Are translation equivalents really equivalent? Evidence from concreteness effects in translation priming

Adel Chaouch-Orozco [ID]
The Hong Kong Polytechnic University, Hong Kong

Jorge González Alonso
Jon Andoni Duñabeitia
Jason Rothman
Universidad Nebrija, Spain; UiT The Arctic University of Norway, Norway

Abstract

Aims and Objectives: Translation equivalents intuitively seem to overlap in meaning. Moreover, the models of the bilingual lexicon often represent the meaning shared between two translations as a holistic node in the semantic network. However, research on semantic representation and processing questions this holistic approach. For instance, abstract words are assumed to be more language-dependent, while concrete words' meanings are seen as more consistent cross-linguistically. The non-cognate translation priming paradigm offers an ideal methodological setting to study semantic overlap (proxied by concreteness) between translations. Priming effects between non-cognate translation equivalents are assumed to emerge due to spreading activation at the semantic level. Hence, a larger semantic overlap between translation prime-target pairs should lead to larger priming effects. Nevertheless, the evidence from previous translation priming studies investigating concreteness displays a blurry picture, potentially reflecting a shared limitation: their relatively small sample sizes. We overcame this problem by analysing the largest translation priming dataset to date.

Methodology: Two hundred Spanish–English highly proficient bilinguals were tested in a bidirectional translation priming experiment employing 314 non-cognate translation equivalents differing in concreteness.

Data and analysis: We analysed response times and error rates employing conservative (generalized) linear mixed-effects models.

Findings: The results showed that concrete translation pairs elicited larger priming effects than abstract ones, evidencing differences in semantic representation between concrete and abstract words. Importantly, the influence of concreteness appeared only in the forward translation direction, suggesting language experience-related differences in meaning representation.

Corresponding author:
Jason Rothman, Department of Language and Culture, UiT The Arctic University of Norway, 9019 Tromso, Norway.
Email: jason.rothman@uit.no

Originality: The present study analysed the largest dataset in the translation priming literature to date, employing a conservative statistical approach to shed light on the effects of concreteness on translation priming.

Implications: Our study spotlights the complexity and non-holistic nature of the bilingual semantic representation of concrete and abstract words. The present findings call for more research to help the current models of the bilingual lexicon implement more nuanced semantic representations.

Introduction

Word meanings map conceptualizations of real-world entities to lexical forms – arbitrary or not (e.g., sound-symbolic words). Theoretical models of the lexicon assume that semantic information is stored at a different level than other types of information (sub-lexical, lexical, and grammatical). However, a question emerges when considering the case of bilingualism: How do bilinguals represent meanings for words that we conventionally refer to as translation equivalents? Is semantic representation shared or separate for these words? The present study investigates the degree of semantic alignment between translation equivalents. To do so, we assess the extent of semantically mediated translation priming in concrete and abstract translation pairs.

Naturally, we tacitly assume that translation equivalents overlap in semantics across languages. That is, we expect 'love' and *amor* (Spanish for 'love') to mean *exactly* the same. However, to what extent this assumption is valid is not a trivial question. There are reasons to suspect that translation equivalents do not entirely match their meanings (e.g., Borghi, 2019; Borghi & Binkofski, 2014; De Deyne et al., in press; Dove et al., 2022; Goddard, 2010; Thompson et al., 2020). Under this view, translation equivalence would be seen more as a collective agreement by which interlingual communication is made possible, rather than an accurate reflection of how these words' semantic information is represented in the minds of bilingual speakers.

A way to approach the study of translation semantic overlap is by exploring word concreteness. Monolingual research suggests that abstract words have more distributed and inconsistent meanings and thus depend more on their linguistic context for interpretation than concrete words do (e.g., Crutch & Warrington, 2005; Hoffman et al., 2013; Pexman et al., 2013). If so, this may result in fewer semantic features shared between abstract translation equivalents (i.e., an overall lower degree of semantic overlap).

This study reanalysed data from Chaouch-Orozco et al. (2022), which tested 200 highly proficient Spanish–English bilinguals in a translation priming paradigm. The stimuli consisted of 314 translation equivalents that varied in concreteness. Anticipating our results, the data suggest that abstract translation equivalents are less aligned semantically than concrete translation pairs. Our analyses also indicate that, despite extensive experience with their second language (L2), our participants' representations of L2 concrete words are not as semantically rich as those of their first language (L1) words.

