STUDIES OF CROSS-LINGUAL LONG-TERM PRIMING

Eva D. Poort[a,1] and Jennifer M. Rodd[a,1]

[a] Department of Experimental Psychology
University College London
26 Bedford Way
London WC1H 0AP
United Kingdom

[1] Corresponding author:
Eva Poort
26 Bedford Way
London WC1H 0AP
United Kingdom
email: eva.poort.12@ucl.ac.uk

[2] email: j.rodd@ucl.ac.uk

This manuscript has not been peer-reviewed.

Abstract

Poort, Warren and Rodd (2016) showed that bilinguals profit from recent experience with an identical cognate in their native language when they encounter the same word in their second language. We conducted two experiments employing the same cross-lingual long-term priming paradigm to determine whether this is also the case for non-identical cognates, as this would indicate they share an orthographic representation in the bilingual lexicon. In Experiment 1, Dutch–English bilinguals read Dutch sentences containing identical cognates (e.g. "winter"–"winter"), non-identical cognates (e.g. "baard"–"beard") or the Dutch translations (e.g. "fiets") of English control words (e.g. "bike"). These words were presented again in an English lexical decision task approximately 19 minutes later. The analysis revealed only weak evidence, based both on *p*-values and Bayes factors, for a small 6-9 ms facilitative priming effect. Experiment 2 aimed to determine whether including interlingual homographs (e.g. "angel"–"angel") in the experiment modulates the size of the priming effect. This time, the analysis revealed no evidence for a priming effect, either based on *p*-values or Bayes factors, in either version of the experiment for either the cognates or the interlingual homographs. In line with previous findings (Poort & Rodd, 2017, May 9), we did find strong evidence for an interlingual homograph inhibition effect and no evidence for a cognate facilitation effect. We conclude that, since the cross-lingual long-term priming effect is largely semantic in nature, the lexical decision tasks we used were not sensitive enough to detect an effect of priming.

Keywords: bilingual, cognates, lexical decision, cross-lingual long-term priming
psychINFO classification code: 2340 (Cognitive Processes)

1 Introduction

It is estimated that half of the world's population is multilingual, so one of the key issues in research on bilingualism is how words from both languages are stored and accessed in the bilingual's lexicon. In particular, there is a great deal of interest in the representation and processing of cognates and interlingual homographs, words that exist in a similar or identical orthographic form in more than one language. Cognates, such as the word "winter" in Dutch and English, also have a similar or near-identical meaning, while interlingual homographs, such as "angel", have dissimilar, unrelated meanings ("angel" means "insect's sting" in Dutch). A great number of experiments have shown that, during word recognition, both readings of a cognate or interlingual homograph are accessed, so it is thought that a bilingual has one integrated lexicon and that access to it is language non-selective.

These experiments have shown that cognates are processed more quickly than words that exist in one language only (i.e. that do not share their form with their translation), like "bike" in English and its translation "fiets" in Dutch. This *cognate facilitation effect* has been observed in a range of tasks in both visual word recognition and in word production (Costa, Caramazza & Sebastian-Galles, 2000; Cristoffanini, Kirsner & Milech, 1986; De Groot & Nas, 1991; Dijkstra, Grainger & Van Heuven, 1999; Dijkstra, Miwa, Brummelhuis, Sappelli & Baayen, 2010; Dijkstra, Van Jaarsveld & Ten Brinke, 1998; Duyck, Van Assche, Drieghe & Hartsuiker, 2007; Font, 2001; Lemhöfer & Dijkstra, 2004; Lemhöfer et al., 2008; Libben & Titone, 2009; Peeters, Dijkstra & Grainger, 2013; Sánchez-Casas, García-Albea & Davis, 1992; Schwartz & Kroll, 2006; Schwartz, Kroll & Diaz, 2007; Van Assche, Duyck, Hartsuiker & Diependaele, 2009; Van Hell & De Groot, 2008; Van Hell & Dijkstra, 2002). In contrast, interlingual homographs are often processed more slowly than control words. The *interlingual homograph inhibition effect* has been reported in experiments investigating both visual and auditory word recognition as well as word production (Dijkstra et al., 1999; Dijkstra et al., 1998; Jared & Szucs, 2002; Lagrou, Hartsuiker & Duyck, 2011; Lemhöfer & Dijkstra, 2004; Schulpen, Dijkstra, Schriefers & Hasper, 2003; Van Heuven, Schriefers, Dijkstra & Hagoort, 2008).

To accommodate these findings, in the Bilingual Interactive Activation plus (BIA+) model (Dijkstra & Van Heuven, 2002), cognates share a single orthographic node and a single semantic node. Resonance between the fully shared orthographic and semantic representations results in faster recognition of cognates relative to single-language control words (Peeters et al., 2013). Interlingual homographs in the BIA+ model are represented by two orthographic nodes (one for each language) and two semantic nodes (one for each language-specific meaning). The two orthographic nodes of the interlingual homograph laterally inhibit each other. Because the two orthographic nodes of an interlingual homograph are identical, this competition is stronger than that between any two other words, resulting in slower reaction times in comparison (Kerkhofs, Dijkstra, Chwilla & De Bruijn, 2006).

Crucially, the size of the cognate facilitation effect is greater for identical cognates like "winter" than for non-identical cognates like "beard" and "baard" (Comesaña et al., 2015; Dijkstra et al., 2010; Duyck et al., 2007; Font, 2001; Van Assche et al., 2009), but researchers differ in their interpretations of what this must mean for how non-identical cognates are stored in the bilingual mental lexicon. In Experiment 1 of their set of three experiments, Dijkstra et al. (2010) asked their Dutch–English bilinguals to make lexical decisions to a set of English words that varied in how orthographically similar to their Dutch translations they were, from not at all similar ("leger"–"army") to somewhat similar ("rijk"–"rich") to identical ("menu"–"menu"). The results showed that as orthographic similarity increased, the lexical decision reaction times decreased. In addition, Dijkstra et al. (2010) observed a steep decline in the reaction times going from non-identical cognates to identical cognates. This led Dijkstra et al. (2010) to conclude that non-identical cognates, in contrast to identical cognates, must consist of two orthographic nodes (one for each language-specific form) connected to a single semantic node (to represent their shared meaning). The lateral inhibition between the non-identical cognate's two orthographic representations cancelled out some of the increased activation due to the shared semantic representation, resulting in a much smaller facilitation effect for the non-identical cognates compared to the identical cognates.

Similarly, Comesaña et al. (2015) only found evidence for a cognate facilitation effect for identical cognates and not for non-identical cognates. In two experiments, Comesaña et al. (2015) asked Catalan–Spanish bilinguals to make Spanish lexical decisions to identical cognates (e.g. "plata"–"plata"), non-identical cognates (e.g. "brazo"–"braç") and Spanish control words (e.g. "abuela, which translates to "àvia" in Catalan; Exp. 1a) or only to non-identical cognates and Spanish control words (Exp. 2). In Experiment 1a, they observed an overall significant facilitation effect of approximately 11 ms for the identical and non-identical cognates compared to the control words, but the follow-up analyses revealed only a significant facilitation effect (of approximately 20 ms) for the identical cognates (the facilitation effect of approximately 9 ms for the non-identical cognates was not significant). In contrast, in Experiment 2, they found a significant inhibition effect of approximately 20 ms for the non-identical cognates compared to the control words. From these results, Comesaña et al. (2015) conclude that non-identical cognates must consist of two orthographic (and two phonological) nodes and that lateral inhibition between these representations resulted in the inhibition effect observed in Experiment 2.

In contrast, Van Assche, Drieghe, Duyck, Welvaert and Hartsuiker (2011) asked their Dutch–English bilinguals to complete an English lexical decision task and found both a categorical effect of cognate status and a continuous effect of orthographic overlap. The identical and non-identical cognates in their experiment were recognised more quickly than the control words, and at the same time, cognates with higher degrees of orthographic overlap (based on Van Orden's (1987) measure) were recognised more quickly than those with lower degrees of overlap. Similar results were found when they used a combined measure of orthographic and phonological overlap based on subjective ratings of similarity and when they embedded the cognates in sentences and tracked their participants' eye movements (Exp. 2). The results of these two experiments, in contrast to Dijkstra et al.'s (2010) and Comesaña et al.'s (2015) results, suggest a more graded nature for the bilingual lexicon and that, perhaps, non-identical cognates share an orthographic representation.

A recent experiment conducted by Poort et al. (2016) offers a unique opportunity to examine how non-identical cognates are represented in the bilingual lexicon from a different angle. Poort et al. (2016) primed identical cognates and interlingual homographs in Dutch sentences (e.g. "Alleen vrouwelijke bijen en wespen hebben een angel", i.e. "Only female bees and wasps have a sting") and presented the same items again approximately 16 minutes later in an English lexical decision task. Their aim was to find out whether a bilingual is affected by recent experience with a cognate or interlingual homograph in their native language when processing such words in their second language. The data revealed that priming was beneficial for the cognates, which were recognised 28 ms more quickly in English when they had been primed in Dutch. In contrast, priming was disruptive for the interlingual homographs, which were recognised 49 ms more slowly in English if the participants had recently come across them in Dutch. This kind of priming, which Rodd, Cutrin, Kirsch, Millar and Davis (2013) and Rodd, Cai, Betts, Hanby, Hutchinson and Adler (2016) call long-term word-meaning priming, is thought to strengthen the connection between a word's form and meaning.

