be, for example, identified in visual word recognition. Multilink claims that the elusiveness of L2-L1 priming effects might lie on the RLAs of L2 words, which are lower than those of L1 words. Given short prime presentations under masked priming conditions, L2 primes may not receive sufficient stimulation or have enough time to process that stimulation and pass that activation on to their L1 translation equivalents.

Multilink, like the RHM, proposes that higher L2 proficiency may change this situation, as this tends to correlate with higher frequency and recency of use, which should, in turn, raise the RLA of L2 lexical representations. As the distance between the RLA and the threshold is shortened, the amount of stimulation—and, therefore, the processing time—needed to activate these words is reduced, increasing the chance of observing priming effects on their translation equivalents. However, proficiency may not be the only factor at play in determining the RLA of these words. Word frequency and (recent) high exposure to the L2 are likely to modulate the RLA, potentially making L2 lexical processing faster for (i) high-frequency words, and/or (ii) speakers that are immersed in or otherwise more frequently exposed to the L2.

While the asymmetry seems to be observed when *unbalanced* bilinguals are tested, no attempt has been made so far to understand the granularity of this factor and its relationship with L2 proficiency. This study attempts to fill that gap by examining a group of L1 Spanish – L2 English late bilinguals living in an L2-dominant environment, differing in degree of active

exposure/use and L2 proficiency. Anticipating the results, the data show significant priming effects for L1 primes. The effect for L2 primes is modulated by L2 active exposure/use, measured as a continuous variable at the individual level. Differently, the effect of L2 proficiency was only found to be significant in an interaction with the frequency of the L2 targets.

Our results raise several questions regarding the nature of cross-language masked priming patterns and the role of methodological factors. In this sense, they highlight the need for more fine-grained measures to tap into individual differences that can serve as proxies of bilingual language use and representation.

### *Factors investigated in the literature as potential modulators of L2-L1 priming*

While there is consensus that L2-L1 priming effects are notably less robust than their L1-L2 counterparts, effect sizes have varied considerably across studies, which range from null effects (e.g., Gollan, Forster, & Frost, 1997; Grainger & Frenck-Mestre, 1998; Xia & Andrews, 2015) to significant L2-L1 priming (e.g., Duyck & Warlop, 2009; Nakayama, Ida & Lupker, 2016; Lee, Jang & Choi, 2018; Lijewska, Ziegler & Olko, 2018). Although these studies have investigated a substantial number of factors potentially involved in L2-L1 priming effects (e.g., L2 proficiency, prime duration, word frequency, or the dominant language in the participants' environment, among others) results are mixed for all of these variables.

#### *Word frequency*

There is substantial evidence indicating that word frequency is a major predictor of the speed of lexical access in both the L1 and the L2 (e.g., Diependaele, Lemhöfer & Brysbaert, 2013;

Brysbaert, Stevens, Mandera & Keuleers, 2016; Brysbaert, Lagrou & Stevens, 2017; Brysbaert, Mandera & Keuleers, 2018). Despite this well-known effect, whereby more frequent words are accessed faster than less frequent ones, the factor has rarely been studied in the translation priming literature—to the extent that the word "frequency" does not even appear in the only currently available meta-analysis on the priming asymmetry (Wen & van Heuven, 2017). What is more, the great majority of studies have used stimuli with word frequencies within the range of, approximately, 3 to 4.3 in the Zipf scale (i.e., between 1 and 24 occurrences per million; see the Methodology section below for an explanation of the Zipf scale), where frequency effects in the access to L2 words are reported to appear (Brysbaert et al., 2017). Thus, frequency could certainly be expected to play a role in masked translation priming effects, and yet it is almost never examined as a factor.

A recent study by Nakayama, Lupker, and Itaguchi (2018) offers relevant insights on what the role of word frequency might be. These authors carried out distributional and frequency-based analyses of response times obtained with very highly proficient bilinguals and high-frequency words in an LDT using L2 primes. The observed 20 ms priming effect was reflected in a shift and a differential positive skewing on the response latency distributions. Furthermore, they observed that the distributional pattern was caused by an interaction of target frequency and the experimental condition (i.e., related vs. control L2 primes). Nakayama et al. argue that these results suggest that high-frequency translation primes (but, crucially, not control primes) are able to mitigate the target frequency effect, whereby less frequent targets are responded to more slowly.

### *L2 Proficiency*

Regarding the influence of L2 proficiency, Dimitropoulou, Duñabeitia, and Carreiras (2011) found that it did not play a key role in their data, given the similar L2-L1 priming effects (between 11 and 14 ms, differences not significant) displayed by three different L2 proficiency groups. More recently, however, Nakayama et al. (2016) report significant L2-L1 (English to Japanese) priming in two experiments with highly proficient bilinguals (mean TOEIC scores: 872 and 917, respectively, out of 990). Importantly, the materials for Experiment 2 were the same as the ones used in a previous study by members of the same cohort, Nakayama, Sears, Hino and Lupker (2013), where no significant L2-L1 priming had been observed with less proficient L2 speakers (mean TOEIC score: 740). To confirm their results in Experiments 1 and 2, Nakayama et al. (2016) conducted a third experiment in which less proficient bilinguals (mean TOEIC score: 710) were tested with the materials of Experiment 1. No significant L2-L1 priming was found this time. Together with the insight provided by regression analyses in the first two experiments, which showed that L2 proficiency modulated the effect size of L2-L1 priming, these results indicate that (very) high proficiency is a crucial factor behind the disappearance of the priming asymmetry. To the extent that high proficiency is a necessary condition, this could explain the discrepancy in results from other studies where lower proficiency groups do not yield the effect.

***Language exposure/use and immersion***

Although the language environment of participants has been discussed and tangentially addressed in the literature, few studies have examined it directly. Zhao, Li, Liu, Fang, and Shu (2011) investigated translation priming in three groups of Chinese-English bilinguals: two groups of low- and high-proficiency participants living in China (i.e., non-immersed) and one high-proficiency group living in an L2-dominant environment. Replicating the priming

asymmetry, L1-L2 priming effects obtained across the board, but L2-L1 priming was observed only for the immersed group. These results, while illuminating, effectively confound two individual-level variables, (high) L2 proficiency and immersion, because the factorial design is incomplete: there is no low-proficiency group in an immersed context.

Sabourin, Brien and Burkholder (2014) tested four groups of English-French bilinguals who had acquired the L2 at different ages (i.e., from birth, 3-5 years, 3-10 years, and 2-29 years). The participants' self-reported L2 proficiency (approximately intermediate) was matched across the early and late bilinguals groups to test how age of acquisition (AoA) could account for the translation priming effects in the L2-L1 direction. Their results showed significant priming only for the simultaneous and early bilinguals, but not for the late bilinguals, providing evidence for the role of AoA on the emergence of the priming asymmetry. Nevertheless, in this study, AoA was determined by the age of immersion in the L2 environment, thus confounding the potential influence of these two factors.