Exploring the bilingual semantic system with translation priming

Translation priming experiments have been a valuable tool to study the inner workings of the bilingual lexicon. In this paradigm, prime and target in the related condition are translation equivalents

(e.g., *amor* and 'love'), whereas control primes and targets are cross-language unrelated pairs (e.g., *amor* and 'hand'). In the standard procedure, priming effects take the shape of faster response times, or increased accuracy, for related prime-target pairs as compared to controls when deciding on the target's lexical status (i.e., in a lexical decision task). The effect is usually explained by resorting to the workings of spreading activation within the lexical-semantic network (Collins & Loftus, 1975). Processing a prime presented as related results in the activation, and subsequent faster recognition, of its translation equivalent target. In this line of explanation, translation priming effects are taken to expose cross-linguistic connectivity and support parallel activation of the two languages during bilingual lexical access, where words from the two languages are activated in parallel.

The priming literature with non-cognate translation equivalents – translation pairs with no overlap at the form level – is abundant. Because these pairs are taken to be connected exclusively at the semantic level, we can expect non-cognate translation priming effects to happen mostly – or at least in part – through semantic mediation (Xia & Andrews, 2015; see also Jiang & Forster, 2001 for an alternative account based on episodic links), making these a good testing ground for the study of bilingual semantic processing. With other factors minimized, priming is assumed to be larger when semantic activation is more robust. Crucially, this should be a function of the degree of semantic overlap between prime and target: with higher overlap, more of the activation reaching the prime can spread over to the target, priming it more.

Importantly, not all models of the bilingual lexicon assume different degrees of semantic alignment among pairs of translation equivalents. The following section discusses two of the most prominent models of bilingual lexical-semantic organization: the Distributed Feature Model (de Groot, 1992; van Hell & De Groot, 1998) and Multilink (Dijkstra et al., 2019).

Models of bilingual semantic representation and processing

The Distributed Feature Model (DFM) and Multilink align in assuming that all nodes within the lexicon are fully interconnected. However, the two proposals contrast in how the semantic system operates, being the DFM distributed in nature, while, Multilink, holistic (we also refer the reader to Dong et al., 2005, for a more developmentally nuanced proposal; and Finkbeiner et al., 2004, for a distributional alternative with a slightly different focus).

For the DFM, meanings are not represented by single units but by activation patterns and connection weights across semantic features within the network. For example, the meaning of 'dog' is not a single semantic unit but a co-activation of more primitive features (i.e., an animal, four-legged, able/prone to bark, and so on). 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 semantic), 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 bear less feature overlap, are processed more slowly.

Multilink is a computational, connectionist model with a tradition in the study of the bilingual lexicon, as it is the continuation of the Bilingual Interactive Activation (BIA) and BIA+ models (Dijkstra & Van Heuven, 2002; van Heuven et al., 1998). Multilink offers a comprehensive account of bilingual word recognition and production, incorporating some of the tenets of another prominent theory, the Revised Hierarchical Model (Kroll et al., 2010; Kroll & Stewart, 1994). In Multilink's view, differences in L1 and L2 word processing are mostly accounted for by the *resting level* or baseline activation of a word, which is independent of its language membership. Crucially, subjective word frequency (i.e., each individual's experience with each particular word) is what

determines the resting level activation of words. In a feature that is relevant for the present study, Multilink is (currently) a localist model and thus envisions bilingual semantic representation as consisting of holistic nodes. In other words, there is only one shared, holistic node between translation equivalent pairs. The model thus leaves no room for non-overlapping, distributed word meaning representations.

We should note here that Dijkstra and colleagues have acknowledged from the start the aspiration for Multilink to incorporate a more nuanced semantic system in future instantiations, potentially embracing distributed representations (Dijkstra et al., 2019, p. 5). However, little is known about how meaning is represented in the bilingual lexicon, and more research is needed to better understand how bilinguals juggle (potential) non-holistic semantic representations for translation equivalents (De Deyne et al., in press). In an effort to increase our understanding of bilingual semantic representation, the present analysis employs word concreteness as a candidate to proxy semantic overlap between translation equivalents. The following section reviews some studies that have attempted to explore bilingual semantic memory by manipulating this factor.