As an adaptive mechanism, long-term word-meaning priming is assumed to make it easier to retrieve the primed meaning again in the (near) future, though potentially at the cost of retrieving another meaning. Because cognates share both their form and meaning, priming was indeed facilitative in Poort et al.'s (2016) experiment. In contrast, because interlingual homographs only share their form and not their meaning, priming was disruptive. If non-identical cognates, like identical cognates, share an orthographic representation, cross-lingual long-term priming should be facilitative. If instead, like Dijkstra et al. (2010) suggest, non-identical cognates consist of two orthographic representations, long-term word-meaning priming should be disruptive for these words, as it was for the interlingual homographs in Poort et al.'s (2016) experiment. Experiment 1 of the current study was therefore designed to determine whether a non-identical cognate behaves more like an identical cognate or an interlingual homograph when using a cross-lingual long-term priming paradigm. After finding only weak evidence for a priming effect based on the results of Experiment 1, the aim of Experiment 2 was to determine whether the size of the priming effect is influenced by the presence of interlingual homographs in the experiment.

2 EXPERIMENT 1

The procedure for Experiment 1 was identical to the procedure employed by Poort et al. (2016), except that we decided not to include any interlingual homographs to allow us to include more identical and non-identical cognates. In the priming phase, Dutch–English bilinguals read Dutch sentences that contained either an identical cognate, a non-identical cognate or the Dutch translation of an English control word (to create a semantic priming control condition). After a filler task of approximately 20 minutes, the identical cognates, non-identical cognates and English controls (from here on referred to as 'translation equivalents') were presented again in isolation in an English lexical decision task.

In line with Poort et al. (2016), we predict facilitative priming for the identical cognates. If we find that priming is facilitative for the non-identical cognates as well, this would indicate that their representation in the bilingual lexicon is similar to that of an identical cognate. In contrast, if we find that priming is disruptive for the non-identical cognates, this would suggest that they are more like interlingual homographs and that they are indeed subject to lateral inhibition. Based on research in the monolingual domain with synonyms (Rodd et al., 2013), we do not expect to find priming for the translation equivalents, as they do not share their form with their Dutch translation. Poort et al. (2016), however, did observe a facilitative though non-significant priming effect for these words (which they called 'semantic controls'), so it may be that the translation equivalents possess some sort of special status in the bilingual lexicon that makes them behave more like cognates than cross-lingual synonyms. Finally, for the unprimed trials, based on the research discussed above, we expect to find a large cognate facilitation effect for the identical cognates compared to the translation equivalents and a smaller effect for the non-identical cognates, or even no effect at all.

2.1 Methods

2.1.1 Participants

Thirty-three London-based Dutch–English bilinguals were recruited through social media and word-of-mouth. All of the participants but one indicated they were native speakers of Dutch (or Flemish). The participant who did not indicate that Dutch was their native language did score very highly on the Dutch version of the LexTALE (96.3%), so they were not excluded from the analysis. Similarly, all participants indicated that they were fluent speakers of English except for one participant, but their score on the English version of the LexTALE was high (88.8%) and they were not excluded. The LexTALE (Lemhöfer & Broersma, 2012) is a simple test of vocabulary knowledge that provides a good indication of a participant's general language proficiency. The participants gave informed consent and were paid £6 for their participation in the experiment.

The data from two participants were excluded due to technical problems. The data from one additional participant were excluded because this participant's performance on the semantic relatedness task (40.2%) was below chance level (50%). The data from a fourth participant were excluded because this participant's percentage correct for the experimental items included in the lexical decision task (86.8%) was more than three standard deviations below the mean of all participants (M = 97.1%, SD = 2.7%).

The remaining 29 participants (10 male; M_{age} = 27.2 years, SD_{age} = 6.0 years) had started learning English from an average age of 7.1 years (SD = 3.2 years) and so had an average of 20.1 years of experience with English (SD = 6.8 years). The participants rated their proficiency as 9.4 out of 10 in Dutch (SD = 0.7) and 8.7 in English (SD = 0.8). A two-sided paired t-test showed this difference to be significant [$t(28) = 4.209$, $p < .001$; $\alpha = .05$]. The self-ratings were confirmed by their high LexTALE scores in both languages, which were also slightly higher in Dutch, though this difference was not significant [Dutch: M = 92.8%, SD = 5.5%; English: M = 90.3%, SD = 7.9%; $t(28) = 1.405$, $p = .171$].

	Dutch word	English word	prime sentence (Dutch original)	prime sentence (English translation)
identical cognate	winter	winter	Van kinds af aan was zijn favoriete seizoen al winter.	Ever since he was a child, his favourite season was winter.
non-identical cognate	baard	beard	De Kerstman heeft een lange, witte baard.	Santa Claus has a long, white beard.
translation equivalent	fiets	bike	Hij ging elke dag met de fiets naar school.	Every day he went to school by bike.
interlingual homograph	angel	angel	Alleen vrouwelijke bijen en wespen hebben een angel.	Only female bees and wasps have a sting.

Table 1: An example of an identical cognate, non-identical cognate, translation equivalent and interlingual homograph with their Dutch prime sentences (along with their English translation). The interlingual homographs were only included in Experiment 2.

	Dutch characteristics				English characteristics				Similarity ratings			
	frequency	log10(frequency)	word length	OLD20	frequency	log10(frequency)	word length	OLD20	meaning	spelling	pronunciation	prime sentence length
identical cognates	37.0 (56.3)	2.9 (0.5)	4.6 (1.1)	1.6 (0.4)	41.5 (54.0)	3.1 (0.5)	4.6 (1.1)	1.6 (0.4)	6.8 (0.2)	7.0 (0.0)	5.9 (0.7)	9.6 (1.6)
non-identical cognates	31.4 (34.8)	2.9 (0.5)	4.8 (1.0)	1.5 (0.3)	44.0 (50.7)	3.1 (0.4)	4.7 (1.0)	1.6 (0.3)	6.9 (0.2)	5.3 (0.6)	5.1 (0.8)	9.5 (1.9)
translation equivalents	29.7 (31.4)	2.9 (0.5)	4.7 (0.9)	1.5 (0.3)	34.2 (31.1)	3.1 (0.4)	4.6 (1.0)	1.6 (0.3)	6.9 (0.2)	1.1 (0.3)	1.1 (0.2)	9.5 (1.5)

Table 2: Means (and standard deviations) for all key matching variables, prime sentence length and the pre-test similarity ratings. Frequency refers to the SUBTLEX word frequency in occurrences per million (see Keuleers et al., 2010, for Dutch and Brysbaert & New, 2009, for English); log10(frequency) refers to the SUBTLEX log-transformed raw word frequency (log10[raw frequency+1]); OLD20 refers to Yarkoni et al.'s (2008) measure of orthographic complexity. Meaning, spelling and pronunciation similarity ratings were obtained through a pre-test and were given on a scale of 1 (not at all similar) to 7 (almost identical). Prime sentence length refers to the length of the Dutch prime sentences shown during the semantic relatedness task.

2.1.2 Materials

2.1.2.1 Words & sentences

We selected a large number of Dutch–English translation pairs with varying degrees of orthographic overlap from Poort et al. (2016), Dijkstra et al. (2010) and Tokowicz, Kroll, De Groot and Van Hell (2002). All words were between 3 and 8 letters long and their frequency in both English and Dutch was between 2 and 600 occurrences per million according the SUBTLEX-NL (Keuleers, Brysbaert & New, 2010) and SUBTLEX-US databases (Brysbaert & New, 2009). From this set, we selected 65 pairs with identical orthographic forms to serve as identical cognates. The remaining items we divided into categories of high and low degrees of orthographic overlap based on the similarity ratings Dijkstra et al. (2010) and Tokowicz et al. (2002) had collected. From these two categories, we selected 80 pairs with a high degree of orthographic overlap to serve as non-identical cognates and 80 with a low degree of overlap to serve as translation equivalents. We wrote prime sentences for these items in Dutch (see Table 1 for examples) that were between 6 and 12 words long, with the target placed as far towards the end of the sentence as possible, since this minimises ambiguity. A second native speaker of Dutch checked the sentences for mistakes and suggested clarifications when the meaning of the sentence was not entirely clear.

To ensure that the items in each condition had the appropriate properties, we then asked a total of 67 Dutch–English bilinguals who did not take part in the main experiment to rate how similar the Dutch and English items in each pair were to each other with respect to their meaning, spelling and pronunciation. We also included a set of identical and non-identical interlingual homographs as fillers, to ensure that the participants would make full use of the rating scales. We presented the participants with the Dutch prime sentences with the target items marked in bold (here underlined, e.g. "De Kerstman heeft een lange, witte baard.") and next to it the English targets (e.g. "beard"). Each pair received ratings from at least 12 participants. Pairs with an average meaning similarity rating of less than 6 on our 7-point scale were discarded. Translation equivalents with an average spelling similarity rating higher than 2 or non-identical cognates with an average spelling similarity rating of less than 4 were also discarded.

The software package Match (Van Casteren & Davis, 2007) was then used to select the 58 best-matching identical cognates, non-identical cognates and translation equivalents. The items were matched based on log-transformed word frequency in Dutch and English (weight: 1.5), the number of letters of the word in Dutch and English (weight: 1.0) and orthographic complexity of the word in Dutch and English using the word's mean orthographic Levenshtein distance to its 20 closest neighbours (OLD20; Yarkoni, Balota & Yap, 2008; weight: 0.5)[1]. Table 2 lists means and standard deviations per word type for each of these measures, as well as prime sentence length and spelling, pronunciation and meaning similarity.

An independent-samples two-tailed Welch's *t*-test showed that there was a small but significant difference between the identical cognates and the translation equivalents in terms of Dutch

---

[1] We decided to weight the matching variables in this order as it has been shown that frequency is a more important predictor of lexical decision reaction times than word length and orthographic complexity (Brysbaert, Stevens, Mandera & Keuleers, 2016; Keuleers, Stevens, Mandera & Brysbaert, 2015; Yarkoni et al., 2008).