Finally, at least two studies have shown the importance of balanced bilingualism when considering cross-language masked priming effects. In Duñabeitia, Perea, and Carreiras (2010), a symmetric priming pattern was reported when testing a group of highly proficient bilinguals (i.e., native speakers of Basque and Spanish). Importantly, although they differed in their frequency of use of the languages in academic contexts, using much more Basque than Spanish, the use in non-academic contexts was almost identical. Likewise, Wang (2013) reported a beneficial effect of balanced bilingualism on the emergence of L2-L1 priming effects when investigating highly-proficient Chinese-English bilinguals living in a bilingual society like Hong Kong. Group 1 consisted of English-dominant bilinguals, whereas Group 2 was formed by bilinguals whose use of (and proficiency in) Chinese and English were equal. Although the L1-L2 translation direction was not tested, preventing us from drawing

conclusions on the priming asymmetry itself, only Group 2 showed significant L2-L1 priming effects.

Language exposure/use and L2 proficiency are both ultimately proxies of L2 subjective word frequency, that is, how often a given speaker has encountered a given word. Although the studies reviewed in this section highlight the relevance of these two factors on the processing of L2 words and the ability of L2 primes to efficiently activate their translation equivalents in priming experiments, the field still needs more fine-grained measures that allow for a better estimation of their role in bi-/multilingual lexical processing. After all, discrete-variable approaches, while providing a good approximation to the presence or absence of certain effects, are likely to miss subtle transitions and non-linear trajectories along the continuum of influence of these factors. Van Hell and Tanner (2012:165), in discussing individual differences in L2 proficiency and their relationship to cross-language lexical activation, argue that

> […] providing a clearer picture of the relationship between cross-language activation effects and individual differences in L2 proficiency requires a move away from group designs and toward designs that allow for more robust statistical modelling of the interaction between individual-level characteristics (e.g., language proficiency) and stimulus-level characteristics (e.g., word cognate status). As previously mentioned, regression-based approaches can model the continuous nature of individual-level variables, like language proficiency.

Furthermore, the fact that all these variables have typically been investigated separately (or, at best, in pairs) potentially obscures important interactions between them (Diependaele et al., 2013). For these reasons, the present study aims to examine the role of L2 exposure/use

and L2 proficiency in cross-language masked priming effects by treating these variables as continuously distributed, in an attempt to reflect their nature more efficiently and weigh their role on the priming asymmetry. If models such as Multilink are right in their assumption that asymmetrical priming patterns have their origin in RLA differences between L1 and L2 words, these differences must be a direct consequence of the individual experience of a given speaker with a given word. This experience, in turn, can only be approached through factors that quantify the relative exposure of the speaker to linguistic contexts potentially containing the word, as well as the relative availability of the word itself. While we examine here two individual-level variables (i.e., active exposure/use and L2 proficiency), we will only deal with one stimulus-level factor: word frequency—albeit effectively represented twice in the design, through the independent contribution of prime and target frequencies. This is not to deny that other properties of the stimulus (e.g., length, orthographic and/or phonological neighborhood size, concreteness, morphological family size) have the potential to affect responses. However, their effects have been consistently proven to be smaller in size and more reduced in scope than those of word frequency (Diependaele et al., 2013).

***The present study***

Sixty L1 Spanish-L2 English sequential bilinguals living in the UK took part in an LDT experiment with masked translation priming. Participants were tested in both translation directions to investigate the priming asymmetry directly. In light of the available literature, we expected to replicate the priming asymmetry, as L1-L2 effects are relatively robust, and our choice of participant profiles and word frequencies did not favor the appearance of L2-L1 priming effects. However, our study was also designed to shed light on the role of three variables, which we quantified and included as continuous predictors: L2 proficiency, amount

of L2 exposure/use, and word frequency. If, as expected, these factors impact the processing of L1/L2 words and consequently the priming effects, we should observe three-way interactions (potentially four-way interactions too) between translation direction, type of prime, and individual- and stimulus-level predictors in the statistical models' outcomes.

## *Method*

### *Participants*

Sixty Spanish-English sequential bilinguals were recruited from the Spanish-speaking communities in three large cities in North and South-West England (see Table 1 for participant characteristics). The data was collected in sound-insulated rooms at a university or teaching institution in each location. To evaluate English proficiency, all participants took the Oxford Quick Placement Test (OQPT; Oxford University Press, University of Cambridge, & Association of Language Testers in Europe, 2001). The test examines English grammar and vocabulary knowledge and consists of 60 multiple-choice questions.[1] The participants' mean score was 50 (*SD* = 4.84, range: 40-60), corresponding to a lower-advanced proficiency according to the OQPT's manual. The scores of English proficiency were normally distributed throughout our sample, as indicated by the exploration of a Q-Q plot and a Shapiro-Wilk test for normality ($p = 0.38$). The participants had started learning English, on average, at the age of 9 (*SD* = 2.9, range: 4-16). A version of the Dominance Scale questionnaire by Dunn and Tree (2009) was employed to collect information regarding the participants' use of English. The questionnaire provides a scale based on the relative use of one language over the other

[1] Geranpayeh (2003) reports an SEM of around 4 for the 60-item OQPT (the one used in the current study), and test-retest reliabilities of around 0.9 during the task's validation procedure.

(Dunn & Tree, 2009:1). The scale ranges from -25 to 25. Following the authors, a score above 5 was considered to reflect greater use of the L1 (Spanish) over the L2 (English), whereas the range between -5 to 5 was considered to reflect an equal use of both languages. Although the scale originally makes reference to "dominance", as it was designed to test simultaneous bilinguals of balanced proficiency, we speak here of "active language exposure/use" instead, which we consider a better reflection of what the scale actually measures—as well as being the variable of interest in our study. Consider, for instance, a 20-year-old late sequential bilingual who has lived in an L2 environment for a year, speaks the L2 at her new home as well as at her new job, and received more than ten years of education in the L2 at a bilingual school in Spain. Such a participant would most probably have a score below -5 in the scale; nevertheless, should we conclude that her L2 is now the dominant language over the L1, despite her having been overwhelmingly more exposed to the L1 for 19 years (95%) of her life? With this example, we hope to highlight the potential misinterpretation that the use of "dominance" can lead to. However, we acknowledge that the term is still operationalized as a function of language use in much work on bilingualism. As Treffers-Daller (2019:1) explains,

> […] language dominance is often seen as relative proficiency in two languages, but it can also be analyzed in terms of language use—that is, how frequently bilinguals use their languages and how these are divided across domains.