The study of translation equivalence through concreteness effects

As we said earlier, abstract meanings have been claimed to depend more on the linguistic context of each language, making abstract words and their meanings more language-specific (e.g., van Hell & De Groot, 1998). In addition, abstract meanings may be more dispersed and consist of a lower number of semantic features (e.g., de Groot, 1989). As such, concrete translation pairs may overlap semantically more than abstract ones (van Hell & De Groot, 1998). These claims have found support in different studies with bilingual populations. For instance, van Hell and de Groot (1998) asked Dutch-English unbalanced bilinguals to produce within- and between-language associates to word cues in both Dutch and English. Their results showed that subjects elicited more translation equivalents as associates when words were concrete, suggesting that concrete translations share a larger number of semantic features. In Tokowicz et al.'s (2002) Dutch–English translation norms, participants gave higher similarity ratings to concrete translations than to abstract ones, indicating that concrete meanings are more similar across languages.

In the realm of translation priming studies, Schoonbaert et al. (2009) tested Dutch–English bilinguals in two priming experiments (1 and 2) manipulating translation direction (L1 to L2 and L2 to L1) and stimulus onset asynchrony (SOA; the time elapsed between the onset of the prime and the onset of the target) (100 milliseconds vs 250 milliseconds). Note that SOA is one of the central manipulations in priming experiments; longer SOAs give more time for primes to be processed and activate targets. Schoonbaert et al. (2009) employed a masking paradigm. Masked presentations (Forster & Davis, 1984) are an important implementation of the traditional priming paradigm. In this type of presentation, primes are presented after a forward mask (e.g., #######) and, occasionally, before a backward, post-prime mask too. This paradigm is believed to tap into more automatic/early processes by making a very briefly presented prime even less salient (Forster et al., 2003).

In Schoonbaert et al. (2009), the prime duration was set at 50 milliseconds in both conditions. However, in the long-SOA condition, a 50-millisecond post-prime blank screen and a 150-millisecond backward mask were presented, which meant that the onset of the prime and the onset of the target were separated by 300 milliseconds. In the short-SOA condition, the blank screen was dropped, and the backward mask was reduced to 50 milliseconds, leaving an SOA of 100 milliseconds. The authors reported a trend towards larger priming for concrete as compared to abstract words. In another study, Ferré et al. (2017) tested highly proficient Spanish–English bilinguals in two masked translation priming lexical decision tasks that crossed SOA (50 milliseconds vs 100 milliseconds) and concreteness manipulations. In Experiment 1, only a 50-millisecond prime

appeared before the target. In Experiment 2, SOA was extended to 100 milliseconds by introducing a 50-millisecond post-prime backward mask. Overall, Ferré and colleagues observed larger priming for concrete words, but only in Experiment 2, suggesting that the 50 milliseconds SOA of Experiment 1 may not be enough time for a complete semantic activation. Chen et al. (2014) reported a lack of concreteness effects on masked translation priming in both translation directions when testing Chinese-English bilinguals with a 50-millisecond prime and a 150-millisecond backward mask (i.e., a 200 milliseconds SOA).

While all the studies discussed earlier employed masked priming, two studies have used overt priming. This point is critical because longer prime presentations may be needed when inspecting semantic processes (see Altarriba & Basnight-Brown, 2007, for a review) such as concreteness. Jin's (1990) Experiment 1 tested Korean–English bilinguals in an overt (i.e., unmasked) translation priming lexical decision task with 150-millisecond primes. Participants were tested in both translation directions. Jin's (1990) results showed larger priming effects for concrete than abstract words. Crucially, this difference was larger in the L1–L2 direction than in the L2–L1, backward one. Finally, Smith et al. (2019) tested Hebrew–English bilinguals in an overt translation priming lexical decision with a 200-millisecond SOA (150-millisecond prime plus 50-millisecond post-prime blank screen) in both translation directions. Contributing to the overall unsettled nature of these phenomena, their data showed no concreteness effects on priming.