OLD20 (p = .030; α = .05). No other differences between the word types on these matching criteria were significant, nor were the differences in prime sentence length significant (all ps > .1). All three word types were significantly different from each other in terms of spelling and pronunciation similarity ratings (all ps < .001), but not with respect to the meaning similarity ratings (p >.1). This confirmed the word types' intended status.

2.1.2.2 NON-WORDS

The 174 non-words in the English lexical decision task comprised 144 English-sounding pseudohomophones (e.g. "mistaik") from Rodd (2000) and 30 Dutch words (e.g., "schaar") of a similar frequency as the target items, selected pseudo-randomly from the SUBTLEX-NL database. The Dutch words were included to ensure participants only responded 'yes' to English words. Pseudohomophones were used instead of regular non-words to encourage relatively deep processing (Rodd, Gaskell & Marslen-Wilson, 2002). The non-words were of a similar length as the target items.

2.1.3 DESIGN AND PROCEDURE

For each participant, half the words of each word type were primed (i.e., appeared during the priming phase) while half were unprimed (i.e., only occurred in the later test phase). Two versions of the experiment were created such that participants saw each experimental item only once but across participants items occurred in both the primed and unprimed conditions. The experiment was created using MATLAB (version R2012a; The Mathworks Inc., 2012) and conducted at the Department of Experimental Psychology at University College London.

The experiment comprised five separate tasks (mean duration in mm:ss): (1) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012; 04:00), (2) a Dutch semantic relatedness task (12:13), (3) an English digit span task (06:39), (4) an English lexical decision task (13:36) and (5) the English version of the LexTALE (Lemhöfer & Broersma, 2012; 03:15). On average, lexical decisions to primed items were made 19 minutes and 27 seconds after they were primed, as measured from the end of the break between the two blocks of the semantic relatedness task to the end of the break of the fourth block (of eight) of the lexical decision task. The five tasks were presented separately, with no indication that they were linked. Responses were recorded via a standard keyboard. At the end of the experiment, participants completed a self-report language background survey in Dutch.

2.1.3.1 DUTCH SEMANTIC RELATEDNESS TASK

This task served to prime the identical cognates, non-identical cognates and translation equivalents. To ensure the participants processed the prime sentences, participants were asked to indicate via button presses whether a subsequent probe was semantically related to the preceding sentence. The probes were either very strongly related or completely unrelated to the sentence. The same native speaker of Dutch that proofread the prime sentences also confirmed that the probes were (un)related to the sentence as specified. The 58 sentences for each word type were pseudo-randomly divided into two sets, matched for all key variables, for use in the two versions of the experiment.

Participants read six practice sentences and two blocks of 44 and 43 experimental sentences presented in different random orders. The order of the blocks was fixed but counterbalanced across participants (making four versions of the experiment in total). Each sentence remained on the screen for 4000ms; each probe until the participant responded or until 3000ms passed. The inter-trial interval was 1000ms. A five-second break was enforced after block 1.

2.1.3.2 DIGIT SPAN TASK

This task served to introduce a delay between priming and testing, while minimising exposure to additional linguistic material. It was conducted in English to minimise any general language switch cost on the lexical decision task. Each string of digits comprised four to eight digits. Each digit was presented for 500ms with 1000ms between trials and between 4000ms and 6000ms (dependent on the string length) for the participants to recall the sequence. Participants saw five practice strings followed by 36 experimental strings divided into two blocks. A 10-second break was enforced after the first block.

2.1.3.3 ENGLISH LEXICAL DECISION TASK

Participants saw all 348 experimental stimuli (58 of each word type plus 174 non-words) and were asked to indicate, by means of button presses, as quickly and accurately as possible, whether it was a real English word or not. A practice block of 24 strings was followed by eight blocks of 44 or 43 experimental stimuli. The target items that a participant had seen in the first (or second) block of the semantic relatedness were always presented in one of the first (or last) four blocks of the lexical decision task, creating two parts for the lexical decision task. The experiment was designed like this to minimise the variation in delay between the two presentations of the same target. The order of the items within blocks was randomised for each participant, as was the order of the blocks within each part. Six fillers were presented at the beginning of each block and the participants had a five-second break at the end of each block. All items remained on the screen until the participant responded, or until 2000ms passed. The inter-trial interval was set at 500ms.

2.2 RESULTS

All analyses were carried out in R (version 3.3.2; R Core Team, 2016) using the lme4 package (version 1-1.12; Bates, Maechler, Bolker & Walker, 2015), following Barr et al.'s (2013) guidelines for confirmatory hypothesis testing and using Type III Sums of Squares likelihood ratio tests to determine significance. All *p*-values were compared against an α of .05 unless stated otherwise. We also report Bayes factors, which were computed using the following formula suggested by Wagenmakers (2007): $BF_{10} = e^{\frac{BIC_{null}-BIC_{alternative}}{2}}$. We follow Jeffrey's (1961) guidelines when interpreting the Bayes factors; however, since these Bayes factors have an uninformative prior, they are not the main focus of our analyses.

Three items were excluded from the analyses (the identical cognate "fort"–"fort", the non-identical cognate "deed"–"daad" and the translation equivalent "ant"–"mier"), as the percentages correct for those items on the lexical decision task (86.2%, 51.7% and 86.2%, respectively) were more than three standard deviations below the mean of all items of the same word type (for the identical cognates: $M = 97.2\%$, $SD = 3.4\%$; for the non-identical cognates: $M = 97.6\%$, $SD = 6.7\%$; for the translation equivalents: $M = 97.4\%$, $SD = 3.7\%$). Excluding these items did not affect the matching of the word types.

2.2.1 DIGIT SPAN TASK

The participants' digit span (i.e. greatest string length recalled with at least 50% accuracy) was within normal limits ($M = 5.8$ digits, $SD = 1.0$, range 4–8 digits), confirming task engagement.

2.2.2 SEMANTIC RELATEDNESS TASK

High accuracy scores ($M = 93.6\%$, $SD = 3.8\%$, range 85.1%–98.9%) confirmed participants had processed the sentence meanings. To determine whether any of the observed effects in the lexical decision task could have been due to differences between the word types at the time of priming a model with the fixed factor word type (3 within-participants/between-items levels: identical cognate, non-identical cognate, translation equivalent) was fitted to the data. The maximal model converged and included a random intercept by participants and by items and a random by-participants slope for word type. This model revealed that the effect of word type was not significant [$\chi^2(2) = 1.225$, $p = .542$, $BF_{10} < 0.001$]. The Bayes factor also provided decisive evidence for the null hypothesis.

2.2.3 LEXICAL DECISION TASK

2.2.3.1 ANALYSIS PROCEDURE

The following analyses were conducted on both the reaction time data and the accuracy data. In all cases, positive effects of priming indicate a facilitative effect of priming.

In the main 3×2 analysis two fixed factors were included: word type (3 within-participant/between-items levels: identical cognate, non-identical cognate) and priming (2 within-participant/within-items levels: unprimed, primed). The random effects structure of this model included an intercept by participants and items, as well as a by-participants slope for word type and priming and their interaction and a by-items slope for priming. For the reaction times analysis, this maximal model did not converge, so we removed the correlations between the random effects to achieve convergence, as recommended by Barr et al. (2013); this was not necessary for the accuracy analysis.

We also conducted three simple effects analyses, to look at the effect of priming on each of the three word types separately. The maximal model converged for the reaction times analysis and included a random intercept by participants and by items and random a by-participants and by-items slope for priming. For the accuracy analysis, we had to remove the correlations between the

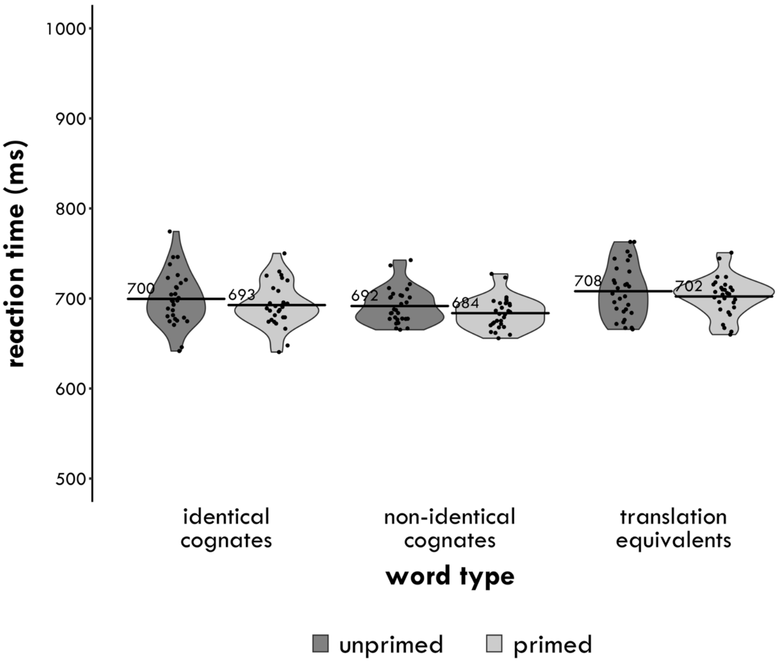

Figure 1: Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by word type (x-axis: identical cognates, non-identical cognates and translation equivalents) and priming (dark grey: unprimed; light grey: primed). Each point represents a participant. The horizontal lines provide the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

random effects to achieve convergence. The *p*-values for these three analyses were compared against a Bonferroni-corrected α of .017.

