The nature of the bilingual lexicon: Investigating cross-language priming and its predictors

PhD in Psycholinguistics

School of Psychology and Clinical Language Sciences

Adel Chaouch Orozco

July 2021

Abstract

A growing consensus in bilingual lexical processing research sees the bilingual lexicon as a non-selective system where words from both languages are activated simultaneously during comprehension and production. Less is known about how exactly words are represented and intertwined at different levels and what factors shape the lexicon. The present dissertation investigates the bilingual lexicon's organization and functioning, employing cross-language visual priming, one of this field's most productive methodological tools. We assess the contribution of individual-, word- and methodology-level factors. The first study explores the effects of second language (L2) proficiency, L2 use and word frequency in translation masked priming. The second study, employing unmasked primes, focuses on investigating L2 use (L2 proficiency is factored out), word frequency, executive control and degree of semantic overlap. Taking advantage of distributional analyses, the third study compares the cognitive processes recruited during masked and unmasked priming experiments. Overall, our results show that word frequency—prime frequency in particular—is a robust predictor of cross-language priming. The contribution of proxies of bilingual experience is more elusive. Some evidence suggests that language use plays a relevant role in the bilingual lexicon's functioning, while the effect of L2 proficiency in our data is negligible. Further, we show that the degree of semantic overlap and executive control modulate the priming effects. Finally, our results indicate that cognitive recruitment differs in masked and unmasked priming experiments. The findings in this dissertation have several implications: 1) Experiential and word-level factors

should be further investigated in this type of studies, especially in a continuous manner and focusing on interactions. 2) The models of the bilingual lexicon would benefit from adopting a distributed view of semantic representation. 3) Studies examining response times should include distributional analyses as they provide unique insights into the cognitive mechanisms at play.

Declaration of authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged. I also declare that the three studies in this thesis report my original work in collaboration with my supervisors and have been either published or submitted, and that I am the first author on these papers.

Adel Chaouch Orozco

Study 1

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(2), 447-474.

Study 2

Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (submitted). What shapes the bilingual lexicon? Insights from a comprehensive cross-language priming study.

Study 3

Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (submitted). Beyond the mean: Shape distribution analysis provides unique insights into how masked and unmasked cross-language priming effects unfold.

Acknowledgements

If there is something I can say about these three years is that I feel a fortunate person. I am going to try to briefly acknowledge why.

I feel truly privileged to have had the opportunity to do research during the last three years. That would not have been possible without the studentship provided by the University of Reading and the School of Psychology and Clinical Language Sciences. I am enormously grateful for this.

I will never forget all that I have learned from Jason Rothman; without him, I would not be the researcher I am today. Thank you for the opportunities you have brought to my life and the confidence you have placed in me during these years. Working with you has been a constant learning and growing process at the personal and professional levels. I am sure I will keep learning from you in the years to come. My deepest gratitude for being always by my side.

I am so lucky for having devoted my professional life to the pursuit of knowledge—in my mind, the highest aspiration a human being can have. More importantly, I have done so surrounded by world-class researchers and amazing people. A huge thanks to Ludovica Serratrice, Ian Cunnings and Christos Pliatsikas for all the moments together and your support and advice.

There are two people I will always feel in debt to. They have been there to help me whenever I have needed them. I could not have learnt the ropes from better people. They are professional referents to me, and there is no way I can pay them back all they have done for me. If I get any close to Jorge González Alonso and Eloi Puig Mayenco, I will be able to confidently say that I am on the path to become an excellent researcher. *Mil gracias, chicos.*

I feel now I am part of a family—a beautiful concept revealing that the bonds bringing us together are not merely professional but also emotional. I have been fortunate to meet and share experiences—sometimes in foreign countries and under *interesting* circumstances—with many exceptional people during these years. Thank you for enriching my life, Yesi, Toms, Dave, Fatih, Fernando, Vince, Sergio, Nico, Hiroki, Jen, Tiffany and Mike.

Part of this dissertation is my mum's too. I would not be here without her insistence and her constant support. Above all, I would not be a researcher if I had not inherited her curiosity for all aspects of life; if there is someone I can discuss everything with, that is she. *Mare, espero que sientas que este proyecto es también tuyo. Gracias por todo. Poquet a poquet…*

Lastly, the person illuminating my life for the last nine years deserves some credit. This adventure would not have been as fun and endurable without you, Mutlu. Thank you for all your constant and unconditional love and support and for making me a better and happier person every single day. We have grown a lot together, and I am sure this is just the beginning. Let's continue the adventure! *Seni çok seviyorum.*

Table of Contents

Chapter 1: Introduction

1.1 General introduction

This dissertation presents three studies investigating the bilingual lexicon, its architecture and functioning, and the factors that shape it. Moving beyond the study of aspects of the lexicon on which a certain degree of consensus has been achieved (e.g., the integrated and non-selective nature of the lexicon), this dissertation focuses on how cross-language words are connected and interact together. Specifically, we examine how individual differences, the characteristics of words, and methodological factors affect cross-language lexical-semantic connections and the underlying cognitive processes operating in the lexicon. To do so, we investigate late sequential bilingual visual word recognition with the cross-language priming methodology. This dissertation has three main goals:

1. Investigating how subjective word frequency and executive control affect cross-language priming.
2. Exploring how the degree of semantic overlap between stimuli pairs impacts visual word recognition and priming effects.
3. Examining how the type of prime presentation influences the cognitive processes recruited during task completion.

The first study reports a translation priming experiment with late sequential bilinguals examining the interplay between word frequency and two proxies of second language (L2) development (L2 use and L2 proficiency). The second study, employing a similar methodology and participants, continues the exploration of individual factors and word-level variables that may influence bilingual word processing. Both translation equivalent and cross-language semantically related pairs are investigated. Compared to the first study, the scope of the second one is expanded by placing a more significant focus on semantic aspects. Also, in a novel manner in the examination of translation and cross-language semantic priming effects, a potential role of executive control is explored. The third study investigates how the different presentation procedures employed in the experiments of the first two studies may give rise to the use of different cognitive mechanisms during the completion of the tasks. In sum, the three studies presented in this dissertation aim to comprehensively explore the factors potentially shaping the way bilinguals represent and visually recognise words in their two known languages. We believe that the findings presented here make a solid contribution to the understanding of the bilingual lexicon and the processes that occur within it during visual word recognition.

1.2 The study of the bilingual lexicon

The study of bilingualism has acquired an independent status from that of a single language. The reason lies in the exclusive questions that knowing two (or more languages) posit about how language representation and processing occurs. In the particular case of the investigation of how bilinguals learn and process words, an immediate question arises: What does knowing words from two different languages imply for how words are recognised? Are words from different languages stored physically together? Does activation of the words in the two languages follow the same lexical-semantic routes? Does activation co-occur cross-linguistically? These are just some of the questions that one may ask when considering how a monolingual lexicon and its functioning differ from a bi-/multilingual one.

Research during the last four decades seems to be reaching an agreement on some of these matters. For instance, evidence has accumulated on a view of the bilingual lexicon as an integrated, non-selective network, where, instead of two potentially independent systems, words from the two languages are stored together, and word activation occurs simultaneously (for a review, see Brysbaert & Duyck, 2010; Dijkstra, Wahl, Buytenhuijs, van Halem, Al-Jibouri, de Korte, Rekké, 2019). Reports of words from one language influencing the processing of words from the other are taken as support for this view. For example, cross-language influences have been reported with interlingual homographs (e.g., Dijkstra, Timmermans & Schriefers, 2000), cognates (e.g., van Hell & Dijkstra, 2002, with isolated words; Duyck, van Assche, Drieghe & Hartsuiker, 2007, in sentential contexts), or cross-language neighbours (Meade, Midgley, Dijkstra & Holcomb, 2018). Although these studies can be taken as evidence for a unitary lexicon, it can also be argued that only access is parallel (i.e., non-selective), while both lexicons remain separated—connections at the lexical level would guarantee parallel activation (see Kroll, van Hell, Tokowicz & Green, 2010). Whether integrated or not—indeed, not an easy question to prove empirically—there is little doubt on the non-selective nature of the bilingual lexicon. However, less is known about other intriguing questions, such as how words from the two languages are connected or how exactly access to meaning occurs for words in the second language. For instance, it is commonly assumed that there is a shared conceptual store for L1 and L2 words (French & Jacquet, 2005; Kroll & Tokowicz, 2005). Nonetheless, it is not clear whether access to concepts occurs equally for L1 and L2 lexical representations (e.g., Kroll & Stewart, 1994). Further, recent studies, like Thompson, Roberts and Lupyan (2020), suggest that L1 and L2 conceptual spaces may not be entirely overlapping and word meanings may not arise from a pre-existing, universal perceptual system. This relativist view spotlights the potential inadequacy of assumed one-to-one semantic mappings between translation pairs (more in line with a universalist approach). Related to this, concreteness may affect cross-language semantic consistency—concrete meanings are assumed to show less variability across-languages (e.g., Thompson et al., 2020). The reason may lie on a higher linguistic context dependency for abstract words (e.g., Crutch & Warrington, 2005; Pexman, Siakaluk & Yap, 2013; and Li, Liang, Qu, Sun, Jiang & Mei, 2021, for neuroimaging evidence), contrary to the richer and more cross-linguistically-constant nature of concrete meanings. Importantly, as we will see below, these word-type differences may have consequences for bilingual semantic memory and bilingual word processing (Tokowicz, Kroll, de Groot & van Hell, 2002; van Hell & de Groot, 1998).

In this line, the current project investigates how words from different languages are represented and connected, focusing on two types of cross-language pairs. i) Non-cognate translation equivalents (e.g., *vaca*, 'cow' in Spanish, and *cow*); that is, cross-language word pairs with a maximal overlap in meaning but not in form, where the opportunity for links established at the lexical level is minimised. ii) Cross-language semantically related pairs (e.g., *vaca* and *milk*), where word-association links are assumed to be non-existent. To study this, we employed the visual cross-language priming methodology, one of the most common ways of exploring connections within the lexicon.

1.3 Cross-language priming and the study of the bilingual lexicon

In a typical visual priming paradigm, a target word is presented before a prime. In critical trials, the prime is a related word (e.g., semantically, morphologically, orthographically), whereas, in the control condition, prime and target are unrelated words. One of the commonly used tasks with this paradigm, and the one employed in our studies, is lexical decision, where participants judge if the target is a real word or not—usually, half of the targets consists of pseudowords. Priming effects result from differences in response times (RT) and error rates between the treatment and control conditions.

The common assumption is that priming effects reflect spreading activation processes (e.g., Collins & Loftus, 1975). Under the spreading activation theory, the lexicon is a network consisting of nodes representing words or concepts related to each other at different levels. When a particular word or concept is stimulated, activation spreads through the network to other related units that also become exited to some extent. Thus, priming effects are usually interpreted as activation transferring from the related prime's processing to that of the target, causing a facilitatory effect.[1] In cross-language priming, the effect has been taken to evidence the interrelation of L1 and L2 words. For instance, if response times to *cow* are faster than to *wood* when these two words are preceded by *vaca* (Spanish for 'cow'), this suggests that activation has spread from *vaca* to *cow* (but not from *vaca* to *wood*), facilitating the processing of the target *cow*.

In cross-language studies, the paradigm has been extensively used with non-cognate translation equivalents and, especially, in masked presentations (Forster & Davis, 1984), where a forward mask (e.g., #####) precedes a prime that is usually shown for less than 60 ms (Forster, Mohan, Hector, Kinoshita & Lupker, 2003). This presentation procedure makes the prime less salient and its processing potentially unconscious. Therefore, masked priming paradigms allow tapping into more automatic processes while avoiding the risk of participants adopting strategic processing (i.e., translating the prime). Notably, this type of presentation has been the choice of preference in much of the research with non-cognate translation equivalents, especially in more recent times.

Although the early priming experiments with non-cognate translation equivalents employed overt prime presentations (e.g., Altarriba, 1992; Keatley & de Gelder, 1992), authors promptly switched to masked paradigms (e.g., Williams, 1994; Grainger & Frenck-Mestre, 1998). Interestingly, under masked presentations, a typical pattern of results began to emerge: Priming effects were larger and more robust in the L1-L2 direction (L1 primes – L2 targets) than in the

[1] Priming effects could also arise from inhibitory effects; that is, the processing of a control prime would interfere with that of the target, delaying response in the task for those pairs (see, e.g., discussion in McNamara, 2005).

opposite one (*masked priming asymmetry*; see Wen & van Heuven, 2017, for a review). This finding has drawn the attention of researchers during the last two decades because of the implications it is supposed to have for the models of bilingual lexical-semantic representation and processing.

1.4 Models of the bilingual lexicon

The Revised Hierarchical Model (RHM; Kroll & Stewart, 1994; Kroll et al., 2010) is one of the first and most successful—in terms of the amount of research and discussion generated—models of the bilingual lexicon. Although the RHM was proposed to account for production and, in particular, translation effects, its tenets have been widely discussed within the cross-language priming literature too. In this model, L1 and L2 words are represented separately at the lexical level while sharing a conceptual store. Crucially, the retrieval of those shared semantic representations does not occur equally for L1 and L2 words. As a result of how late sequential bilinguals usually learn L2 words (i.e., via the L1 translation equivalents), the model assumes direct access to meanings for L1 lexical representations, with weaker access for L2 words. Moreover, L2 words would be strongly connected to L1 words at the lexical level—their semantic route being indirect and mediated by the L1 translation equivalents. Hence, the RHM proposes an asymmetrical configuration of the bilingual lexicon, with qualitative differences in place. Crucially, connection weights may change with L2 development, leading to more robust lexical-semantic links for L2 words.

Notably, priming asymmetries between translation directions can be explained by these asymmetrical connections, as long as one assumes that priming effects are semantically mediated—especially in the case of non-cognate words (Xia and Andrews, 2015). In such a scenario, processing an L1 prime would guarantee sufficient activation of the shared conceptual node, which would stem down to the L2 lexical representation of the target, resulting in priming. In the opposite direction, access for the L2 primes to the semantic store would be neglected (at least to some extent), disallowing for L2-L1 priming effects to appear. Importantly, if this account is on the right track, the model makes a precise prediction: Increased L2 development would lead to larger L2-L1 priming effects because of stronger L2 lexical-semantic connections.

In contrast to the verbal nature of the RHM, the Bilingual Interactive Activation model (BIA; Dijkstra & van Heuven, 1998) and its successor, the BIA+ (Dijkstra & van Heuven, 2002), are computational, connectionist models of visual word recognition. The recent Multilink (Dijkstra et al., 2019), following the same computational fashion, aims at providing a comprehensive model of bilingual lexical processing while overcoming the intrinsic limitations in the scope of the RHM and the BIA+ by adopting assumptions from both views. Multilink presupposes an integrated and non-selective lexicon, where the differences between L1 and L2 words are only quantitative. Divergent L1/L2 word processing results from the confluence between the impact of differential connection weights at different levels and a baseline activation parameter that determines the ease for a word to reach a recognition threshold. Notably, this baseline activation (*resting level activation*; RLA) would be dependent on subjective word frequency (i.e., the frequency of a particular individual encountering a particular word). That is, each individual's experience with each word (irrespective of its language membership) would influence the speed of processing of that word (i.e., the time taken to reach a threshold of recognition).

Although measuring such a slippery construct is no easy task (but see, e.g., Balota, Pilotti & Cortese, 2007), subjective word frequency can be operationalized in different ways. For instance, the most immediate candidate for proxying the factor would be standard word frequency. However, despite the success in applying modern frequency measures to predict mono-/bilingual processing (e.g., Brysbaert, Mandera & Keuleers, 2018), we consider that such measures are limited by their own nature—they are overgeneralizations. Further, in the study of bilingual populations, relying on frequency variables based on L1 corpora may be problematic. For instance, it is an empirical question whether the predictive effectiveness of the variable is similar for immersed and non-immersed bilinguals. After all, immersed bilinguals' experience with the L2 should be more similar to that of the native speakers living in the society where the language at stake is the majority language.

Moreover, although standard frequency would be a good candidate for proxying subjective frequency—perhaps as a baseline indicator—the amount of exposure to and active use of the L2 should also serve—if not equally, at least similarly—as an approximation to subjective experiences with words. The logic behind this claim is simple: Increased L2 use would result in more opportunities for encountering L2 words. Yet, increased experience with the L2 would not be the only way of measuring bilingual experience and subjective word frequency; L2 proficiency could also serve as a proxy. Nevertheless, in our opinion, the latter measure should, in principle, be less relevant—and, therefore, a worse predictor—of subjective word frequency. Passed a certain proficiency threshold, enough for an L2 speaker to communicate efficiently, the factor should not *per se* be reflective of bilingual experience. Even though high L2 proficiency correlates with more L2 use on many occasions. This logical interdependence between L2 use and proficiency—and the relative ease of measuring the factor—makes proficiency a commonly employed proxy of subjective word frequency and bilingual experience. The present project investigates how subjective word frequency can be operationalized more efficiently and explores the interplay between its potential proxies.

Another fundamental aspect of the RHM's and Multilink's proposals that is crucial to this project is the adoption of holistic semantic representations. However, this is not the only way such representations can be conceptualised. The Distributed Feature Model (DFM; de Groot, 1992; van Hell & de Groot, 1998)—and distributed views of semantic representation, in general—assume non-overlapping meanings.[2] For the DFM, semantic representations consist of semantic features (e.g., a dog is an *animal*, a *pet*, and it *has four legs*). Further, connection weights across conceptual features determine the meaning of words. Notably, being a connectionist model, activating a word's features results in

[2] Dijkstra et al. (2019) acknowledged that other alternatives to a holistic view may be considered in future instantiations of Multilink.

parallel (partial) activation of other words sharing those features across the network. Thus, if *dog* becomes activated, *cat* would be expected to do so as well, at least to some extent, as both a dog and a cat are animals, pets, and have four legs.[3]

Whereas Multilink and the DFM share an underlying cognitive mechanism (i.e., spreading activation), the DFM's more nuanced view about meaning representation and the role given to connection weights between features make different predictions concerning how semantic memory influences processing speed and priming effects. For instance, the DFM can account for translation priming effects in the following way: because the feature overlap between translation equivalents is believed to be large, activation at the semantic level would be stronger, leading to enhanced facilitation and faster responses. Further, it can predict frequency effects on priming. For instance, more frequent L2 related primes would have richer feature representations, resulting in a more considerable ability to stimulate their target counterparts at the semantic level, leading to faster processing and increased priming (see Finkbeiner, Forster, Nicol & Nakamura, 2004, for a similar account).

Importantly, the three models discussed here pretty much align on the general role that experience-related factors may have on cross-language priming effects (albeit the underlying mechanisms triggered by the predictors differ in each model). However, there is a fundamental difference between the models

[3] We are focussing on discussing the DFM, a feature-based view. However, alternative ways of understanding semantic representations include some where features are not relevant, like in models arising from the distributional hypothesis (Harris, 1970) (see Kumar, 2021, for further discussion).

when it comes to predicting priming effects with concrete and abstract words. As noted, concrete words' meanings are assumed to be richer (i.e., more detailed) and consistent across languages than abstract ones. Thus, the DFM would predict larger priming for concrete translation equivalent pairs. The stronger stimulation provided by activating a higher number of shared features in concrete related trials would facilitate target processing further than with abstract pairs, where fewer features are shared. Note that the RHM and Multilink would remain agnostic as per an effect of concreteness on translation priming.

In sum, Multilink represents the state-of-the-art computational model of the bilingual lexicon, and simulations are showing to correlate well with empirical data. In addition, the model has great research productivity, generating new hypotheses to test. An excellent example of this latter point is the need to understand subjective word frequency in bilingual speakers better. In this sense, investigating potential proxies for the predictor such as L2 proficiency, L2 use and standard word frequency and their interplay can shed light on the nature of the factor. In addition, the RHM's and Multilink's adoption of holistic semantic representations contrasts with alternative distributed views of meaning and contradicts some recent findings on the nature of translation equivalents' meanings (e.g., Thompson et al., 2020).

Finally, an important aspect of the present project not addressed by the models concerns the role of executive functioning on priming effects. As we discuss below, bilinguals with better switching abilities may be able to suppress competition from non-target word candidates in priming studies, or they can be better at keeping language sets in memory and sticking to the rules of the task.

To examine these questions, the first and second studies in the present dissertation investigate cross-language priming with an eye on individual differences and stimuli-level factors. Crucially, the experiments in both studies have slightly different presentation procedures, allowing us investigating the effect of this methodological manipulation on cognitive involvement. The first study examines the masked priming asymmetry phenomenon, focusing on the role played by L2 proficiency, L2 use and word frequency. The second study employs overt primes, and its design provides a thorough investigation of L2 use and word frequency while also addressing the impact that the degree of semantic overlap and executive control have on cross-language priming. The third study explores the effect of manipulating the prime presentation (masked vs unmasked) on the recruitment of cognitive mechanisms during late sequential bilingual visual word recognition. Thus, this dissertation provides a comprehensive examination of the predictors (at the individual-, word-, and methodological-level) influencing cross-language priming effects and the bilingual lexicon's architecture and functioning.

The following sections discuss how the study of the present factors of interest has been approached in cross-language priming studies with late sequential bilinguals.[4]

1.5 Potential modulators of cross-language priming effects

[4] More detailed state of affairs for each specific topic are provided in each study's chapters.

Even though L2 proficiency holds a prominent status in all models of bilingual word comprehension and production, the factor has been scarcely investigated in cross-language priming studies, and the results about its role are inconclusive. For instance, in what can be considered the most exhaustive examination of L2 proficiency in masked priming asymmetry research, Dimitropoulou, Duñabeitia and Carreiras (2011) observed similar relatively small priming effects in three different L2 proficiency groups, concluding that the factor did not play a significant role in their data. Contrary to that, Nakayama and colleagues claimed that L2 proficiency was crucial for the emergence of L2-L1 masked priming in light of the results of two studies with Japanese-English bilinguals (Nakayama, Sears, Hino & Lupker, 2013; Nakayama, Ida & Lupker, 2016).

The first study presented here aimed at shedding light on these conflicting results by investigating the predictor in a continuous manner, contrary to the categorical use of the variable by Dimitropoulou et al. and Nakayama et al. Importantly, it did so while also exploring the interplay between L2 proficiency, L2 use and word frequency.

The comprehensive study of predictors of L2 experience (e.g., L2 use) has also received relatively little attention in previous cross-language priming studies. This is despite BIA+ and Multilink specifically stating the theoretical importance of the amount of L1/L2 use in shaping RLA—while L2 experience can also be assumed to be relevant for both the RHM's and the DFM's predictions. In a study with immersed and non-immersed highly proficient bilinguals, Zhao, Li, Liu, Fang and Shu (2011) observed significant L2-L1 masked priming with the immersed group. Their results suggested that immersion (i.e., active exposure to the L2) plays a crucial role in enhancing L2 primes' ability to facilitate target processing. Language dominance has also been shown to be crucial for the priming asymmetry to be obtained (e.g., Duñabeitia, Perea & Carreiras, 2010; Wang, 2013).

Interestingly, despite the great attention that word frequency has received in the visual word recognition literature, the examination of the factor's role on cross-language priming has been rare. To the best of our knowledge, only Nakayama, Lupker and Itaguchi (2018) investigated the predictor directly. Their results with masked L2-L1 data and Japanese-English bilinguals suggested a relevant role of the predictor in translation priming. In particular, significant priming effects were obtained when trials consisted of both high-frequency primes and low-frequency targets. The authors concluded that L2-L1 priming arose only in optimal conditions—when prime processing was faster, and there was more time for it to be completed (i.e., when responding to less frequent, more difficult targets).

The effect that the degree of semantic alignment has on cross-language priming has been examined in several studies, yet again, whether the type of word pair affects the size of cross-language priming is unclear. For instance, Schoonbaert, Duyck, Brysbaert and Hartsuiker (2009) tested Dutch-English bilinguals and showed that the effect of concreteness on the asymmetrical priming patterns was negligible. However, the asymmetry was larger for translation equivalents compared to cross-language semantically related pairs. Similarly to

Schoonbaert and colleagues' results, a lack of concreteness effect has been replicated in other studies (e.g., Chen, Liang, Cui & Dunlap, 2014; Smith, Walters & Prior, 2019). As per the finding with cross-language semantic priming, Basnight-Brown and Altarriba (2007) observed larger priming effects for translation pairs than for cross-language semantically related pairs when testing Spanish-English bilinguals. Notably, this occurred with 100 ms unmasked primes (Experiment 1). With 50 ms masked primes (Experiment 2), the significant effect remained for translation pairs but not for the semantically related pairs.

Finally, executive control (EC) may play a relevant role in cross-language priming effects. To the best of our knowledge, this hypothesis has not been tested yet with translation equivalents or cross-language semantically related pairs, but some recent studies seem to point towards an involvement of EC in this type of experiments. For instance, in a priming study with English-French homographs and highly proficient bilinguals, Friesen and Haigh (2018) observed a significant effect of EC on priming. The authors concluded that enhanced executive functioning might better control the potential competition from non-target cross-language words. A similar finding was reported by Freeman, Blumefeld and Marian (2017) in a study of L2 phonological priming examining L1 phonotactic constraints. Therefore, it is possible that EC is involved in regulating the within- and between-language competition during target word selection or keeping language sets in memory and sticking to the rules of the task.

In sum, all the above factors have been suggested to regulate the (bilingual) lexicon's organization and functioning. Cross-language priming, assumed to reflect spreading activation between units (i.e., words and concepts) within the network, offers an ideal tool for investigating these predictors' influence on lexical-semantic processing. However, some authors have argued that spreading activation may not be the only cognitive process involved in priming. As we discuss in the next section, unmasked paradigms may rely on episodic memory recruitment too. Further, the variables investigated and discussed in this dissertation may impact the likeliness of that episodic memory involvement. These hypotheses may have consequences for the conclusions that can be drawn from priming studies. Crucially, to explore these predictions, one should employ distributional analyses instead of the more traditional mean-based ones.