At least five studies have thus attempted to explore the role of semantic overlap (proxied by concreteness) on translation priming, with mixed findings. At this point, it is necessary to consider potential reasons for the disparities between these results. First, the different bilingual experiences of the subjects in previous studies may account for at least some of the divergencies in the observed priming patterns. However, the use of different measures and scales to assess the participants' L1/L2 experience makes comparisons difficult. Second, prime duration and SOA could also be thought to modulate the presence of concreteness effects. However, the evidence is too mixed for us to be able to draw conclusions in this regard. Third, although non-significant, the trending results in Schoonbaert et al. (2009) may indicate that, regardless of prime duration and SOA, an effect could be better captured and characterized if larger datasets were available.

Sample size and its connected issues are a common problem in translation priming studies, and in chronometric research more generally. First, because response times are noisy measures, effect sizes in this type of study tend to be small (e.g., Brysbaert, 2019). Second, creating non-cognate translation stimuli while controlling for factors that have been reported to affect processing speed (e.g., number of letters, frequency, orthographic neighbourhood, and many others) is particularly challenging – especially in languages sharing many cognate words – which makes it difficult to raise the number of observations by having more stimuli. Brysbaert and Stevens (2018) recently suggested that a minimum of 1,600 observations are necessary to interpret differences of about 15 milliseconds. Importantly, none of the four studies discussed above are even close to that number – the most extensive dataset having less than 800 observations per condition. To answer recent calls for better statistical practices, we need larger, more statistically reliable datasets to better understand the role of concreteness as a proxy of semantic overlap between translation equivalents.

The present study aims to make a significant contribution in this regard. We re-analysed a translation priming dataset from Chaouch-Orozco et al. (2022). Their study followed up on a previous investigation on the role of language experience on translation priming effects (Chaouch-Orozco et al., 2021), with a more comprehensive scope. Chaouch-Orozco et al. (2022) tested 200 highly proficient Spanish–English unbalanced bilinguals, using 314 non-cognate translation equivalent pairs, providing the largest dataset in the translation priming literature to date. Their study solely focused on the role of L1/L2 experience on priming, which was manipulated through immersion. Importantly for our purposes, their word stimuli were obtained from a continuum of concreteness

Table 1. Participant characteristics.

Age	29 (5.7; 19–40)
LexTALE	89 (5.4; 80–100)
LSBQ	9.5 (5.8; –2.3 to 21.6)

Note. Mean values (standard deviation; range). 'LSBQ' column shows composite L2 use score across contexts (home, social, etc.). LSBQ: Language and Social Background Questionnaire.

values, but this factor was only used as a covariate in their analyses and failed to show a significant modulation of the interaction between immersion and priming effects. Given the large sample size, Chaouch-Orozco et al.'s (2022) dataset has the potential to overcome some of the limitations of previous studies, making it a valuable source of information on the role of concreteness in bilingual lexicon-semantic representation and processing.

The present study provides a systematic analysis of concreteness in this dataset, where we expected larger priming effects in concrete trials due to increased activation at the semantic level. This would be in line with the previous literature suggesting that concrete translation pairs are more semantically aligned than abstract ones.

Method

Participants

Two hundred highly proficient, late sequential Spanish–English bilinguals (164 females and 36 males) were tested in two lexical decision experiments, one for each translation direction (see Table 1 for participant's characteristics). The main focus of Chaouch-Orozco et al. (2022) was to investigate how immersion in a second language affected priming effects. For this purpose, participants were divided into two equally large groups manipulating this factor (i.e., immersed vs non-immersed). Thus, half of the participants were living in Spain and the other half in the United Kingdom (mean length of immersion: 6 years; range: 1–21). Importantly, L2 proficiency, measured with the LexTale test (Lemhöfer & Broersma, 2012), was controlled across participants, being all subjects highly proficient in their L2 (*M*: 89/100; range: 80–100). Moreover, the Language and Social Background Questionnaire (LSBQ; Anderson, Mak, Chahi & Bialystok, 2018) was used to gather information about the participants' linguistic backgrounds. The LSBQ provides a composite measure of relative L1/L2 use; the mean score differed significantly across both groups ($p < .05$) – the immersed group reported more use of the L2 than the non-immersed group (LSBQ score: 14.6 vs 4.6, respectively). All participants started learning English after the age of six in primary school. Overall, their results failed to show the expected effect of immersion (and/or L2 proficiency) on priming. More importantly for the present study, however, their data showed that concreteness did not modulate the effect immersion had on priming. For this reason, the present study will collapse across immersion levels and include all 200 participants in its main analysis. To be on the safe side, the interaction between immersion and concreteness was also explored in a parallel analysis here.