*Materials*

Fifty pairs of Spanish-English non-cognate translation equivalents were used in the experiment (see Table 2 for sample stimuli). To avoid the concreteness effect found in different studies (e.g., Finkbeiner, Forster, Nicol & Nakamura, 2004; Schoonbaert et al., 2009), whereby

abstract words are responded to more slowly than concrete words, only concrete nouns were used. As shown in Table 3, the Spanish words had a standardized mean frequency of 4.01 (*SD* = 0.43, range: 2.72-4.9) on the 1 to 7 Zipf scale (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The standardized mean frequency of English words was 3.97 (*SD* = 0.34, range: 2.94-4.92) (Table 3). In the Zipf scale, word frequencies between 1 and 3 are considered low, whereas those between 4 and 7 are considered high frequencies (see Van Heuven et al., 2014, for details). Word frequencies for the English items were extracted from the SUBTLEX-UK database (Van Heuven et al., 2014), and the ones for the Spanish words from SUBTLEX-ESP (Cuetos, Glez-Nosti, Barbón & Brysbaert, 2011). Word frequencies were normally distributed in the English stimuli, as indicated by the exploration of a Q-Q plot and a Shapiro-Wilk test for normality ($p = 0.36$). This was not true, however, of the Spanish stimuli. Although this is not ideal, we were limited by the small amount of translation pairs at our disposal (recall that these had a relatively low-frequency) and the need for our participants to know the L2 words. For this (and other) reason(s), we chose a statistical method—linear mixed modeling (see below)—that can accommodate deviations from normality in both independent and dependent variables.

Additionally, 50 nonwords were created in both languages to make the lexical decision possible. Spanish nonwords were created by substituting one letter from real words while respecting the phonotactics of the language. The English nonwords were created using the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002).[2] All nonwords were

[2] Applying different methodologies to create our nonword stimuli might have caused divergences in the baseline difficulty to reach a lexical decision in each translation direction. In our data, then, an effect of nonword construction should have translated into not only (equal or) shorter RTs for L2 targets as compared to L1 targets, but also lower accuracy rates in the L1-L2 experiment as compared to the L2-L1one. However, the latter was not found: as we will see, our participants were faster and more accurate when responding to L1 targets than to L2 targets.

phonologically and orthographically plausible in Spanish and English, respectively. The complete list of stimuli is provided in Appendix A.

Four stimulus lists (two in each language) of 50 word and 50 nonword targets were created. In one of the lists, half of the target words were preceded by their translation equivalents and the other half by control primes. The translation equivalents from those pairs in the baseline condition of each list were scrambled to serve as control primes, paying attention to keep the pair semantically unrelated. In the other list, the order was inverted, so that across both lists all the words were preceded by their translation equivalents and control primes. Each list began with sixteen practice items. All words were matched in frequency and word length.

*English-Spanish translation task*

To ensure that responses to the L2 words were not arbitrary, participants completed an English to Spanish translation task with the English items. Only answers identical to the translation pairs used in the experiment were counted as correct. All the items had a minimum 65% rate of correct answers, and the correct answer was given on average 88% of the time. Although a 65% rate of correct responses might seem a low cut-off, some of the answers provided were synonyms of the expected translations, even though they did not count as correct answers. More importantly, in the post-task debriefing, many of the subjects reported knowing the translation of certain English words but having been unable to recall them during the translation task. Their incapacity to remember the translation at that point, or the fact that they chose to provide a synonym to the target translation, would not necessarily entail an insensitivity to those English primes during the experimental task.

*Test of familiarity with Spanish words*

The degree of familiarity with the Spanish materials was normed across the first 29 participants, roughly half, to control for the fact that the materials used in the lexical decision experiments were created using European Spanish, spoken by 20 of those first 29 participants. The other nine subjects spoke other varieties of Spanish. The test used a 7-point Likert scale, where 1 represented no knowledge of the word, and 7 described a word that was known and frequently employed in the participant's variety of Spanish. No items were removed from the experiments due to a lack of familiarity, since all the words' mean scores were higher than the cut-off value of 4, well above what could be considered unfamiliarity with a given lexical item.[3]

*Procedure*

The experiment was programmed using the PsychoPy v1.8 software (Peirce, 2007). Each trial began with a 500 ms forward mask (#########), followed by a 60 ms prime (in lowercase letters). Immediately after, the target (in uppercase letters) appeared and remained on screen until the participant's response. Stimuli were presented on a white screen in a 44-point black Arial font. Participants were asked to judge whether the targets were real words or not by pressing on a keyboard, '0' for NO or '1' for YES, as quickly and accurately as possible. They were not informed about the presence of the primes. During a post-experiment debriefing, participants were asked about their awareness of any word-like material other than the target words in the course of the experimental trials.

The tasks were presented in the following order: The OQPT was the first test to be administered, since a score below 40 (i.e., equivalent to intermediate proficiency in the

[3] To ensure that dialectal differences had no effect on the results, we conducted post-hoc analyses running the models with the interaction of Dialect (coded binarily as Castilian vs. non-Castilian Spanish), Prime Type and Target Language as a fixed factor. The interaction was non-significant.

Common European Framework of Reference for Languages, CEFR) was used as exclusion criteria to participate in the study. Then, the experimental tasks were conducted. After completing them, the participants did the English words translation task and the familiarity task.

***Results***

Following Baayen and Milin (2010), responses to experimental trials with latencies below 200 ms and above 5000 ms were removed from the dataset (1 observation), on the assumption that those latencies would be too short to reflect a conscious judgment of the targets or too long to ensure that conscious strategies are not involved in the decision. Eighteen data points were removed due to glitches in the presentation, and 100 data points were excluded because of a problem during the counterbalancing of the critical condition for one of the subjects in the L1-L2 direction. After removing incorrect responses and responses to nonwords, the dataset contained a total of 5881 observations.

An exploratory analysis of the RT distribution was performed by transforming the latencies to obtain inverse Gaussian, log-normal and Box-Cox distributions. The exploration of Q-Q plots and the results of Shapiro-Wilk tests for both translation directions showed that the inverse Gaussian transformation provided a slightly better correction of the distribution's skewness than did the other two (inverse Gaussian: $p = 1$; Box-Cox: $p = 0.99$; log-normal: $p = 0.73$).

Analyses of the error rates for word targets and the transformed RTs for correct responses to word targets were conducted using (generalized) linear mixed-effects models (LME; Baayen, 2008; Baayen, Davidson, & Bates, 2008) in *R* (version 3.3.1) (R Core Team, 2016) with the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). A theory-driven model was

used for both the accuracy and response latency analyses. The model included the following factors: Target Language (Spanish or English), Prime Type (Related or Control), Proficiency (modelled as a continuous variable quantified by English placement test scores), Language exposure/use (modelled as a continuous variable quantified by the Dominance Scale questionnaire) and Prime and Target Frequency (Zipf values)[4]. Sum contrasts were used for categorical variables. Proficiency, Language exposure/use, and Prime and Target Frequency were scaled and centred, and converted to *z*-units. This model thus contained the main effect of Target Language, the interaction between Target Language and Prime Type, and three- and four-way interactions between Target Language, Prime Type and the stimulus- and individual-level factors. (See Appendix B for the complete models and rationale).