Finally, we conducted three pairwise comparisons on the unprimed data only, comparing the identical cognates, non-identical cognates and translation equivalents against each other. Again, the maximal model converged for these three models and included a random intercept by participants and by items and a by-participants random slope for word type. The *p*-values for these three analyses were also compared against a Bonferroni-corrected α of .017.

2.2.3.2 Reaction times

Lexical decision reaction times are shown in Figure 1. Reaction times (RTs) faster than 300ms or slower than 1500ms were discarded, as they were assumed to be the result of accidental key presses or participants losing focus for a moment (2.8% of the data). RTs for incorrect trials were also discarded (1.5% of the remaining data). The RTs were inverse-transformed (inverse-transformed RT = 1000/raw RT) as the assumptions of normality and homoscedasticity did not hold for the main 3×2 analysis. As a last step in the data processing stage, we removed inverse-transformed RTs more than three standard deviations above or below a participant's mean inverse-transformed RT (0.2% of the remaining data). Finally, it should be noted that the graph in Figure 1 displays the harmonic participant means, while the effects reported in the text are derived from the estimates of the fixed effects provided by the model.

In the 3×2 analysis, the main effect of word type was not significant [$\chi^2(2) = 4.123$, $p = .127$, $BF_{10} = 0.002$], while the 7 ms facilitative effect of priming was marginally significant [$\chi^2(1) = 3.261$, $p = .071$, $BF_{10} = 0.074$]. The interaction between word type and priming was also not significant [$\chi^2(2) = 0.226$, $p = .893$, $BF_{10} < 0.001$]. All Bayes factors provided strong or even decisive evidence for the null hypothesis.

The three simple effects analyses revealed that the effect of priming was not significant for either the identical cognates [$\chi^2(1) = 0.852$, $p = .356$, $BF_{10} = 0.039$, $\Delta = 6$ ms], the non-identical cognates [$\chi^2(1) = 2.053$, $p = .152$, $BF_{10} = 0.070$, $\Delta = 9$ ms] or the translation equivalents [$\chi^2(1) = 0.701$, $p = .403$, $BF_{10} = 0.036$, $\Delta = 5$ ms]. The Bayes factors again provided strong evidence for the null hypothesis.

Finally, none of the three pairwise comparisons was significant [identical cognates vs non-identical cognates: $\chi^2(1) = 0.615$, $p = .433$, $BF_{10} = 0.034$, $\Delta = -7$ ms; identical cognates vs translation equivalents[2]: $\chi^2(1) = 0.569$, $p = .451$, $BF_{10} = 0.034$, $\Delta = 8$ ms; non-identical cognates vs translation equivalents: $\chi^2(1) = 2.636$, $p = .105$, $BF_{10} = 0.094$, $\Delta = 20$ ms]. (Positive effects indicate a reaction time advantage for the first-named word type over the second-named word type.) The Bayes factors again provided strong evidence for the null hypothesis.

2.2.3.3 ACCURACY

Lexical decision task accuracy is shown in Figure 2. In line with the trimming procedure for the reaction times, any trials with RTs faster than 300ms or slower than 1500ms were removed. We used the bobyqa optimiser instead of the default Nelder-Mead optimiser as more complex models in our experience are more likely to converge when the bobyqa optimiser is used. Overall, the analyses on the accuracy data revealed a similar pattern of results as the analyses on the reaction time data.

In the the 3×2 analysis, the main effect of word type was not significant [$\chi^2(2) = 1.199$, $p = .549$, $BF_{10} < 0.001$], nor was effect of priming [$\chi^2(1) = 0.006$, $p = .937$, $BF_{10} = 0.014$, $\Delta = 0.02\%$]. The interaction between word type and priming was also not significant [$\chi^2(2) = 1.745$, $p = .417$, $BF_{10} < 0.001$]. All Bayes factors provided strong or even decisive evidence for the null hypothesis.

Unsurprisingly, the simple effect of priming was not significant for any of the three word types separately either [identical cognates: $\chi^2(1) = 0.339$, $p = .560$, $BF_{10} = 0.029$, $\Delta = 0.1\%$; non-identical cognates: $\chi^2(1) = 1.669$, $p = .196$, $BF_{10} = 0.057$, $\Delta = 0.2\%$; translation equivalents: $\chi^2(1) = 0.074$, $p = .786$, $BF_{10} = 0.026$, $\Delta = 0.1\%$]. Again, the Bayes factors provided strong to very strong evidence for the null hypothesis.

Finally, none of the three pairwise comparisons was significant [identical cognates vs non-identical cognates: $\chi^2(1) = 0.036$, $p = .850$, $BF_{10} = 0.025$, $\Delta = 0.1\%$; identical cognates vs translation equivalents: $\chi^2(1) = 0.184$, $p = .668$, $BF_{10} = 0.027$, $\Delta = 0.1\%$; non-identical cognates vs translation

[2] There was a small but significant difference between the identical cognates and translation equivalents in terms of the Dutch OLD20 measure. As the pairwise comparison between these two word types was not significant, we felt it unnecessary to re-run the analysis with the Dutch OLD20 measure included as a covariate.

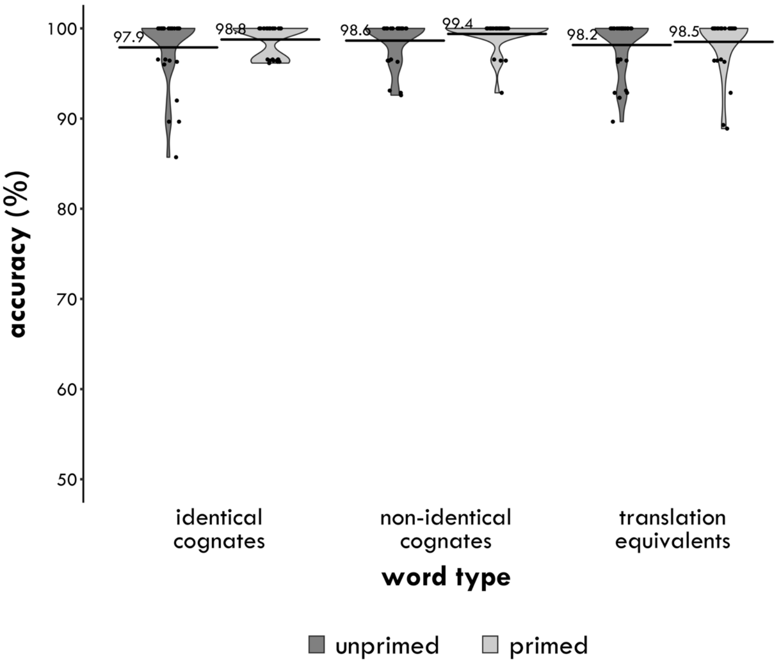

Figure 2: Participant means of the lexical decision accuracy data (in percentage) by word type (x-axis: identical cognates, non-identical cognates and translation equivalents) and priming (dark grey: unprimed; light grey: primed). Each point represents a participant's mean accuracy. The violin is a symmetrical density plot rotated by 90 degrees. The horizontal line and number provide the mean across all participants in that condition.

equivalents: $\chi^2(1) = 0.047$, $p = .828$, $BF_{10} = 0.026$, $\Delta = 0.1\%$]. (Positive effects indicate an accuracy advantage for the first-named word type over the second-named word type.) Again, the Bayes factors provided very strong evidence for the null hypothesis.

2.3 Discussion

In contrast to our predictions, we found little evidence for a cross-lingual long-term word-meaning priming effect. The main effect of priming was much smaller than expected (7 ms) and only marginally significant. When looking at the identical and non-identical cognates separately, although priming was facilitative in both cases, the size of the effect was only 6 ms for the identical cognates and 9 ms for the non-identical cognates. The Bayes factors in all cases provided strong (or even decisive) evidence for the null hypothesis. In addition, we also did not find evidence for a cognate facilitation effect in the unprimed trials, either for the identical cognates (a facilitative effect of 7 ms) or for the non-identical cognates (an effect of 20 ms). Again, the Bayes factors provided strong evidence for the null hypothesis.

With regards to the latter finding, this may not be as surprising as it seems, as recent research suggests that the cognate facilitation effect is significantly smaller (even non-existent) when the experiment includes non-target language words (Poort & Rodd, 2017, May 9), as our experiment did. However, the fact that we did not find evidence of priming is unexpected given Poort et al.'s (2016) results. The main difference between their experiment and the current experiment was that they also included interlingual homographs. As these are quite difficult items for bilinguals, it might be the case that the participants taking part in their experiment strategically relied more on their recent experience with these words to inform their decisions. In contrast, all of the items included in our experiment were relatively easy words, so our participants may not have felt the need to rely on their recent experience with these words as much. Alternatively, it could have been the case that the interlingual homographs in Poort et al.'s (2016) experiment encouraged deeper, more semantic processing. Experiment 2 was designed to determine whether priming is indeed affected by the presence of interlingual homographs in the experiment.

3 Experiment 2

Experiment 2 was preregistered as part of the Center for Open Science's Preregistration Challenge (cos.io/prereg). Our stimuli, data and processing and analysis scripts can be found in our project on the Open Science Framework (osf.io/dytp7). The preregistration can be retrieved from osf.io/33r86 (Poort & Rodd, 2017, January 24). Where applicable, deviations from the pre-registration will be noted.

To find out whether the presence of interlingual homographs in the stimulus list modulates the cross-lingual long-term priming effect, two versions of the experiment were created. Both versions included identical and non-identical cognates in the semantic relatedness task and identical and non-identical cognates, translation equivalents, pseudohomophones and some Dutch words in the lexical decision task. We decided not to prime the translation equivalents in this experiment, as this allowed us to include more items of the critical word types without increasing the experiment's duration and we had no theoretically motivated predictions for this word type. In addition to these stimuli, one version of the experiment also included interlingual homographs in both the semantic relatedness task and the lexical decision task. To reflect the fact that this version included interlingual homographs, we call it the +IH version; the version that did not include any interlingual homographs is termed the –IH version.