Materials

In all, 314 non-cognate translation equivalents were used to create the critical stimuli (see the complete list in Chaouch-Orozco et al., 2022; see Table 2 for stimuli characteristics) and were employed in each translation direction. Since translation pairs were extracted from a continuum of frequency

Table 2. Stimuli characteristics.

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

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

and concreteness values, these two factors were introduced as covariates in the original analysis. None of them modulated the effect of immersion on priming effects, so they were not further discussed. Across the two languages, words were matched for concreteness, frequency, length, and orthographic neighbourhood. 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 ratings by Brysbaert et al. (2014; range: 1.04–5). Note that, because these were translation equivalent pairs and to avoid employing different measures, only the value from the English word in each pair was used. Next, two different methods for categorizing each pair as concrete or abstract were employed. First, following Reilly and Desai (2017), the stimuli were divided into thirds, classifying 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. The results showed minimal differences between the two treatments. Consequently, the median-split method was chosen because it allowed to keep the whole set of stimuli (compared to giving up a third of the observations). Word frequencies were obtained from the SUBTLEX corpora (British English: Spanish: Cuetos et al., 2011; van Heuven et al., 2014). Another 314 pseudowords in each language were created with the software *Wuggy* (Keuleers & Brysbaert, 2010) to generate 'no' responses in the lexical decision task. These pseudowords were paired with 314 words in each language 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. These lists were counterbalanced across participants – half of the participants saw one list in each translation direction and the other half the other two lists. 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.

Procedure

The experiments were presented online with *Gorilla Experiment Builder* (Anwyl-Irvine et al., 2020). Given the limits that online presentation poses to the experimenter's role in 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 to finish each session, estimated with the durations of the pilot experiments. Attention checks were implemented, and their presentation was pseudorandomised so participants could not anticipate when they would appear. Participants had to press 'B' on the keyboard within 2 seconds from the onset of the instructions. There was a check every 20 trials, approximately, 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 consisted of a fixation point (500 milliseconds), followed by the prime (200 milliseconds), immediately followed by the target, which remained on the screen until the participants made the lexical decision. They were asked to respond as fast and as accurately as possible. The experiments were divided into blocks of 40 trials, and participants could rest between them.

Data analysis

The code for the present study can be found in Supplemental Material; the data can be found in the first author's OSF repository (https://doi.org/10.17605/OSF.IO/4H63C). Both the accuracy (error rate) and the response time (RT) analyses considered only trials involving word stimuli. RT analyses were run only on correct-response trials. Following Baayen and Milin (2010), RTs below 200 milliseconds and above 5,000 milliseconds were removed from the analysis. We employed inverse-Gaussian transformations on the RTs after confirming they offered the best fit through visual (quantile-quantile plots) and Shapiro–Wilk tests (inverse Gaussian: $p = .42$; BoxCox: $p = .33$; log-normal: $p = .08$). All continuous variables were scaled, centred, and z-transformed, whereas sum contrasts were applied to categorical variables.

We employed (generalized) linear mixed-effects models (Baayen et al., 2008) in *R* (version 3.6.1; R Core Team, 2021) with the *lme4* package (Bates et al., 2015) for the analysis of both error rates and RTs. We used Complex Random Intercepts (CRI) for an optimal trade-off between processing times and computational power and convergence and overfitting issues (Scandola & Tidoni, 2021). We fitted maximal models in all analyses. In the case of non-convergence, we removed the CRI that explained the least variance until a maximal converging model was obtained (Barr et al., 2013; Brauer & Curtin, 2018; Scandola & Tidoni, 2021). We applied further criticism to the converging models, removing standardized outliers above 2.5 *SD* and checking assumptions (Baayen & Milin, 2010).