1.6 Distributional analyses and the nature of priming effects

Ever since the beginning of chronometrical research in psychology, an overwhelming majority of studies measuring response times (RT) have relied on analyses of the mean to examine the cognitive processes believed to be reflected by response latencies (Balota & Yap, 2011). Although there are good theoretical and practical reasons for the ubiquity of such an approach, it also presents limitations worth discussing. Beginning with the advantages, the mean—any measure of central tendency, in fact—is believed to offer a good representation of the entire data. In addition, it is easy to calculate and rather stable—mean estimates are relatively replicable across studies. Nevertheless, there are significant caveats to its use. First, means are highly sensitive to outliers and skewness. To put a remedy on that, transforming distributions and removing outliers is standard practice—despite the lack of agreement on how that should be done (Cousineau & Chartier, 2010). Nevertheless, almost invariably, RT distributions present a long tail in one of its extremes. More importantly, on many occasions, the experimental effect one is aiming to observe locates in that tail.

Thus, mean-based analyses are not exempt from weaknesses. However, some of these drawbacks can be overcome by exploring RT distributions. There are several ways of performing distributional analyses. In the third study, we opted by fitting a mathematical function to the raw data as the primary analysis. The ex-Gaussian is one of the most commonly employed functions, as its parametric density estimators can be closely related to those of empirical distributions (Van Zandt, 2000). This distribution consists of a normal (Gaussian) and an exponential distribution. In the normal distribution, the μ and σ parameters reflect the mode and the variance, respectively, whereas τ reflects the mean, the standard deviation and the overall skewness in the exponential distribution.

As we show in the third study, the mean hides changes in the distributions' shapes. That is, two distributions may have the same mean values but different forms. Notably, the different shapes (caused by the effect of the treatment condition) have been argued to reflect distinct cognitive processes operating in priming studies. In this sense, a few studies in the last decade have investigated the effect that the experimental condition has on distributions under masked and unmasked priming paradigms and what that may reveal about the processes involved (e.g., Gomez, Perea & Ratcliff, 2013).

For instance, a relatedness effect (i.e., priming effect) in lexical decision tasks with masked primes may originate from a headstart (e.g., Forster et al., 2003). The processing of a related prime activates some of the features of the target. In other words, part of the prime's processing is transferred to that of the target. Crucially, if that is the only mechanism involved, the time taken to respond should not affect the size of the priming effects. Hence, an equal shift across the distributions (for the related and control condition) is expected. This would be reflected by a change in the μ parameter. Alternatively, distributions may not only shift but also skew, showing a change in both μ and τ. According to Ratcliff & McKoon (1988), such a pattern is to be expected under unmasked presentations and would indicate the ability of primes to create episodic memory traces and joint retrieval cues along with the targets. The familiarity of the prime-target cue would influence how fast participants respond to the lexical decision. The greater the familiarity between prime and target, the faster the response, and the larger the priming effect. Importantly, evidence on the cue familiarity accumulates over time; this is why the priming effect is expected to be larger in the longer quantiles (i.e., the distributions shift and skew). Note that Bodner & Masson (2001, 2003) reported a skewness occurring with masked primes too, contrary to the expectation that unconsciously processed primes cannot create memory traces. The authors employed masked identity and semantic/associative priming and found that the priming effects were affected by the ratio of critical trials in the lists. Priming was larger when the number of critical trials was increased (0.80). This finding led the authors to suggest that the enhancement of the prime's utility allows for the prime to create an episodic record despite potential unawareness by the participants (but see, e.g., Kinoshita, Mozer & Forster, 2011).

To the best of our knowledge, Balota, Yap, Cortese & Watson (2008) were the first to investigate distributions with masked and unmasked primes and lexical decision tasks (see, e.g., Kinoshita et al. 2011; Pollastek, Perea & Carreiras, 2005; Ratcliff, Gomez & McKoon, 2004, for alternative approaches). In a series of experiments, they examined how the presentation procedure affected semantic priming effects and their distributions by manipulating the masking of the prime and the presentation of blank screens with different durations after the primes. They observed distributional shifting only (i.e., a change in μ), irrespective of the type of presentation, suggesting a headstart account. In a similar study, Gomez et al. (2013) reported a shift in the distributions with masked presentations, but in contrast to Balota et al., they observed a shift and a change in the distributions with unmasked primes, in line with a compound cue effect.

Although no study has attempted to specifically investigate distributional differences under masked and unmasked presentations with cross-language priming, Nakayama, Lupker and Itaguchi (2018) examined RT distributions with non-cognate translation masked primes. They were interested in exploring the masked priming asymmetry. The authors reported significant effects for μ and τ, which suggested that masked primes could create a memory trace retrieved during target processing. Further, in a more conventional regression analysis, they observed that the effect was driven by the more frequent L2 primes and the less frequent L1 targets. These results raise an intriguing question about the potential role of prime frequency and the speed of processing the prime in the prime's capacity to create a joint retrieval cue along with the target. Moreover, Nakayama et al.'s approach highlighted the importance of adopting new perspectives to examine *old* problems. In the case of distributional analyses, these can be combined with mean-based ones to provide a more exhaustive investigation of the issues at hand.

The third study here aimed at shedding light on the conflicting results from these three studies while, in a novel manner, investigating the effect that manipulating prime presentation has on cross-language priming—instead of monolingual semantic priming. Our data consisted of both non-cognate translation equivalent and cross-language semantically related word pairs. Although, as noted, it is not clear what is the amount of semantic involvement during translation priming in lexical decision tasks, considering that we employed non-cognate words (with reduced opportunities to exhibit lexical connections), we assumed that priming would be semantic, at least to a large extent (see Xia and Andrews, 2015, for a similar argument). Moreover, because lexical connections are assumed to be non-existent, relatedness at the semantic level must be the primary driver of priming effects between cross-language semantically related pairs. Thus, the comparison between relatedness effects in monolingual semantic priming and the two types of cross-language word pairs in the present experiments seems pertinent. Also, in the first two studies, we were interested in exploring the role that different individual- and word-level variables may play in cross-language priming. In the third study, we also examine whether these variables impacted the type of distributions emerging from critical manipulation. Finally, the third study provides an excellent opportunity to compare the results obtained from distributional and more traditional mean-based analyses of the same datasets.

1.7 This dissertation

The present dissertation aims at providing an exhaustive exploration of the factors and mechanisms involved in cross-language visual priming with late sequential bilinguals. The project has three main goals:

1. Investigating how subjective word frequency and executive control affect cross-language priming.
2. Exploring how the degree of semantic overlap between stimuli pairs impacts visual word recognition and priming effects.
3. Examining how the type of prime presentation influences the cognitive processes recruited during task completion.

To this end, the first two studies are devoted to the examination of potential predictors of translation and cross-language semantic priming in Spanish-English bilinguals. The first study explores the masked translation priming asymmetry focusing on the role of subjective word frequency (proxied by L2 proficiency, L2 use and standard word frequency). This is the first attempt in the field to combine the investigation of these three factors measured continuously while also examining the potential interplay of the predictors.

The second study offers a more comprehensive investigation of cross-language priming effects. Employing unmasked primes, it continues exploring subjective word frequency and its role, this time proxied by L2 use and word frequency—while (high) L2 proficiency is kept constant across participants. This way, we could isolate the potential contribution of L2 use in a group of participants at the latest stage of L2 acquisition. Moreover, we were also interested in examining how the degree of semantic overlap influences the ability of the related primes to activate their target counterparts. As discussed, both the RHM and Multilink adopt a holistic way of bilingual semantic representations that contrasts with the distributed approach of the DFM. Thus, observing effects of semantic overlap would suggest that assuming holistic representations may be oversimplistic. Also, we were interested in exploring how executive control affects bilingual visual word recognition in light of recent findings in support of the involvement of this type of mechanisms.

The third study investigates how prime presentation (masked vs unmasked) affects the type of cognitive processes involved in non-cognate translation and cross-language semantic priming. Ours is the first study investigating this matter while employing this type of stimuli. There is some degree of consensus on how semantic priming occurs with masked primes: Being the primes processed unconsciously, they would not be able to create an episodic memory trace to form a joint retrieval cue along with the target (but see discussed contradicting results and alternative hypotheses). Thus, their benefit would only stem from offering a headstart to the processing of the target. On the contrary, with overt primes, a compound cue effect is predicted. However, it is not clear whether i) the same would be observed in a cross-language design and ii) the individual- and word-level factors investigated in this project will affect the ability of the prime to create that joint retrieval cue.

In sum, this dissertation provides a thorough investigation of late sequential bilingual visual word recognition by exploring cross-language priming effects and some of the factors that may affect them. The ultimate goal is to inform the theoretical models on bilingual lexical-semantic representation and processing and to better understand the variables that modulate the bilingual lexicon and the processes taking place within it during visual word recognition.

Chapter 2: Individual differences in bilingual word recognition: The role of experiential factors and word frequency in cross-language lexical priming[5]

Abstract

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.

[5] 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*(2), 447-474.

2.1 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).

2.2 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).

2.2.1 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.

2.2.2 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.

2.3 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.

2.3.1 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.

2.3.2 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.

2.3.3 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 Singapore. 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).

2.4 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.

2.5 Method

2.5.1 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.[6] 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 (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

[6] 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.

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.

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

2.5.2 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 modelling (see below)—that can accommodate deviations from normality in both independent and dependent variables.

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 2. Sample stimuli.

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

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).[7] All nonwords were phonologically and orthographically plausible in Spanish and English, respectively. The complete list of stimuli is provided in Appendix A.

[7] 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.

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.

2.5.3 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.

2.5.4 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.[8]

2.5.5 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

[8] 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.

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 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.

2.6 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)[9]. 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).

[9] 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 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.

2.6.1 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.

	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

Table 4. Mean RTs (in milliseconds; standard errors), error rates (%), and priming effects (in milliseconds) in the LDT. Note: * = $p < .01$.

Following Matuschek, Kliegl, Vasishth, Baayen & Bates (2017), we carried backward-selection and employed the likelihood ratio test criterion to obtain a more parsimonious model. 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. 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).

	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 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.

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.

2.6.2 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.

	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

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.

2.7 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.

2.8 Conclusion 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.

Chapter 3: What shapes the bilingual lexicon? Insights from a comprehensive cross-language priming study[10]

Abstract

A growing consensus sees the bilingual lexicon as an integrated, nonselective system, yet the interplay between speaker- and word-level factors and their role in shaping the lexicon's functioning is less well understood. This study investigates bilingual lexical-semantic representation and processing employing cross-language visual priming. We focus on the role of word frequency and second language (L2) use—as continuous proxies of subjective frequency—and relative semantic overlap. We investigate two novel factors in cross-language priming studies: a word's number of associates and the engagement of executive control. We tested 200 highly-proficient Spanish-English bilinguals differing in L2 use with 400+ word pairs. Results show a robust role of prime frequency, yet are less conclusive regarding the impact of L2 use at high proficiency. The data support a distributed view of bilingual semantic representations and suggest executive control is involved in cross-language priming.

3.1 Introduction

3.1.1 General introduction

Over the past 40 years, empirical evidence has led to relative consensus on some aspects of bilingual lexical organisation. For example, the bilingual lexicon is

[10] Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (submitted). What shapes the bilingual lexicon? Insights from a comprehensive cross-language priming study.

taken to be integrated (words from both languages are likely stored together; e.g., Meade, Midgley, Dijkstra & Holcomb, 2017; van Heuven, Dijkstra & Grainger, 1998). Moreover, there is now little doubt that first (L1) and second language (L2) words are simultaneously activated during lexical access, potentially competing for selection (although see Navarrete, Del Prato, Peressotti & Mahon, 2014, for an updated challenge to the competition account). Nonselective access has been shown to occur in a large number of studies examining comprehension and/or production of isolated words (e.g., Kroll, Bobb & Wodniecka, 2006; van Hell, Dijkstra & Grainger, 1998). The most compelling evidence comes from sentence comprehension studies, where the context could potentially constrain selection to only one of the languages (e.g., Duyck, van Assche, Drieghe & Hartsuiker, 2007; although restrictions depending on the degree of semantic constraint and the reading stage are noted, e.g., van Assche, Duyck & Hartsuiker, 2012).

An enduring question connecting some of these debates is how access to meaning occurs for L2 words, or more broadly, how conceptual representations are organized in the bilingual lexicon. It is generally assumed that the conceptual store is shared cross-linguistically (e.g., French & Jacquet, 2005; Kroll & Tokowicz, 2005), although such a postulation comes with its own theoretical challenges. For example, if all words activate a series of nodes encoding conceptual features, how much do these sets overlap for translation equivalents? (e.g., Tayler, 1976; Thompson, Roberts & Lupyan, 2020). Is the degree of semantic/conceptual overlap higher for concrete than for abstract word pairs? If so, how does this affect L2 semantic memory? (e.g., Tokowicz, Kroll, de Groot &

van Hell, 2002; van Hell & de Groot, 1998). Monolingual research suggests that abstract words have more distributed and inconsistent meanings, and thus depend on the linguistic context more than concrete words (e.g., Crutch & Warrington, 2005; Hoffman, Ralph & Rogers, 2013; Pexman, Siakaluk & Yap, 2013; see also Li, Liang, Qu, Sun, Jiang & Mei, 2021, for neuroimaging evidence in bilinguals). This may result in fewer semantic features to share with their translation equivalents (assuming a distributed view; e.g., Van Hell & de Groot, 1998), which would in turn make the connections between these pairs more reliant on association (mediated by co-occurrence) than by the co-activation of semantic features.

The present study investigates some of these potential misalignments in semantic representations across languages. Crucially, we do so with an eye on the effects of individual differences in the amount of L1/L2 use, word frequency, and two novel factors in this type of studies. First, we address the role of semantic richness (e.g., Pexman et al., 2013), operationalized through a word's number of associates (NoA). This factor has been previously reported to affect response times in visual word recognition with monolinguals—the larger the number of associates, the faster the word is processed (e.g., Duñabeitia, Avilés & Carreiras, 2008). The effect is assumed to stem from the multiple activation at the orthographical, phonological and semantic levels provided by associates, which results in increased stimulation of the target word. Second, we explore potential effects of executive control on cross-language priming effects, which would arise from individual differences in the participants' abilities to keep mental sets in memory and inhibit competition from non-target word candidates.

3.1.2 Models of bilingual lexical-semantic representation and processing

Two of the most prominent models in bilingual lexico-semantic processing and representation are Multilink (Dijkstra, Wahl, Buytenhuijs, van Halem, Al-Jibouri, De Korte & Rekké, 2019) and the Distributed Feature Model (DFM; de Groot, 1992; van Hell & de Groot, 1998). Each of these models focuses on different aspects of bilingual word representation and processing, providing different, and sometimes complementary, frameworks to investigate bilingual word recognition.

Multilink, developed in a localist-connectionist fashion, is a comprehensive computational model of word recognition and production. For this model, 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 word has been encountered by a particular speaker). Subjective frequency is of course not directly observable, but it may be proxied by different measurable factors (stimulus/word-level or individual-level), such as corpus word frequency or active language exposure/use. The sensitivity of RLA to changes in amount and type of exposure, for L1 as well as L2 speakers, makes Multilink particularly suited to account for developmental effects. Moreover, Multilink crucially assumes cross-language related words to be connected only through shared semantic/conceptual nodes, and that competition between words (from the same or a different language) occurs at the decision level and not via lateral inhibition.

Central to the design of the present study, Multilink assumes holistic conceptual representations. Although this may be a temporary implementation (see Dijkstra et al., 2019: 5; see also Chuang, Bell, Banke & Baayen, 2021), it has some implications for the kind of predictions the model makes. For example, priming between translation equivalents cannot be modulated by different degrees of semantic overlap across pairs. Likewise, no effects are expected as a function of concreteness, because the differences between concrete and abstract words are not understood as a difference in semantic richness (i.e., amount of activated conceptual features).

A distributed view, on the other hand, predicts modulations of cross-language connectivity as a function of misalignments in semantic representation. Whereas the Distributed Feature Model (DFM) shares Multilink's assumption that all nodes are fully interconnected, it differs in how the conceptual system operates. Meanings are not represented by single units but by activation patterns and connection weights across conceptual features within the network—e.g., DOG is not a single conceptual unit, but a co-activation of more primitive features such as {animal}, {pet}, etc. This has consequences for both within- and cross-language lexical processing. Simply put, the larger the overlap in features between two words (both sublexical and conceptual), the faster the processing. Consequently, the DFM can account for things like cognate status effects, whereby cognate translations, which share features at all levels, are processed faster, or concreteness effects, whereby abstract word pairs, which have less feature overlap, are processed more slowly. Moreover, it also predicts weaker L2-L1 priming as compared to L1-L2 priming, one of the most recurrent findings in the literature (see Wen & van Heuven, 2017). The DFM holds that, in a bilingual priming task, L2 prime words, which are semantically less detailed (i.e., activate fewer features), have more difficulties activating the comparatively richer semantic representations of L1 words (see Finkbeiner, Forster, Nicol & Nakamura, 2004, for a similar account).

In sum, whereas the DFM is more suited to account for potential effects of semantic overlap, Multilink makes more precise predictions with regards to L2 developmental factors influencing RLA, like word frequency and language experience. The present study investigates the predictions of these two models in areas where they are at odds. We focus on two conditions where the degree of semantic overlap may differ (translation equivalents and cross-language semantic associative pairs), and on the potential contribution of two proxies of subjective word frequency: language use and standard word frequency.

3.1.3 Cross-language priming and semantic overlap

Translation- and cross-language semantic priming tasks are common experimental contexts to study bilingual word processing and representation. The usual understanding of these priming effects is in terms of spreading activation, a fundamental notion in interactive activation theories of semantic processing (e.g., Collins & Loftus, 1975), where nodes connect words, concepts and sublexical features (e.g., phonemes, graphemes) at different levels, and activation spreads through the network modulated by association strength and degrees of relatedness between units. Cross-language priming experiments exploit interactive activation processes to gain insight into architectural properties of the bilingual lexicon. In the most typical instance of this method, where it is combined with a lexical decision task (LDT), participants are presented with a prime word followed by a string of letters on which participants make a lexical decision (i.e., "Is this a real word?"). In the critical condition, prime and target are related at some level of interest (e.g., semantically, morphologically, orthographically), while in the control condition they are unrelated. Significantly different mean response times (RTs) between the two conditions indicate priming effects, that is, a given amount of (pre)activation spread from a related prime to a target.

The priming literature with noncognate translation equivalents (i.e., translation pairs with no overlap at the form level; e.g., English *dog* and Spanish *perro*) is abundant, especially in combination with prime masking manipulations (Forster & Davis, 1984). Studies have employed cross-language semantically related words (e.g., English *cow* and Spanish *oveja*, 'sheep'; e.g., de Groot & Nas, 1991; Ferré et al., 2015; Perea, Duñabeitia & Carreiras, 2008), with some examining both translation- and cross-language semantic priming (e.g., Basnight-Brown & Altarriba, 2007; Guasch, Sánchez-Casas, Ferré & García-Albea, 2011;

Kiran & Lebel, 2007). Similarly important to understand the nature of semantic representations, concreteness effects have also received some attention (Schoonbaert et al., 2009; Chen et al., 2014, Ferré et al., 2017, Smith, Walters & Prior, 2019). Finally, some studies have combined all of the above by investigating translation and cross-language semantic priming while manipulating concreteness (e.g., Jin, 1990; Schoonbaert et al., 2009; Smith et al., 2019). Schoonbaert et al. (2009) tested Dutch-English bilinguals in translation and semantic priming LDTs while manipulating concreteness and stimulus onset asynchronies (SOAs), maintaining prime presentation at 50 ms throughout. They observed asymmetrical priming effects depending on the target language, with L1-L2 priming always being larger than L2-L1 priming. In addition, these asymmetries were larger with translation pairs as compared to cross-language semantically related pairs, but concreteness seemed to have no overall effect. Schoonbaert et al. concluded that differences in the processing of L1 and L2 words are quantitative only. Similarly, Smith et al. (2019) employed two overt cross-language priming experiments (150 ms primes) to test translation equivalents (Experiment 2) and semantically related pairs (Experiment 4). Their participants were Hebrew-English bilinguals (likely less exposed to the L2 than the participants in Schoonbaert et al.). Asymmetries were observed with both types of stimuli, but no effect of concreteness was found in this study either.

Aside from differences in design, there are some general patterns of findings in the literature. First, priming effects are asymmetrical in most cases: L1-L2 priming is larger than L2-L1 priming. Second, this asymmetry is less consistently found with semantically related words than with translation equivalents. This might be explained by the fact that some of these studies employed balanced bilinguals (a factor that has been reported to at least attenuate the priming asymmetries, e.g., Duñabeitia, Perea & Carreiras, 2010; Wang, 2013), or because semantic effects tend to be less robust, especially in subliminal (masked) presentation conditions. Finally, there seems to be little support for a role of concreteness on cross-language priming effects, even though it would be predicted by some theories—most notably the DFM.

3.1.4 Individual differences on cross-language priming effects

While central to Multilink and the architecture of many other interactive-activation models, subjective frequency (the main modulator of RLA) cannot be measured directly. Much research has been explicitly or implicitly devoted to finding a reliable way to approach its measurement through secondary, directly observable factors. The most intuitive operationalization of this question is to collect frequency estimates in a norming study (e.g., Balota, Pilotti & Cortese, 2007). However, Brysbaert and Cortese (2011) showed that modern frequency measures outperform these ratings. In fact, *standard* frequency measures have been extensively employed in the study of mono-/bilingual word processing, showing that frequency is one of the major predictors of speed of lexical processing (e.g., Brysbaert, Mandera & Keuleers, 2018; Diependaele, Lemhöfer & Brysbaert, 2013).

Despite the robustness of frequency effects, this factor has largely been forgotten in the study of cross-language priming. To the best of our knowledge, only two studies have specifically and directly manipulated the frequency of stimuli. Nakayama, Lupker, and Itaguchi (2018) explored L2-L1 noncognate translation priming effects with high-proficiency Japanese-English bilinguals living in Japan. Their results showed a significant three-way interaction between prime type, prime frequency and target frequency, suggesting that L2-L1 priming effects were larger for less frequent L1 targets (i.e., more processing-costly trials) in the presence of the most frequent L2 primes. Chaouch-Orozco et al. (2021) explored the role of prime and target frequency by investigating masked translation priming effects with Spanish-English late bilinguals living in the United Kingdom who had different levels of L2 use and proficiency. They used low-to-moderate frequency words, but obtained no conclusive results with respect to this factor. Contrary to Nakayama et al. (2018), the effect of L2 prime frequency was only marginally significant (perhaps due to the relatively low number of observations and the range of frequencies employed), whereas target frequency only influenced priming effects in a complex interaction with task-related variables. The present study constitutes a further step in the study of word frequency effects in bilingual lexical processing, with a larger, better controlled frequency range. Anticipating our results, the present data support a fundamental role of prime frequency, among other factors.

Although, intuitively, subjective word frequency should be at least partly correlated with standard frequency measures, stimulus-level variables are not the only way to approach this construct—perhaps not even the most appropriate in certain circumstances (see Chaouch-Orozco et al., 2021). Standard word frequency estimates are, by nature, overgeneralizations, that is, they do not take into account individual patterns of language exposure/use, and therefore are not sensitive to individual variance. We are not arguing that traditional frequency measures are not valid or appropriate, quite the contrary. After all, the word *cloud* in any given language is surely going to be more frequent than *walrus*, which in turn will be a reliable predictor of lexical processing speed. However, when it comes to the study of cross-language priming in bilingual populations, we believe we should try to edge towards ways of capturing and exploring individual variation in exposure. Investigating L2 use is one such way.[11]

Previous studies have employed similar conceptualizations (e.g., Ibrahim, Cowell & Varley, 2017, with L1/L2 use) to proxy subjective frequency and RLA. However, to our knowledge, Chaouch-Orozco et al. were the first to do so while treating L2 use continuously and exploring the interaction with word frequency. In this regard, the main finding in Chaouch-Orozco et al. (2021) was that L2 use shaped L2-L1 priming: Increased active exposure and use of the L2 allowed for greater benefit from L2 related primes. Notably, no effect of L2 proficiency was observed, allowing Chaouch-Orozco and colleagues to weigh the different contributions of proficiency and language use—which, we believe, must be clearly differentiated, all the more so in the study of bilingual lexical-semantic

[11] Note that we are interested in language use in both the L1 and the L2. Since we are investigating L2 development, we will refer more often to the construct as *L2 use*; however, we will also employ the terms *L1 use* or *language use* when that helps to better understand the role of this factor.

processing. The lack of L2 proficiency and frequency effects led Chaouch-Orozco et al. to highlight the importance of L2 exposure/use when attempting to proxy subjective word frequency.

Although individual differences in L2 development can generate straightforward hypotheses about cross-language priming, effects can also be modulated by individual differences in executive control (EC), as recent studies have begun to show. For instance, Friesen and Haigh (2018) investigated cross-language priming with English-French homographs in a population of highly proficient bilinguals. They employed a color Stroop task to obtain a measure of EC, which they argued could regulate the ability to inhibit cross-language competitors. They observed that poorer inhibitory control was associated with negative priming when participants processed the L1 more slowly. Similar support for a role of EC on controlling cross-language competition was found in Freeman, Blumenfeld and Marian (2017) with a Stroop arrows task and an L2 phonological priming LDT, examining L1 phonotactic constrains. As these and several other studies across various paradigms suggest (e.g., Pivneva, Mercer & Titone, 2014; Linck, Hoshino & Kroll, 2008), EC may be involved in suppressing within- and between-language competitors when selecting the target word, as a result of the nonselective nature of bilingual lexical access. Despite this emerging evidence, to our knowledge no studies have attempted to examine the role of EC on the recurrent cross-language priming asymmetries.