The random structure of this initial model included random intercepts for subjects, primes and targets (Feldman, Milin, Cho, Moscoso del Prado Martín & O'Connor, 2015), and random slopes for subjects within Target Language, Prime Type, Target Frequency, Prime Frequency, and the interaction between Target Language and Prime Type; as well as random slopes for primes and targets within Target Language, Prime Type and the interaction between the two factors.

***Response time analysis***

Table 4 provides a summary of error rates, mean RTs, and priming effects (calculated as the difference between mean RTs to control and critical trials) for correct responses to word targets. Following Matuschek, Kliegl, Vasishth, Baayen & Bates (2017), we carried backward-selection and employed the likelihood ratio test criterion to obtain a more parsimonious model.

[4] Due to concerns about the potential collinearity between some independent variables, tests were conducted to examine the correlation between Prime and Target Frequency, and L2 proficiency and Language exposure/use. While the second pair of variables did show some correlation (Prime/Target freq.: $r = .02$, $p = 0.25$; L2 Prof./Exposure: $r = .21$, $p < .001$), this was not a strong one.

The reason for this is that, given our relatively small sample size, models with complex random structures might not be supported by the data (Matuschek et al., 2017, p. 307). Thus, during backward-selection, we iteratively removed the random slopes that accounted for the least amount of variance from the model, until convergence was achieved. Full code for the analyses can be found in the first author's GitHub repository (https://github.com/chaouch-orozco/Individual-factors-in-bilingual-word-recognition/blob/master/Ind_diff_bil_wor_rec_script_R.2.R). The final model had the fixed effects specified above as well as random intercepts for Subject, Prime, and Target (Table 5 for the full model summary). Exploration of this model's residuals through Q-Q plots showed that the residuals did not follow a normal distribution in the longer latencies. Therefore, as suggested in Baayen and Milin (2010), we applied further model criticism by excluding those observations with absolute standardized residuals above 2.5 SDs (116 observations were removed, 2% of the total).

The interaction between Target Language and Prime Type indicated that translation primes elicited faster responses to targets as compared to control ones (i.e., a priming effect) in the L1-L2 direction. A significant interaction between Target Language (Spanish), Prime Type and Language exposure/use, indicated that those participants with a higher degree of active L2 use benefited more from the L2 related primes during the processing of the L1 targets. The interaction of Target Language (English), Prime Type and Target Frequency showed that RTs were significantly faster for more frequent L2 targets in both the Related and the Control condition. A significant four-way interaction between Target Language (English), Prime Type (Related), Proficiency and Target Frequency was observed, indicating that RTs for low-frequency L2 English words preceded by L1 Spanish related primes were significantly slower for less proficient bilinguals. Also, three marginally significant interactions were observed. First, that of Target Language (English) and Proficiency, showing faster RTs for more

proficient participants. Second, a Target Language (Spanish), Prime Type (Control) and Prime Frequency interaction, indicating that, in the control condition of the L2-L1 direction (Spanish targets), responses were slower with more frequent L2 English primes. Third, a Target Language (Spanish), Prime Type (Related) and Prime Frequency interaction, suggesting the opposite effect: faster RTs for more frequent L2 related primes.

Awareness of the prime was included as a post-hoc factor. Although unexpected at 60 ms prime duration, 24 participants reported having seen some characters on the screen during the prime presentation time window, that is, between the forward mask and the target word, on at least one trial. Importantly, this only happened during the L1-L2 task (Spanish primes), and most of the subjects reported only one occurrence. The reason to include prime awareness in the analysis, then, instead of excluding these participants from the study altogether, was that LME models allow us to control for and estimate the influence of similar factors without discarding the data. To investigate the influence of prime awareness on participant responses, we carried out an analysis including Awareness in an interaction with Target Language and Prime Type. The results showed that this factor did not significantly modulate priming effects. Given this outcome, we consider it reasonably safe to keep all the participants in the analysis.

***Accuracy analysis***

Accuracy was dummy-coded as 1 (correct) or 0 (incorrect) and generalized linear mixed-effects models with a binomial family were fit to the error data. In this case, the initial model, which was the same as in the response time analysis, did not converge. We thus proceeded to simplify its random structure applying the same backward-selection method. In the final model, the fixed effects were the same as in the model for the RT analysis. The random structure contained intercepts for Subject, Prime and Target, and slopes for Subject within Target Language, Prime

Type, Target Frequency, Prime Frequency, and the interaction between Target Language and Prime Type. It also contained slopes for Prime within Target Language and the interaction between Target Language and Prime Type, as well as for Target within Target Language, Prime Type, and the interaction between Target Language and Prime Type.

Table 6 provides the summary of the model, which shows a significant effect of the three-way interaction between Target Language (Spanish), Prime Type (Related and Control), and Target Frequency. This effect shows significantly lower accuracy rates for low-frequency L1 targets in both Prime Type conditions. That is, overall, participants were less accurate with less frequent L1 targets. Furthermore, a significant four-way interaction between Target Language (Spanish), Prime Type (Control), Proficiency, and Prime Frequency was observed. In the L2-L1 translation direction, the frequency of the L2 primes affects less and more proficient bilinguals differently. Whereas for the less proficient participants a potential inhibitory effect of the control primes is larger when these are less frequent, the effect is the opposite for the more proficient bilinguals. This finding is intriguing, but it is hard to attribute it confidently to a single (group of) factor(s) or combination thereof—e.g., cognitive, methodological. Given the inherent difficulty to interpret higher-order interactions, the overall small differences in error rates across the four data subsets (lower-proficiency/low-frequency: 1.07; lower-proficiency/high-frequency: 2.78; higher-proficiency/low-frequency: 2.59; higher-proficiency/high-frequency: 0.69), and the fact that error rate analyses have typically received far less attention in this type of studies, we are cautious in interpreting this result and will not comment on it further.

***Discussion***

In this study, we conducted a masked translation priming lexical decision task, testing late sequential Spanish-English bilinguals immersed in an L2-dominant environment. Overall, our data do replicate the priming asymmetry in general terms, but provide a fairly more nuanced picture, as (i) the priming effects were numerically similar in both translation directions, and (ii) the main effect of Prime Type was significant only in the L1-L2 direction, albeit modulated by Language exposure/use in the L2-L1 direction (i.e., participants with increased active exposure and use of the L2 showed larger priming effects). Furthermore, we observed a complex interaction between Target Language, Prime Type, Proficiency and Target Frequency: the less proficient bilinguals responded more slowly to low-frequency L2 English words in the related condition (i.e., when preceded by their Spanish translations).