The maximal models included main effects and interactions of interest as fixed effects (Brauer & Curtin, 2018). Thus, the model for the main analysis included language (i.e., target language; English vs Spanish), prime type (related vs control), and concreteness (i.e., concrete vs abstract) and their interactions as fixed effects. Subject, prime, and target were included as random intercepts (Feldman et al., 2015), whereas any predictor and interaction that varied within subject, prime, and target were included as a random slope. Full-CRI structures were specified with random intercepts by each grouping factor for subjects, primes, and targets. The full specification of the maximal and non-converging models can be found in the OSF repository linked above.

Results

Response times

Table 3 summarizes RTs and error rates in all conditions (see the summary of the final model in Supplemental Material). The final model revealed the main effects of language ($\beta = -0.05$, $t = -3.46$, $p < .001$), and prime type ($\beta = -0.16$, $t = -27.67$, $p < .001$), indicating that RTs were faster to Spanish targets and in the related condition (i.e., an overall priming effect). The interaction between language and prime type was significant ($\beta = -0.04$, $t = 4.66$, $p < .001$), showing that priming was larger in the L1–L2 direction. Prime type and concreteness interacted significantly ($\beta = -0.04$, $t = -4.92$, $p < .001$). This interaction reflected larger priming effects with concrete words. However, a significant three-way interaction between language, prime type, and concreteness ($\beta = 0.03$,

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

	Related		Control		Priming
	RT	Error rate	RT	Error rate	
Concrete					
L1 to L2	672 (2.4)	1.1	768 (3.1)	3.2	96*
L2 to L1	654 (2.2)	0.6	718 (2.5)	1.4	64*
Abstract					
L1 to L2	688 (2.5)	1.6	759 (2.9)	3.4	71*
L2 to L1	668 (2.6)	1.0	729 (2.5)	2.0	61*

*$p < .05$.

$t=2.11$, $p<.05$) revealed that the larger priming with concrete words was, in fact, driven by responses in the L1–L2 direction.[1] In other words, the effect of concreteness on priming only appeared when primes were in L1 Spanish (L1–L2: 25 milliseconds; L2–L1: 3 milliseconds). This was assessed in two independent models with subsets of the data by translation direction (target language). In the L1–L2 direction, the interaction between prime type and concreteness was significant ($\beta=-0.05$, $t=-4.68$, $p<.001$), whereas in the L2–L1 direction the interaction did not reach significance.

To further inspect this finding and to discard any potential confounds, we ran independent analysis were group (immersed vs non-immersed), prime frequency, and target frequency were included as co-variates. In both analyses including group and prime frequency, the same three-way significant interaction was obtained ($\beta=0.03$, $t=2.04$, $p<.05$ and $\beta=0.04$, $t=2.32$, $p<.05$, respectively). In the analysis with target frequency, the interaction was nonsignificant. However, prime type and concreteness still interacted significantly ($\beta=-0.04$, $t=-5,29$, $p<.001$). Importantly, none of the covariates in any of the models significantly modified the pattern of results obtained in the main model with regard to the effect of concreteness on priming.

Accuracy analysis

Correct or incorrect responses were dummy-coded (1 or 0, respectively). We used generalized linear mixed-effects models with a binomial family to fit the error data. Significant effects were obtained only for the main effect of language ($z=0.52$, $t=2.60$, $p<.01$) and prime type ($z=0.92$, $t=6.29$, $p<.001$), indicating more accurate responses to Spanish targets as well as to related targets.[2] No effect or interaction with concreteness was significant (all ps $>.05$). However, note that accuracy analyses are usually not the focus in this type of studies, as they are less sensitive to experimental manipulations.

Discussion

This study investigated whether the degree of semantic overlap between translation equivalents predicts non-cognate translation priming effects. To address this question, we proxied semantic overlap with word concreteness. Concrete and abstract meanings have been argued to differ in their semantic robustness and uniformity across languages – concrete meanings are assumed to be richer and more consistent overall. In line with our predictions, concrete translation equivalents yielded larger priming effects than abstract pairs, indicating stronger semantic activation from prime to

target in these pairs, potentially due to a larger semantic overlap. However, this effect was only obtained in the L1 Spanish–L2 English direction.