This study employs a Dimensional Change Card Sort (DCCS; Zelayo, 2006), which has been argued to tap into inhibitory control (Bialystok & Martin, 2004; Kirkham, Cruess & Diamond, 2003) and the ability to switch between different mental sets (Miyake et al., 2000) and competing rules (Ramscar et al., 2013). We obtained two different indices from the task: local switch and mixing costs (see Yang, Hartanto & Yang, 2018). While local costs may reflect the ability to inhibit interference, mixing costs are assumed to reflect the ability to monitor and keep two competing sets in memory (Yang et al., 2018). These skills may be relevant in cross-language priming experiments, where participants decide on the target's lexical status after having to identify the appropriate language membership from two competing translation equivalents and potentially inhibit further competitors.

3.2 The present study

In light of the above discussion, the present study expands on the main line adopted by Chaouch-Orozco et al. (2021) to investigate bilingual lexical processing, with a larger sample, a larger set of word pairs, and a more systematic exploration of L1/L2 use and word concreteness. We conduct two experiments addressing different aspects of L1 and L2 lexico-semantic representations and the mechanisms underlying their interactions. We focus on the contributions of word frequency and L2 use, dissociating this latter factor from the intimately related construct of L2 proficiency, which we keep constant across participants. Finally, we offer the first dataset to combine an examination of word- and speaker-level variables with the role of EC in bilingual lexical priming.

Both experiments in the present study employ LDTs with overt priming. In Experiment 1, we investigate the degree of semantic overlap between translation equivalents by manipulating concreteness within noncognate translation pairs. Despite the evidence suggesting that priming effects in these tasks are largely semantic in nature (e.g., Xia & Andrews, 2015), it is possible that associations at the lexical level account for at least some of the variance—especially in lexical decisions (see Perea et al., 2008). Experiment 2 dissociates lexical and semantic contributions by employing cross-language semantically related pairs with three levels of semantic overlap, instead of direct translation equivalents. Even if smaller in size (as compared to translation priming), the presence of significant effects in this experiment would strongly support a semantic pathway account of bilingual lexical priming. Moreover, Experiment 2 allows us to investigate the role of a word's number of associates (NoA) in cross-language priming effects. In both experiments, we investigate the role of individual differences in executive control. In particular, we pay attention to the extent to which these modulate priming effects, potentially as a consequence of better EC (as measured by the DCCS) allowing for better competitor inhibition and monitoring of the language sets at hand during a bilingual task.

We tested 200 highly proficient L1 Spanish L2 English sequential bilinguals, who differed in their degree of L2 exposure and use. Each subject saw a total of 314 translation pairs and 112 cross-language semantically related pairs. We believe that this large number of observations, together with an unusually conservative analysis (e.g., we set $\alpha = .025$ for main effects and $\alpha = .01$ for interactions), will constitute one of the most reliable datasets to date on this subject. Moreover, we answer calls for sufficiently powered studies in bilingualism (see Brysbaert, 2019, 2020; Brysbaert & Stevens, 2018), while acknowledging the inherent difficulties regarding participant recruitment and stimuli development in this type of studies.

3.3 Experiment 1 – Lexical decision task

Experiment 1 is a LDT with overt translation priming. In order to focus on activation beyond sublexical levels, the task employs noncognate translation equivalents, whose relationship is fundamentally semantic. Crucially, these word pairs differ in their degree of concreteness, which may result in lower semantic overlap between more abstract translation equivalents. L2 use and word frequency are treated as continuous variables in order to better understand potential nuances in their effects on the processing of L1 and L2 words. We entertain the following hypotheses:

> **H1.** In light of current theories, amount of active L2 exposure/use should modulate L2 word processing speed, affecting both overall response latencies to L2 target words and priming effects, especially in the L2-L1 direction. Participants with more use of one language should respond faster to targets in that language, whether this is caused by higher RLA or by richer representations with higher degrees of semantic overlap.

H2. Similarly, higher word frequency should lead to faster processing or richer semantic representations. In both cases, we expect more benefit in terms of target processing facilitation. As a result, although not central here, translation priming asymmetries should be modulated by this factor.

H3. We expect to replicate a general concreteness effect, with faster responses to concrete words. Furthermore, if distributed-connectionist accounts of bilingual semantic processing are on the right track, priming should be larger with concrete word pairs as a result of increased semantic overlap between translation equivalents.

H4. With regards to EC effects, if greater monitoring competence allows for faster decisions when a competing word (i.e., a related prime) is also activated in a non-target language, priming effects should be larger for participants with smaller (local and mixing) switching costs.

3.3.1 Method

3.3.1.1 Participants

200 Spanish-English sequential bilinguals (see Table 7 for participant characteristics) took part in two LDT experiments 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 UK. Most of the participants in Spain were

completing a degree in English studies at different universities, the majority of them in cities where Spanish was the only societal language. Participants in the UK had a more diverse professional background and lived in various cities, mainly in London. All participants reported not using a third language on a daily basis. L2 proficiency was kept constant across participants to isolate the effect of L2 use, and was assessed with the LexTALE test (Lemhöfer & Broersma, 2012). A minimum score of 80/100 was required to participate in the study. A two-sample *t*-test showed that both groups differed significantly in their LexTALE score (see Table 7 for averages). However, further exploration with a parsimonious mixed-effects model showed that the factor, treated continuously across the whole population, did not significantly modulate RTs nor priming effects. For this reason, we decided to continue the analysis as planned.

Group	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)

Table 7. Participant characteristics. Mean values (standard deviation; range).

Language use information was collected through the Language and Social Background Questionnaire (LSBQ; Anderson, Mak, Chahi & Bialystok, 2018), which provides a fine-grained, context-dependent and dynamic measure of relative L1/L2 use. Recruiting participants from two different immersion settings allowed us to have enough variability in the distribution of the L2 use variable. Mean values differed significantly between these groups ($p < .05$). 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. From each participant, we obtained two measures of EC, local and mixing costs, from a DCCS task (more information in Appendix C).

Participants were tested in two sessions at least seven days apart. Task order for Session 1 was as follows: first direction of the translation priming LDT (Experiment 1) – LSBQ – second direction of the LDT (Experiment 1) – DCCS. Order of LDT priming direction for Experiment 1 (L1-L2 vs. L2-L1) was counterbalanced across participants. Task order in Session 2 was the following: semantic LDT (Experiment 2) – picture-word matching task – translation task. Participants were recruited online and compensated with £20 (or the equivalent in euros) for their participation.

3.3.1.2 Materials

314 noncognate translation equivalent pairs were used in each translation direction (see Appendix A for the stimuli list and Table 8 for stimuli characteristics). Out of these, 191 were abstract and 123 concrete (see Appendix C for the procedure to determine concreteness). English word frequencies were obtained from SUBTLEX$_{UK}$ (van Heuven et al., 2014), whereas Spanish

frequencies were extracted from SUBTLEX$_{ESP}$ (Cuetos, Glez-Nosti, Barbón & Brysbaert, 2011). Mean values between languages did not differ significantly. Words in both languages were also controlled for length and orthographic neighbourhood.

	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)

Table 8. Stimuli characteristics in Experiment 1. Mean values (standard deviation and ranges). Concreteness for Spanish words is taken to approximate the English words' values.

To generate "no" responses, 314 pseudowords were created for both translation directions with the software Wuggy (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), such that, for each language, in one of the lists, half of the words were preceded by their translation equivalents and the other half by control primes, which were obtained from 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 neighbourhood. Each list began with 16 practice items.

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 (see Appendix C for a more detailed account). 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. 39 (abstract) words showed an accuracy below 80% and were removed from the dataset.

3.3.1.3 Procedure

All experiments were created and presented online using Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 2020). Given the limitations of online data collection for experiment monitoring, data quality control (attention checks) and exclusion criteria were implemented to guarantee subjects' constant attention during the tasks. Failing to meet these criteria resulted in exclusion from the study.

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 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. They were also encouraged not to complete the sessions at night or when they felt tired. In sum, we paid special attention to simulating, to the extent possible, lab testing conditions.

3.3.2 Results

In Appendix C, we provide an expanded version of the data analysis, offering a more detailed description of the methods and their justifications.

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 the remaining 270 word pairs. Incorrect responses and pseudoword trials, as well as RTs below 200 ms (4 in total) and above 5000 ms (80 in total) were removed (see Baayen and Milin, 2010). An inverse Gaussian distribution was fit to the RT data. Sum contrasts were employed for categorical variables, and all continuous independent variables were scaled, centred, and converted to z units.

In all experiments, error rates and response times were analyzed employing (generalized) linear mixed-effects models (Baayen, Davidson & Bates, 2008) in *R* (version 3.6.1; *R* Core Team, 2019) with the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). The following procedure was applied in all experiments. We followed Brauer and Curtin (2018) in including main effects and interactions of interest as fixed effects in the analyses for both accuracy and RTs. We also 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. The method minimizes Type-I error risk. Additionally, in line with our commitment to reduce the risk of false positives and draw more robust conclusions, we established the significance level for main effects at .025 and for interactions at .01. Model assumptions were inspected, and further criticism was applied by removing observations with absolute standardized residuals above 2.5 *SD*.

The factors included in the models in Experiments 1 and 2 were prime type (related vs. control), concreteness (concrete vs. abstract), language (i.e., translation direction), L2 use (i.e., LSBQ score), and frequency (prime frequency in half of the models, and target frequency in the other half), as well as interactions of interest as discussed below. The random-effects structure included any predictor and interaction that varied within subject (i.e., language, prime type, prime frequency, and target frequency), prime (prime type), or target (prime type). A full-CRI structure was specified with random intercepts by each grouping factor for subjects, primes, and targets (see Feldman, Milin, Cho, Moscoso del Prado Martín & O'Connor, 2015, for the inclusion of primes and targets as random intercepts).

Age (and its interactions with prime type, concreteness and language) was the only post-hoc factor. Ideally, both groups should have had similar mean ages, but that was not the case in the present study. The reason is related to the demographics of each group, and almost inevitable by design. We were interested in exploring the effect of prolonged exposure to, and use of, the L2. For this reason, the UK group consisted mainly of migrants, which tend to be at least in their mid-twenties. The upper age limit for participating in the study was 40 years. While we could have established a lower cut-off, this would in turn have impacted our variable of interest, L2 use. Matching the mean age of this group (32 years; range: 22 – 40) was not feasible given other constraints and aims of our design, which required a large number (100) of highly proficient L2ers who were late bilinguals. The proficiency criterion forced us to resort to university students and English-language professionals of the English language in the Spain-based group, which lowered the mean age to 26 (range: 19 – 39). This difference was significant across groups. Although L2 use and age showed a moderate correlation (r = 0.51, p = .001), all variance inflation factors in the final models were below 2, indicating no collinearity (Zuur, Ieno & Elphick, 2010).

3.3.2.1 Accuracy analysis

Table 9 summarises RTs and error rates in all conditions in Experiment 1. Appendix B provides tables with summaries of the effects reaching or approaching significance in each of the models (also for Experiment 2). 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. Simplified versions of the models employed in the RT analysis were used; random structures were reduced to random intercepts for subjects, primes, and targets, and the interactions of prime type with those factors.[12] A significant effect of prime type (i.e., a priming effect) was found in all models. This was the only significant effect to arise in this analysis, along with the interaction of prime frequency and prime type. Both effects speak to the strength of these findings, which are replicated in the RT analysis.

Concrete words					
	Related		Control		
	RT	Error rate	RT	Error rate	Priming
L1 to L2	672 (2.4)	1.2	768 (3.1)	3.2	96*
L2 to L1	654 (2.2)	0.6	718 (2.5)	1.7	64*
Abstract words					
	Related		Control		
	RT	Error rate	RT	Error rate	Priming
L1 to L2	688 (2.5)	1.5	759 (2.9)	2.9	71*
L2 to L1	668 (2.3)	1	729 (2.5)	2.2	61*

Table 9. Mean response times (RTs; in milliseconds; standard errors), error rates (%), and priming effects (in milliseconds) in Experiment 1. Note: *$p < .01$.

3.3.2.2 Response time analysis

To avoid increasing model complexity, especially in the random structure, and to facilitate the interpretation of results, prime and target frequency were never included in the same model, and analyses were performed separately for the datasets of each priming direction. Four main models were thus fitted.

[12] We acknowledge that this procedure may have increased the risk of Type-I errors. However, accuracy analyses tend to be less sensitive to experimental manipulations, and were not central to the present study—as is common in the relevant literature.

The effect of prime type was significant in all models, indicating slower RTs for control primes. This indicates that participants benefitted from the presence of a related prime when processing the targets in both translation directions.

The effect of concreteness (faster RTs for concrete words) was significant in both translation directions, but only in the analyses that included target frequency. This result replicates previous findings of a concreteness effect in single word processing. Furthermore, the interaction between prime type and concreteness was significant in the L1-L2 direction: as expected, priming effects were larger with concrete words. The effect was mainly driven by faster RTs in the related condition for concrete words. This finding confirms our hypothesis of larger priming effects with concrete stimuli (H3), albeit only for L2 targets. The faster responses for L1 related primes with concrete words suggest that semantic mediation between that type of translation equivalents occurs faster or more efficiently than for abstract pairs.

The standard effect of target frequency was observed in both translation directions, with slower responses to less frequent targets. Moreover, a series of significant three- and four-way interactions between prime type, concreteness, L2 use, and target frequency provided insights into the roles of these factors. Figure 1 shows the overall effect of concreteness on L1-L2 priming effects, reflected by larger differences between the related and control conditions (three left-hand panes vs. three right-hand panes). Importantly, this is driven by the related condition—note how the lines for the control condition are at similar levels for both concrete and abstract words—which suggests that the effect reflects larger benefits for concrete related primes. There is also a general effect of L2 use on responses to L2 targets: responses tend to be faster with increased L2 use (cf. lines for the control condition, where prime influence is assumed to be minimal). However, this effect of L2 use is attenuated in the related condition, where, except with low-frequency concrete targets, RTs remain largely stable independent of L2 use. This pattern is particularly clear in the abstract condition.

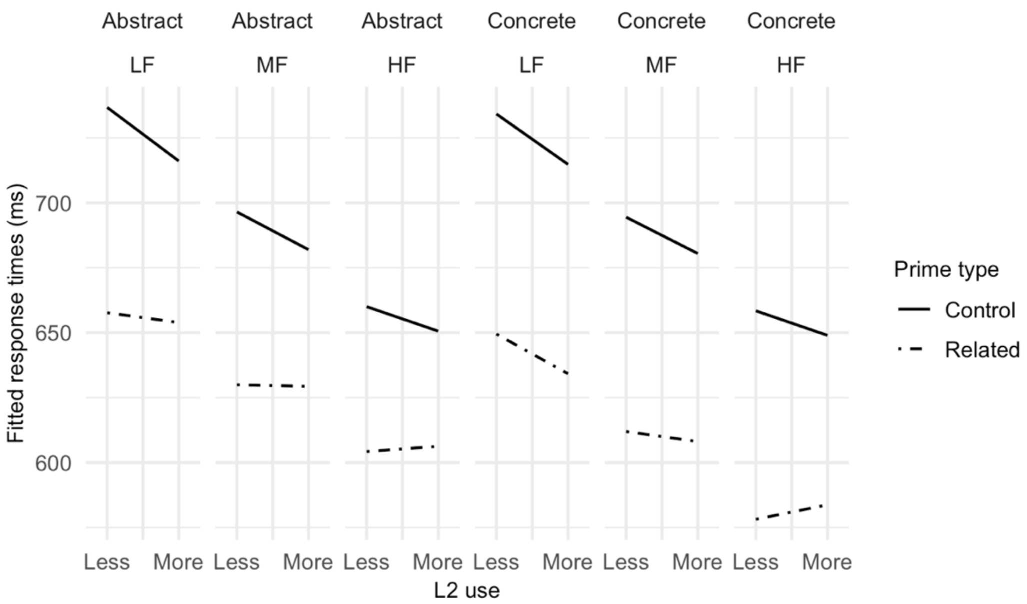

Figure 1. Plot of the interaction between prime type, condition, L2 use, and target frequency in the L1-L2 direction. LM (low target frequency); MF (medium target frequency); HF (high target frequency).

A tension between L1 and L2 use effects is of interest here. It seems that participants with more L1 use were able to respond to L2 related targets in the abstract condition as fast as participants with more L2 use. When L2 targets were concrete and less frequent, the benefit associated with more L1 use seemed to vanish. To be clear, those participants with more L1 use still benefitted from the related primes (see difference between control and related lines), but not more than participants with more L2 use. With more frequent L2 targets, the impact of L1 use resurfaces. In sum, this complex interaction suggests a relevant role of language use: increased L1 use leads to a larger benefit from the L1 related primes (i.e., more L1 use counteracts the overall slower responses to L2 targets), except with L2 targets that are concrete and less frequent.

Prime frequency interacted significantly with prime type in both directions. This effect evidenced larger priming with more frequent primes, driven by faster RTs to frequent related primes. A significant three-way interaction between prime type, L2 use, and prime frequency (see Figure 2) showed a similar pattern to the interaction with target frequency (also in the L1-L2 data). Although increased L2 use led to faster responses to L2 targets (note that the main effect is not significant), this did not happen with more frequent L1 related primes, where the effect of L2 use disappears. Therefore, increased L1 use only led to greater benefit from the more frequent related primes.

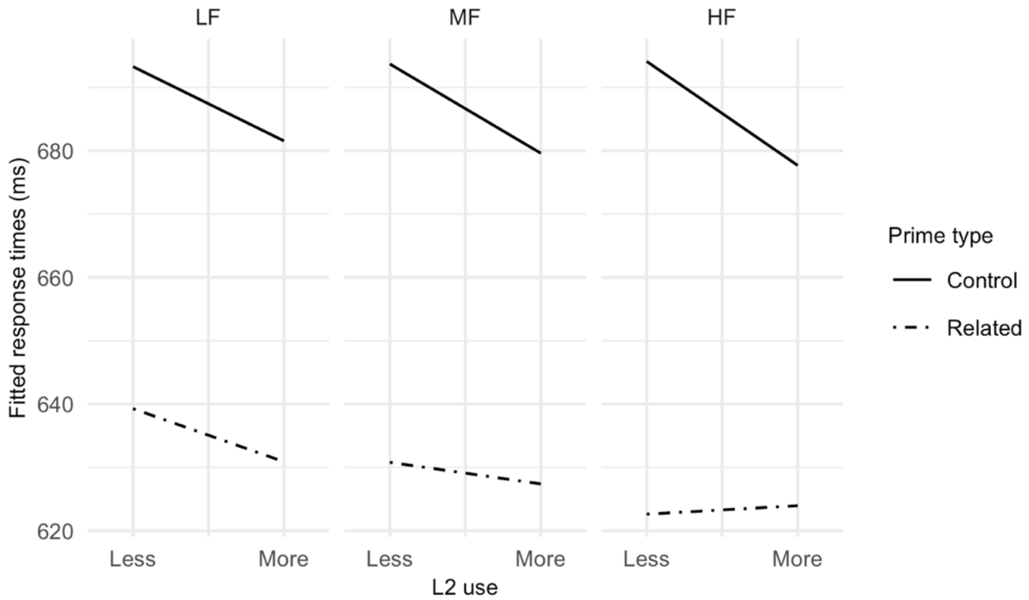

Figure 2. Plot of the interaction between prime type, L2 use, and prime frequency in the L1-L2 direction. LM (low target frequency); MF (medium target frequency); HF (high target frequency).

Finally, age reliably emerged as a significant predictor in the L1-L2 direction, with slower RTs in older participants. This is potentially related to declining speed/efficiency in the recognition and integration of visual information, as well as in the processes linked to response preparation and the motor skills required to physically provide a response. Unexpectedly, a significant interaction between age and prime type was observed. In the L2-L1 direction, priming effects were larger for younger participants.[13] However, it is not clear whether this reflects a larger *overall* effect of age in the related condition, or an attenuation of the main effect of age in the control condition. Both hypotheses could have implications for how age impacts the neurocognitive mechanisms associated with cross-language

[13] Age also interacted significantly with concreteness. The effect was found this time in the L1-L2 direction, and it revealed that the standard effect of age (i.e., slower responses with older age) was larger for abstract words. Note that this effect did not involve the critical manipulation of prime type.

spreading activation, or the way bilinguals deal with cross-language competition during word selection. The following analysis on the role of EC provides insight into this second possibility.

Maximal models were fitted to the dataset, including the main effects and interactions of interest of language, prime type, and switch cost index (local or mixing) as well as (prime or target) frequency. The reason to not include L2 use, age, and concreteness in these analyses was twofold. First, previous analyses showed multicollinearity issues when those factors were included. Second, by comparing different models, we ensured that the EC indices did not interact with any of these factors (see more details in Appendix C). The main effect of local costs was significant in the model with target frequency and marginally significant with prime frequency, whereas for mixing costs the effect was non-significant. The double interactions of prime type with local and mixing costs, respectively, were not significant. However, both local and mixing costs entered into triple interactions with prime type and (prime and target) frequency in all models. Smaller local and mixing costs led to faster RTs, but the effect was attenuated in the related condition when words were more frequent. This likely reflects the smaller room for improvement in the easier trials, where the compounded effects of frequency and relatedness bring RTs closer to a lower limit (a ceiling/floor effect).

3.3.3 Summary

The results of Experiment 1 contribute to our understanding of the role of word frequency and L2 use on translation priming effects. Regarding L2 use (H1), this factor influenced priming, although only under certain circumstances. In the L1-L2 direction, participants with more L1 use benefitted more from the related primes when primes and/or targets were more frequent. This finding partially meets our expectations: a more active use of a language allows for larger benefits from related primes. Nevertheless, the fact that we do not observe the same in the L2-L1 direction for participants with more L2 use is intriguing and warrants further examination.

Notably, prime frequency effects are robust, and obtain in both translation directions (H2). Furthermore, the present data offer evidence of an effect of concreteness on translation priming (H3), although only in one translation direction. This finding supports differential semantic overlap as a function of concreteness, in line with distributional models of semantic memory.

Finally, the analyses including local and mixing costs could be interpreted as supporting H4, that is, participants with enhanced EC may be better at dealing with the competition introduced by the prime. As expected, the data suggest that task switching ability correlates with larger priming effects, which might be due to better conflict monitoring in resolving the tension between the activation of the prime and the target. Furthermore, our results show an attenuation of this effect with more frequent related primes. It is possible that, because of the higher frequency (and saliency) of the more frequent related targets (and the higher facilitation provided by related primes), having a relatively better or worse ability to discriminate between word candidates has negligible effects on the task—a ceiling/floor effect. An alternative interpretation may place the locus of the effect on the lexical decision itself, rather than on target word recognition. Participants with enhanced EC would be faster at deciding on the target's lexical status. Therefore, individual differences in switching would only become less relevant when participants respond to targets in the presence of more frequent primes.

Although the results of Experiment 1 already speak to the relevance of the semantic component in the emergence of priming effects, in Experiment 2 we continued to explore cross-language priming under conditions where (i) direct lexical associations are not expected, (ii) the degree of semantic overlap is much reduced, and (iii) the number of associates (NoA) of the target word differs.

3.4 Experiment 2 – Semantic LDT

Experiment 2 was procedurally similar to Experiment 1, but stimuli consisted of cross-language semantic associative pairs (e.g., *river*-PUENTE, Spanish 'bridge'). We hypothesize the following:

> **H5.** If priming in lexical decision tasks is semantically mediated—as suggested by the results of Experiment 1—and if cross-language semantic associates are connected in the lexicon (through shared conceptual features), we should expect to observe priming effects in both directions.

H6. Those effects should be modulated by language use, to the extent that this correlates with higher RLA/richer semantic representations, speeding up cross-language activation of word pairs.

H7. If the number of associates is particularly relevant for abstract words and indeed facilitates visual word recognition of these items, we expect priming effects to be modulated by NoA.

H8. As in Experiment 1, we expect priming effects to be influenced by EC. Since the conflict between semantic associates is likely to be smaller than between translation equivalents (because the latter are typically mutually exclusive, as they refer to the same entities), it is possible that EC mechanisms are less actively involved.

H9. If higher NoA entails the activation of more competitors, and task switching costs can predict how efficiently speakers deal with that competition, participants with better EC will show larger priming effects in the high NoA condition.

3.4.1 Method

3.4.1.1 Participants

Participants in this experiment were the same as in Experiment 1.

3.4.1.2 Materials

Critical trials in each translation direction contained 112 noncognate cross-language semantic associative pairs (see Table 10 for stimuli characteristics). These pairs were obtained from a free association task in Spanish conducted on a separate group of 100 participants with the same linguistic profile as our participants in the UK Group. List composition followed the procedure in Experiment 1.

	Spanish	English
Frequency	3.1 (0.7; 1.3-4.7)	4.6 (0.6; 3.5-6.3)
Length	5.0 (1.3; 3-8)	5.0 (1.3; 2-8)

Table 10. Stimuli characteristics in Experiment 2. Length indicates number of characters. Frequency is expressed in terms of the 1-7 Zipf scale, where 1 is the lowest and 7 the highest frequency (see Appendix C for details).

Word frequency could not be matched between Spanish and English stimuli. Spanish words were overall less frequent than English ones (3.16, *SD*: 0.65, range: 1.28 – 4.76 vs 4.54, *SD*: 0.64, range: 3.34 – 6.36). However, recall that the main aim of this experiment was to determine whether cross-language semantic priming would obtain for these participants outside of translation equivalent pairs, and whether priming was modulated by the strength of the associative connections as measured by the number of associates between related pairs (see below for further discussion).

Due to the high number of repeated target words in the free association task, we repeated 13 primes (out of 112) in each translation direction in the stimuli for the present experiment. Inspection of a potential effect of this repetition showed almost identical RTs and priming effects—in direction, significance and effect size.[14]

3.4.1.3 Procedure

The presentation procedure in this experiment was identical to that of Experiment 1.