Recall that one of the goals of the present study was to shed light on the role that L2 proficiency and, somewhat novelly, active language exposure/use at the individual level play in translation priming effects, by treating them both as continuous predictors. Doing so allows for a more fine-grained understanding of each factor's weight. With respect to L2 proficiency, a central factor in Multilink and especially the RHM, we do not observe an effect directly modulating priming in either translation direction. However, our data do show that, when less proficient bilinguals had to respond to less frequent L2 targets, their responses were slower only in the Related condition. Therefore, the L2 proficiency measure was able to account for some differences in the processing of the low-frequency L2 related targets, potentially closing or widening the gap in priming effects by modulating the speed of related trials with respect to a (presumably constant) unrelated baseline. More deterministic in our data, however, is the role of language exposure/use. This factor directly interacted with Prime Type (and Target Language), conditioning priming effects in the L2-L1 direction. Recall that this is the direction of interest in most previous studies, as translation priming effects have been less reliably found

across the board. In our study, those participants showing a higher active exposure/use to the L2 showed larger priming effects.

Despite the less salient role of L2 proficiency in our data as compared to that of language exposure/use, we cannot conclude that this predictor plays no significant part in shaping masked translation priming effects. Although there were methodological reasons for doing so, the range of L2 proficiencies covered in this study (i.e., upper intermediate to advanced) prevents us from making conclusive claims in this regard. Alternatively, and especially considering that any potential factors involved in such complex phenomena may have nonlinear trajectories, we would have needed to test a broader range of L2 proficiencies (e.g., low to high)—although the feasibility of such manipulations is directly conditioned (and directly conditions) the frequency range of the stimuli.

At the time of the experiment, all participants had been living in the UK for five years on average (SD = 3.62, range: 0.75-13). Observing the broadness of this range, and given that lexical attrition is a well-documented phenomenon, one might argue that the Spanish of some of these participants might have attrited to some extent. To test this hypothesis, we conducted a post-hoc analysis of the L2-L1 task (where the targets were Spanish words) including the interaction between Length of Immersion and Target Language as a fixed factor, as well as the three-way interaction between Length of Immersion, Target Language, and Prime Type. The outcome of this model contained non-significant effects for all of these interactions, suggesting that participants' responses in Spanish were not dependent on their time living in an L2-dominant environment.

With respect to the effects of word frequency, we observe that Target Frequency significantly interacted with Target Language (English), Prime Type (Related) and Proficiency. As reported above, RTs to L2 targets were significantly slower in two contexts: with control primes overall and, when L2 target frequency was low, for less proficient as

compared to more proficient participants. This result suggests that, when responding to less frequent L2 targets (i.e., in longer/more difficult trials), only the high-proficiency participants benefitted from the presence of the L1 related primes. It would be problematic to argue that this outcome is due to the inability of the primes to be processed. Given the linguistic profile of our participants, it should not be difficult to process the L1 primes (even the less frequent ones). Indeed, such difficulties should have had a larger impact on those bilinguals who had the largest potential for attrition, that is, those on the upper ends of the proficiency and active L2 exposure/use scales. However, this is not the case in our data. Alternatively, one could also argue that the less proficient bilinguals did not know the low-frequency L2 targets. This is unlikely, since the accuracy rates for low- and high-proficiency bilinguals were numerically similar and high (96% vs 97%). Therefore, lack of knowledge of the lower-frequency L2 words does not seem to explain the slower latencies in the Related condition for less proficient bilinguals. This significant interaction thus remains an open question and should be further investigated if it were found to replicate in future data sets.

Returning to our most novel result, the modulation of L2-L1 priming effects by Language exposure/use, it should be noted that this finding does not provide a reliable way to adjudicate between the RHM and Multilink, since their predictions largely overlap here. For the RHM, a larger amount of active L2 exposure/use should bring about stronger L2 lexico-semantic connections, which would in turn enhance L2-L1 priming effects. Alternatively, Multilink would predict the L2 lexical representations of bilinguals with more L2 exposure/use to have higher RLAs, facilitating their processing and increasing the likelihood of observing L2-L1 priming effects—because, in short, they should be more effective primes.

The RHM and Multilink explain the differences in L1/L2 lexical processing by resorting to different conceptualizations of the operations underlying cross-language effects in the bilingual lexicon. Those models of the lexicon lead to different predictions about how

words are (differently) processed depending on an array of experience-level factors (e.g., frequency, language membership, learning context), which, in many cases, predict the most common pattern of L1/L2 differences: L1 words tend to be processed faster than L2 words. However, the factor that ultimately shapes lexical processing—word form- and semantic-level variables such as word length, concreteness being equal—might in fact be the same: subjective frequency. In that sense, the present results suggest that, in our data, language exposure/use was a better proxy for subjective frequency than L2 proficiency. In fact, it might fare even better than a stimulus-level variable such as (corpus) word frequency—although we also find effects for these two predictors, showing that their validity as proxies cannot be disregarded.

Here we should note that, for Multilink, all these factors might potentially affect cross-language priming effects, as all of them approach subjective frequency to some degree. Similarly, although not originally specified by the RHM, a deterministic role of exposure/use is not necessarily incompatible with its tenets. For instance, bilinguals who are exposed and use their L2 more (on a scale) might have available more entrenched L2 word-meaning connections, whose strength would be independent of how proficient they are in the L2 overall. This point is of significant consequence not least because proficiency is often measured as a categorical variable, predicated on a relative standard model (i.e., how one fares juxtaposed against an idealized standard, often a monolingual one). By its very nature, proficiency measures are only able to test a subset of knowledge a truly competent speaker would have, which is more or less attainable and/or is a greater or lesser proxy for what it seeks to uncover depending very much on context (e.g., Norris & Ortega, 2012; Rothman & Iverson, 2010). At the end of the day, especially in light of these models, opportunity for links within an individual's mental lexicon is of primary importance. And so, it is not clear how or if L2 proficiency measured as typically done can faithfully proxy for actual competencies (grammatical and/or performative), even if, in many cases, they will ultimately overlap.

Therefore, it is worth looking into and taking more seriously measures that are more fine-grained proxies for actual opportunities that should, reasonably, correlate with greater linking. This discussion is accentuated under two conditions, both of which apply in our study: (i) at so-called higher levels of proficiency, where a threshold of specific knowledge has been attained to test relatively high on measures we currently have but which do not necessarily say anything about real-world abilities in the language per se, and (ii) under conditions of increased potential exposure such as immersion, where individual differences in how immersion is capitalized on might nevertheless have some determinism.