Based on Poort et al.'s (2016) findings and the results of Experiment 1, we make the following predictions regarding the priming effect: (1) we will find evidence for long-term word-meaning priming in the +IH version, (2) we will not find evidence for long-term word-meaning priming in the –IH version and (3) priming in the +IH version will be facilitative for the identical cognates and disruptive for the interlingual homographs. Provided that the priming manipulation works, the same explanations hold in case we find a facilitative or disruptive effect of priming for the non-identical cognates. If priming is facilitative for the non-identical cognates, this would suggest that they are stored in the bilingual lexicon much the same way an identical cognate is. In contrast, if priming is disruptive for the non-identical cognates, this would indicate that they are more like interlingual homographs than identical cognates and that they may indeed be subject to processes of lateral inhibition.

For the unprimed trials, we make the following predictions: (1) we will not find a cognate facilitation effect for the identical or non-identical cognates in either version of the experiment but (2) we will find evidence for an interlingual homograph inhibition effect in the +IH version. The predictions for the identical and non-identical cognates follow from the results of Experiment 1 and Poort & Rodd's (2017, May 9) findings discussed previously. The prediction that we will find evidence for an interlingual homograph inhibition effect also follows from Poort and Rodd's (2017, May 9) experiments, as well as many other experiments (e.g. Dijkstra et al., 1998), that have shown that inhibition for interlingual homographs is found most consistently when the experiment includes non-target language words.

3.1 METHODS

3.1.1 PARTICIPANTS

Given our uncertainty surrounding the size of the priming effect, we decided to recruit at least 32 participants per version, consistent with Experiment 1. In the end, because we excluded a number of participants while recruitment was still on-going, a total of 74 participants were recruited through Prolific Academic (Damer & Bradley, 2014), social media and personal contacts resident in the Netherlands. All participants were between the ages of 18 and 50, held the Dutch or Belgian nationality and were resident in the Netherlands or Belgium at the time of the experiment. They were also all native speakers of Dutch and fluent speakers of English and did not have any language disorders. They gave informed consent and were paid for their participation in the experiment. Participants recruited through prolific Academic were paid £8, while participants recruited through other means were given the choice either to receive a €10 gift card or to donate €10 to charity.

Excluding participants was carried out in two stages. First, while testing was still on-going, we excluded and replaced participants who scored less than 80% correct on the semantic relatedness task, less than 80% on the lexical decision task and/or less than 50% on either of the two language proficiency measures. We also excluded and replaced participants if their average delay between the two presentations of the primed items was more than 30 minutes. Four participants that met these criteria were excluded and four new participants tested in their stead. Due to a technical error, the data from three other participants could not be used, so these participants were also replaced. Finally, it should be noted that although we also intended to exclude participants whose internet connection was not fast or stable enough, our software frequently filtered out participants' whose internet connection appeared to be fine when tested with an online internet speed test. For this reason, we decided to disable the internet connection filter early on during testing. This did not affect the accuracy of the reaction time measurements, as these were calculated locally on each participants' own computer.

Second, after testing had finished and a total of 67 useable datasets had been gathered, 33 in the –IH version and 34 in the +IH version, we compared each participant's performance on the words (identical cognates, non-identical cognates, English and, depending on the version, identical interlingual homographs) included in the lexical decision task to the grand mean of all participants who completed that version to determine whether any more participants needed to be excluded. All participants performed within three standard deviations from their version's mean (–IH version: M = 96.8%, SD = 2.2%; +IH version: M = 95.1%, SD = 3.1%), so none were excluded at this stage.

The 67 included participants (21 males; M_{age} = 23.4 years, SD_{age} = 6.4 years) had started learning English from an average age of 7.4 (SD = 3.3 years) and so had an average of 16.0 years of experience with English (SD = 6.7 years). The participants rated their proficiency in Dutch as 9.5 out of 10 (SD = 0.6) and in English as 8.7 (SD = 0.9). A two-sided paired t-test showed this difference to be significant [$t(66) = 8.246, p < .001; \alpha = .05$]. In contrast, the participants scored slightly higher on the English LexTALE than the Dutch version, but this difference was not significant [Dutch: M = 87.3%, SD = 7.5%; English: M = 88.7%, SD = 8%; $t(66) = -1.636$ $p = .107$]. An independent-samples two-tailed Welch's t-test showed that the participants in the +IH version had scored higher on the Dutch LexTALE (M = 89.4%, SD = 6.3%) than the participants in the –IH version [M = 85.2%, SD = 8.1%; $t(60.3) = -2.368, p = .021; \alpha = .05$]. No other differences between the two versions were significant (all ps > .08).

3.1.2 Materials

3.1.2.1 Words & sentences

A second pre-test, identical in design to the first, was conducted on a set of 80 identical interlingual homographs selected from the same sources and according to the same criteria as the stimuli for Experiment 1. Twenty-three participants who did not take part in either of the two experiments or the first pre-test completed the second pre-test. The participants were again presented with the Dutch prime sentence with the target items marked in bold (here underlined, e.g. "Alleen vrouwelijke bijen en wespen hebben een <u>angel</u>", i.e. "Only female bees and wasps have a <u>sting</u>") and next to it the English targets (e.g. "angel"). Each item received ratings from at least 11 participants. We discarded any interlingual homographs for which at least one participant had indicated that they did not know the word either in Dutch or English and any that had received an average meaning similarity rating of more than 2. We decided to keep three such items, however, as only one or two participants had given them a high meaning similarity rating of 7, whereas all other participants had given them a rating of 1 or 2.

Fifty identical interlingual homographs were selected from the set of pre-tested items to match 50 identical cognates and 50 non-identical cognates selected from the stimuli included in Experiment 1. We chose to use a subset of 50 of the 58 items of these two word types as we could achieve better matching this way. We also selected 50 English control words (from the total set of items pre-tested for Experiment 1) to serve as matched fillers in the lexical decision task. The four word types were matched on English log-transformed frequency (weight: 1.5), word length (weight: 1.0) and OLD20 (weight: 0.5), using the software Match (Van Casteren & Davis, 2007). Table 3 lists means and standard deviations per word type for each of these measures (in both Dutch and English), as well as prime sentence length and the spelling, pronunciation and meaning similarity ratings obtained from the pre-tests. Table 4 additionally lists the number of items of each word type included in the semantic relatedness and lexical decision task per version of the experiment. We made minor changes to some of the prime sentences for the identical and non-identical cognates, to ensure that no target item appeared in any other sentence than its own prime sentence.

	Dutch characteristics				English characteristics				Similarity ratings			
	frequency	log10(frequency)	word length	OLD20	frequency	log10(frequency)	word length	OLD20	meaning	spelling	pronunciation	prime sentence length
identical cognates	39.0 (59.4)	2.9 (0.5)	4.5 (1.0)	1.6 (0.4)	44.8 (57.0)	3.1 (0.5)	4.5 (1.0)	1.6 (0.3)	6.9 (0.2)	7.0 (0.0)	5.9 (0.7)	9.5 (1.6)
non-identical cognates	32.2 (35.4)	2.9 (0.5)	4.7 (1.0)	1.5 (0.3)	45.4 (51.6)	3.2 (0.4)	4.6 (0.9)	1.5 (0.3)	6.9 (0.2)	5.3 (0.6)	5.0 (0.7)	9.3 (1.9)
interlingual homographs	39.2 (82.9)	2.7 (0.7)	4.3 (0.9)	1.3 (0.4)	56.2 (107)	3.0 (0.7)	4.3 (0.9)	1.5 (0.3)	1.2 (0.3)	7.0 (0.0)	5.5 (0.8)	9.2 (1.8)
English controls	—	—	—	—	30.8 (26.5)	3.0 (0.4)	4.5 (0.9)	1.6 (0.3)	6.9 (0.2)	1.1 (0.2)	1.1 (0.2)	—

Table 3: Means (and standard deviations) for all key matching variables, prime sentence length and the pre-test similarity ratings. Frequency refers to the SUBTLEX word frequency in occurrences per million (see Keuleers et al., 2010, for Dutch and Brysbaert & New, 2009, for English); log10(frequency) refers to the SUBTLEX log-transformed raw word frequency (log10[raw frequency+1]); OLD20 refers to Yarkoni et al.'s (2008) measure of orthographic complexity. Meaning, spelling and pronunciation similarity ratings were obtained through two pre-tests and were given on a scale of 1 (not at all similar) to 7 (almost identical). Prime sentence length refers to the length of the Dutch prime sentences shown during the semantic relatedness task.

			Number of items per stimulus type							
	N		identical cognates	non-identical cognates	interlingual homographs	translation equivalents	pseudo-homophones	Dutch words	task duration	average delay
Exp. 1	28	semantic relatedness task	58	58	0	58	–	–	12:13	19:27
		lexical decision task	58	58	0	58	144	30	13:36	
Exp. 2: –IH	33	semantic relatedness task	50	50	0	0	–	–	07:13	17:13
		lexical decision task	50	50	0	50	125	25	14:02	
Exp. 2: +IH	34	semantic relatedness task	50	50	50	0	–	–	10:20	17:54
		lexical decision task	50	50	50	50	166	34	17:08	

Table 4: Overview of the types and numbers of stimuli included in each version of Experiment 2, as well as durations of the different tasks and of the delay between priming and testing. *N* is the number of participants included in the analysis.