3.4.2 Results

Data cleaning and analysis (including model determination and selection) in this task followed the protocols described for Experiment 1 above. Some unexpected properties of the data should be discussed up front, to the extent that they account for certain choices in the analysis.

As mentioned above, Spanish words were on average less frequent than their English counterparts. On the one hand, this might have resulted in underestimated L1-L2 priming since less frequent primes might have had more difficulty activating their L2 targets—a prediction that all relevant theories would make. This could be amplified by the fact that L2 targets are of high frequency and thus potentially processed faster, reducing "room for improvement" for primes. This could have the opposite effect in the L2-L1 direction: priming effects might be overestimated.

[14] A caveat from the norming study employed to obtain the cross-language semantically related pairs is that it prevented us from exploring concreteness effects within this dataset, since, in some cases, responses in the norming study resulted in mixed pairs with one concrete and one abstract word (e.g., eye-view).

Finally, when introducing prime and target frequency in the models, multicollinearity issues arose, which are known to increase Type-I and Type-II errors alike (e.g., Grewal, Cote & Baumgartner, 2004), complicating the exploration of potential interactions between the factors involved. For this reason, and in line with our main research focus, we decided to include L2 use alone in this analysis. Leaving out concreteness and frequency allowed us to fit a model with the whole dataset from both translation directions, including language, prime type, L2 use, age and NoA (categorically operationalized as high vs. low), as well as interactions of interest between these factors, as fixed effects. The random structure from the models in Experiment 1 was modified accordingly.

3.4.2.1 Accuracy Analysis

Table 11 summarises error rates and RTs in all conditions. The analysis of accuracy followed the one in Experiment 1. Significant main effects of prime type (i.e., priming effects) and age (i.e., more incorrect responses with increased age) were observed.

	Related		Control		
	RT	Error rate	RT	Error rate	Priming
L1 to L2	671 (2.5)	1.7	695 (2.7)	2.7	24*
L2 to L1	629 (2.4)	1.1	652 (2.1)	1.4	23*

Table 11. Mean response times (RTs; in milliseconds; standard errors), error rates (%), and priming effects (in milliseconds) in Experiment 2. Note: **p* < .01.

3.4.2.2 Response time analysis

All variance inflation factors in the final model were below 2, indicating non-collinearity of predictors. Prime type significantly influenced RTs, which were faster in the related condition, suggesting a facilitatory effect of cross-language semantically related primes in both translation directions. This finding provides important evidence for the ability of LDTs with overt priming to capture bilingual effects of a lexical-semantic nature, and supports the idea that the priming effects in Experiment 1 have a semantic component. The effect of language (i.e., priming direction) was also significant, indicating that responses were slower in the L1-L2 direction. The interaction between language and prime type was not significant.

The priming asymmetry (larger L1-L2 than L2-L1 priming) has been a recurring theme in the literature on bilingual lexical processing. Our results here are symmetrical across directions, which may be explained in several ways. One is that the effects are indeed symmetrical for this population. Another, more methodological, is that the mismatch in the L1 and L2 frequencies reported above has levelled the playing field. While our experiment cannot adjudicate between those two accounts, the present data reinforces the idea that L1-L2 priming effects are generally robust, as they obtained here even with primes that were significantly less frequent than their corresponding targets—i.e., in the less favourable possible context.

Further, we observed two marginally significant interactions. First, between prime type and age. This replicates the finding of Experiment 1, although this time the effect does not differ between priming directions. Second, a four-way interaction between prime type, language, L2 use, and NoA. This interaction shows that L2 use does not modulate priming effects in the L1-L2 direction. In the L2-L1 direction, however, priming effects are larger with more L2 use for words with many associates, whereas the effect is the opposite for words with low NoA (i.e., larger priming effects with more L1 use).

Finally, our EC measures did not account for a significant amount of variance in the data, neither as main effects nor in interaction with NoA.

3.4.3 Summary

The results of Experiment 2 replicate previous findings on translation priming (at least in the L1-L2 direction), suggesting that (i) translation priming has a semantic component, and (ii) LDTs are sensitive to semantic effects (H5). Effects of L2 use (H6) are rather modest, approaching significance only in the L2-L1 direction, and there only in interaction with NoA. Similarly, this last factor fails to show a substantial impact on priming effects (H7). The interaction between L2 use and NoA in the L2-L1 task may be explained by the number of competitors activated by the higher NoA of L1 targets. Assuming an interactive activation view, higher activation in L1 words for participants with more L1 use entails the activation of a larger cohort of associates for each L1 target.[15] The combined effect of these associates and the competition of the related L2 prime might have delayed word recognition for these participants, reducing priming effects for those L2 primes. Although only marginally significant, we consider this effect to be worthy of further investigation.

[15] Note that a similar account has been proposed for words with high orthographic neighbourhood density (Müller, Duñabeitia & Carreiras, 2010), a variable not controlled here.

Lastly, the lack of EC-related effects, both by themselves and in interaction with NoA (H8 and H9), may be due to the cross-language semantic LDT not being sensitive enough to reliably capture EC and NoA effects, as it seemed to occur with the contributions of language use. Alternatively, NoA may indeed play a negligible role (if at all) in the mechanisms underlying cross-language priming.

3.5 General discussion

The present study reports on data from two experiments that address different aspects of bilingual semantic representation and processing. We investigated cross-language priming effects, placing special attention on how they may be shaped by bilingual experience—operationalized here through individual- and stimulus-level factors (language use and word frequency). In addition, we explored what might be an important factor in a comprehensive characterization of cross-language priming: the role of executive control in regulating the multiple levels of competition and conflict inherent to these tasks.

In Experiments 1 and 2, we employed an overt priming LDT with translation equivalents and cross-language semantic associates, respectively. Through a DCCS task, we also obtained switching cost indices that allowed us to investigate the relationship between EC and cross-language priming effects. Overall, we observed significant priming effects in both Experiments 1 and 2, although these were larger with translation equivalents. Taken together, the results of Experiments 1 and 2 suggest that priming with cross-language semantically related pairs is less sensitive to individual differences in bilingual experience than priming between translation equivalents. Moreover, we found robust support for a major involvement of prime frequency in cross-language priming, whereas the effects of language use and concreteness are most apparent in the L1-L2 direction and with translation equivalents only. Finally, we found some evidence of local and mixing costs significantly modulating translation priming effects (Experiment 1); however, it is not entirely clear what the locus of the effects is—i.e., whether these factors have an impact on word recognition or on lexical decision.

Next, we discuss the study's main hypotheses and how they relate to the present results, while offering an interpretation in light of some of the most relevant theories within bilingual semantic processing and representation.

3.5.1 L2 use

The findings from Experiment 1 regarding L2 use, while suggestive, are not entirely conclusive. First, contrary to what we predicted, the amount of L1/L2 use did not significantly modulate overall RTs in any priming direction. Second, in line with the predictions of models of bilingual lexical-semantic processing, we expected language use to increase the likelihood of priming effects. In the present data, this is only true of the L1-L2 direction. None of the models presented above, the DFM and Multilink, predict this asymmetry. For Multilink, the speed of lexical processing depends on RLA. Taken at face value, this stipulation entails that participants with more L2 use in the present study *must* have shown significant differences in the size of the L2-L1 priming effects. That is, increased

L2 use should have led to higher RLA in L2 primes, resulting in faster processing and greater benefit. Although we focus on the predictions of Multilink here, as the model is more suited to account for L2 development, similar arguments could be made regarding the assumptions of the DFM.

3.5.2 Word frequency

Language experience (and, more specifically, subjective frequency) was also proxied here through the stimulus-level factor of word frequency. Importantly, we inspected prime and target frequency separately since the results in Nakayama et al. (2018) and Chaouch-Orozco et al. (2021) suggested that both measures may play different roles in cross-language priming effects. First, contrary to Nakayama et al., we did not observe a significant interaction between target frequency and prime type. A possible reason for this is that target presentation was longer in our experiments. This reduction of time pressure may also flatten typical differences between less and more frequent targets, whereby the longer processing of low-frequency targets leaves a larger "window of opportunity" for priming effects.

Finally, we observed a robust effect of prime frequency. Priming was larger with more frequent primes, an effect that was entirely driven by the related primes. Nakayama et al.'s and Chaouch-Orozco et al.'s results pointed towards a similar effect, but the present data are much more conclusive in that respect—recall that this is one of the few studies exploring the role of prime frequency, and the first to do so in a systematic manner. For Multilink, this finding would be explained by higher RLA with increased frequency. The DFM would resort to the richer semantic representations assumed for more frequent primes, which would enhance cross-language activation through shared conceptual features. Whatever framework is on the right track, the present results highlight the relevance of word frequency, and the need for future studies to keep examining its role in bilingual lexical processing, especially attending to the continuous nature of this predictor.

3.5.3 Concreteness

Although the present data showed that concreteness modulated translation priming effects, this only happened in the L1-L2 direction. As discussed, Multilink assumes holistic shared conceptual nodes between translation equivalents, which makes any effect of concreteness hard to explain for the model. The DFM accommodates the finding straightforwardly: the higher degree of semantic overlap between concrete translation equivalents would lead to increased stimulation at the semantic level, resulting in larger priming effects for those words. However, why the effect was absent in the L2-L1 direction remains unclear. After all, even if the facilitation produced by L1 primes is weaker due to poorer semantic specification, there is no reason not to expect the same for L2 primes. A way of solving the puzzle is by arguing that L2 concrete representations do not develop to the same extent that L1 ones do, suggesting qualitative differences in representation even at high levels of proficiency, which would predict a directional asymmetry. This would in turn flatten differences between concrete and abstract words in the L2, limiting the scope of concreteness effects. In fact, when we look at global priming effects in the two directions and for both word types, we find striking similarities in the size of the effect between L1-L2 priming with abstract words and L2-L1 priming for both abstract and concrete words.

3.5.4 Executive control

Finally, this study incorporated switching costs into the study of cross-language priming effects. Due to the very nature of semantic priming experiments, executive function may be relevant when deciding between competing word candidates from both the target and the non-target language. A greater capacity to monitor competing mental sets, such as a non-target language, may help participants deal with such tasks.

Our results point towards an effect of executive function in this type of studies, which may be attenuated in those scenarios where the trials are easier and thus switching abilities are less in demand (e.g., when responding to targets in the L1 or when words are more frequent). However, it is not entirely clear whether the effect arises during word candidate selection or at a higher level, when deciding on the word's lexical status. We are optimistic about the potential of this new approach to improve our understanding of the mechanisms at play during cross-language priming, especially in those tasks that involve candidate selection among a variably large cohort.

3.6 Concluding remarks

The present study aimed at comprehensively exploring some of the factors underpinning cross-language priming effects, bilingual lexical-semantic representations and processing more generally. The present data are rather conclusive with regards to the effects of some of these factors, and highlight exciting future directions. We report a consistent effect of prime frequency on the size of translation priming effects, indicating that the factor deserves much more attention in this type of studies, especially given the importance of understanding individual variation in L2 development. In this respect, the observed asymmetries in the effects of language use and concreteness are puzzling and seem to open the possibility of persistent qualitative differences in the way L1 and L2 lexical-semantic representation and processing develops. Notably, along with the findings of previous studies, the concreteness effect reported in Experiment 1 supports a distributed view of bilingual semantic memory, which, in our opinion, would enrich Multilink's coverage in explaining bilingual lexical-semantic representation and processing. Thus, we submit that bilingual lexical research would greatly benefit from incorporating distributional semantic analyses as part of its toolkit. Such methodologies have proved promising in monolingual research (Jones, Dye & Johns, 2017; Mandera, Keuleers & Brysbaert, 2017), and hold the potential to open new epistemological pathways into which the field would be wise to venture. We maintain that, despite the inherent difficulties in applying such methods in bilingualism research, it is worth taking up the challenge. Finally, in light of the present results, it would be of great interest to explore the role that executive function and age play on the modulation of cross-language priming effects and, potentially, on shaping the neurocognitive mechanisms argued to subserve these interactions.

Chapter 4: Beyond the mean: Shape distribution analysis provides unique insights into how masked and unmasked cross-language priming effects unfold[16]

Abstract

Priming effects in lexical decision tasks (LDT) are assumed to occur due to spreading activation from prime to target, providing a head start to the target's processing, especially under masked presentations (Forster et al., 2003). However, unmasked priming could arise from accumulated evidence with respect to the familiarity of a prime-target cue (Ratcliff & McKoon, 1988). Despite evidence for these hypotheses coming (primarily) from mean-based analyses, doing so is beyond their inherent capabilities. Conversely, examining the distributions' shapes can provide key insights. Herein, we analyze reaction times (RT) distributions from three (masked and unmasked) cross-language visual priming LDTs. Results with masked primes support a head start account, whereas a compound cue effect was obtained with unmasked primes. Thus, our findings highlight the informative uniqueness of distributional analyses and suggest that: (i) episodic memory is involved in cross-language unmasked, but not in masked priming and (ii) word frequency and concreteness affect accumulation of evidence on the compound cue's familiarity.

[16] Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (submitted). Beyond the mean: Shape distribution analysis provides unique insights into how masked and unmasked cross-language priming effects unfold.

4.1 Introduction

4.1.1 General introduction

Word priming has been a reliable tool to inform theories on the architecture and functioning of the (mono-/bi-/multilingual) lexicon. In these paradigms, two words (or pseudowords) with varying degrees of formal, morphosyntactic or semantic similarity are presented consecutively. The first item is known as the *prime*, the second as the *target*. Responses to the target word or pseudoword are collected (e.g., in a lexical decision task—"Is this string of letters a word?"). In the related condition, the hypothesized level of relation between prime and target is instantiated. This is then contrasted with a control condition where this relationship is (assumed to be) absent. The term *priming effect* refers to significant differences between the related and control conditions in response times, error rates, or other measures (e.g., modulations of the N400 component in electrophysiological data). These effects are often assumed to stem from spreading activation from related primes to the targets within the network (Collins & Loftus, 1975; Dijkstra, Wahl, Buytenhuijs, van Halem, Al-Jibouri, De Korte & Rekké, 2019). Herein, we consider in juxtaposition an alternative hypothesis, the joint retrieval or compound cue theory (Ratcliff & McKoon, 1988).[17] Under the latter view, prime and target form a joint retrieval cue that speeds up the decision on the target.

[17] Inhibitory effects caused by processing of the control prime interfering processing of the target could also explain priming effects in addition to facilitation from related primes (see, e.g., discussion in McNamara, 2005).

An important addition to the implementation of priming methods in lexical processing research was the introduction of masked presentations (Forster & Davis, 1984), where the prime is shown for a very brief time (usually less than 60 ms; Forster, Mohan, Hector, Kinoshita & Lupker, 2003) and is preceded by a mask (e.g., #####). Both modifications are aimed at making the prime less salient and its processing *unconscious*. Consequently, it is assumed that masking the prime allows tapping into early automatic processing, potentially overcoming the risk of the participants incurring into strategic processes—as is believed to occur with unmasked presentations, the norm in earlier studies (Forster et al., 2003).

In cross-language priming research, where prime and target belong to different languages, early studies made use of longer Stimulus Onset Asynchronies (SOA) in overt (unmasked) priming paradigms, switching to masked presentations in more recent times (see Altarriba & Basnight-Brown, 2007, and Wen & van Heuven, 2017, for reviews). The preference for masking the prime has been more prominent in translation priming experiments than in those employing cross-language semantically related pairs (Altarriba & Basnight-Brown, 2007). The reason is that activation at the semantic level is considered to require longer SOAs.[18]

The vast majority of lexical-semantic priming studies base their analyses on mean response latencies. This is potentially problematic because the approach reduces response time (RT) distributions to a singular, unique value. As we discuss in the next section, doing so results in the potential loss of valuable

[18] However, note that the masking procedure has been widely employed with non-cognate translation equivalents. Priming, in such pairs, should arguably entail a semantic component, too (Xia & Andrews, 2015).

information reflecting qualitative differences in how facilitation occurs in masked as compared to unmasked priming paradigms. In this light, the present study contributes to the overall literature by re-analysing the data from two studies employing masked Chaouch-Orozco, González Alonso and Rothman (2021) and unmasked Chaouch-Orozco, González Alonso, Duñabeitia and Rothman (submitted) presentations, exchanging their original mean RT analyses for RT distributional ones. Doing so reveals something that could not be appreciated without this approach: the benefit from masked and unmasked primes in cross-language experiments stems from different cognitive processes. Thus, the present contribution is twofold. First, we offer the first dataset testing cross-language stimuli (translation equivalents and semantically related pairs) analysed in a distributional manner and intending to examine the cognitive processes behind masked and unmasked priming. Second, our approach further innovates by examining potential effects of language experience, word frequency and concreteness on the type of cognitive accounts supporting cross-language priming. Results bear witness to the importance of considering not only traditional measures of central tendency but also distributional patterns because they offer relevant information that the mean alone cannot provide (e.g., Balota & Yap, 2011). Implementing distributional analyses in chronometrical studies, we submit, is a necessary step towards a more fine-grained understanding of language processing, especially in bilingualism research where this approach has been rarely employed.

4.1.2 RT distributional analysis

The quantitative study of mental processes can be traced back to the emergence of mental chronology, which assumes that response times—in combination with an explanatory theoretical paradigm—can generate useful inferences about the cognitive mechanisms involved in completing a given experimental task. An overwhelming majority of the analyses within psychological studies rely on mean RT performance (Balota & Yap, 2011). Besides the ease of their calculation and their high stability, the rationale for focusing on mean RTs is based on the general agreement that measures of central tendency provide good representations of the entire data set (e.g., see discussion in Ratcliff, 1993). As we review below, however, this is not always the case.

For instance, RT distributions are almost invariably non-negatively (left) skewed, presenting a long tail to the right (i.e., towards longer latencies). Crucially, the effect one aims to examine might originate precisely in the tail of the distribution, suggesting that more nuanced ways of analysing the data than solely focusing on the mean may be desirable. Moreover, in practice RT data are often transformed to meet the assumptions of some commonly used statistical analyses. These transformations correct the skewness of the distribution, which inevitably leads to the loss of potentially relevant information. Complicating things further, trimming the data (removing outliers) is a common practice, despite the lack of consensus on a standard procedure to perform such a delicate operation (Cousineau & Chartier, 2010). Outliers do not only highly influence the mean but also the standard deviation, which reflects the spread of the distribution.

One way of overcoming these drawbacks is “moving beyond the mean” and exploring RT distributions (Balota & Yap, 2011). One of the general methods to do so is to fit a mathematical function to the data. Parametric density estimators assume the form of the empirical distribution, and the most common one for the density of RT data is the ex-Gaussian function (van Zandt, 2000), which consists of a normal (Gaussian) and an exponential distribution. The ex-Gaussian function has three parameters: μ and σ reflect the mode and the variance of the Gaussian component, respectively, whereas τ reflects the rate of the exponential component, representing the distribution's skewness.

Importantly, besides the advantage of providing an optimal fit to RT data, the ex-Gaussian function can also be related to analyses based on the mean, as the sum of μ and τ reflects that value. Further, inaccessible through mean RT analyses, fitting the ex-Gaussian function to raw RT data can reveal the effect of the treatment condition on the shape of the distributions. For example, the primary assumption in mean RT analyses is that the variable of interest causes a shift in the distribution. This is illustrated in Figure 3, which contains the density plots of two distributions of simulated response latencies where only a change in μ has occurred. In contrast, the density plots in Figure 4 show changes that go beyond a shift, involving also a change in the distribution spread (i.e., an effect in τ). Note that, in both figures, the mean response latency in the control condition is roughly the same (800 ms), indicating that any difference in the shape of the resulting distributions would go unnoticed in analyses of mean RTs. This fact would not be of much importance had such distributional variances not been suggested to reflect differences in the cognitive mechanisms involved in priming experiments.

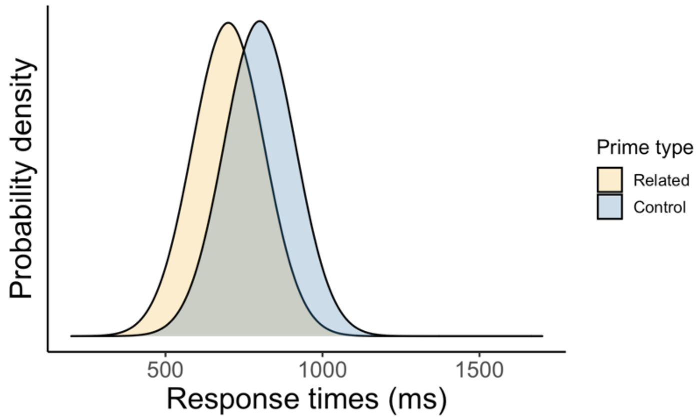

Figure 3. Density plot of simulated RTs in the related and control conditions with only a change in μ.

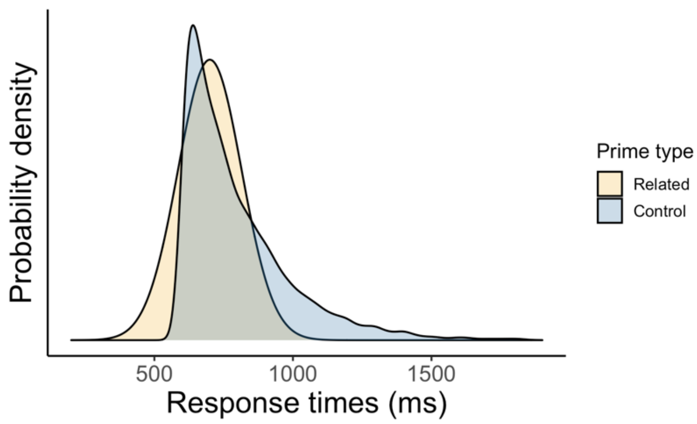

Figure 4. Density plot of simulated RTs in the related and control conditions with a change in μ and in τ.

4.1.3 Theories of the cognitive processes underlying masked/unmasked priming

As discussed in Gomez, Perea, and Ratcliff (2013), when a prime is masked, priming effects may arise from the prime's early pre-activation of some of the targets' features, providing a head start (e.g., Forster et al., 2003). In other words, the prime initially speeds up processing of the target. Notably, that boost occurs only once—at the target's onset, making it independent from the ultimate length of the trial. Hence, this account predicts an equal priming effect across the distribution (i.e., a shift; relatedness would affect the μ parameter). On the other hand, a different pattern may be observed in the distributions with unmasked primes. For instance, the relatedness effect may be driven by both a shift and a skew (i.e., a change in μ and τ). Such an outcome would be in line with the compound cue account (Ratcliff & McKoon, 1988), where prime and target create a joint retrieval cue that improves the quality of information used to provide a response. Succinctly, the theory assumes that the more closely related two words are in memory, the larger the familiarity of these two words' joint cue would be. Evidence for this familiarity would accumulate over time before making a lexical decision, resulting in the benefit provided by the familiarity having a greater impact on the tail of the distribution. In other words, priming effects would be larger when more information about the familiarity of the cue is gathered, that is, in longer trials. Note that Bodner and Masson (2001, 2003) argued that masked primes too can create episodic memory traces recruited during target processing. They based this claim on results from masked identity and semantic/associative priming. In both cases, priming size depended on the ratio of critical trials in the list. The priming effect was larger when the proportion of critical trials was higher (0.80). According to the authors, such effect could only be expected if the prime, despite potential unawareness, creates an episodic record (see Balota, Yap, Cortese & Watson, 2008, for a similar conclusion regarding the utility of highly masked primes under degraded conditions; and Kinoshita, Mozer & Forster, 2011, for conflicting evidence and an alternative explanation).

To the best of our knowledge, only two studies comparing semantic priming effects in masked and unmasked LDTs have tested these two accounts employing a distributional analysis (for similar explorations with different paradigms, see, e.g., Kinoshita et al. 2011; Pollatsek, Perea & Carreiras, 2005; Ratcliff, Gomez & McKoon, 2004). Both studies investigated this in the context of monolingual semantic priming. In Experiment 2, Balota et al. (2008) manipulated prime duration and SOA (200 ms + 50 ms blank screen vs 1000 ms + 250 ms blank screen). In the analysis, they fitted an ex-Gaussian function to the raw data. Independently of the type of presentation, they obtained a significant relatedness effect for μ, indicating only a shift in the distribution, in line with the head start account. This effect was replicated in subsequent experiments with shorter presentations. In Experiment 5, they employed a 150 ms prime preceded by a mask and followed by a 650 ms blank screen, whereas in Experiment 7 they presented a mask followed by a 42 ms prime without a blank screen afterwards. Gomez et al. (2013) investigated semantic priming effects in two masked and unmasked experiments where primes were presented without a subsequent blank screen for 56 ms and 200 ms, respectively. They carried out a distributional analysis by employing different fits of the diffusion model (Ratcliff, 1978)—fitting a model is another way of inspecting an RT distribution (see discussion in

Balota et al., 2008). Their results showed distributional shifting under masking conditions and a shift and a skew for unmasked primes.

The partly mixed results across these two studies are not conclusive regarding the influence that the relatedness of the prime has on shaping the distributions under masked and unmasked conditions. In particular, whereas both studies reported only a shift with masked primes, they found conflicting results with unmasked presentations. Balota et al. showed only a shift, whereas Gomez et al. also reported a change in the rate of accumulation of information.