Our study adds to the few previous investigations on how concreteness impacts translation priming. Overall, this significant effect of concreteness aligns with Ferré et al.'s (2017) and, particularly, Jin's (1990) findings. Our study replicates Jin's first report on the asymmetric impact of concreteness on translation priming across translation directions. It remains an open question, however, why the effect did not appear in other studies (Chen et al., 2014; Schoonbaert et al., 2009; Smith et al., 2019). In support of the present data, our experiment uses the largest dataset to date, with 15,000 observations per condition, dramatically increasing the number employed in previous studies.

At this point, we should note that observing a concreteness effect with non-cognate stimuli – irrespective of the target language – suggests that abstract translation equivalents are less aligned semantically than concrete translation pairs and that the bilingual semantic system is non-localist in nature. If translation equivalents shared equally overlapping semantic nodes, we should not have obtained the current differences between concrete and abstract priming effects. Thus, our results support the predictions of the Distributed Feature Model and challenge those of localist accounts such as Multilink, highlighting the need for these to implement more nuanced semantic systems. Importantly, however, the concreteness effect in our data is asymmetric with respect to translation direction. This language dependency suggests the existence of persistent differences in the way bilinguals represent the meaning of concrete and abstract L1/L2 words. Notably, the DFM is not suited to explain this asymmetry, as it remains agnostic regarding potential cross-language differences in the robustness of meanings brought about by bilingual experience. To accommodate our findings, we resort to the Dual Coding Theory (DCT; Paivio, 1991, 2010 see Pavlenko, 2012, for a similar account).

According to the DCT, concrete and abstract words are represented qualitatively differently. While both concrete and abstract meanings operate within a linguistic system, activation of concrete words also spreads into the sensorimotor system, which encodes non-verbal encounters with the real-world entities these words' meanings represent. This way, the DCT provides a rationale for the idea of concrete meanings being richer than abstract ones: Extra sensorimotor information would make these representations more complex.

Crucially, non-verbal experiences may be language-dependent in late sequential bilinguals like those in the present study (Jared et al., 2013, p. 388; Pan et al., 2021). The reason lies in the quantitatively and qualitatively different experiences these bilinguals have with their first and second languages. Contrary to the acquisition of the first language, their L2 is – usually – learned in formal contexts (i.e., classrooms) and is considerably deprived of real-life experiences (e.g., Pavlenko, 2017). As such, this type of bilingual would be much more exposed to linguistic than sensory information in their second language. Due to this divergent non-verbal experience in the two languages, L1 concrete words would be semantically richer than L2 concrete words for late sequential bilinguals. In the case of abstract words, however, differences would not be that pronounced given that both L1 and L2 abstract words primarily rely on linguistic information to carve out their meanings – precisely the type of information that is just as abundant in formal L2 learning contexts. Following this line of thought, we argue that the asymmetry of our concreteness effect emerged because the less detailed semantic specifications of the L2 concrete primes could not activate the semantically richer L1 targets to the same extent as the L1 primes did with the L2 targets. Likewise, the comparable size of priming for the abstract translation pairs in both directions would be explained by a similar integration/representation of L1 and L2 abstract meanings in the bilingual lexicon.

In light of the above, we conclude by noting that it would be worth addressing the interactions between concreteness and the multifaceted nature of bilingual experience (e.g., different linguistic profiles and language combinations) in future investigations. We strongly believe that this line of

research has the potential to be a fruitful approach to better understand the extent to which translation equivalents overlap in meaning and how that affects their processing.

ORCID iD

Adel Chaouch-Orozco https://orcid.org/0000-0003-0102-1754

Supplemental material

Supplemental material for this article is available online.

Notes

1. Note that the same significant interaction, with even a larger *p*-value, was obtained when concreteness was treated continuously ($\beta=0.02$, $t=3.18$, $p=.002$) – the rest of effects being comparable as well. However, we decided to report the analysis with the dichotomized variable to make the results more easily interpretable and the conclusions clearer.
2. The overall error rate for Spanish and English nonwords was 4.0 and 10.1, respectively, in line with the overall higher accuracy when responding to Spanish word targets. Note that this shows that the participants were not responding 'word' on every trial.

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., Keyvani Chahi, A., & 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., 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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

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

Borghi, A. M. (2019). Linguistic relativity and abstract words. *Paradigmi*, *37*(3), 429–448.

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts* (Vol. 2). Springer.

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), 389–411.

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*, Article 16.

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

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.