Independent-samples two-tailed Welch's *t*-tests showed that the differences between the identical and non-identical cognates, the identical interlingual homographs and the translation equivalents on the matching criteria and prime sentence length were not significant (all *p*s $> .4$; $\alpha = .05$). An analysis of the meaning similarity ratings confirmed that the identical and non-identical cognates both differed significantly from the interlingual homographs, as intended (both *p*s $< .001$), but not from each other ($p = .396$). In addition, the identical cognates and interlingual homographs were significantly different from the non-identical cognates in terms of spelling similarity ratings (both *p*s $< .001$), but not from each other ($p = .106$). The translation equivalents were also significantly different from the identical cognates, non-identical cognates and the interlingual homographs in terms of spelling similarity ratings (all *p*s $< .001$). In terms of meaning similarity ratings, the translation equivalents did not differ from the identical and non-identical cognates (both *p*s $> .4$), but they did differ from the interlingual homographs ($p < .001$). All word types differed significantly from each other in terms of pronunciation similarity ratings (all *p*s $< .03$).

3.1.2.2 Non-words

Each version of the experiment included the same number of non-words as words. In the –IH version, the 150 non-words comprised 125 English-sounding pseudohomophones and 25 Dutch words. In the +IH version, the 200 non-words comprised 166 pseudohomophones and 34 Dutch controls. The pseudohomophones were selected from Rodd (2000) and the ARC non-word and pseudohomophone database (Rastle, Harrington & Coltheart, 2002). A number of additional pseudohomophones needed to be newly created to ensure each word was matched to a non-word of the same length. These items had similar properties to the pseudohomophones from the aforementioned sources. The Dutch words were of a similar frequency as the target items, selected pseudo-randomly from the SUBTLEX-NL database. In both versions, the non-words were matched word-for-word to a target in terms of word length.

3.1.3 Design and procedure

Only the identical and non-identical cognates and the identical interlingual homographs were primed. Of those, for each participant half the words of each word type were primed (i.e., appeared during the priming phase) while half were unprimed (i.e., only occurred in the later test phase). Two versions of the experiment were created such that participants saw each experimental item only once but across participants items occurred in both the primed and unprimed condition. The experiment was created and conducted using Gorilla online experimental software (Evershed & Hodges, 2016).

The experiment comprised five separate tasks (mean duration in mm:ss): (1) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012; 01:58), (2) a Dutch semantic relatedness task (–IH version: 07:13; +IH version: 10:20), (3) a Towers of Hanoi task (with instructions presented in English; –IH version: 07:15; +IH version: 03:23), (4) an English lexical decision task (–IH version: 14:02; +IH version: 17:08) and (5) the English version of the LexTALE (Lemhöfer & Broersma, 2012; 01:46). The design of each task (except for the Towers of Hanoi task) was the same as in

Experiment 1. On average, lexical decisions to primed items were made 17 minutes and 54 seconds after they were primed in the –IH version and 17 minutes and 13 seconds in the +IH version. A two-sided independent-samples Welch's *t*-test revealed that the difference in delay of 40.7 seconds was not significant [$t(58.2) = 1.232$, $p = .223$; $\alpha = .05$]. The five tasks were presented separately with no indication that they were linked. At the start of the experiment, the participants completed a self-report language background survey in Dutch.

3.1.3.1 DUTCH SEMANTIC RELATEDNESS TASK

The 50 sentences in each condition were pseudorandomly divided into two sets, matched for all key variables, for use in the two versions of the experiment. Participants read four (–IH version) or six (+IH version) practice sentences and two blocks of 25 (–IH version) or 37 or 38 (+IH version) experimental sentences presented in different random orders. The order of the blocks was fixed but counterbalanced across participants. Each sentence remained on the screen for 4000ms; each probe until the participant responded or until 2500ms passed. The inter-trial interval was 1000ms. A 30-second break was enforced after the first block and two (–IH version) or three (+IH version) fillers were presented at the start of the second block.

3.1.3.2 THE TOWERS OF HANOI TASK

This task served to introduce a delay, while minimising exposure to additional linguistic material. The Towers of Hanoi is a puzzle in which disks of progressively smaller sizes must be moved from one peg to another. There are two simple rules: (1) only one disk may be moved at a time and (2) a larger disk may not be placed on top of a smaller disk. The goal is to move the disks from the starting peg to the finish peg in as few moves as possible.

To ensure the priming delay in the –IH and +IH versions was of roughly equal duration, we set a time limit on the Tower of Hanoi task based on estimated durations of the semantic relatedness and lexical decision tasks in both versions. Participants in the –IH version were given six minutes and participants in the +IH version were given two minutes. Participants completed as many puzzles within the time limit as they could, starting with a puzzle with three disks and three pegs. Each subsequent puzzle had the same number of pegs but one disk more than the previous puzzle.

3.1.3.3 ENGLISH LEXICAL DECISION TASK

Participants saw all 300 (–IH version) or 400 (+IH version) experimental stimuli and were asked to indicate, by means of button presses, as quickly and accurately as possible, whether they were a real English word or not. A practice block of 24 or 32 strings was followed by eight blocks of 37 or 38 (–IH version) or 50 (+IH version) experimental stimuli. Five (–IH version) or six (+IH version) fillers were presented at the beginning of each block and a 30-second break followed the end of each block.

3.2 Results

This section reports the results of our planned confirmatory analyses. We do not report any exploratory analyses. All *p*-values were compared against an α of .05 unless stated otherwise. Detailed results of all analyses for Experiment 2 can be found in the Excel document analysisResults.xlsx in the Analysis scripts component of our OSF project (osf.io/6qnfc).

Three items (the identical cognate "fruit"–"fruit", the non-identical cognate "koord"–"cord" and the translation equivalent "kruid"–"herb") were excluded from the analyses, as the percentages correct on the lexical decision task for those items (88.1%, 81.5% and 85.1%, respectively) were more than three standard deviations below the overall mean of all items of the same word type (for the identical cognates: *M* = 88.4%, *SD* = 2.9%; for the non-identical cognates: *M* = 86.1%, *SD* = 3.7%; for the translation equivalents: *M* = 86.4%, *SD* = 3.5%). Excluding these items did not affect the matching of the word types.

3.2.1 The Towers of Hanoi Task

In the –IH version, where participants were given six minutes to complete as many puzzles as they could, the participants completed on average 2.7 puzzles (mode: 3, range: 0-4). In the +IH version, participants were given two minutes and so completed fewer puzzles, with an average of 1.4 (mode = 1, range: 0-3).

3.2.2 Semantic relatedness task

As can be seen in Figure 3, high accuracy (–IH version: *M* = 94.0%, *SD* = 4.2%, range: 82.0%–98.0%; +IH version: *M* = 92.2%, *SD* = 4.6%, range 80.0%–98.7%) confirmed participants had processed the sentence meanings. To determine whether any of the observed effects for the identical and non-identical cognates in the lexical decision task could have been due to differences between these two word types at the time of priming, we conducted a 2×2 analysis with factors word type (2 within-participants/between-items levels: identical cognate, non-identical cognate) and version (2 between-participants/within-items levels: –IH, +IH). The maximal model converged for this analysis and included a random intercept by participants and items as well as a by-participants random slope for word type and a by-items random slope for version. This analysis revealed that neither the main effect of word type [$\chi^2(1) = 0.852$, $p = .356$, $BF_{10} = 0.027$, $\Delta = 0.7\%$] nor the main effect of version [$\chi^2(1) = 1.149$, $p = .284$, $BF_{10} = 0.031$, $\Delta = 0.6\%$] was significant. The interaction was also not significant [$\chi^2(1) = 1.277$, $p = .259$, $BF_{10} = 0.033$]. The Bayes factors provided strong or even very strong evidence for the null hypothesis.

In addition, we conducted three pairwise comparisons on the data of the +IH version, comparing the identical cognates, non-identical cognates and interlingual homographs against each other. The maximal model also converged for these three analyses and included a random intercept by participants and items and a by-participants random slope for word type. The *p*-values for pairwise comparisons were compared against a Bonferroni-corrected α of .017. Again, these analyses revealed no significant differences between the word types [identical cognates vs non-identical cognates: $\chi^2(1) = 0.114$, $p = .736$, $BF_{10} = 0.026$, $\Delta = -0.3\%$; identical cognates vs

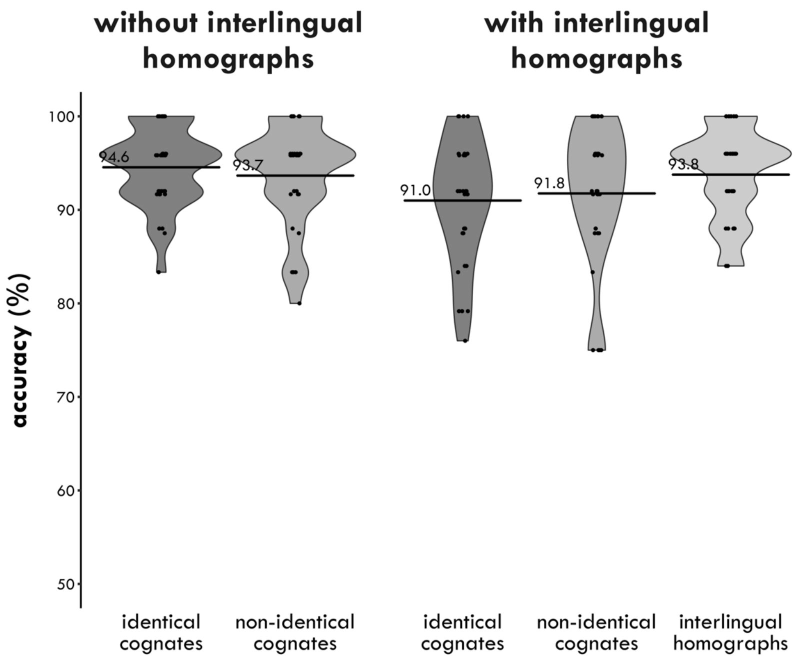

Figure 3: Participant means of the semantic relatedness accuracy data (in percentage) by word type (x-axis: identical cognates, non-identical cognates, interlingual homographs). The right panel pictures the –IH version; the left panel pictures the +IH version. Each point represents a participant. The horizontal line provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

interlingual homographs: $\chi^2(1) = 0.004$, $p = .950$, $BF_{10} = 0.024$, $\Delta = 0.1\%$; non-identical cognates vs interlingual homographs: $\chi^2(1) = 0.150$, $p = .699$, $BF_{10} = 0.026$, $\Delta = 0.4\%$]. (Positive effects indicate an advantage of the first-named word type over the second-named word type.) The Bayes factors also provided very strong evidence for the null hypothesis.