To our knowledge, no such attempts to investigate RT distributions comparing masked and unmasked paradigms have been made in bilingualism research. The present study aims at filling this gap by exploring data from two cross-language priming studies with lexical decision tasks. Chaouch-Orozco et al. (2021) investigated masked translation priming, whereas Chaouch-Orozco et al. (submitted) studied overt (unmasked) translation as well as cross-language semantic priming. The use of bilingual stimuli in the data sets analysed here constitutes a further increase in scope with respect to Balota et al. (2008) and Gomez et al. (2013). These bilingual stimuli were of two types: translation equivalent pairs (e.g., *dog-PERRO*, Spanish 'dog') and cross-language semantically related pairs (e.g., *milk-QUESO*, Spanish 'cheese'). Arguably, translation priming may involve a semantic component, especially when observed between non-cognate translation equivalents (i.e., pairs without overlap at the form level), where the opportunities for priming to arise from sublexical activation are minimal. At least to some extent, then, the types of cognitive processes recruited during monolingual semantic priming may be related to those involved in translation priming, providing grounds for a relevant comparison. This is less controversially true for cross-language semantically related pairs.

The only previous study (that we are aware of) to analyse RT distributions in the cross-language priming literature is Nakayama, Lupker, and Itaguchi (2018). Although not specifically interested in differences between masked and unmasked paradigms, Nakayama et al. (2018) inspected the distributions of masked translation priming data, with primes presented for 47 ms and followed by an "extra" mask ("&&&&&") for 23 ms. Their focus was on the priming asymmetry, whereby priming effects are typically weaker or absent with second language (L2) primes and first language (L1) targets, in contrast to generally robust findings in the opposite direction (Wen and van Heuven, 2017, for review). Nakayama et al. (2018) fit an ex-Gaussian function to the observed data and also performed a quantile analysis. The results revealed significant effects of μ and τ (i.e., a shift and a change in the shape of the distributions). In addition, the authors explored the effect of prime and target frequency, which is also central to the present study. Linear mixed-effects models evidenced that priming in the longer trials was driven by high-frequency L2 primes when the L1 targets were less frequent (i.e., more difficult to respond to). Interestingly, their findings contradict the observations in Balota et al. and Gomez et al. in their masked experiments, and the expectation that prime awareness would be necessary to create an episodic memory trace that generates a change in rate in addition to the shift. Thus, these results can be seen as support for Bodner and Masson's arguments on the possibility of prime retrieval occurring during target processing in masked paradigms, with that possibility being even more likely when the speed of (or time for) processing the prime increases. After all, it is possible that the prime's speed of processing, or the time available to retrieve the prime's information, has an impact on how fast the evidence on the familiarity of the cue accumulates.

All in all, given the scarcity of research in this topic and the generally mixed results the present study seeks to contribute to the literature employing RT distributional analyses to explore masked/unmasked cross-language priming data. We do so with an eye on the roles of an individual difference variable (L2 use) and two word-level factors (word frequency and concreteness), which were the focus of the original studies and may impact the distributions in interesting ways, either through main effects or by modulating the size, shape or nature of the standard priming effects in interaction with these.

4.2 The present study

We re-analyse here two openly available data sets containing cross-language priming data, focusing on their distributions instead of mean performance, intending to provide complementary insights to those already gained with the more traditional analyses of mean RTs conducted in the original studies. The masked data (Dataset 1) was obtained from Chaouch-Orozco et al. (2021). The authors were interested in exploring the contributions of three speaker- and word-level factors measured continuously: L2 proficiency, L2 use, and word frequency. To that end, they tested 60 Spanish-English bilinguals living in the UK, with L2 proficiency ranging from upper-intermediate to advanced and differing degrees of L2 use. The frequency of stimuli ranged from low to medium. The presentation procedure consisted of a mask of hash signs shown for 500 ms, followed by a 60 ms prime, immediately followed by the target. As the authors were interested in translation priming asymmetries (i.e., differences in the size of priming effects as a function of target language), participants were tested in two separate lexical decision tasks (L1 prime to L2 target and L2 prime to L1 target). Their results revealed a significant priming effect in the L1-L2 direction. Further, there was a significant effect of L2 use in L2-L1 priming; those participants with more active exposure to and use of the L2 benefitted more from the presence of L2 related primes (i.e., they showed larger L2-L1 priming effects). Interestingly, the effects of L2 proficiency were negligible. Word frequency had a minor impact on the priming patterns, too: a marginally significant effect of L2 prime frequency, and an effect of L2 target frequency within a complex (higher-order) interaction.

The unmasked data (Dataset 2 and 3) were obtained from Chaouch-Orozco et al. (submitted). The study continued the line of Chaouch-Orozco et al. (2021) more comprehensively, while factoring out the L2 proficiency variable. The authors tested 200 highly proficient Spanish-English bilinguals with varying degrees of L2 use. One of the foci of the study was placed on investigating how the degree of semantic overlap between cross-language pairs affected priming effects. For that purpose, 314 translation equivalent pairs (split between concrete and abstract words) and 112 cross-language semantic associative pairs were employed in Experiments 1 and 2, respectively. Words from a wide range of frequencies were used. In both experiments, participants were shown a fixation point for 500 ms, followed by a 200 ms prime preceding the target. The results showed robust effects of prime type (i.e., priming effects), prime frequency in both directions (i.e., larger priming with more frequent primes), and concreteness (i.e., larger priming with concrete words, although only in the L1-L2 direction). In comparing the results of the original studies with those in the present one, we aim at showing the advantages of combining traditional mean-based and distributional analyses for a better understanding of bilingual visual word recognition.

Predictions are not entirely straightforward given the scarcity of comparable previous studies and the mixed results therein. On the one hand, Balota et al. (2008) and Gomez et al. (2013) observed distributional shifting for masked primes, suggesting a head start for the processing of a related target. Contrary to that, Nakayama et al.'s (2018) results showed increased benefit from the prime throughout quantiles (i.e., a compound cue effect), suggesting that masked primes can also create an episodic memory trace despite unawareness. For this reason, it was not entirely clear what to expect from the masked priming data drawn from Chaouch-Orozco et al. (2021). Likewise, previous results differ as to whether relatedness causes a change in the shape of the RT distributions under unmasked conditions: it did so in Gomez et al. (2013), but not in Balota et al. (2008).

Moreover, given the results in the two original studies from which the datasets were obtained, we expected L2 use (in Dataset 1) and word frequency and concreteness (in Dataset 2) to influence the parameters' estimates, especially

μ, as this would indicate effects more in line with what was obtained in mean-based analyses. Nevertheless, opportunities for the prime to be processed and create a memory trace can increase under certain circumstances. How fast a prime is processed, or how much time is allowed for it to be processed, may determine its ability to create a joint cue along with the target, and thus affect the rate of accumulation of information. For this reason, the accumulated evidence on the cue familiarity can be larger with more frequent primes or less frequent targets, resulting in larger priming effects across quantiles as a function of word frequency. Note that this may occur even under masked priming conditions, as in Nakayama et al. (2018) and as argued by Bodner and Masson.

Likewise, if increased L2 use results in stronger activation of L2 words and faster processing of those words (see, e.g., *Multilink*; Dijkstra et al.,2019), we should also observe combined effects of prime type and L2 use on the rate of evidence accumulation. Lastly, the level of concreteness of the translation equivalent pairs may influence the familiarity of a potential joint cue, too. If we assume that abstract translation equivalents have a relatively lower degree of semantic overlap, with semantic representations in such pairs being comparatively less aligned, we might expect lower familiarity values for abstract compound cues than for concrete ones. This should, in turn, create larger priming effects on τ for concrete words.

4.3 Data analysis

The following analyses were performed by fitting an ex-Gaussian function to the three datasets. Recall that the function is captured by three parameters, μ, σ, and τ. Parameter estimation was carried out in R (version 3.6.1; R Core Team, 2019), with the function *timefit* from the *retimes* package (Massidda, 2013), which employs the maximum likelihood method. Repeated measures ANOVAs on these estimates with prime type (related and control) and the factors of interest in each dataset were performed. Also, as Balota et al. (2008) recommend, quantile analyses were conducted to compare function fits and the resulting distributions. To do this, mean RTs were calculated grouped by participant, prime type, and quantile (0.1, 0.3, 0.5, 0.7 and 0.9). Repeated measures ANOVAs on these means with prime type and quantile were performed.

4.3.1 Results

4.3.1.1 Dataset 1 - Masked translation priming

4.3.1.1.1 Ex-Gaussian analysis

The data were obtained from Chaouch-Orozco et al. (2021). The stimuli consisted of translation equivalent pairs, with frequencies ranging from low to medium. Recall that the authors observed a significant effect of L2 use on masked priming effects in the L2-L1 direction. Importantly, the factor did not interact with (prime or target) frequency. Consequently, for the present analysis, we first obtained parameter estimates for the raw data from the two translation directions, without considering word frequencies. Then we run separate ANOVAs with prime type and the interaction between prime type and L2 use.

In the L1-L2 direction, there was a significant main effect of prime type for μ, $F(1, 114) = 4.1$, $p < .05$. The effects for σ and τ were non-significant. This indicated a shift in the RT distribution as a function of prime type. Further, the main effect of L2 use was non-significant for the three parameter estimates. Finally, the interaction between prime type and L2 use reached significance for σ, $F(1,114) = 4.46$, $p < .05$, revealing a change in the spread of the Gaussian component of the distributions as a function of condition and L2 use. In the L2-L1 direction, where the effect of L2 use was found in the original study, no significant effects were obtained for the interaction between prime type and L2 use for the three parameters.

As shown, the effect of L2 use on priming effects was relatively modest in these analyses, significant only in the L1-L2 direction and only for σ. For this reason, we decided to perform further analyses, collapsing across L2 use and exploring the potential effects of prime and target frequency on the resulting distributions. We ran new ANOVAs on this data set with prime type and the interaction of prime type and (prime and target) frequency.

In the analysis with prime frequency, the main effect of prime type was significant for μ in the L1-L2 direction, $F(1, 232) = 9.53$, $p < .01$, reflecting a distributional shift. The effect did not reach significance for σ and τ. A marginally significant interaction between prime type and prime frequency was observed for μ, $F(2, 232) = 2.96$, $p = .054$. These results confirmed the shift in the distribution caused by prime-target relatedness when L2 use was employed in the previous ANOVAs. In the L2-L1 direction, only a marginally significant effect of prime type for μ was observed, $F(1, 236) = 3.72$, $p = .06$. This marginal effect indicating a shift in the distribution is not a particularly new insight, since the numerical priming effects in both directions were very similar in the original study (39 ms vs 38 ms).

In the analysis of target frequency, the main effect of prime type was again significant in the L1-L2 direction for μ, $F(1, 232) = 11.92$, $p < .001$. The interaction between prime type and target frequency was significant for that estimate, too, $F(2, 232) = 3.9$, $p < .05$, indicating a distributional shift for the effect of relatedness that is larger with less frequent targets. The effect of prime type approached significance for σ, $F(1, 232) = 3.47$, $p = 0.06$. No significant effects arose for τ. Finally, in the L2-L1 direction, there was a significant effect of prime type for μ, $F(1, 236) = 6.1$, $p < .05$. No main effect or interaction was significant for σ. In the analysis for τ, the interaction between prime type and target frequency approached significance, $F(2, 236) = 2.72$, $p = 0.07$, indicating a larger change in the spread of the RT distributions for low-frequency L1 targets.

4.3.1.1.2 Quantile analysis

As in the ex-Gaussian analysis, we first explored the effect of L2 use in the two translation directions. We run ANOVAs on the mean RTs for each participant in each quantile with prime type, L2 use and quantile (0.1, 0.3, 0.5, 0.7, 0.9). We applied Greenhouse-Geisser corrections for violations of the sphericity assumption. In both translation directions, neither the interaction between prime type and quantile nor the interaction between prime type, L2 use and quantile reached significance, indicating and confirming that the effect of prime type does not increase in higher quantiles (i.e., there is no distributional shifting).

Again, further ANOVAs were carried out with prime type and the interaction between prime type and (prime and target) frequency. The quantile factor did not significantly interact with neither prime type nor prime type and (prime or target) frequency in any direction, indicating no changes in the shape of the distributions.

4.3.1.1.3 Summary

Overall, the ex-Gaussian and quantile analyses revealed a consistent distributional shift in the L1-L2 direction as a function of prime type. A similar shift in the L2-L1 direction was only observed in the ex-Gaussian analysis when the interaction with target frequency was introduced in the model (see Figures 5 and 6). Importantly, no significant effects were obtained for τ in any analysis. Therefore, we can conclude that under masked translation priming conditions, priming effects consisted of a distributional shift only, indicating that the prime provides a head start to the target's processing. Finally, the present results are somewhat puzzling regarding the impact of L2 use on the priming effects, which was significant in the L2-L1 direction in the original study. The reason for these mixed results might lie on the different dependent variables (function parameter estimates vs means) and statistical methods (linear mixed-effects models vs ANOVA) employed in the two analyses. This is, in and of itself, a good methodological argument for approaching our datasets with different analytical tools.

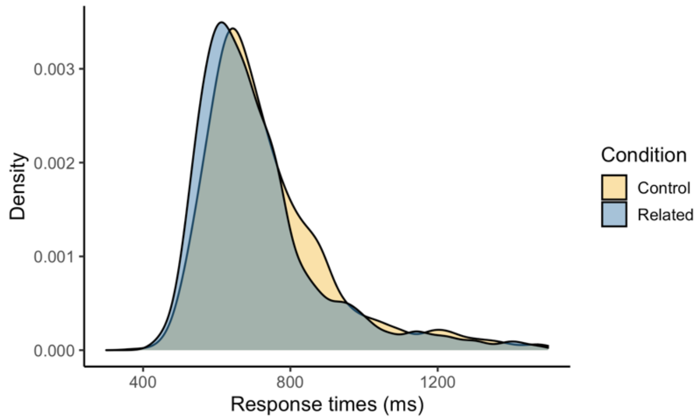

Figure 5. Density plot of RTs in the L1-L2 direction in the related and control conditions with masked primes.

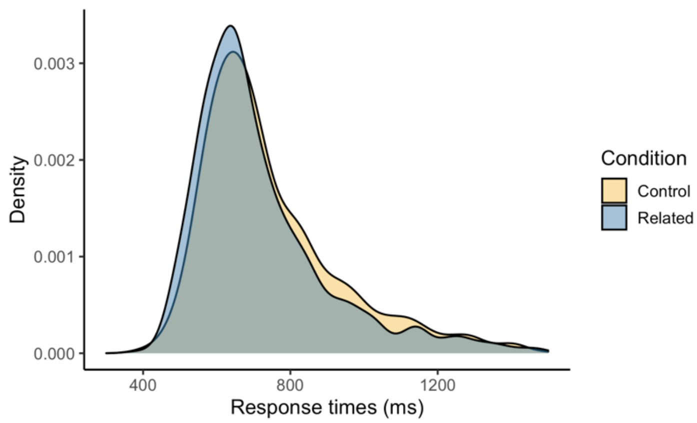

Figure 6. Density plot of RTs in the L2-L1 direction in the related and control conditions with masked translation primes.

4.3.1.2 Dataset 2 - Unmasked translation priming

4.3.1.2.1 Ex-Gaussian analysis

These data were obtained from Experiment 1 in Chaouch-Orozco et al. (submitted). The stimuli consisted of concrete and abstract translation equivalents of different frequencies (low to high). An initial exploration of the effect of concreteness, with models for each translation direction including prime type and the interaction between prime type and concreteness, replicated the significant effect observed in the original study. The interaction between prime type and concreteness was significant for μ, $F(2, 1524) = 8.65$, $p < .001$, only in the L1-L2 direction, reflecting a larger priming effect for concrete words that remained constant across quantiles. No significant interactions for σ and τ were observed. Given this outcome, subsets were created for the two translation directions and for concrete and abstract words. The present analysis focused on the effects of prime and target frequency. Different analyses were carried out for each subset exploring the effects of prime type and the interaction of prime type and (prime and target) frequency.

In the analysis of concrete pairs in the L1-L2 direction, prime type showed a significant effect for μ, $F(1, 760) = 124.05$, $p < .001$, σ, $F(1, 760) = 12.99$, $p < .001$, and τ, $F(1, 760) = 60.08$, $p < .001$. These effects indicated not only a shift in the distributions but also a change in the rate of accumulation of evidence for the effect of prime type. Moreover, the interaction between prime type and prime frequency reached significance for μ, $F(2, 760) = 10.22$, $p < .001$, and τ, $F(2, 760) = 15.72$, $p < .001$, suggesting both larger priming effects with more frequent primes and an increase of this difference in the higher quantiles. For abstract pairs, the main effect of prime type was significant for μ, $F(1, 760) = 32.72$, $p < .001$, σ, $F(1, 760) = 9.98$, $p < .001$, and τ, $F(1, 760) = 26.11$, $p < .001$. Again, this indicates both a shift and a skew in the distribution as a function of prime type.

Also, the interaction between prime type and prime frequency was significant for μ, $F(2, 760) = 18.38$, $p < .001$, suggesting that more frequent primes yielded larger priming effects, with the rate of accumulation of evidence remaining stable. For concrete pairs in the L2-L1 direction, a main effect of prime type arose for μ, $F(1, 768) = 71.49$, $p < .001$, σ, $F(1, 768) = 24.35$, $p < .001$, and τ, $F(1, 768) = 26.61$, $p < .001$, indicating a shift and a change in the spread of the RT distributions. A significant interaction between prime type and prime frequency was observed for μ, $F(2, 768) = 4.03$, $p < .05$, and τ, $F(2, 768) = 4.48$, $p < .05$. For abstract pairs, the effect of prime type was significant for μ, $F(1, 768) = 50.42$, $p < .001$, σ, $F(1, 768) = 10.83$, $p < .01$, and τ, $F(1, 768) = 19.62$, $p < .001$, indicating a distributional shift and a change in the shape of the distributions for this factor. As for the interaction between prime type and prime frequency, it reached significance for μ, $F(2, 768) = 8.34$, $p < .001$, and τ, $F(2, 768) = 8.32$, $p < .001$. Both in concrete and abstract pairs, then, the significant interaction between prime type and prime frequency for μ and τ suggested that more frequent primes yielded larger priming effects which were even larger in higher quantiles.

In the L1-L2 direction, the analysis with target frequency showed a significant main effect of prime type for the three estimates, both in concrete pairs, μ, $F(1, 760) = 129.9$, $p < .001$, σ, $F(1, 768) = 15.58$, $p < .001$, and τ, $F(1, 760) = 42.21$, $p < .001$, and with abstract words, μ, $F(1, 760) = 64.13$, $p < .001$, σ, $F(1, 760) = 6.98$, $p < .01$, and τ, $F(1, 760) = 22.61$, $p < .001$, suggesting a shift and a change in the spread of the RT distributions with both word types. Similarly, the interaction between prime type and target frequency reached significance for μ and τ with concrete and abstract pairs (μ, $F(2, 760) = 29.42$, $p < .001$; τ, $F(2, 760) = 27.33$, $p < .001$; μ, $F(2, 760) = 23.06$, $p < .001$; τ, $F(2, 760) = 29.42$, $p < .001$). Similar results were obtained in the L2-L1 direction. In both concrete and abstract pairs, there was a main effect of prime type for the three parameter estimates (concrete: μ, $F(1, 768) = 71.49$, $p < .001$, σ, $F(1, 768) = 24.35$, $p < .001$, and τ, $F(1, 768) = 26.61$, $p < .001$; abstract: μ, $F(1, 768) = 50.42$, $p < .001$, σ, $F(1, 768) = 10.83$, $p < .001$, and τ, $F(1, 768) = 19.67$, $p < .001$); and an interaction between prime type and target frequency for μ and τ (concrete: μ, $F(2, 768) = 4.03$, $p < .05$; τ, $F(2, 768) = 4.48$, $p < .05$; abstract: μ, $F(2, 768) = 8.34$, $p < .001$; τ, $F(2, 768) = 8.32$, $p < .001$). Overall, these effects show not only that the effect of prime type was larger in the higher quantiles, but also that this was further modulated by target frequency (i.e., larger priming in the higher quantiles with less frequent L1 targets).

4.3.1.2.2 Quantile analysis

As in the previous analysis, we first explored the effect of concreteness. In the L1-L2 direction, a significant interaction between prime type and concreteness was observed, $F(1, 7632) = 29.74$, $p < .001$, indicating a larger priming effect for concrete words. Furthermore, prime type, concreteness, and quantile interacted significantly, $F(1, 7632) = 5.34$, $p < .05$. The interaction revealed that the increase in the size of the prime type effect was larger for concrete words. In the L2-L1 direction, neither prime type and concreteness nor prime type, concreteness and quantile interacted significantly.

In the analysis with prime frequency, L2 targets and concrete words, quantile significantly interacted with prime type, $F(1, 3812) = 57.53$, $p < .05$, reflecting increasing priming effects in higher quantiles. Also, quantile interacted with prime type and prime frequency, $F(1, 3812) = 19.47$, $p < .001$. This interaction revealed that the increase in the effect of prime type in longer trials was larger with high-frequency primes. For abstract pairs, only the interaction between prime type and quantile reached significance, $F(1, 3812) = 21.37$, $p < .001$. Thus, we observed an increase of the priming effect in the higher quantiles that was not modified by prime frequency. In the L2-L1 direction, the interaction between prime type and quantile was significant with concrete translation pairs, $F(1, 3852) = 25.67$, $p < .01$, but this was not true of the three-way interaction between prime type, prime frequency and quantile. Again, this suggested an overall larger effect of prime type as a function of quantile. Finally, when looking at the abstract pairs, quantile significantly interacted with both prime type, $F(1, 3852) = 38.88$, $p < .05$, and prime type and prime frequency, $F(1, 3852) = 11.84$, $p < .01$, suggesting larger priming effects in higher quantiles that increased even more with more frequent primes.

In the analysis with target frequency, L2 targets and abstract words, only the interaction between quantile and prime type was significant, $F(1, 3812) = 42.8$, $p < .05$, reflecting an increase in the size of the priming effect in higher quantiles independently of target frequency. With abstract words, the three-way interaction also reached significance [prime type x quantile, $F(1, 3812) = 25.68$, $p < .01$; prime type x target frequency x quantile, $F(1, 3812) = 5.26$, $p < .01$]. This showed that the effect of quantile on the priming effect was larger for less frequent targets. In the L2-L1 direction, for concrete words, only the interaction between prime type and quantile was significant, $F(1, 3852)$, = 17.94, $p < .001$. This means that the effect was not modified by target frequency. Finally, for abstract words in the L2-L1 direction, only prime type and quantile interacted significantly, $F(1, 3852) = 26.21$, $p < .01$.

4.3.1.2.3 Summary

The results from the ex-Gaussian and quantile analyses were clear with respect to the type of benefit provided by the prime in this dataset. The effect of relatedness (i.e., priming effect) was larger in the higher quantiles in all conditions (in both translation directions and irrespective of concreteness). This is evidenced by the significant interactions observed for τ in the ex-Gaussian analysis and by the significant interactions between prime type and quantile in the quantile analysis. Hence, results with both concrete and abstract words were compatible with a compound cue account. Furthermore, when looking in more detail at the concreteness effect, the present data revealed that concrete translation equivalents elicited larger priming effects only in the L1-L2 direction, as in Chaouch-Orozco et al. (submitted). Moreover, as predicted, the effect of concreteness was overall larger in longer trials, suggesting that the familiarity of the joint retrieval cues was stronger for concrete than for abstract word pairs.

Less conclusive, however, is the role of prime and target frequency. In the ex-Gaussian analysis, both prime and target frequency clearly influenced the rate of accumulation of information (with the only exception of the effect of prime frequency with abstract words in the L1-L2 data). In line with our predictions, larger priming effects in higher quantiles were observed for more frequent primes and less frequent targets (see Figure 7 for the effect of prime frequency in the L2-L1 direction). This finding suggests that the speed of (or opportunity for) prime processing is a determining factor on the ability of primes and targets to create a compound cue. Nevertheless, these results were not consistently replicated in the quantile analyses. There, the effect of frequency on the shape of the distribution was only observed when considering the frequency of L1 concrete primes, L2 abstract primes, and L2 abstract targets. All in all, the results of analyses conducted on Dataset 2 suggest that under unmasked translation priming conditions, primes can create an episodic memory trace that aids the lexical decision. This effect is very likely to be larger under optimal conditions for the primes to create a memory cue, that is, when prime processing is faster or when there is more time for it to be completed (i.e., in more difficult trials).

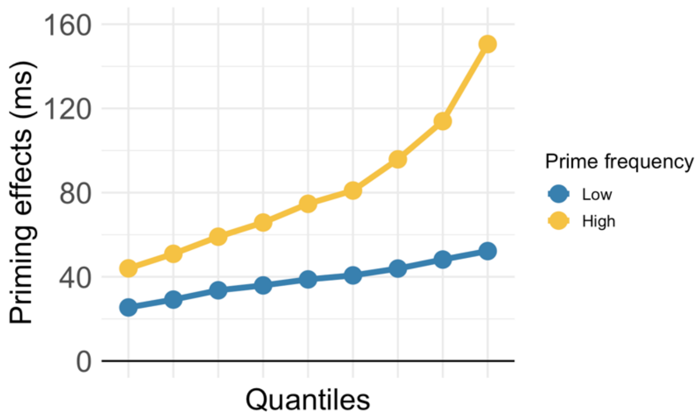

Figure 7. Plot of unmasked translation priming effects across quantiles in the L2-L1 direction with low and high frequency primes. Each point represents a quantile. Note that nine quantiles, 0.1 to 0.9, were employed for smoother curves.

4.3.1.3 Dataset 3 - Unmasked cross-language semantic priming

4.3.1.3.1 Ex-Gaussian and quantile analysis

These data were obtained from Experiment 2 in Chaouch-Orozco et al. (submitted). The stimuli consisted of cross-language semantically related pairs obtained from a free association norming study. Word frequency and concreteness were not manipulated in these stimuli, and L2 use appeared not to impact the results in the original study; therefore, in the present analysis we run ANOVAs on the three parameter estimates with only prime type as a predictor.

Overall, the main effect of prime type is modest in both translation directions. For L2 targets, the effect was only marginal for τ, $F(1, 368) = 2.75$, $p = .0098$). This is similar in the L2-L1 direction, this time with μ, $F(1, 368) = 2.73$, $p = 0.099$, in addition to τ, $F(1, 368) = 2.89$, $p = .0090$).