Chaouch-Orozco, A., González Alonso, J., Duñabeitia, J. A., & Rothman, J. (2022). The elusive impact of L2 immersion on translation priming. *Studies in Second Language Acquisition*. Advance online publication. https://doi.org/10.1017/S0272263122000249

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.

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.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

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: Frecuencias de las palabras espanolas basadas en los subtitulos de las peliculas [Frequencies of Spanish words based on movie subtitles]. *Psicológica*, *32*(2), 133–144.

De Deyne, S., Brysbaert, M., & Elgort, I. (in press). In I. Elgort, A. Siyanova-Chanturia, & M. Brysbaert (Eds.), *Cross-language influences in second language acquisition and processing: Interdisciplinary insights and perspectives*. John Benjamins.

de Groot, A. M. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 824–845.

de Groot, A. M. (1992). Bilingual lexical representation: A closer look at conceptual representations. *Advances in Psychology*, *94*, 389–412.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *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. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(4), 657–679.

Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, *8*(3), 221–238.

Dove, G., Barca, L., Tummolini, L., & Borghi, A. M. (2022). Words have a weight: Language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research*, *86*, 2451–2467.

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*, Article 111.

Ferré, P., Sanchez-Casas, R., Comesana, M., & Demestre, J. (2017). Masked translation priming with cognates and noncognates: Is there an effect of words' concreteness? *Bilingualism: Language and Cognition*, *20*(4), 770–782.

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

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

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

Goddard, C. (2010). The natural semantic metalanguage approach. In B. Heine, & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis*. (pp. 459–484) Oxford University Press.

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.

Jared, D., Poh, R. P. Y., & Paivio, A. (2013). L1 and L2 picture naming in Mandarin–English bilinguals: A test of bilingual dual coding theory. *Bilingualism: Language and Cognition*, *16*(2), 383–396.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*(1), 32–51.

Jin, Y. S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills*, *70*(3_suppl), 1139–1154.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373–381.

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.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *45*(3), 255–287.

Paivio, A. (2010). Dual coding theory and the mental lexicon. *The Mental Lexicon*, *5*(2), 205–230.

Pan, X., Xiong, A., Jouravlev, O., & Jared, D. (2021). Conceptual representations in bicultural bilinguals: An ERP approach. *Bilingualism: Language and Cognition*, *24*(5), 932–944.

Pavlenko, A. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*(6), 405–428.

Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current Opinion in Psychology*, *17*, 74–78.

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*, Article 723.

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

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

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*. https://doi.org/10.31234/osf.io/kfhgv

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

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 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., 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.

Xia, V., & Andrews, S. (2015). Masked translation priming asymmetry in Chinese-English bilinguals: Making sense of the Sense Model. *Quarterly Journal of Experimental Psychology*, *68*(2), 294–325.

Author biographies

Adel Chaouch-Orozco is a postdoctoral fellowship at the Brain, Language, and Cognition Laboratory at The Hong Kong Polytechnic University (Hong Kong). His work primarily focuses on how individual differences modulate bilingual lexical-semantic representation and processing, combining behavioural studies with computational language models and network science tools.

Jorge González Alonso is a Senior Researcher at the Nebrija Research Center in Cognition (CINC), Spain, and the AcqVA Aurora Research Center, Norway, where he leads two funded projects investigating the neurocognition of multilingual language learning and its interaction with core cognitive processes like attention and memory.

Prof. Dr. Jon Andoni Duñabeitia is a researcher in cognitive science of language working as a Full Professor at the Universidad Nebrija, where he is the Director of the Centro de Investigación Nebrija en Cognición (CINC). He is also Professor II at the Department of Language and Culture of the UiT Arctic University of Norway. His main research interests focus on the study of neurobiological and cognitive substrates of multilingualism and reading throughout the life cycle. Since Prof. Duñabeitia began his research career, he has published more than 130 articles in major international scientific journals, and he is one of the leading figures among Spanish experimental psychologists.

Jason Rothman is a Professor of Linguistics at UiT The Arctic University of Norway and an Adjunct Professor of Cognitive Science at Universidad Nebrija, Madrid. His work spans younger to older bi-multilingual individuals examining language acquisition, linguistic processing and the links between linguistic experience, domain-general cognitive functions and brain structure and function.