3.2.3 Lexical decision task

3.2.3.1 Analysis procedure

The following analyses were conducted on both the reaction time data and the accuracy data. In all cases, positive effects of priming indicate facilitative effect of priming.

In the main 2×2×2 analysis, three fixed factors were included: word type (2 within-participant/between-items levels: identical cognate, non-identical cognate), priming (2 within-participant/within-items levels: unprimed, primed) and version (2 between-participants/within-items levels: –IH, +IH). The random effects structure of this model included an intercept by participants and items, as well as a by-participants slope for word type and priming and their interaction and a by-items slope for priming. For both the reaction times and accuracy analysis, this maximal model did not converge and we had to remove the correlations between the random effects to achieve convergence.

In addition, as planned, in the –IH version, we conducted a 2×2 analysis comparing the effect of priming for the identical and the non-identical cognates; in the +IH version, we ran three 2×2 analysis comparing the effect of priming for the identical and non-identical cognates, the identical cognates and interlingual homographs and the non-identical cognates and interlingual homographs. The random effects structure for these models included an intercept by participants and by items and a by-participants slope for word type, priming and their interaction, as well as a by-items slope for priming. Again, the maximal model did not converge for either the reaction times or the accuracy analysis, but a model without correlations between the random effects did. The *p*-values for these four analyses were compared against a Bonferroni-corrected α of .013.

We then conducted five simple effects analyses, looking at the effect of priming on each of the three word types in the two versions. The maximal model converged for the reaction times analysis and included a random intercept by participants and by items and random a by-participants and by-items slope for priming. For the accuracy analysis, we again had to remove the correlations between the random effects to achieve convergence. The *p*-values for these five analyses were compared against a Bonferroni-corrected α of .01.

Finally, we conducted nine pairwise comparisons on the unprimed data only, comparing the identical cognates, non-identical cognates and translation equivalents against each other in the –IH version and the identical cognates, non-identical cognates, interlingual homographs and translation equivalents against each other in the +IH version. The maximal model converged for these three models for both the reaction times and accuracy analysis and included a random intercept by participants and by items and a by-participants random slope for word type. The *p*-values for these nine analyses were compared against a Bonferroni-corrected α of .006.

3.2.3.2 REACTION TIMES

Lexical decision reaction times are shown in Figure 4. As in Experiment 1, reaction times (RTs) faster than 300ms or slower than 1500ms were discarded (1.8% of the data), as were RTs for incorrect trials (3.4% of the remaining data). The RTs were again inverse-transformed (inverse-transformed RT = 1000/raw RT) as a histogram of the residuals and a predicted-vs-residuals plot for the main 2×2×2 analysis showed that the assumptions of homoscedasticity and normality were violated. We then removed any inverse-transformed RTs more than three standard deviations above or below a participant's mean inverse-transformed RT for all experimental items (0.3% of the remaining data). Finally, it should again be noted that the graph in Figure 4 displays the harmonic participant means, while the effects reported in the text are derived from the estimates of the fixed effects provided by the model.

In the 2×2×2 analysis, non-identical cognates were recognised 3 ms more quickly than identical cognates, but this difference was not significant [$\chi^2(1) = 0.188$, $p = .665$, $BF_{10} = 0.014$], nor was the less than 0.1 ms negative effect of priming [$\chi^2(1) < 0.001$, $p = .999$, $BF_{10} = 0.013$]. Participants in the +IH version responded 27 ms more slowly on average than participants in the –IH version, but this difference was not significant either[3] [$\chi^2(1) = 2.254$, $p = .133$, $BF_{10} = 0.039$]. The interaction

[3] There was a small but significant difference between the two versions with respect to the Dutch LexTALE scores. As the main effect of version was not significant, nor were any of the interactions that included version, we decided not to re-run the analysis with the Dutch LexTALE scores included as a covariate.

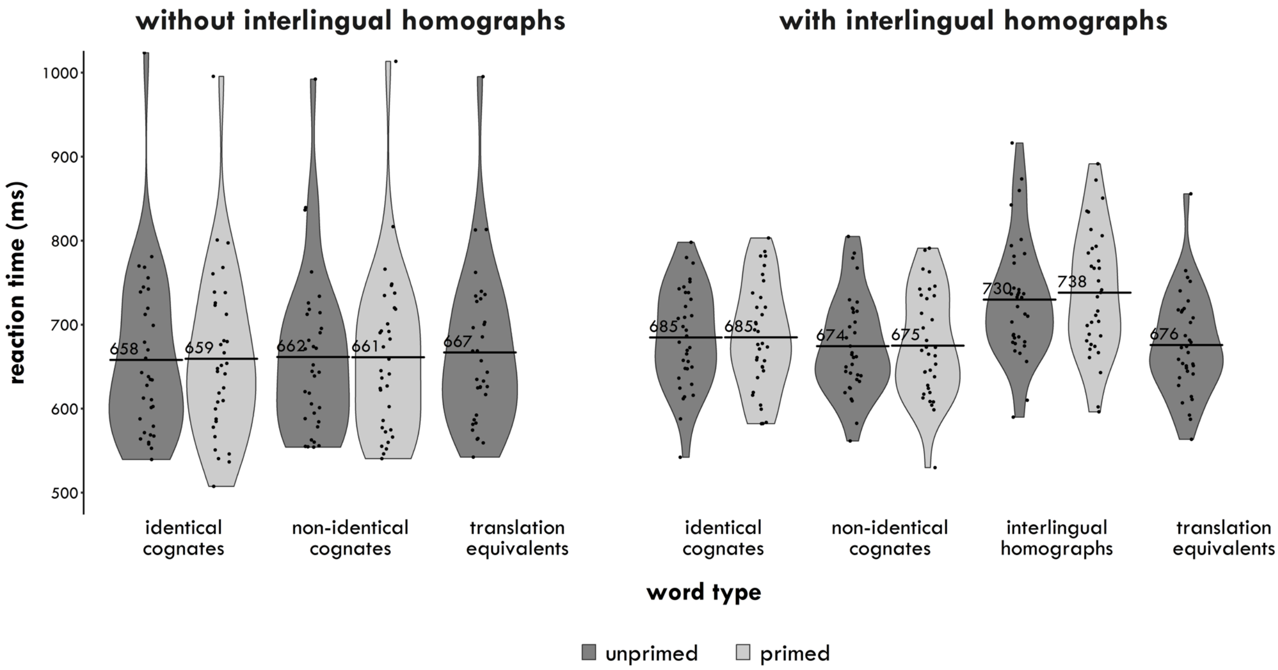

Figure 4: Harmonic participant means of the inverse-transformed lexical decision reaction time data (in milliseconds) by word type (x-axis: identical cognates, non-identical cognates, interlingual homographs, translation equivalents) and priming (dark grey: unprimed; light grey: primed). The right panel pictures the –IH version; the left panel pictures the +IH version. Each point represents a participant. The horizontal line provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

between word type and priming was marginally significant [$\chi^2(1) = 3.117$, $p = .077$, $BF_{10} = 0.060$], but none of the other two-way interactions or the three-way interaction was significant (all ps > .8, all BF10s = 0.013). All Bayes factors provided strong or even very strong evidence for the null hypothesis.

None of the interactions between word type and priming in the four 2×2 analyses were significant, neither in the –IH version nor in the +IH version, and all Bayes factors provided very strong evidence for the null hypothesis (all ps > .3, all BF10s < 0.03).

The effect of priming for the identical cognate in the –IH version was not significant [$\chi^2(1) = 0.023$, $p = .880$, $BF_{10} = 0.026$, $\Delta = -0.9$ ms] nor was the effect of priming for the non-identical cognates [$\chi^2(1) < 0.001$, $p = .988$, $BF_{10} = 0.025$, $\Delta = 0.1$ ms]. Both Bayes factors provided very strong evidence for the null hypothesis. The effect of priming was also not significant for the identical cognates in the +IH version [$\chi^2(1) < 0.001$, $p = .982$, $BF_{10} = 0.025$, $\Delta = -0.1$ ms], nor was it significant for the non-identical cognates [$\chi^2(1) = 0.034$, $p = .855$, $BF_{10} = 0.025$, $\Delta = -1$ ms] or the interlingual homographs [$\chi^2(1) = 1.084$, $p = .298$, $BF_{10} = 0.045$, $\Delta = -9$ ms]. Again, the Bayes factors provided strong or even very strong evidence for the null hypothesis.