In the quantile analysis with L2 targets, the interaction between prime type and quantile reached significance, $F(1, 1846)$, = 4.91, $p < .01$, indicating that the effect of prime type is larger in the higher quantiles. The same effect was found in the L2-L1 direction, $F(1, 1846)$, = 4.62, $p < .01$.

4.3.1.3.2 Summary

The results obtained in the two analyses of Dataset 3 converged in an effect of prime type that increased towards the higher quantiles of the RT distributions (i.e., a change in the shape of these distributions). These results align with what we observed in the analysis of Dataset 2. However, the effect with cross-language semantically related pairs is less robust than that obtained with translation equivalents, as it was only significant in the quantile analysis, but marginally significant in the fits to the ex-Gaussian function. The plot of priming effects across quantiles for these data (Figure 8) provides some helpful visualization of the findings, revealing how priming effects were larger in longer trials. Additionally, the figure shows that cross-language semantically related primes only gave a mild head start (i.e., priming effects remained smaller across the first quantiles). This is a logical scenario considering that cross-language semantically related word pairs are, comparatively, only weakly related.

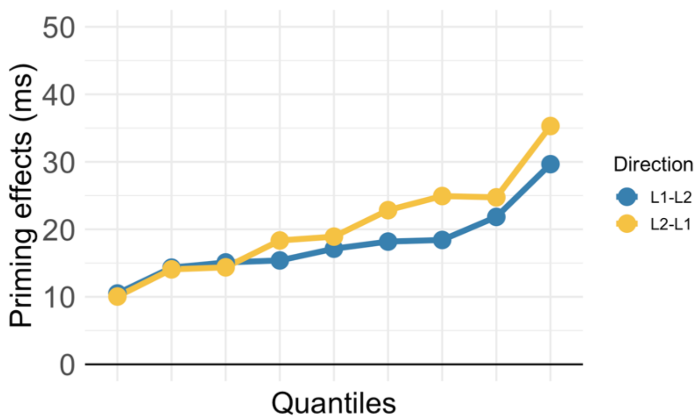

Figure 8. Plot of cross-language semantic unmasked priming effects across quantiles in the two translation directions. Each point represents a quantile. Note that nine quantiles, 0.1 to 0.9, were employed for smoother curves.

4.4 Discussion

The present study investigated RT distributions obtained from three bilingual lexical decision experiments with masked and unmasked (overt) priming, focusing on two hypotheses about the nature of priming effects. On the one hand, the pre-activation account explains priming as the activation of some of the target's features as a result of prime processing, thus providing a head start to the processing of the target word. In other words, some of the prime's processing is transferred to that of the target, speeding up responses in the related condition as opposed to a control condition where the lack of overlap between prime and target makes prime activation "unusable" in the process of target word recognition. Such a scenario would be reflected on the RT distributions by an equal shift across quantiles. On the other hand, under a compound cue account, primes can create episodic memory traces that, in the case of related primes, aid lexical decision by improving the quality of the information required to respond to the target word. The larger the familiarity of the joint cue created between the prime and the target, the shorter the response latencies. Crucially, this process is characterised by an accumulation of evidence leading to higher familiarity values, resulting in larger priming effects in trials involving longer response times.

In the analysis of Dataset 1, we inspected masked translation priming data with 60 ms primes. Datasets 2 and 3 were obtained from two experiments with unmasked 200 ms primes employing translation equivalents and cross-language semantically related pairs, respectively. Overall, the results of ex-Gaussian and quantile analyses were conclusive with regards to how relatedness affects the RT distributions under masked and unmasked presentations (although differences are occasionally found between the two approaches). While masked priming effects were mainly driven by a stable distributional shifting across quantiles (i.e., a head start), those observed under overt (unmasked) priming procedures included changes in the spread of the distributions too, a pattern that is more compatible with the predictions of the compound cue account. Next, we discuss the specifics of these results, which raise interesting questions about the role of the factors we investigated in the original studies and speak to the value of employing distributional analyses as a complementary tool when dealing with chronometric data.

With masked primes, the effects obtained were relatively small and restricted to the μ parameter in the ex-Gaussian analysis. Notably, this effect was more consistent in the L1-L2 direction. Such an outcome is not entirely unexpected, considering the variable nature of masked translation priming in general, and particularly in the L2-L1 direction (Wen & van Heuven, 2017). Note that, in the original study, although numerically similar to the effect observed in the opposite direction, the L2-L1 priming effect did not reach significance. Therefore, both the current and the original analyses (employing mean RTs and linear mixed-effects models) converged on robust priming in the L1-L2 direction as well as less reliable effects with L1 targets (i.e., a masked translation priming asymmetry).

Our results aligned with those obtained by Balota et al. (2008) and Gomez et al. (2013), and stand in contrast to what Nakayama et al. (2018) reported. The present data suggests that masked primes cannot create an episodic memory trace (i.e., they cannot form a joint cue with the targets). This was the case even when factoring in prime and target frequency, which showed non-significant effects.

Arguably, under optimal conditions (i.e., when prime frequency is high and target frequency is low), there are more opportunities for the joint retrieval cue to be created. Although we investigated word frequency, like Nakayama et al., we did not proceed in the exact same manner. The present analysis did not attempt to inspect the combined effects of prime and target frequency because these factors presented a low correlation in our data (r = 0.3), and the number of observations was relatively small. It would be interesting for future research to try to replicate Nakayama et al.'s (2018) findings while exploring the role of these two factors simultaneously with variations in the experimental design.

Turning our attention to unmasked translation and cross-language semantic priming (Datasets 2 and 3), results are conclusive with respect to the type of account (i.e., head start or compound cue) they support. Overall, our results were in line with the joint retrieval theory. That is, priming effects were increasingly larger across quantiles. The evidence, however, was relatively less strong with semantically related cross-language pairs (the effect was marginally significant in the ex-Gaussian analysis, albeit significant in the quantile analysis). Whereas Balota et al. (2008) observed only distributional shifting, Gomez et al. (2013) reported a shift and a skew with unmasked primes. Both studies employed within-language priming. The present results are in line with those reported by Gomez and colleagues (i.e., the priming effect was larger in the higher quantiles), but it should be noted that the datasets analysed here used cross-language word pairs.

Finally, visual inspection of the quantile plot for cross-language semantic priming confirmed the findings revealed by the ANOVAs while also showing that a timid head start might be driving the effect. This relatively small initial boost is not unexpected considering that we are dealing with semantic relations—comparatively weaker per se as they do not include any link at the (sub-)lexical level, and have variable degrees of semantic overlap—between words from different languages. That is, a cross-language related prime would only activate a comparatively smaller number of the target's semantic features. Further, recall that Gomez et al. employed monolingual stimuli. Thus, these differences in the type of stimuli used can explain the variation in the size of priming effects in the present data (24 ms and 23 ms) and Gomez et al. (2013) (44 ms), and the more robust evidence of a change in the spread of the distributions in the latter study. Therefore, while priming onset (and the onset of the creation of the prime-target cue, Ratcliff & McKoon, 1988: 392) would have been similar in both studies, the within-language related joint cues in Gomez et al. (2013) would have had higher familiarity than the between-language cues in the present data, resulting in faster responses and larger priming effects.

We now turn our attention to the individual- and word-level factors under investigation in the present study. Importantly, although there were no relevant effects of L2 use in the data, both concreteness (in the L1-L2 data) and frequency (in both translation directions) had an impact not only on the priming effects but also on the increase in effect size towards the higher quantiles (although some differences were observed between the different analyses).

The negligible effect of L2 use in the masked translation priming data contrasts with what was observed in Chaouch-Orozco et al. (2021), where participants with more L2 use showed larger priming effects. The only effect of language use in the present data was obtained for L2 targets (not in the L2-L1 direction, as in the original study). Further, language use affected the σ parameter, which is related to the standard deviation of the Gaussian component and not immediately interpretable since the two hypotheses we focused on are more in line with effects for μ and τ. It is likely that the different statistical approaches employed in the two studies explain at least part of these divergences.

In the case of concreteness (explored in the unmasked translation data), we expected that the compound cue familiarity would be stronger for concrete translation equivalents because of a larger degree of semantic similarity. This would lead to a larger increase of the priming effects across quantiles for this type of words. The results of the quantile analysis partially supported this hypothesis. L1-L2 priming was larger for concrete words, potentially because of a stronger activation of features from prime to target, and the effect grew larger with time. This purported influence of concreteness and degree of semantic overlap on cue familiarity adds a novel dimension to the original retrieval theory, and calls for further research to prove its validity. Likewise, as suggested in Chaouch-Orozco et al. (submitted), the lack of an impact of concreteness on the priming effects in the L2-L1 direction raises questions regarding potential qualitative differences in the representation of L1 and L2 words, and should also be further investigated.

Focusing now on the effect of word frequency (with unmasked translation priming data, where the factor was relevant), as with concreteness, the results did not completely converge across the ex-Gaussian and quantiles analyses. Both prime and target frequency significantly affected priming for μ and τ (with the only exception being the frequency of L1 abstract primes). However, the results from the quantile analysis were intriguing: frequency modulated a change in spread of the distributions only in some specific conditions: when the frequency of L2 concrete primes, L1 abstract primes, or L2 abstract targets was involved. Nevertheless, there was strong evidence overall that priming was larger with more frequent primes or less frequent targets, and that the interactions with word frequency also led to a larger effect in higher quantiles. These results supported our hypothesis regarding the role that word frequency and the speed of (or time available for) prime processing may play on the possibility of compound cue creation during this type of experiments. Given the present findings and the robust influence of prime frequency in Chaouch-Orozco et al. (submitted), we can confidently say that word frequency is a crucial modulator of translation priming (at least in unmasked paradigms). Moreover, this predictor leads to both distributional shifting and skewness (see Balota & Spieler, 1999, for a similar frequency effect in LDTs). As already argued in Chaouch-Orozco et al. (2021, submitted), the role of prime and target frequency (both categorically and continuously operationalised) on priming effects should be further investigated in cross-language priming studies.

Finally, besides the primary goal of showing how RT distributions can provide useful information about the cognitive processes recruited during lexical decisions under masked/unmasked priming conditions, the present study also attempted to highlight the advantages of employing this type of analysis in parallel with traditional mean-based approaches. In this sense, although the present findings and those in Chaouch-Orozco et al. (2021, submitted) do not entirely converge, there is still a notable degree of overlap. Whereas the effect of L2 use on L2-L1 masked priming effects was not replicated here (since it appeared, quite unexpectedly, in the L1-L2 data here), the same was not true for the impact of concreteness and prime frequency—although, also unexpectedly, an effect of target frequency was observed here too. As discussed, these mismatches may be expected given that each analysis makes use of different dependent variables (e.g., mean RTs vs μ, σ, and τ), methods (linear mixed-effects models vs ANOVAs), and practices (e.g., removing outliers or not, transforming response times). Furthermore, distributional analyses may need (even more) observations than typical mean approaches to capture significant effects. To keep exploring these divergences while approaching the study of empirical data from new points of view, we believe future research would benefit from incorporating distributional analyses, which not only provide information beyond the reach of mean RT performance but can serve as benchmarks for more traditional mean-based explorations.

4.5 Concluding remarks

While the use of response times, assumed to reflect the time course of cognitive processes, is ubiquitous in the study of language processing, researchers tend to focus only on the mean in their analyses. Although there are good reasons for proceeding that way, there is also an inherent loss of information when we trade off the intrinsic fidelity to the participants' behaviour preserved in the observed RT distributions for the simplicity, robustness and convenience of mean values. The present data join an increasing number of studies turning the focus of attention to the entire RT distribution, and acknowledging the informativeness of the typical long-tailed skewness—regarded as highly problematic in traditional mean-based analyses. As is generally the case with innovations in psycholinguistics research, distributional analyses are comparatively more common—although by no means prevalent—within the monolingual literature. Notably, this way of exploring the data is almost absent from bilingualism research. In that sense, we hope that studies like this one help communicate the advantages of this analytical method and consolidate its use, if not as the main analysis procedure, at least as part of the statistical toolkit available to bilingualism researchers employing chronometric data.

Chapter 5: Conclusion and future directions

In this chapter, we come back to the main three goals that motivated the present dissertation:

1. Investigating how subjective word frequency and executive control affect cross-language priming.
2. Exploring how the degree of semantic overlap between stimuli pairs impacts visual word recognition and priming effects.
3. Examining how the type of prime presentation influences the cognitive processes recruited during task completion.

5.1 The role of subjective word frequency and executive control on cross-language priming

A much-debated aspect of bilingual lexicon research is whether the representation and processing of L1 and L2 words are *qualitatively* or just *quantitatively* different. For the RHM, L2 words' direct connections to the conceptual store are weak and semantic access is primarily mediated by the L1 translation equivalents. The reason lies in the *qualitatively* different way words in the two languages are learned (i.e., via L1 translation equivalents in the case of L2 words). Weak L2 lexical-semantic connections would be especially prominent at low proficiencies, where experience with the L2 has not been sufficient for the L2 words to establish strong links with the conceptual store and be more independent from their L1 translation equivalent pairs. For Multilink, however, L2 words have access to concepts even from early on in the acquisition process. Further, frequency is a much more deterministic factor on how L2 words are represented and processed than the way of learning vocabulary—for which Multilink remains agnostic, in fact. Therefore, according to the model, no mechanism other than subjective word frequency—which is language-independent—is needed to account for the potential processing differences between L1 and L2 words. Lastly, the DFM resorts to quantitative-only distinctions between L1 and L2 words' processing. In this model, the underlying mechanism to explain activation dynamics in priming experiments is the degree of semantic overlap between stimuli pairs.

The predictions made by the three models mostly overlap as per the role of L2 proficiency, L2 use and word frequency. Succinctly, more experience with a given language would lead to faster responses with related pairs when the primes are in that language. For example, increased L2 proficiency, L2 use, or L2 word frequency (all operationalizations of L2 experience) may raise L2 words RLA (according to Multilink), lead to stronger L2 lexical-semantic connections (according to the RHM), or guarantee richer semantic representations for L2 words, potentially increasing the overlap with their L1 pairs (according to the DFM). Under all scenarios, the results would be larger L2-L1 priming effects because of the greater capacity of the prime to stimulate the target's processing. Thus, because the three models make the same prediction (i.e., larger priming with increased L2 experience), the present data are inconclusive as per the cognitive mechanism behind the way L1 and L2 words are processed and how that is reflected on cross-language priming.

However, the present results make a substantial contribution to understanding the role of the three experience-related factors under examination. The findings can be summarised as follows: i) Overall, L2 proficiency seems to have little impact on the masked priming data in the first study (recall the factor was not investigated in the second study). ii) L2 use modulated L2-L1 priming in the first study, but its contribution in the second study occurred in the L1-L2 data only, and it was modest (i.e., only in combination with target frequency). iii) Prime frequency was a robust modulator of priming effects in the second study, but its role was only marginal in the L2-L1 direction in the first study.

One of the main findings in the first study was the absence of an L2 proficiency effect, contrary to the models' predictions. We argued that the way the variable was measured (i.e., with a test assessing formal knowledge of the language) could explain the minor impact of the factor. Further, given the significant influence of L2 use, we concluded that active use of and experience with a language is a much more deterministic modulator of how words are represented and processed than how well the grammar rules of that language are known. However, there are caveats to this conclusion. First, albeit in a complex interaction with target language and frequency, there was an effect of L2 proficiency, indicating that the factor was able to explain some of the data. Second, the range of L2 proficiencies we employed (upper-intermediate to advanced) could not discard a more relevant role of the predictor at earlier stages of L2 acquisition. In other words, perhaps, our participants were too proficient for the predictor to yield a more robust effect. However, the range of proficiencies we could test was constrained by the low-to-medium frequency words employed. That is, participants at lower proficiencies would not have known the stimuli.

In light of the findings reported in this dissertation, L2 use seems to play an important role in L1 and L2 word processing, although the exact contribution of the factor remains unclear. In the first study, where primes were masked and very briefly presented, the results aligned with what we predicted, and language use only affected the L2-L1 direction. That is, more active L2 use led to increased L2-L1 priming. It may be the case that for rapidly processed L1 primes, differences in L1 use are not so relevant. Interestingly, the second study observed the opposite pattern: only increased L1 use led to larger priming effects in the L1-L2 direction (and in interaction with frequency). Although finding a significant effect of L1 use in the L1-L2 data was not unexpected, the question of why increased L2 use—which, arguably, can be more deterministic than increased L1 use when investigating bilingual lexical processing—did not yield larger L2-L1 priming effects remained unexplained and calls for further investigation. Overall, the effect of language use in the second study was not as strong as in the first one. Such a finding suggests that, at high proficiencies, differences in the amount of L2 use may not be as relevant as we predicted them to be.

As per word frequency, the effect of prime frequency is straightforward and robust in the second study: More frequent related primes in both languages elicited faster responses than less frequent ones. As discussed, this finding does not shed light on which of the theoretical models' assumptions are on the right track, but it does highlight the essential role that experience with the prime has on cross-language priming. In that sense, despite the lack of effect when primes were masked—which may be due to the short presentation and the impossibility of frequency influencing the targets' processing facilitation—the significant effect of prime frequency speaks of the factor being the most influential of the three predictors studied here, at least within the type of population and stimuli investigated in this dissertation. Therefore, in light of the present results, it seems that standard measures of word frequency are more suited to proxy subjective word frequency. Nevertheless, more research is needed to determine the contribution of the different experiential factors to the construct of subjective word frequency, especially when they are considered in interaction.

Finally, our results regarding executive control point towards a relevant role of the predictor in this type of experiments, potentially due to the need to monitor competing mental sets when completing the tasks. Interestingly, switching abilities were less relevant with more frequent pairs, suggesting that those faculties are less demanded in optimal processing conditions (i.e., when primes and targets are more frequent and rapidly processed). We are optimistic about the prospects of investigating executive control in future cross-language priming studies. For instance, it would be interesting to examine whether the influence of executive functioning occurs during the selection of the word candidate or at a higher level, when deciding on the target's status.

5.2 The role of the degree of semantic overlap between pairs on cross-language priming

We now address the second main goal of the present dissertation. According to the RHM and Multilink, bilingual semantic representations are holistic, and translation equivalents overlap in meaning. This is a controversial assumption in light of, for example, a relativist view of the perceptual system or a distributional semantics approach, where meaning depends on context. Contrary to the RHM's and Multilink's localist view, in a feature-based fashion, the DFM assumes that meaning consists of semantic features and processing is influenced by connection weights across those features. Therefore, if priming involves a semantic component, stronger semantic activation would result in larger priming. Crucially, that stronger activation would occur in prime-target pairs where the degree of features overlap is larger. We tested this hypothesis in Experiment 1 from the second study manipulating concreteness, as abstract meanings are believed to be more disperse and less consistent across languages. Further, in Experiment 2 from the same study, we employed cross-language semantically related pairs, where the degree of semantic alignment is at a minimum. We expected the size of the priming effects to vary as a function of the amount of semantic overlap; priming would be larger with concrete translation equivalent pairs and lower with cross-language semantically related pairs. The results partially supported these predictions. First, we observed an effect of concreteness with the translation equivalent pairs. However, as it occurred with the effect of L2 use, it only appeared in the L1-L2 data. Any model predict such an outcome, and it may reflect that L2 concrete representations and their connections do not develop to the same extent as L1 ones do, resulting in a flattening of the concreteness effect in the L2-L1 direction. In support of this tentative explanation, the size of the priming effects was numerically similar for abstract words in the L1-L2 direction and both concrete and abstract stimuli in the L2-L1 data. With cross-language semantically related pairs (Experiment 2), priming effects were overall smaller than in Experiment 1 with translation equivalents. This finding speaks of the importance of the degree of semantic alignment in cross-language priming effects.

Overall, the data from the second study highlights the need for further investigation on the nature of bilingual semantic representations. Both the RHM and Multilink assume holistic representations, for which the present findings are problematic. Even if word-association links are assumed between translation equivalents—which would make the comparison between concrete and abstract translation and cross-language semantic priming irrelevant—assuming a one-to-one mapping at the semantic level between translation pairs seems incompatible with the present effect of concreteness. Therefore, we consider that i) future research should continue examining bilingual semantic representation and ii) the models of the bilingual lexicon would benefit from adopting a distributed view of meanings.

5.3 The role of the type of prime presentation on cognitive recruitment in cross-language priming lexical decision tasks

The third main question this dissertation addressed was whether presenting primes in a masked or unmasked manner affects the nature of the priming effects. Crucially, this is the first time this has been explored employing cross-language stimuli. In addition, given that we used the datasets from the first and second studies, where some individual- and stimuli-level variables were explored, we also examined the contribution of these factors on the type of processes taking place during cross-language priming experiments.

Overall, the results are clear: With masked presentations, a headstart is observed. When related primes are masked and shortly presented, the type of benefit they provided comprises activation transferring from the prime to the target. This was reflected by a distributional shift, where priming did not depend on the response time for each trial. In contrast, with overt primes, the priming effect increased across time, suggesting that information about the familiarity between the prime and the target was recruited before responding. This result was reflected by both a distributional shift and skewness.

Further, prime and target frequency did not play a role in modulating the ability of the prime to create a joint retrieval cue with masked primes. That is, the distributional pattern did not change with the frequency of those primes. Importantly, this was not the case in unmasked presentations. The data from the second study revealed that, overall, the compound cue effect was larger at higher frequencies, which supports the critical role of the factor in lexical processing. Concreteness, which was expected to influence the joint cue familiarity—concrete pairs would show a higher similarity—had a significant effect on the change in the shape of the distribution in the L1-L2 data. That is, the distributional skewness was larger with concrete words. To our knowledge, this is the first time that concreteness has been investigated in light of the joint retrieval theory. Further, as it occurred with the priming effects in the second study, there was no effect of concreteness on the ability of L2 primes and L1 targets to form a compound cue. As previously argued, this may be due to a poorer semantic specification of L2 concrete words. Finally, L2 use (and L2 proficiency) did not affect the RT distributions of neither the data from the first study nor that of the second study. The outcome with overt primes confirms the negligible impact that L2 use had on the second study, aligning with what was shown in the mean-based analysis. As per the results with masked primes (first study), observing an effect of L2 use (or prime frequency, for instance) would have been surprising. After all, with masked presentations, non-awareness is the reason why masked primes are believed not to be able to form an episodic memory trace. Therefore, even if the masked primes were processed faster because they had a higher RLA—following Multilink's tenets—that would not have made them more salient.

Overall, the distributional analyses contributed to understanding the underlying cognitive mechanisms in cross-language priming studies, mainly supporting the previous reports with monolingual semantic priming. Notably, the third study spotlights the importance of employing this type of analyses, given that they provide unique information beyond the reach of mean-based implementations. Moreover, the current approach's validity was supported by the relative overlap between the results in the three studies. We encourage both mean-based and distributional analyses to be carried out simultaneously to get a comprehensive understanding of what the data at hand in chronometrical studies reflects.

5.4 Future directions

The three studies included in this dissertation provide a comprehensive investigation of the cognitive factors operating during bilingual visual word recognition and shaping the bilingual lexicon's architecture and functioning. One of the foci was placed on how the experience with the L1 and the L2 shapes lexical-semantic representation and processing. In that sense, we offer robust evidence on the role that prime frequency plays in cross-language priming. Although Multilink stipulates more clearly the relevance of frequency of use being a primary modulator of the lexicon, it cannot be concluded that the factor would not be relevant under the paradigms proposed by the RHM and the DFM. The same can be said about the role of active experience with the L2, whose importance in light of the present results seems clear but could be attenuated when bilinguals are at the latest stages of L2 acquisition. Intuitively, the amount of experience a bilingual accumulates with the L2—irrespective of how that is operationalized—must play an essential role in configuring the bilingual lexicon. However, more research is needed to better understand how different proxies of bilingual experience influence L1/L2 word processing and how those predictors interact together. Research directed to this goal would benefit from employing continuous and nuanced operationalizations along with large datasets that allow the investigation of interactions between factors. Continuing with individual differences, the second study of this dissertation contributes to the investigation of a novel predictor in cross-language priming studies. In light of the present results, executive control and its role when dealing with competing language sets and candidate selection deserve further attention in this field of research.

Another significant contribution of the present dissertation lies in the examination of the nature of bilingual semantic representations. The results outlined here, along with previous reports on the effect of concreteness on cross-language priming, spotlight the necessity of the models of adopting more detailed conceptual representations. In this line, monolingual research has explored how meaning can be represented in the brain, and many different approaches have emerged in the last decades. In fact, this is a vibrant field of investigation, where many divergent views about how to represent meaning compete and also complement each other. Those views include feature-based models, semantic similarity measures obtained from free association tasks, semantic network research, and distributional approaches based on natural language processing techniques. In our opinion, bilingualism research would greatly benefit from incorporating this type of perspectives to gain a better understanding of the organization of bilingual semantic memory. Notably, when adding bilingual experience to the equation, the already fascinating question of what meaning is is enriched with new exciting inquiries about how concepts are represented across languages, which factors influence those representations, whether or not translation equivalents' meanings align or how partially overlapping concepts coexist. The time has come to start pursuing these avenues of research.

Finally, we would like to advocate for the necessity of chronometrical research—especially in the realm of bilingual studies—to incorporate distributional analyses into their statistical toolkit. We consider that mean-based approaches should constitute the primary analyses, but hopefully we have shown the relevance of inspecting distributions, which can provide unique insights about human cognition beyond the reach of the mean.

5.5 A final word

Whereas much has been achieved during the last 40 years of research on the bilingual lexicon, the peculiarities of the complex mechanisms assumed to operate within the system remain unclear. In this sense, we consider that the empirical motivations behind this project are relevant for advancing the field. We hope that the results presented here—intriguing in many senses—are revealing about the role of some of the factors shaping the bilingual lexicon while having the capacity of sparking the researchers' interest in the myriad of aspects that remain unknown about how bilinguals use and understand words in their two languages.