On the other hand, stimulus-level factors such as word frequency have been shown to function as reliable proxies when investigating lexical processing, to the point that frequency has been highlighted as the single most critical variable influencing lexical decision time (Brysbaert, Buchmeier, Conrad, Jacobs, Bölter & Böhl, 2011:1). However accurate this measure has proven to be—aside from debates on which types of corpora better capture its effects—one should not overlook the fact that (i) L1 corpora are far from ideal sources of language use when one is interested in studying lexical retrieval in the L2; and (ii) by their very nature, frequency counts assume equal word frequencies across speakers of a given language and are, thus, inherently imperfect approximations to the concept of *subjective* frequency. Thus, to understand how the speed of L2 lexical access is determined and to account for its variability, it is crucial to first identify which other factors—especially those bearing upon each individual's language experience—might be at play.

### *Conclusions and Future directions*

The typically reported translation priming asymmetry presumably reflects a relative inability of the L2 primes to stimulate (noncognate) translation equivalent targets under masked priming

conditions. Several factors have been suggested to underlie the asymmetry. The data in the present study is compatible with a deterministic role of subjective word frequency in bilingual lexical processing. The present findings might help explain divergent results found in the literature with respect to the role that different individual- and stimulus-level factors have on the priming asymmetry. In particular, conflicting results might reflect how accurately the predictors under examination proxied subjective frequency in those studies. For instance, we have argued that L2 proficiency might not be the most appropriate candidate to gauge the relative frequency with which an L2 word is encountered and used by each individual. Instead, the present data point towards active language exposure/use as a more efficient approximation to individual encounters with each word.

Moving forward with the present program, we are currently working in what we believe are the necessary next steps in characterising and tapping L2 subjective frequency. First, we are preparing a follow-up translation priming study, which will employ a more nuanced operationalization of active language exposure/use. Detailed language history questionnaires and the comparison of immersed and non-immersed L2 speakers will allow us to better estimate how the amount (and context) of L2 use affect bilingual lexical processing. Second, by examining populations with differential exposure to the L2, we will test the predictability of traditional frequency measures—extracted from L1 corpora—when bilingual populations of different types and in different contexts are investigated. Our goal is to contribute to building better approximations to what is ultimately a major factor in the online recruitment of lexical representations: subjective word frequency. Finally, we will test a larger population and employ a larger set of words. Having a larger sample size along with a simpler design will contribute to overcome the shortcoming of potentially low statistical power in the present study.

We consider that addressing the above issues is a necessary step in the integration of current theories of mental linguistic representation and processing, particularly at the lexical level. By doing so, we hope to contribute to a better understanding of bi-/multilingual lexical processing, inclusive of related questions pertaining to native vs. nonnative differences and the role that input quantity and quality play in shaping the observable spectrum of linguistic competencies.

**Acknowledgements**

We wish to acknowledge the following funding: Adel Chaouch-Orozco is supported by a PhD studentship from the University of Reading (ref GS18-009), and the Language Learning Dissertation Grant. Jorge González Alonso and Jason Rothman are supported by a grant for the HeLPiNG project via Tromsø Forskningsstiftelse (TFS), 2019-2023. We wish to thank Jon Andoni Duñabeitia for helpful discussion.

**Appendix A.** Complete list of stimuli.

Table A1. Prime and target words and nonwords for the L1-L2 and L2-L1 tasks.

| **Spanish translation equivalent** | **English translation equivalent** | **Spanish nonword (target)** | **English nonword (target)** |
|---|---|---|---|
| abuelo | grandpa | colle | drurch |
| acero | steel | ciela | phrague |
| algodón | cotton | vuenta | sazz |
| anillo | ring | ation | skeigh |
| ascensor | lift | masura | twourse |
| avellana | hazelnut | humi | scroothe |
| bandeja | tray | sopu | cheuth |

| | | | |
|---|---|---|---|
| bandera | flag | atuela | sourge |
| bosque | forest | cirta | pheap |
| camión | lorry | curcel | sneese |
| cangrejo | crab | barne | screathe |
| caramelo | candy | lorte | strarsh |
| carpeta | folder | cuadrí | trube |
| cartera | wallet | barvo | splarce |
| cebolla | onion | volor | gnerf |
| cepillo | brush | ganda | blogue |
| cereza | cherry | apesor | fras |
| cesta | basket | lati | phuiff |
| cicatriz | scar | patu | thwoche |
| cuchara | spoon | cerdu | vargue |
| cuchillo | knife | buerno | scroute |
| deberes | homework | alomna | wrirque |
| edredón | duvet | aje | tharc |
| enano | dwarf | coreza | graith |
| escudo | shield | gote | gube |

| | | | |
|---|---|---|---|
| espada | sword | corredar | splaunch |
| flecha | arrow | comote | splync |
| freno | brake | coimán | rount |
| galleta | cookie | cruma | bromb |
| gusano | worm | botellu | snuin |
| hacha | axe | gako | croosh |
| herida | wound | arbista | thrarse |
| lápiz | pencil | deoda | filk |
| maíz | corn | bruba | chautch |
| manta | blanket | cuñaya | plac |
| masa | dough | josa | throurth |
| moneda | coin | osu | slawn |
| nuez | walnut | almotada | bloothe |
| paraguas | umbrella | climo | knafe |
| payaso | clown | cina | freigthth |
| regla | ruler | cecatriz | shruise |
| seda | silk | sope | jief |
| sobrino | nephew | gangrejo | scrorque |

| | | | |
|---|---|---|---|
| teclado | keyborad | chatal | cruge |
| tela | fabric | foesta | wealt |
| tijeras | scissors | aldei | phrein |
| tinta | ink | artol | flis |
| trigo | wheat | bafete | guelch |
| uva | grape | bloqui | phrelf |
| vela | candle | burso | phrip |

Table A2. Spanish and English primes for nonword targets in L1-L2 and L2-L1 tasks.