References

Altarriba, J., & Basnight-Brown, D. M. (2007). Methodological considerations in performing semantic-and translation-priming experiments across languages. *Behavior Research Methods, 39*(1), 1-18.

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*(1), 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*(1), 388-407.

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.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & cognition*, 29(4), 639-647.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 128(1), 32.

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(4), 495-523.

Balota, D. A., & Yap, M. J. (2011). Moving beyond the mean in studies of mental chronometry: The power of response time distributional analyses. *Current Directions in Psychological Science,* 20(3), 160-166.

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.

Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 65(4), 229.

Bodner, G. E., & Masson, M. E. (2001). Prime validity affects masked repetition priming: Evidence for an episodic resource account of priming. *Journal of Memory and Language*, 45(4), 616-647.

Bodner, G. E., & Masson, M. E. (2003). Beyond spreading activation: An influence of relatedness proportion on masked semantic priming. *Psychonomic Bulletin & Review*, 10(3), 645-652.

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(3), 389.

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. (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(1), 16.

Brysbaert, M. (2020). Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition*, 26 August 2020. DOI: 10.1017/S1366728920000437.

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., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, 64(3), 545-559.

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. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, 1(1), 9.

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.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904-911.

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

Chaouch-Orozco, A., González Alonso, J., & 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(2), 447-474.

Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (submitted). What shapes the bilingual lexicon? Insights from a comprehensive cross-language priming study.

Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (submitted). Beyond the mean: Shape distribution analysis provides unique insights into how masked and unmasked cross-language priming effects unfold.

Chen, B., Liang, L., Cui, P., & Dunlap, S. (2014). The priming effect of translation equivalents across languages for concrete and abstract words. *Acta psychologica*, 153, 147-152.

Chuang, Y. Y., Bell, M. J., Banke, I., & Baayen, R. H. (2021). Bilingual and multilingual mental lexicon: a modeling study with Linear Discriminative Learning. *Language Learning*, 71(S1), 219-292.

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.

Cousineau, D., & Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1), 58-67.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, 128(3), 615-627.

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. (1992). Bilingual lexical representation: A closer look at conceptual representations. In *Advances in psychology* (Vol. 94, pp. 389-412). North-Holland.

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.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45(4), 1246-1258.

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

Duñabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin & Review*, 15(6), 1072-1077.

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., 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.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, 57(2), 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(4), 663.

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.

Faraway, J. J. (2002). *Practical regression and ANOVA using R*. Bath: University of Bath Press.

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.

Ferré, P., Guasch, M., García-Chico, T., & Sánchez-Casas, R. (2015). Are there qualitative differences in the representation of abstract and concrete words? Within-language and cross-language evidence from the semantic priming paradigm. *Quarterly Journal of Experimental Psychology*, 68(12), 2402-2418.

Ferré, P., Sánchez-Casas, R., Comesaña, M., & Demestre, J. (2017). Masked translation priming with cognates and noncognates: Is there an effect of words' concreteness?. *Bilingualism: Language and Cognition*, 20(4), 770.

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. (2004). Category size effects revisited: Frequency and masked priming effects in semantic categorization. *Brain and language*, 90(1-3), 276-286.

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.

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

Foy, J. G., & Foy, M. R. (2020). Dynamic Changes in EEG Power Spectral Densities During NIH-Toolbox Flanker, Dimensional Change Card Sort Test and Episodic Memory Tests in Young Adults. *Frontiers in Human Neuroscience*, 14, 158.

Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2017). Cross-linguistic phonotactic competition and cognitive control in bilinguals. *Journal of Cognitive Psychology*, 29(7), 783-794.

Friesen, D. C., & Haigh, C. A. (2018). Cross-language associative priming is influenced by language proficiency and executive control. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 72(4), 264.

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.

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.

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different?. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1731.

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In *Advances in Psychology* (Vol. 83, pp. 207-220). North-Holland.

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

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. *Marketing Science*, 23(4), 519-529.

Guasch, M., Sanchez-Casas, R., Ferre, P., & García-Albea, J. E. (2011). Effects of the degree of meaning similarity on cross-language semantic priming in highly proficient bilinguals. *Journal of Cognitive Psychology*, 23(8), 942-961.

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, 45(3), 718-730.

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.

Jin, Y. S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills*, 70(S3), 1139-1154.

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an organizing principle of the lexicon. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 67, pp. 239-283). Academic Press.

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

Kinoshita, S., Mozer, M. C., & Forster, K. I. (2011). Dynamic adaptation to history of trial difficulty explains the effect of congruency proportion on masked priming. *Journal of Experimental Psychology: General*, 140(4), 622.

Kiran, S., & Lebel, K. R. (2007). Crosslinguistic semantic and translation priming in normal bilingual individuals and bilingual aphasia. *Clinical Linguistics & Phonetics*, 21(4), 277-303.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, 6(5), 449-467.

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

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(2), 119.

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., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325-343.

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.

Li, H., Liang, Y., Qu, J., Sun, Y., Jiang, N., & Mei, L. (2021). The effects of word concreteness on cross-language neural pattern similarity during semantic categorization. *Journal of Neurolinguistics*, 58, 100978.

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.

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *The Mental Lexicon*, 3(3), 349-374.

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

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, 92, 57-78.

Massidda D 2013. *retimes: reaction time analysis. R package version 0.1.2*. https://CRAN.R-project.org/package=retimes.

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.

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(1), 70-85.

Müller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: What is shared, what is different? *Psychophysiology*, 47, 455-466.

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.

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B.Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, 76, 253-272.

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.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265-278.

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.

Perea, M., Dunabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language*, 58(4), 916-930.

Pexman, P. M., Siakaluk, P. D., & Yap, M. J. (2013). Introduction to the research topic meaning in mind: semantic richness effects in language processing. *Frontiers in Human Neuroscience*, 7, 723.

Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 787.

Pollatsek, A., Perea, M., & Carreiras, M. (2005). Does conal prime CANAL more than cinal? Masked phonological priming effects in Spanish with the lexical decision task. *Memory & cognition*, 33(3), 557-565.

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

R Core Team (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2), 59.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, 111(1), 159.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological review*, 95(3), 385.

Reilly, M., & Desai, R. H. (2017). Effects of semantic neighborhood density in abstract and concrete words. *Cognition*, 169, 46-53.

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.

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.

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. DOI: 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(5), 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(1), 157-176.

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, 4(10), 1029-1038.

Tokowicz, N., Kroll, J. F., De Groot, A. M., & Van Hell, J. G. (2002). Number-of-translation norms for Dutch—English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, & Computers*, 34(3), 435-451.

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

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

Van Hell, J. G., & De Groot, A. M. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3), 193-211.

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., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3), 458-483.

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.

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 Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7(3), 424-465.

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.

Yang, H., Hartanto, A., & Yang, S. (2018). Bilingualism confers advantages in task switching: Evidence from the dimensional change card sort task. *Bilingualism: Language and Cognition*, 21(5), 1091-1109.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, 1(1), 297-301.

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.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3-14.

Appendices

First study

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.

Second study

Appendix A. Complete list of stimuli.

Table A.1. Prime and target words and pseudowords, and concreteness in Experiment 1.

Spanish translation equivalent	English translation equivalent	Spanish pseudoword (targets)	English pseudoword (targets)	Concreteness
ley	law	ler	baw	abstract
año	year	abe	croding	abstract
amor	love	anur	wixth	abstract
lío	mess	túo	pess	abstract
odio	hate	edia	hamp	abstract
daño	harm	gazo	hask	abstract
tía	aunt	lúa	aste	abstract
jefe	boss	joñe	bomp	abstract
lado	side	bano	sipe	abstract
frío	chill	flúo	chall	abstract
noche	night	gorre	nimes	abstract
gemelo	twin	necilo	knin	abstract
fiesta	party	deusta	manty	abstract
lástima	pity	víntima	moty	abstract
mezcla	mix	mechra	mox	abstract
tamaño	size	vanazo	rize	abstract

viaje	trip	guije	spip	abstract
enfado	anger	envate	arver	abstract
sueño	dream	ruebo	bleam	abstract
manojo	bunch	sacozo	budes	abstract
suciedad	dirt	suriadal	dort	abstract
prisa	haste	briga	haits	abstract
mitad	half	pital	harf	abstract
vida	life	nila	libe	abstract
lugar	place	rumar	plawn	abstract
brillo	glow	chirro	prash	abstract
rugido	roar	dusedo	rour	abstract
olor	smell	ecor	skell	abstract
trabajo	work	pradaño	wolt	abstract
ejército	army	exáncito	angy	abstract
ajedrez	chess	ajegrua	chend	abstract
multitud	crowd	ductitul	croif	abstract
fin	end	fen	ews	abstract
hecho	fact	horro	farp	abstract
ajetreo	hustle	ajebleo	huggle	abstract
mañana	morning	pafala	perning	abstract
sonrisa	smile	sarcisa	smale	abstract
choque	clash	chehue	spash	abstract
presión	strain	cremión	strail	abstract

huelga	strike	hesiga	strind	abstract
tregua	truce	chegué	trurn	abstract
nivel	level	nibal	tuvel	abstract
rebote	bounce	depose	bouths	abstract
soborno	bribe	soponlo	clibe	abstract
hito	feat	lico	fout	abstract
caída	plunge	meína	plurse	abstract
oración	prayer	osacuas	praire	abstract
sequía	drought	señida	prought	abstract
subida	raise	duveda	ralps	abstract
alcance	scope	armange	scode	abstract
capítulo	chapter	camónumo	shalter	abstract
verdad	truth	hordad	sluth	abstract
gripe	flu	fribe	spu	abstract
tarea	task	facia	tage	abstract
dueño	owner	huejo	uffer	abstract
invierno	winter	infiarmo	warter	abstract
juventud	youth	juvendid	yeath	abstract
ansia	craving	ancio	ac	abstract
zumbido	buzz	zórnido	burf	abstract
encanto	charm	endisto	chawl	abstract
diseño	design	simejo	denifs	abstract
aumento	surge	eisanto	fleed	abstract

resaca	hangover	demasa	foleover	abstract
hambre	hunger	fimbre	henser	abstract
broma	prank	frema	outsheek	abstract
revés	setback	repís	pedback	abstract
ráfaga	flurry	díjaga	plerry	abstract
elogio	praise	ecobia	pralps	abstract
perfil	profile	ponfil	progale	abstract
alivio	relief	acegio	reroof	abstract
guión	script	quión	scrimb	abstract
reunión	meeting	teulión	soating	abstract
tristeza	sadness	prosteña	sumness	abstract
tiempo	weather	tienzo	weinter	abstract
creencia	belief	pleangia	berieu	abstract
cambio	change	campia	chathe	abstract
engaño	deceit	esvazo	decoal	abstract
deleite	delight	deciose	detisse	abstract
ayuda	help	aelta	doduanty	abstract
vistazo	glimpse	tastaño	glitzed	abstract
invitado	guest	osminado	gurnt	abstract
locura	madness	nevura	macless	abstract
amigo	friend	arezo	pove	abstract
búsqueda	pursuit	térqueda	purgoot	abstract
altura	height	etrura	reight	abstract

regaño	scolding	devaho	scumping	abstract
sentido	sense	dindido	serbs	abstract
escasez	shortage	escalot	shuntage	abstract
rasgo	trait	resno	trarf	abstract
lesión	injury	reción	uncupy	abstract
vacío	void	tadúo	voir	abstract
anchura	width	enfrora	yeam	abstract
coartada	alibi	coínnado	atipo	abstract
placer	pleasure	grader	clealure	abstract
frialdad	coldness	crieldad	cortless	abstract
deceso	demise	demico	denite	abstract
clamor	outcry	dragor	eattry	abstract
ensayo	essay	encaua	empay	abstract
vuelo	flight	guero	flinge	abstract
bondad	goodness	rondal	gan	abstract
fantasma	ghost	vintalma	ghoms	abstract
puñado	handful	tufalo	handcal	abstract
farsa	hoax	garga	hoaf	abstract
ingesta	intake	insonta	intive	abstract
tumulto	mayhem	turuzno	magwem	abstract
tropiezo	misstep	trobuejo	misgrap	abstract
hipoteca	mortgage	bitoneta	modlfage	abstract
popurrí	medley	poduchá	mutley	abstract

ruido	noise	tiado	noits	abstract
nómina	payroll	jánina	paydods	abstract
retirada	retreat	decenada	replout	abstract
milagro	miracle	sicaclo	silatre	abstract
amenaza	threat	alacava	squeat	abstract
pandilla	gang	serdilla	gank	abstract
éxito	success	áhilo	suybess	abstract
susurro	whisper	pucucho	swismer	abstract
hambruna	famine	fimpluna	taline	abstract
belleza	beauty	tebreja	wemming	abstract
premio	award	prodio	abail	abstract
activo	asset	atnizo	ampet	abstract
primo	cousin	brico	coomin	abstract
derrota	defeat	rechola	defoul	abstract
trama	plot	brada	drot	abstract
esfuerzo	effort	esluenjo	effall	abstract
diosa	goddess	gaisa	gommess	abstract
pista	clue	minta	grue	abstract
cosecha	harvest	cocigra	hanrest	abstract
prueba	proof	prieva	preaf	abstract
raza	breed	laba	bried	abstract
entierro	burial	endiatro	felial	abstract
pereza	laziness	semepa	fowiness	abstract

jugada	gamble	nudana	gattle	abstract
chisme	gossip	chosbe	gostup	abstract
culpa	guilt	celba	goult	abstract
salud	health	talid	heanse	abstract
demencia	insanity	decensio	inmitaty	abstract
riesgo	jeopardy	reusno	junkainy	abstract
salto	jump	taldo	junt	abstract
medida	measure	sereda	moolure	abstract
retrato	portrait	degraso	porstaim	abstract
reino	realm	toino	reapt	abstract
asunto	affair	acosto	shidecut	abstract
alma	soul	amba	soal	abstract
sigilo	stealth	dipito	steanse	abstract
calor	warmth	macor	warque	abstract
consulta	query	conguara	whety	abstract
consejo	advice	conciño	adhace	abstract
negocio	business	nedomia	bereness	abstract
cielo	heaven	miero	hooden	abstract
brote	outbreak	grete	leep	abstract
molestia	nuisance	polintia	nuisudes	abstract
atajo	shortcut	acage	rurge	abstract
parecido	likeness	manalido	takeness	abstract
llegada	arrival	plecana	ancipal	abstract

cierre	closure	cuelle	closand	abstract
gente	people	garte	daople	abstract
muerte	death	miante	deeth	abstract
entrega	delivery	enchega	demitoly	abstract
caza	hunt	mafa	hund	abstract
hoy	today	hol	logay	abstract
deber	duty	vejer	rury	abstract
fuente	source	hiente	shouch	abstract
ocio	leisure	osia	toesure	abstract
apoyo	support	adoez	aud	abstract
siglo	century	riplo	cerrugy	abstract
valor	courage	galir	cootage	abstract
marca	brand	canca	crand	abstract
ingreso	income	incrito	incert	abstract
olvido	oblivion	utrido	ospetion	abstract
lema	motto	loda	petto	abstract
cuenta	account	coista	acceine	abstract
cena	dinner	pona	panner	abstract
sed	thirst	sey	thisle	abstract
vista	view	hesta	yiew	abstract
boda	wedding	roga	geakness	abstract
mal	evil	ral	eryl	abstract
resto	leftover	sento	loleoper	abstract

mes	month	med	mopse	abstract
lucha	struggle	ligra	struttle	abstract
pena	sorrow	leta	dorrop	abstract
fallo	failure	bacho	forture	abstract
peso	weight	ceto	wought	abstract
carga	burden	marla	bermen	abstract
ira	wrath	ula	wrass	abstract
nana	lullaby	gara	funkavy	abstract
risa	laughter	sina	linchter	abstract
paso	step	laro	stup	abstract
moda	fashion	cona	tushion	abstract
fe	faith	ci	faire	abstract
baja	casualty	dapa	jush	abstract
capa	layer	mava	tager	abstract
nariz	nose	jaréz	nuse	concrete
oso	bear	eno	lear	concrete
sol	sun	sod	sep	concrete
abrigo	coat	allibo	roat	concrete
cisne	swan	cisde	drap	concrete
tacón	heel	tagón	hool	concrete
peine	comb	piole	cacs	concrete
puño	fist	pubo	fims	concrete
percha	hanger	pastra	hanker	concrete

rey	king	rez	kint	concrete
luna	moon	lura	moop	concrete
bruja	witch	bruba	datch	concrete
jabón	soap	jajón	sein	concrete
río	river	lúo	sover	concrete
clavo	nail	llavo	norn	concrete
reloj	watch	telop	wamps	concrete
leche	milk	reche	misk	concrete
búho	owl	lího	orm	concrete
nieve	snow	jaipe	whew	concrete
cadena	chain	cavena	chail	concrete
uva	grape	ubo	grame	concrete
charco	puddle	tarno	moddle	concrete
nido	nest	vimo	nugs	concrete
foca	seal	feda	sool	concrete
bolsillo	pocket	bolpillo	sicket	concrete
muro	wall	cino	wams	concrete
pájaro	bird	májaro	birk	concrete
cabra	goat	mabra	goot	concrete
cuerno	horn	cuergo	hoil	concrete
miel	honey	mial	honem	concrete
cuello	neck	ciello	nids	concrete
horno	oven	forgo	oken	concrete

cuchara	spoon	cucrara	sleen	concrete
dinero	money	dicero	soney	concrete
camarero	waiter	cararero	waider	concrete
cartera	wallet	cancera	gallet	concrete
guante	glove	guaste	glink	concrete
mofeta	skunk	poñeta	glone	concrete
traje	suit	chave	goam	concrete
mechero	lighter	petrero	latcher	concrete
bufanda	scarf	tufanda	scalf	concrete
oveja	sheep	oveza	shoon	concrete
caracol	snail	caracod	snain	concrete
ballena	whale	lachena	snite	concrete
flecha	arrow	plecha	andow	concrete
barba	beard	barla	bearn	concrete
esquina	corner	encaina	cark	concrete
payaso	clown	payuso	clode	concrete
tobillo	ankle	bodillo	arwhe	concrete
manzana	apple	marzana	apste	concrete
cajón	drawer	mabón	draxer	concrete
ojo	eye	ezo	eys	concrete
araña	spider	ecafa	flider	concrete
mujer	woman	muver	goman	concrete
imán	magnet	idín	madnet	concrete

pezón	nipple	mejón	napple	concrete
cuadro	painting	cultro	paunting	concrete
llave	key	claje	tox	concrete
cebolla	onion	cegolla	unoon	concrete
toro	bull	relo	bams	concrete
silla	chair	rilla	chasp	concrete
vestido	dress	tentido	cluss	concrete
ataúd	coffin	atail	coddin	concrete
puente	bridge	peinte	crorks	concrete
puerta	door	pierta	foor	concrete
langosta	lobster	vangosta	losster	concrete
cama	bed	cafo	ped	concrete
gatillo	trigger	narillo	prigger	concrete
calabaza	pumpkin	calabava	pullcin	concrete
lápiz	pencil	lópiz	puncil	concrete
trigo	wheat	chiso	sneat	concrete
paraguas	umbrella	paraguar	umblella	concrete
ala	wing	ado	wint	concrete
aguacate	avocado	abuacate	axocado	concrete
pepino	cucumber	sedeno	cucolser	concrete
erizo	hedgehog	elepo	hedgerig	concrete
avestruz	ostrich	avestriz	ostript	concrete
conejo	rabbit	corejo	rabbot	concrete

escudo	shield	espudo	shoard	concrete
ventana	window	tontana	windig	concrete
iglesia	church	inhesia	chorth	concrete
dedo	finger	hodo	fanger	concrete
bosque	forest	burque	hunest	concrete
hada	fairy	gapa	lamey	concrete
tijeras	scissors	higiras	plundors	concrete
ducha	shower	gurra	shamer	concrete
falda	skirt	dalda	skipe	concrete
ardilla	squirrel	arnilla	sprirrul	concrete
espejo	mirror	escezo	suppar	concrete
sombra	shadow	sostra	bradow	concrete
aguja	needle	aduza	neeble	concrete
hombro	shoulder	fomplo	choolder	concrete
pulgar	thumb	pulmar	thurl	concrete
armario	wardrobe	asparia	wardrolk	concrete
camarera	waitress	calarera	waseress	concrete
caja	box	paña	boc	concrete
mano	hand	mado	habs	concrete
pelo	hair	pono	hact	concrete
mesa	table	moma	lable	concrete
pulpo	octopus	pumzo	octovis	concrete
cereza	cherry	ceseja	shippy	concrete

morsa	walrus	mersa	wamnus	concrete
pato	duck	labo	dack	concrete
huevo	egg	duejo	erv	concrete
burro	donkey	lullo	monrey	concrete
queso	cheese	quido	sheese	concrete
goma	rubber	gedas	dubber	concrete
reina	queen	toina	snoup	concrete
maleta	suitcase	parita	moanpind	concrete
valla	fence	harra	ferks	concrete
pavo	turkey	mapo	furkey	concrete
lobo	wolf	lopo	woit	concrete
boca	mouth	reca	mouch	concrete
pollo	chicken	locho	pricken	concrete
rana	frog	laga	snaw	concrete
hilo	thread	vido	squead	concrete
vela	candle	dola	banble	concrete
cocina	kitchen	coreta	ketchen	concrete
bala	bullet	hara	bellet	concrete
cara	face	cajo	filt	concrete
loro	parrot	bero	pandot	concrete
codo	elbow	mogo	elpow	concrete
casa	house	mada	touse	concrete

Table A.2. Cue and generated words from the norming study, and their translation equivalents.

Cue word	**Cue word translation equivalent**	**Generated word**	**Generated word translation equivalent**
place	lugar	casa	house
dress	vestido	falda	skirt
pencil	lápiz	goma	rubber
trip	viaje	maleta	suitcase
sadness	tristeza	alegría	joy
nest	nido	pájaro	bird
arm	brazo	codo	elbow
today	hoy	mañana	tomorrow
walrus	morsa	foca	seal
milk	leche	vaca	cow
week	semana	mes	month
horn	cuerno	toro	bull
table	mesa	silla	chair
honey	miel	abeja	bee
cheese	queso	cabra	goat
herd	rebaño	oveja	sheep
love	amor	odio	hate

chill	frío	calor	warmth
spell	hechizo	bruja	witch
seal	foca	morsa	walrus
burden	carga	peso	weight
fairy	hada	cuento	tale
prank	broma	chiste	joke
scam	estafa	engaño	hoax
charm	encanto	belleza	beauty
harvest	cosecha	trigo	wheat
deer	ciervo	caza	hunt
wrath	ira	rabia	rage
soap	jabón	ducha	shower
snow	nieve	invierno	winter
support	apoyo	ayuda	help
turkey	pavo	pollo	chicken
ring	anillo	boda	wedding
miracle	milagro	fe	faith
shoe	zapato	cordón	lace
hell	infierno	cielo	heaven
month	mes	año	year
truth	verdad	mentira	lie
life	vida	muerte	death
view	vista	ojo	eye

joke	chiste	risa	laughter
smell	olor	nariz	nose
door	puerta	ventana	window
danger	peligro	cuidado	care
sun	sol	luna	moon
sorrow	pena	tristeza	sadness
dog	perro	mascota	pet
hunger	hambre	sed	thrist
needle	aguja	hilo	thread
bridge	puente	río	river
swan	cisne	pato	duck
threat	amenaza	miedo	fear
portrait	retrato	cuadro	painting
scolding	regaño	enfado	anger
strike	huelga	derecho	right
dream	sueño	cama	bed
forest	bosque	árbol	tree
goat	cabra	queso	cheese
winter	invierno	verano	summer
ostrich	avestruz	huevo	egg
smile	sonrisa	boca	mouth
right	derecho	ley	law
bee	abeja	miel	honey

umbrella	paraguas	lluvia	rain
warmth	calor	frío	chill
people	gente	multitud	crowd
pet	mascota	perro	dog
faith	fe	dios	god
jeopardy	riesgo	peligro	danger
magnet	imán	nevera	fridge
nose	nariz	cara	face
rage	rabia	ira	wrath
tale	cuento	hada	fairy
shark	tiburón	diente	tooth
comb	peine	pelo	hair
hate	odio	amor	love
key	llave	puerta	door
morning	mañana	hoy	today
plunge	caída	daño	harm
chain	cadena	oro	gold
eye	ojo	vista	view
ankle	tobillo	pierna	leg
spoon	cuchara	tenedor	fork
shield	escudo	espada	sword
noise	ruido	molestia	nuisance
axe	hacha	madera	wood

slip	resbalón	caída	plunge
thumb	pulgar	dedo	finger
death	muerte	vida	life
harm	daño	herida	wound
donkey	burro	caballo	horse
river	río	agua	water
cow	vaca	leche	milk
junk	basura	olor	smell
sheep	oveja	lana	wool
profile	perfil	lado	side
moon	luna	sol	sun
nuisance	molestia	dolor	pain
pity	lástima	pena	sorrow
suit	traje	corbata	tie
finger	dedo	uña	nail
candle	vela	cera	wax
heel	tacón	zapato	shoe
joy	alegría	sonrisa	smile
army	ejército	guerra	war
design	diseño	moda	fashion
corner	esquina	calle	street
thread	hilo	aguja	needle
elbow	codo	brazo	arm

lie	mentira	verdad	truth
jump	salto	altura	height
chair	silla	mesa	table

Table A.3. Pseudowords in Experiment 2.

English pseudowords	**Spanish pseudowords**
ringe	anur
wixth	nila
noits	causto
libe	hol
junt	hordad
denifs	pafala
tade	ula
logay	racio
jol	edia
prip	miante
sluth	macor
warter	sescira
quam	garte
hask	feda
bleam	mollo
perning	grazo
holl	quido
plawn	mial
mopse	reche
chawl	moma

henser	hesta
dorrop	pil
bermen	ezo
wrass	aduza
rarl	lura
darser	apeja
silatre	mersa
sumness	gaca
squeat	vido
hamp	rilla
deeth	gapa
scumping	mabra
moty	flúo
warque	infiarmo
plurse	gazo
lou	hodo
progale	oveza
junkainy	cucrara
nuisudes	chave
faire	idín
daople	dola
seel	mada
daw	burque

alk	encaina
fanger	cavena
cluss	lúo
sheese	bipurón
shoon	harra
honem	paraguar
woak	espudo
sleen	bodillo
chie	decitro
spip	tiado
misk	taldo
jonk	simejo
whew	asadrua
goam	resnatán
foor	endapa
lable	ruebo
yiew	invielto
hanrest	rumar
heff	med
sep	endisto
strind	fimbre
eys	leta
skell	marla

nugs	seciglo
hoil	sicaclo
madnet	prosteña
weer	alacava
banble	devaho
tager	víntima
neeble	meína
spoll	ponfil
hunest	reusno
tove	polintia
crorks	ci
outsheek	tentido
porstaim	mesala
puncil	ñabato
drap	guije
moop	truste
bea	jaipe
aud	pierta
wamnus	cocigra
cark	dedago
cag	hesiga
squead	ecor
furkey	vimo

chail	cuergo
chasp	cuelvo
sein	febrijo
angy	amillo
hool	peinte
smale	frema
tox	degraso
sover	lópiz
lamey	cisde
goot	adoez
chark	mapo
aub	jajón
boogty	exáncito
umblella	tagón
nale	sarcisa
shoard	claje
thurl	nedura
chall	jaréz
ostript	pulmar
ped	avestriz
ancle	sandota
monrey	lullo
cacs	piole

Appendix B. Tables with the summaries of the significant or marginally significant effects in the final models.

Tables B (1 to 8) report intercept and significant or marginally significant factors included in the final models in Experiment 1 for the analysis of RTs and their coefficients, standard errors, t-values, and p-values. Table B.9 does so for the results in Experiment 2.

Table B.1. Intercept and significant or marginally significant factors included in the final model in Experiment 1 in the L1-L2 direction with Prime frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.54	0.02	-102.92	< 0.001
Prime type	-0.16	0.01	-23.22	< 0.001
Age	0.05	0.02	2.42	0.02
Concreteness by Age	-0.01	0.004	-3.22	< 0.01
Prime type by Concreteness	-0.01	0.01	-5.39	< 0.001
Prime type (Related) by Prime frequency	-0.02	0.01	-3.12	< 0.01
Prime type (Related) by L2 use by Prime frequency	0.01	0.002	5.59	< 0.001
Prime type	0.008	0.003	2.23	0.026

(Related) by Concreteness by L2 use by Prime frequency

Table B.2. Intercept and significant or marginally significant factors included in the final model in Experiment 1 in the L1-L2 direction with Target frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.54	0.01	-115.46	< 0.001
Prime type	-0.16	0.002	-61.30	< 0.001
Concreteness	-0.02	0.003	-9.18	< 0.001
Target frequency	-0.08	0.001	-57.14	< 0.001
Age	0.05	0.02	2.42	0.02
Prime type by L2 use	0.01	0.003	3.92	< 0.001
Concreteness by Age	-0.01	0.004	-3.54	< 0.001
Prime type by Concreteness	-0.05	0.01	-9.13	< 0.001
Prime type by Concreteness by Age	0.02	0.01	2.24	0.03
Prime type (Control) by L2 use by Target frequency	0.004	0.002	2.19	0.03
Prime type (Related) by	0.01	0.002	4.68	< 0.001

L2 use by Target frequency				
Prime type (Related) by Concreteness by Target frequency	-0.02	0.004	-4.97	< 0.001
Prime type (Related) by Concreteness by L2 use by Target frequency	0.01	0.004	2.71	< 0.01

Table B.3. Intercept and significant or marginally significant factors included in the final model in Experiment 1 in the L2-L1 direction with Prime frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.58	0.02	-103.14	< 0.001
Prime type	-0.13	0.01	-20.90	< 0.001
Prime type by Age	0.02	0.01	3.27	< 0.01
Prime type by Concreteness	-0.02	0.01	-2.06	0.04
Prime type (Related) by Prime frequency	-0.02	0.01	-2.66	0.01

Table B.4. Intercept and significant or marginally significant factors included in the final model in Experiment 1 in the L2-L1 direction with Target frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.59	0.02	-106.76	< 0.001
Prime type	-0.13	0.01	-20.90	< 0.001
Concreteness	-0.03	0.01	-93.40	< 0.001
Target frequency	-0.06	0.004	-13.62	< 0.001
Prime type by Age	0.02	0.01	3.32	0.01
Prime type by Concreteness	-0.02	0.01	-2.37	0.02

Table B.5. Intercept and significant or marginally significant factors included in the final model in Experiment 1 with concrete stimuli and Prime frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.57	0.01	-113.43	< 0.001
Prime type	-0.17	0.01	-25.23	< 0.001
Language	-0.05	0.02	-3.04	< 0.01
Prime type by Age	0.02	0.01	2.18	0.03
Prime type by Language	0.05	0.01	5.33	< 0.001
Prime type (Related) by Prime frequency	-0.02	0.01	-2.63	< 0.01
Prime type (Related) by Language (English) by L2 use by Prime frequency	0.01	0.003	4.93	< 0.001

Table B.6. Intercept and significant or marginally significant factors included in the final model in Experiment 1 with concrete stimuli and Target frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.57	0.01	-117.95	< 0.001
Prime type	-0.17	0.01	-30.27	< 0.001
Language	-0.06	0.02	-4.15	< 0.001
Target frequency	-0.07	0.01	-12.14	< 0.001
Prime type by Age	0.02	0.004	4.40	< 0.001
Prime type by Language	0.05	0.01	4.54	< 0.001
Prime type (Control) by Language by Target frequency	0.03	0.01	2.23	0.03
Prime type (Control) by Language by Target frequency	0.04	0.01	3.25	< 0.01
Prime type (Related) by Language (English) by L2 use by Target frequency	0.02	0.003	5.66	< 0.001

Table B.7. Intercept and significant or marginally significant factors included in the final model in Experiment 1 with abstract stimuli and Prime frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.55	0.01	-110.56	< 0.001
Prime type	-0.13	0.02	-20.15	< 0.001
Language	-0.04	0.02	-2.18	0.03
Age	0.04	0.02	2.17	0.03
Prime type by Age	0.01	0.01	2.08	0.04
Prime type (Related) by Prime frequency	-0.02	0.01	-3.13	< 0.01
Prime type by Language by Age	0.02	0.01	3.02	< 0.01
Prime type (Related) by Language (English) by L2 use by Prime frequency	0.06	0.002	2.59	< 0.01

Table B.8. Intercept and significant or marginally significant factors included in the final model in Experiment 1 with abstract stimuli and Target frequency for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.55	0.01	-118.13	< 0.001
Prime type	-0.13	0.01	-16.42	< 0.001
Language	-0.05	0.01	-3.83	< 0.001
Target frequency	-0.07	0.004	-19.16	< 0.001
Age	0.04	0.02	2.19	0.03
Prime type by Age	0.01	0.004	3.36	< 0.001
Prime type by Language (English) by L2 use	0.01	0.004	2.60	< 0.01
Prime type by Language by Age	0.03	0.01	3.34	< 0.001
Prime type (Related) by Language (English) by L2 use by Target frequency	0.01	0.003	1.99	0.05

Table B.9. Intercept and significant or marginally significant factors included in the final model in Experiment 2 for the analysis of RTs and their coefficients, standard errors, t-values, and p-values.

	Coefficient	Std. Error	t-value	p-value
Intercept	-1.64	0.02	-108.56	< 0.001
Prime type	-0.04	0.01	-9.34	< 0.001
Language	-0.10	0.02	-6.18	< 0.001
Prime type by Age	0.01	0.003	2.45	0.02
Prime type (Related) by Language (Spanish) by L2 use by NoA	0.01	0.01	2.33	0.02

Appendix C. Expanded methods.

Experiment 1

Method

Participants

200 Spanish-English sequential bilinguals (see Table 1 for participant characteristics) took part in two LDT experiments 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 UK. Most of the participants in Spain were completing a degree in English studies, but they were recruited from different universities (most of them from cities were only Spanish was the societal language).[19] Participants in the UK were living in different cities across the country, but mainly in London, and their professional background was more varied. All participants stated not using a third language in their daily life. All participants were recruited online and compensated with £20 (or the equivalent in euros) for their participation. As discussed above, we were interested in isolating the effect of L2 use from that of L2 proficiency, thus, this latter factor was kept constant across participants. L2 proficiency was assessed employing the LexTALE test (Lemhöfer & Broersma, 2012); a minimum score of 80/100 was required to participate in the study. Further L2 proficiency assessment was conducted with some of the tasks included in the experimental sessions (see below). The mean LexTale score was 89.82 (*SD* = 5.64; range: 80-100) and 88.23

[19] Besides Spanish, there are four other co-official languages (i.e., Catalan, Basque, Galician, Aranese). Also, some other regional languages are recognised but do not hold official status.

(*SD* = 5.06; range: 80-100) for the Spain Group and the UK Group respectively. A Two-Sample t-test showed that both groups differed significantly. However, further exploration with a minimal mixed-effects model showed that the factor, treated continuously across the whole population, did not significantly modified RTs nor priming effects. Therefore, as planned, we decided to continue the analysis without further modifications. Language use information was gathered with the Language and Social Background Questionnaire (LSBQ; Anderson, Mak, Chahi & Bialystok, 2018). The questionnaire provides a composite score, reflecting the degree of bilingualism, which we employ as our measure of L2 use. On average, participants had a mean score of 9.66 (*SD* = 5.80, range: -2.26 – 21.61; higher values indicate increased L2 use). The interest of the present analysis was to investigate the effect of L2 use on priming effects. In that sense, recruiting participants from these two populations allowed us to have enough variability in the distribution of the L2 use variable. Mean values differed significantly for both groups ($p < .05$). The mean score in the Spain Group was 4.65 (*SD* = 3.03, range: -2.26 – 11.42); whereas, in the UK Group, it was 14.57 (*SD* = 2.97, range: 6.12 – 21.61). The LexTALE and LSBQ scores were not correlated ($r = -0.11$, $p < .001$). All participants stated having started learning English in the school and never before the age of six (this was a requirement for participating in the study). The mean length of immersion in the UK Group was 75 months (*SD* = 45.59, range: 12 – 254). Only four participants in the Spain Group reported previous immersion experience (crucially, not in the 12 months prior to the experiment).

Participants were tested in two sessions at least seven days apart. Task order for Session 1 was as follows: first direction of the translation priming LDT (Experiment 1) – LSBQ – second direction of the LDT (Experiment 1) – DCCS. Order of LDT priming direction for Experiment 1 (L1-L2 vs. L2-L1) was counterbalanced across participants. Task order in Session 2 was the following: semantic LDT (Experiment 2) – picture-word matching task – translation task. Participants were recruited online and compensated with £20 (or the equivalent in euros) for their participation.

Materials

314 noncognate translation equivalent pairs were used in each translation direction (see Appendix A for the stimuli list and Table 2 for stimuli characteristics). Out of these, 191 were abstract and 123 concrete. Concreteness for each translation equivalent pair was established with the following procedure. First, each English word within each pair was given a concreteness value from the concreteness ratings by Brysbaert, Warriner & Kuperman (2014). Because this were translation equivalents pairs, and to avoid employing different measures, only the value from the English pair was used. Next, we used two different methods for categorizing each pair as concrete or abstract. First, we followed Reilly & Desai (2017), who divided their stimuli into thirds and classified as concrete and abstract only the thirds with the highest and lowest values respectively. The second method consisted of dividing the words into concrete and abstract with a median-split. Two different models were fitted for each of

these variables obtained from the two approaches. Their results showed minimal differences, crucially, not affecting the significant effects. Consequently, we opted for the median-split method because it allowed us keeping the whole set of stimuli (compared to giving up approximately a third of the observations).[20] The mean frequency of English words was 4.30 (*SD* = 0.69, range: 2.14 – 6.35) in the Zipf scale; the Spanish stimuli had a mean frequency of 4.30 (*SD* = 0.69, range: 2.50 – 6.14) (differences were non-significant). English word frequencies were obtained from SUBTLEX$_{UK}$ (van Heuven et al., 2014), whereas Spanish frequencies were extracted from SUBTLEX$_{ESP}$ (Cuetos, Glez-Nosti, Barbón & Brysbaert, 2011). Mean values between languages did not differ significantly. Words in both languages were also controlled for length and orthographic neighbourhood.

To generate "no" responses, 314 pseudowords were created for both translation directions with the software Wuggy (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), such that, for each language, in one of the lists, half of the words were preceded by their translation equivalents and the other half by control primes, which were obtained from scrambling the related primes in the other list. We ensured that control pairs remained orthographically and semantically unrelated. The words in each list

[20] The fact that results with the median-split were very similar when compared with a quite conservative approach as the one used by Reilly & Desai (2017) speaks of the validity of employing median-split in future studies (although comparisons between methods are always recommended).

were matched for frequency, word length, and orthographic neighbourhood. Each list began with 16 practice items.

Picture-word matching task

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. Below the pictures, two words in English appeared, one of them was the correct match, and the other, a control word (orthographically and semantically unrelated to the matching translation).[21] First, a fixation point was presented for 250 ms; a blank screen followed for 100 ms. After that, the picture appeared for 500 ms, followed by the two words, which stayed there until the participants selected the matching word. Participants were asked to respond as accurate and fast as possible. The position (right or left) of the matching word was randomised in each trial. The lowest participant accuracy was 89%, which confirmed their high L2 proficiency. With respect to the stimuli, only five (concrete) words showed an accuracy lower than 80%; they were removed from the dataset.

Translation recognition task

[21] It could be argued that a translation task would be a better test of the knowledge of the translation equivalents in the L2; however, we considered that recalling the translation is a different process—arguably, much more effortful, especially, under time pressure—than recognising that a word correctly matches with a picture depicting an object, which, inevitably also evidences knowledge of that concept-(L2) word (and, by extension, L1 translation) relationship.

Because, in general, abstract words cannot be depicted by pictures, following the same rationale as in the previous task, we employed a translation recognition to test the participants' knowledge of this type of stimuli words. A fixation point was presented for 500 ms; a blank screen of 200 ms followed. After that, a word in English appeared for 300 ms. Before a blank screen of 200 ms, a Spanish word was presented, and stayed on the screen until the participants decided whether that was the translation equivalent of the previous English word. In half of the trials, the Spanish word was the translation, and, in the other half, a control word (orthographically and semantically unrelated). Participants were asked to respond as accurate and fast as possible. 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.

Procedure

All experiments were created and presented online using the Gorilla Experiment Builder (www.gorilla.sc) (Anwyl-Irvine, Massonnié, Flitton, Kirkham & Evershed, 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. Not meeting these criteria resulted in the exclusion from the study. 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 pseudorandomised so participants could not know when they would appear. Participants had to press “B” on the keyboard within 2 seconds from the instructions’ onset. There was a check every approximately twenty trials, and 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.

Each trial began with a fixation point presented for 500 ms. Then, the prime appeared for 200 ms, and it was followed by the target, which remained on the screen until the participants’ decision on whether that was a real word. 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 as accurate as possible. The experiment was blocked by language. The order of these two language blocks was counterbalanced by participant. Each block was further divided into 15 blocks (with approximately 40 trials each). Participants were allowed to rest between these 40-trials blocks. Participants were asked to avoid any distractions during the completion of the sessions, and to ensure their vision was corrected if needed. They were also encouraged not to complete the sessions at night or when they felt tired. All in all, we were determined and paid special attention to simulate—as much as we could given the circumstances during lockdowns—the conditions of testing on a lab.

Results

Besides the five participants already excluded for low accuracy on the translation recognition task, two more participants were removed for the same reason after inspecting the LDT data. Thus, analysis continued with the remaining 193 participants (96 in the Spain group and 97 in the UK group), and with the remaining 270 word pairs. Incorrect responses and pseudoword trials were removed. RTs below 200 ms (4) and above 5000 ms (80) were removed (see Baayen and Milin, 2010). The final dataset contained 97,931 observations. Response times' distribution was explored, evidencing the usual skewness with this measure in this type of studies. We transformed the latencies to obtain inverse Gaussian, log-normal, and BoxCox distributions. After visual inspection (Q-Q plots) and Shapiro-Wilk tests, the 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. All continuous independent variables were scaled, centered, and converted to z units.

Error rates and response times were analysed employing (generalized) linear mixed-effects models (Baayen, Davidson & Bates, 2008) in *R* (version 3.6.1; R Core Team, 2019) with the *lme4* package (Bates, Maechler, Bolker & Walker, 2015). We followed Brauer & Curtin (2018) and included main effects and interactions of interest as fixed effects in both the analyses for accuracy and RTs. For the specification of the random effects and model selection, we also followed Scandola & Tidoni (2021). The authors show that computational times are linked with converging and overfitting issues. Consequently, in cases of high model complexity (as is the case with our models) and relatively low computational power (what could be understood as standard equipment—for instance, the present analysis was performed with a MacBook Pro "M1" 8-Core with 8 GB of RAM, 2020), they recommend to employ Complex Random Intercepts (CRI). In a full-CRI model, as the ones used here, (complex) random slopes (with many interactions) are replaced by different random intercepts for each grouping factor. In the present models, that included random intercepts for any predictor and interaction that varied within subject and stimuli, and random intercepts by each of those grouping factors for those predictors and interactions. This way, an optimal trade-off between maximal random structure specification, convergence, and computational power is achieved. Importantly, with this method, Type-I errors' risk is minimised. Additionally, and speaking of our commitment to reduce Type-I errors inflation and drawing robust conclusions, we established the significance level for the main effects at .025, and for interactions at .01 (Scandola & Tidoni, 2021, p. 15).

Four main models were run. First, prime and target frequency were never included in the same model. This was done not to increase model complexity, especially in the random structure, and for the ease of results' interpretation. The main analysis was performed on datasets split by translation direction (i.e., directly obtained from the two language blocks). The reason for not to join data from both datasets was twofold. Again, we did not want to increase model complexity. Also, the analysis of interest was that of concreteness, L2 use, and frequency. However, we decided to run four more (parallel and confirmatory) models, this time, with data split by concreteness and grouped by language. This was done to explore (and confirm) potential effects of language (i.e., testing the cross-language priming asymmetry). A complete description of the analysis, including all the models, can be found on the GitHub's repository of the corresponding author (*anonymised for review*). The factors included in the models were condition (related or control), concreteness (concrete and abstract; only in the analyses split by language), language (i.e., translation direction; L1-L2 and L2-L1; only in the analyses split by concreteness), L2 use (i.e., LSBQ's composite score), and frequency (prime frequency in half of the models, and target frequency in the other half). As mentioned, interactions of interest were also included. The random effects structure included any predictor and interaction that varied within subject (language, condition, prime frequency, and target frequency) and prime (condition) and target (condition). A full-CRI structure was specified with random intercepts by each of those grouping factors for subjects, primes, and targets (see Feldman, Milin, Cho, Moscoso del Prado Martín & O'Connor, 2015, for the inclusion of prime and targets as random intercepts).

Age (and the interactions with condition, and concreteness and language) was the only post-hoc factor. Although, ideally, both groups should have had similar mean ages, that was not the case in the present study. The reason is related to the composition and characteristics of both groups. Recall that the L2-immersed group consisted of people living in the UK. One of the requirements to be part of this group was to have been living in the UK for at least 12 months. We were interested in exploring the effect of long exposure to and use of the L2, wanting to ensure that people would have been sufficiently exposed to the L2. Thus, the UK group consisted mainly of migrants, which tend to be people at least in their mid-twenties. The upper age limit for participating in the study was 40 years old. Arguably, we could have established a lower cut-off, but, again, that would have had an impact on our variable of interest, which was L2 use. Trying to match the mean age of such group (32 years old; range: 22 – 40), turned out to be practically impossible due to the linguistic profile we were targeting (high L2 proficients) and the number of participants needed (~100). Thus, as previously described, we resorted to university students (comparatively younger) and English professionals. The mean age in the Spain group was 26 years old (range: 19 – 39). This difference was significant across groups.

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 of 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, Levy, Scheepers & Tily (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, p. 13), and (ii), given our large number of observations, a more parsimonious approach (Matuschek, Kliegl, Vasishth, Baayen & Bates, 2017) did not seemed as necessary. Finally, although L2 use and age showed a moderate correlation (r = 0.51, p = .001), all variance inflation factors in the final models were below 2, indicating no collinearity (Zuur, Ieno & Elphick, 2010).

Importantly, significance levels were established at .25 for main effects and .01 for interactions.

Experiment 2

Materials

Trials in each translation direction consisted of 112 noncognate cross-language semantic associative pairs. These pairs were obtained from a previous free association task with 100 participants with the same linguistic profile as the participants in the present UK Group. Following Nelson, Evoy & Schreiber (2004), participants were presented with Spanish cue words and asked to write the first word that came to mind that was semantically related or associated with that word. After data cleansing (e.g., removing any non-noun and idiosyncratic responses, correcting typos), the number of associates for each cue word was calculated (i.e., unique target words for each cue word). Cue words in the norming study served as targets in the semantic LDT, whereas target words in the norming study were primes in the semantic LDT. Cue and target words were translated into English. This way, four lists were created, counterbalancing condition, language, and NoA. As in the standard LDT, scrambled related primes in one of the lists served as control primes in the other. Attention was paid to ensure that control pairs were not semantically or orthographically related.

Word frequency could not be matched between Spanish and English stimuli. Spanish words were overall less frequent than English ones (3.1, *SD*: 0.7, range: 1.3 – 4.7 vs 4.6, *SD*: 0.6, range: 3.3 – 6.3). However, recall that the main aim of this experiment was to determine whether cross-language semantic priming would obtain for these participants outside of translation equivalent pairs, and whether priming was modulated by the strength of the associative connections as measured by the number of associates between related pairs.

Due to the high number of repeated target words in the free association task, we repeated 13 primes (out of 112) in each translation direction in the stimuli for the present experiment. Inspection of a potential effect of this repetition showed almost identical RTs and priming effects—in direction, significance and effect size.[22]

Procedure

The presentation procedure in this experiment was identical to that of Experiment 1.

Results

The analysis of this task followed the one employed in Experiment 1, except that not including concreteness and frequency (see below) allowed us to fit a model with the whole dataset, including language and its interactions.

First, we address the potential effect of frequency. As mentioned, Spanish stimuli was of lower frequency than the English one. This might have impacted the results in the following way. On the one hand, in the L1-L2 direction, less

[22] A caveat from the norming study employed to obtain the cross-language semantically related pairs is that it prevented us from exploring concreteness effects within this dataset, since, in some cases, responses in the norming study resulted in mixed pairs with one concrete and one abstract word (e.g., eye-view).

frequent primes might have had more difficulty to activate their L2 targets (no matter which theoretical framework we assume); that would result in smaller priming effects. This could be amplified by the fact that L2 targets are of high proficiency (i.e., processed faster), giving less time to the L1 primes to assist. The effect would be the opposite in the L2-L1 direction: priming effects might be overstated. Further, when introducing prime and target frequency in the models, a collinearity arose, disallowing the exploration of potential interactions of the factors. Therefore, in this analysis, we only proxied language use with the LSBQ composite score.

The same words that were discarded in Experiment 1 due to low accuracy in the translation task and were part of this task were removed here too, as well as pseudowords and incorrect responses, and RTs below and above 200 ms and 5000 ms respectively. The final dataset contained 38505 observations. All continuous variables employed in the analysis were scaled, centered, and converted to *z* units. Sum contrasts were applied to the categorical variables. Given the skewness of the RT distribution, transformations were applied. Inverse Gaussian offered the best correction and was used in the analysis (inverse Gaussian: 0.91; BoxCox: 0.90; log-normal: 0.02).

Both the analysis of response latencies and error rates were analogue to those of Experiment 1. As in Experiment 1, we fitted a model that contained the main effects (language, condition, L2 use, age, and NoA—treated categorically) and interactions of interest as fixed effects. The random structure included random intercepts for the predictors and interactions that varied within subject (language and condition), prime (condition), and target (condition), and random intercepts by each of those grouping factors for those predictors and interactions. Again, significance levels were established at .25 for main effects and .01 for interactions.

Dimensional Change Card Sort

Materials and procedure

Four test pictures were created manipulating shape (star and diamond) and colour (red and blue). One of these four pictures was presented on the screen. Below them, two options were shown according to the relevant dimension (shape or colour) for making a decision on that trial. Participants had to select the correct option according to the specific dimension on the trial. They had to press A or L on the keyboard according to the position of the correct match, which was randomised. The task began with two demonstration phases for each dimension, each followed by a non-mixed block with that dimension. Then, a new practice phase preceded the critical block, were no-switch and switch trials were included. A switch trial occurred every three no-switch ones. Each demonstration phase started with four practice trials, whereas the critical block included eight practice trials. Sixteen trials were included in each non-mixed block, whereas 48 no-switch and 16 switch trials appeared in the critical block.

Results

Incorrect answers were removed from the dataset. No trials were below and above 200 ms and 5000 ms respectively. Responses from blocks 1 and 2 (i.e., non-mixed blocks) were collapsed into the same level (non-mixed). The RTs distribution was skewed, and transformations were obtained. The Inverse Gaussian transformation was selected, as it provided the better correction for the skewness—along with BoxCox transformation (inverse Gaussian: p = .72; BoxCox: p = .72; log-normal: p < .001).