| Spanish primes for English nonword targets | English primes for Spanish nonword targets |
|---|---|
| aceite | apple |
| ajedrez | armoury |
| almacén | army |
| almohada | ash |
| almuerzo | beach |
| árbol | blaze |
| autobús | blood |
| avena | branch |
| bañera | bubble |
| barrio | burden |
| batería | cane |
| baúl | cape |
| bebida | cello |

| | |
|---|---|
| boleto | chair |
| borde | cinnamon |
| botella | clock |
| bóveda | cloud |
| cabaña | copper |
| cable | cousin |
| cajón | cross |
| calabaza | cruise |
| callejón | cucumber |
| camilla | dish |
| canasta | flat |
| capilla | fork |
| capítulo | glance |
| carpa | grid |
| carreta | gypsy |
| carrito | harbour |
| carro | juice |
| castillo | junk |

| | |
|---|---|
| celda | kettle |
| cemento | kite |
| cera | knight |
| cerveza | laptop |
| chivo | mall |
| cobre | marble |
| código | mortgage |
| cohete | mosque |
| cola | muffin |
| colchón | oyster |
| colina | pen |
| comité | portrait |
| correo | razor |
| crucero | river |
| cuadro | rubber |
| cuartel | saviour |
| cuerda | shell |
| cuna | sign |

desayuno signal

**Appendix B.** Theory-driven model and description of the rationale behind each fixed-effect's inclusion.

RT ~ Target Language + Target Language : Prime Type + Target Language : Proficiency + Target Language : Language Exposure/use + Target Language : Prime Type : Prime Frequency + Target Language : Prime Type : Target Frequency + Target Language : Prime Type : Proficiency + Target Language : Prime Type : Language Exposure/use + Target Language : Prime Type : Prime Frequency : Target Frequency + Target Language : Prime Type : Prime Frequency : Proficiency + Target Language : Prime Type : Prime Frequency : Language Exposure/use + Target Language : Prime Type : Target Frequency : Proficiency + Target Language : Prime Type : Target Frequency : Language Exposure/use

- Target Language: to test whether responses were overall faster in one language or the other.
- Target Language by Prime Type interaction: to test priming effects in the two translation directions.
- Target Language by Proficiency: to test the role of Proficiency in the overall responses.
- Target Language by Language exposure/use: to test the role of Language exposure/use in the overall responses.
- Target Language by Prime Type by Proficiency: to test the (potentially different) role of Proficiency as a modulator of priming effects across translation directions.
- Target Language by Prime Type by Language exposure/use: to test the (potentially different) role of L2 exposure/use as a modulator of priming effects across translation directions.
- Target Language by Prime Type by Prime Frequency: to test the role of Prime Frequency on priming effects across both translation directions.
- Target Language by Prime Type by Target Frequency: to test the role of Target Frequency on priming effects across both translation directions.

- Target Language by Prime Type by Prime Frequency by Target Frequency: to test the potentially different effects of the interaction between Prime and Target Frequency as modulators of priming effects across both translation directions.
- Target Language by Prime Type by Prime Frequency by Proficiency: to test the role of the interaction between Prime Frequency and Proficiency as a potential modulator of priming effects across both translation directions.
- Target Language by Prime Type by Prime Frequency by Language Exposure/use: to test the role of the interaction between Prime Frequency and L2 exposure/use as a potential modulator of priming effects across both translation directions.
- Target Language by Prime Type by Target Frequency by Proficiency: to test the role of the interaction between Target Frequency and Proficiency as a potential modulator of priming effects across both translation directions.
- Target Language by Prime Type by Target Frequency by Language Exposure/use: to test the role of the interaction between Target Frequency and L2 exposure/use as a potential modulator of priming effects across both translation directions.

**References**


Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

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*(4), 390–412.

Baayen R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & Cognition*, 35(5), 953–965.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Brenders, P., Van Hell, J., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, 109(4), 383-396.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental psychology*, 58(5), 412-424.

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(3), 359–371.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, 20(3), 530-548.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45-50.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in psychology*, 7, 1116.

Brysbaert, M., Verreyt, N., & Duyck, W. (2010). Models as hypothesis generators and models as roadmaps. *Bilingualism: Language and Cognition,* 13(3), 383–384.

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society,* 13(4), 212–214.

Clahsen, H., Balkhair, L., Schutter, J. S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research, 29*(1), 7–31.

Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 82(6), 407–428.

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 308–325). New York: Oxford University Press.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language*, 41(3), 365–397.

Council of Europe (2011). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council of Europe.

Cuetos, F., Glez-Nosti, M., Barbon, A., & Brysbaert, M. (2011). Subtlex-ESP: Spanish word frequencies based on film subtitles. *Psicológica,* 32, 133–143.

De Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The Processing of Interlexical Homographs in Translation Recognition and Lexical Decision: Support for Non-selective Access to Bilingual Memory. *The Quarterly Journal of Experimental Psychology,* 53A(2), 397–428.

De Groot, A. M. B., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30(1), 90–123.

Diependaele, K., Lemhöfer, K, & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, 66(5), 843–863.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language*, 41(4), 496–518.

Dijkstra, T., & Rekké, S. (2010). Towards a localist-connectionist model for word translation. *The Mental Lexicon,* 5(3), 403–422. Special issue on *Methodological and analytic frontiers in lexical research*, edited by G. Jarema, G. Libben, & Ch. Westbury.

Dijkstra, T., & Van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A.M. Jacobs (Eds), *Localist connectionist approaches to human cognition* (pp. 189–225). Mahwah, NJ: Lawrence Erlbaum.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to system. *Bilingualism: Language and Cognition*, 5(3), 175–197.

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

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Masked translation priming effects with low proficiency bilinguals. *Memory & Cognition*, 39(2), 260–275.

Dunn, A. L., & Tree, J. F. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition*, 12(3), 273–289.

Duñabeitia, J. A., Dimitropoulou, M., Uribe-Etxebarria, O., Laka, I., & Carreiras, M. (2010). Electrophysiological correlates of the masked translation priming effect with highly proficient simultaneous bilinguals. *Brain Research*, 1359, 142–154.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, 57(2), 98–107.

Duñabeitia, J. A., Dimitropoulou, M., Morris, J., & Diependaele, K. (2013). The role of form in morphological priming: Evidence from bilinguals. *Language and Cognitive Processes*, *28*(7), 967–987.

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language. New evidence from Dutch-French bilinguals. *Experimental Psychology*, 56(3), 173–189.

Feldman, L.B., Milin, P., Cho, K.W., Moscoso del Prado Martín, F., & O'Connor, P.A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience, 9*:111.

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

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 680–698.

Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *UCLES Research Notes, 12*, 8–10.

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2), 303–315.

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

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1122–1139.

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

Jiang, N. (2015). Six decades of research on lexical representation and processing in bilinguals. In J. Schwieter (Ed), *The Cambridge Handbook of Bilingual Processing* (pp. 29–85). Cambridge, MA: Cambridge University Press.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, 42(3), 627-633.

Ko, Y., Wang, M., & Kim, S. Y. (2011). Bilingual reading of compound words. *Journal of Psycholinguistic Research*, 40, 49-73.

Kroll, J. F., & 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.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 531–553). New York, NY: Oxford University Press.

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(3), 373–381.

Lee, Y., Jang, E., & Choi, W. (2018). L2-L1 Translation Priming Effects in a Lexical Decision Task: Evidence From Low Proficient Korean-English Bilinguals. *Frontiers in Psychology,* 9(267), 1–10.

Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and cognitive processes,* 19(5), 585–611.

Lijewska, A., Ziegler, M., & Olko, S. (2018). L2 primes L1 – translation priming in LDT and semantic categorisation with unbalanced bilinguals. *International Journal of Bilingual Education and Bilingualism,* 21(6), 744–759.

Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 480–487.

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.

McClelland, J. L., & Rumelhart, D. E. (1981). An Interactive Activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review,* 88(5), 375–407.

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(5), 1001–1022.

Nakayama, M., Lupker, S. J., & Itaguchi, Y. (2018). An examination of L2-L1 noncognate translation priming in the lexical decision task: insights from distributional and frequency-based analyses. *Bilingualism: Language and Cognition,* 21(2), 265–277.

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(8), 949–981.

Norris, J.M. & Ortega, L. (2012). Assessing learner knowledge. In S.M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 591–607). New York: Routledge.

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

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods,* 162(1–2), 8–13.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology,* 55A(4), 1339–1362.

Rothman, J. & Iverson, M. (2010). 'Independent multilingualism normative assessments, where art thou?' In M. Cruz-Ferreira (Ed.), *Multilingual Norms* (pp. 33–51). Berlin: Peter Lang.

R Core Team. (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/.

Sabourin, L., Brien, C., & Burkholder, M. (2014). The effect of age of L2 acquisition on the organization of the bilingual lexicon: Evidence from masked priming. *Bilingualism: Language and Cognition*, 17(3), 542–555.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition*, 37(5), 569–586.

Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics*, 5(1), 375-393.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic bulletin & review*, 9(4), 780–789.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62(2), 148–171.

Van Heuven, W. J. B., Schriefers, H., Dijkstra, T., & Hagoort, P. (2008). Language Conflict in the Bilingual Brain. *Cerebral Cortex*, 18(11), 2706–2716.

Van Heuven, W. J. B., 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(6), 1176–1190.

Wang, M. (2010). Bilingual compound processing: The effects of constituent frequency and semantic transparency. *Writing Systems Research*, 2(2), 117-137.

Wang, X. (2013). Language dominance in translation priming: Evidence from balanced and unbalanced Chinese–English bilinguals. *The Quarterly Journal of Experimental Psychology*, 66(4), 727-743.

Wen, Y., & Van Heuven, W. J. B. (2017). Noncognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Pyschonomic Bulletin & Review*, 24(3), 879–886.

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.

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.

Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 801–806). Austin, TX: Cognitive Science Society.

**Tables**

Table 1. Participant characteristics. Mean values (standard deviations and ranges).

| | |
|---|---|
| Age (years) | 30 (5.1; 19–39) |
| Self-reported English proficiency (max: 10) | 5.6 (0.6; 4.5–7) |
| Oxford Quick Placement Test scores (max: 60) | 50 (4.8; 40–60) |
| Language Exposure/Use | 12 (6.4; -2–24) |
| Age of acquisition (years) | 9 (2.9; 4–16) |
| Time living in the UK (years) | 5 (3.6; 0.75–13) |

Table 2. Sample stimuli.

| **L1-L2** | | | |
|---|---|---|---|
| **Translation prime** | **Control prime** | **Word target** | **Nonword target** |
| *flecha* ‘arrow’ | *cereza* ‘cherry’ | ARROW | SMOUNT |
| **L2-L1** | | | |
| **Translation prime** | **Control prime** | **Word target** | **Nonword target** |
| onion | pencil | *CEBOLLA* ‘onion’ | TUNGO |

Table 3. Stimuli characteristics. Average frequency (Zipf scale, range: 1-7), and length (in characters), with standard deviations and ranges in parentheses.

| | **Spanish** | **English** |
|---|---|---|
| **Frequency** | 4 (0.4; 2.7–4.9) | 4 (0.3; 2.9–4.9) |
| **Length** | 6 (1.3; 3–8) | 5.4 (1.2; 3–8) |

Table 4. Mean RTs (in milliseconds; standard errors), error rates (%), and priming effects (in milliseconds) in the LDT.

| | Related | | Control | | |
|---|---|---|---|---|---|
| | **RT** | **Error Rate** | **RT** | **Error Rate** | **Priming** |
| **L1 to L2** | 718 (5.3) | 2.4 | 757 (6.9) | 3.1 | 39* |
| **L2 to L1** | 721 (5.1) | 1.5 | 759 (6.7) | 1.8 | 38 |
| **Difference** | | | | | 1 |

Note: * = $p < .01$

Table 5. Intercept and significant or marginally significant factors included in the final model for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

| | **Coefficient** | **Std. Error** | **t-value** | **p-value** |
|---|---|---|---|---|
| Intercept | -1429.35 | 21.12 | -67.68 | < 0.001 |
| Target Language (English) by Prime Type | -59.02 | 14.20 | -4.16 | < 0.001 |
| Target Language (Spanish) by Prime Type by Language exposure/use | 34.10 | 11.45 | 2.98 | < 0.01 |
| Target Language (English) by Prime Type (Control) by Target Frequency | -49.11 | 17.49 | -2.81 | < 0.01 |
| Target Language (English) by Prime Type (Related) by Target Frequency | -42.20 | 19.34 | -2.18 | < 0.05 |
| Target Language (English) by Prime Type (Related) by | 23.63 | 10.43 | 2.27 | < 0.05 |

| | | | | |
|---|---|---|---|---|
| Proficiency by Target Frequency | | | | |
| Target Language (English) by Proficiency | -38.46 | 19.53 | -1.97 | 0.053 |
| Target Language (Spanish) by Prime Type (Control) by Prime Frequency | 19.83 | 11.58 | 1.71 | 0.088 |
| Target Language (Spanish) by Prime Type (Related) by Prime Frequency | -29.09 | 16.18 | -1.80 | 0.072 |

Table 6. Intercept and significant or marginally significant factors included in the final model for the analysis of accuracy and their coefficients, standard errors, z-values, and p-values.

| | Coefficient | Std. Error | z-value | p-value |
|---|---|---|---|---|
| Intercept | 4.62 | 0.40 | -27.83 | < 0.001 |
| Target Language | 1.54 | 0.66 | 2.33 | < 0.05 |
| Target Language (Spanish) by Prime Type (Control) by Target Frequency | 1.22 | 0.48 | 2.55 | < 0.05 |
| Target Language (Spanish) by Prime Type (Related) by Target Frequency | 1.25 | 0.49 | 2.54 | < 0.05 |
| Target Language (Spanish) by Prime Type (Control) by Proficiency by Prime Frequency | 0.73 | 0.31 | 2.34 | < 0.05 |
| Target Language (English) by Proficiency | 0.42 | 0.24 | 1.78 | 0.08 |
| Target Language (English) by Prime Type (Related) by Target Frequency | 0.82 | 0.43 | 1.92 | 0.06 |

| | | | | |
|---|---|---|---|---|
| Target Language (English) by Prime Type (Control) by Language exposure/use by Prime Frequency | -0.35 | 0.19 | -1.84 | 0.07 |