In the –IH version, none of the pairwise comparisons were significant [identical cognates vs non-identical cognates: $\chi^2(1) = 0.278$, $p = .598$ $BF_{10} = 0.021$, $\Delta = 4$ ms; identical cognates vs translation equivalents: $\chi^2(1) = 1.699$, $p = .192$, $BF_{10} = 0.042$, $\Delta = 10$ ms; non-identical cognates vs translation equivalents: $\chi^2(1) = 0.801$, $p = .371$, $BF_{10} = 0.027$, $\Delta = 7$ ms]. (Positive effects indicate a reaction time benefit for the first-named word type over the second-named word type.) The Bayes factors also provided strong or even very strong evidence for the null hypothesis that there were no differences between the word types. In the +IH version, in contrast, all of the pairwise comparisons that involved the interlingual homographs were significant [identical cognates vs interlingual homographs: $\chi^2(1) = 28.43, p < .001$ $BF_{10} = 2.7{\cdot}10^4$, $\Delta = 53$ ms; non-identical cognates vs interlingual homographs: $\chi^2(1) = 35.30$, $p < .001$, $BF_{10} = 8.4{\cdot}10^5$, $\Delta = 62$ ms; translation equivalents vs interlingual homographs: $\chi^2(1) = 37.05$, $p < .001$, $BF_{10} = 2.0{\cdot}10^6$, $\Delta = 61$ ms]. The Bayes factors provided decisive evidence for the alternative hypothesis that there were differences between the interlingual homographs and the other word types. None of other pairwise comparisons in the +IH version were significant [identical cognates vs non-identical cognates: $\chi^2(1) = 1.460$, $p = .227$ $BF_{10} = 0.037$, $\Delta = -10$ ms; identical cognates vs translation equivalents: $\chi^2(1) = 1.225$, $p = .268$, $BF_{10} = 0.033$, $\Delta = -8$ ms; non-identical cognates vs translation equivalents: $\chi^2(1) = 0.028$, $p = .867$, $BF_{10} = 0.018$, $\Delta = 1$ ms]. In this case, all Bayes factors provided strong or even very strong evidence for the null hypothesis that there were no differences between the word types.

3.2.3.3 Accuracy

Lexical decision task accuracy is shown in Figure 5. In line with the trimming procedure for the reaction times, any trials with RTs faster than 300ms or slower than 1500ms were removed. Again, we used the bobyqa optimiser instead of the default Nelder-Mead optimiser. Overall, the results for the accuracy data are essentially the same as those for the reaction time data.

In the 2×2×2 analysis, there was a benefit of 0.1% for the non-identical over the identical cognates, but this effect was not significant [$\chi^2(1) = 0.301$, $p = .583$, $BF_{10} = 0.014$], nor was the 0.2% facilitative effect of priming [$\chi^2(1) = 0.835, p = .361, BF_{10} = 0.019$]. The main effect of version was significant [$\chi^2(1) = 4.216, p = .040, BF_{10} = 0.102$], with participants in the +IH version making 0.5% fewer errors. In contrast, all Bayes factors provided moderately to very strong evidence for the null hypothesis. None of the two-way interactions or the three-way interaction was significant and all Bayes factors again provided very strong evidence for the null hypothesis (all *p*s > .2, all BF10s < 0.03). Because the effect of version was significant and there was a small but significant difference in the Dutch LexTALE scores between the two versions, we decided to re-run the 2×2×2 analysis including this variable as a covariate[4]. Note that this should be considered an exploratory analysis as we had not preregistered it. The effect of the covariate Dutch LexTALE scores again was not significant [$\chi^2(1) = 0.325$, $p = .569$, $BF_{10} = 0.015$], but with it included the effect was version was only marginally significant, though still of the same size [$\chi^2(1) = 3.665$, $p < .056$, $BF_{10} = 0.078$]. The priming by version interaction was also marginally significant [$\chi^2(1) = 3.264$, $p = .071$, $BF_{10} = 0.064$]. Both Bayes factors still provided strong or very strong evidence for the null hypothesis. None of the other results had changed.

[4] We had to reduce the random effects structure to an intercepts-only model to achieve convergence when the Dutch LexTALE scores were included as a covariate.

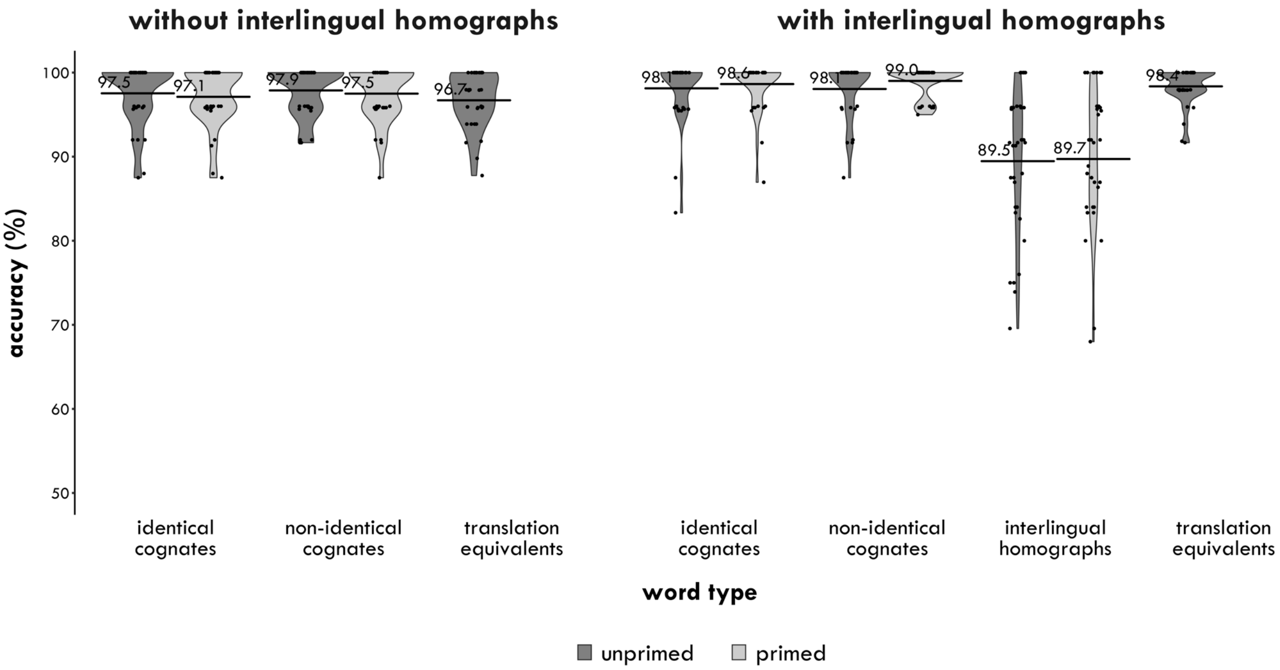

Figure 5: Participant means of the lexical decision accuracy data (in percentage) by word type (x-axis: identical cognates, non-identical cognates, interlingual homographs, translation equivalents) and priming (dark grey: unprimed; light grey: primed). The right panel pictures the –IH version; the left panel pictures the +IH version. Each point represents a participant. The horizontal line provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

None of the interactions between word type and priming in the 2×2 analyses were significant, neither in the –IH version nor in the +IH version, and all Bayes Factors provided strong or even very strong evidence for the null hypothesis (all *p*s > .1, all BF10s < .05).

None of the simple effects of priming were significant in the –IH version [identical cognates: $\chi^2(1) = 0.092$, $p = .762$, $BF_{10} = 0.026$, $\Delta = -0.1\%$; non-identical cognates: $\chi^2(1) = 0.030$, $p = .862$, $BF_{10} = 0.025$, $\Delta = -0.1\%$]. None of the simple effects of priming were significant in the +IH version either [identical cognates: $\chi^2(1) = 0.254$, $p = .615$, $BF_{10} = 0.028$, $\Delta = 0.1\%$; non-identical cognates: $\chi^2(1) = 1.004$, $p = .316$, $BF_{10} = 0.041$, $\Delta = 0.3\%$; interlingual homographs: $\chi^2(1) = 0.033$, $p = .855$, $BF_{10} = 0.025$, $\Delta = 0.2\%$]. All Bayes factors provided strong or even very strong evidence for the null hypothesis.

Finally, none of the pairwise comparisons were significant in the –IH version [identical cognates vs non-identical cognates: $\chi^2(1) = 0.242$, $p = .623$ $BF_{10} = 0.020$, $\Delta = -0.2\%$; identical cognates vs translation equivalents: $\chi^2(1) = 0.474$, $p = .491$, $BF_{10} = 0.022$, $\Delta = 0.4\%$; non-identical cognates vs translation equivalents: $\chi^2(1) = 1.286$, $p = .257$, $BF_{10} = 0.034$, $\Delta = 0.6\%$]. (Positive effects indicate and advantage for the first-named word type over the second-named word type.) The Bayes factors also provided strong or even very strong evidence for the null hypothesis. In the +IH version, however, all of the pairwise comparisons that involved the interlingual homographs were again significant [identical cognates vs interlingual homographs: $\chi^2(1) = 27.95$, $p < .001$, $BF_{10} = 2.1 \cdot 10^4$, $\Delta = 4\%$; non-identical cognates vs interlingual homographs: $\chi^2(1) = 25.46$,

$p < .001$, $BF_{10} = 5.9·10^3$, $\Delta = 4\%$; translation equivalents vs interlingual homographs: $\chi^2(1) = 25.45$, $p < .001$, $BF_{10} = 5.9·10^3$, $\Delta = 4\%$]. In this case, the Bayes factors provided decisive evidence for the alternative hypothesis that there were differences between the interlingual homographs and the other word types. None of other pairwise comparisons in the +IH version were significant [identical cognates vs non-identical cognates: $\chi^2(1) = 0.012$, $p = .912$ $BF_{10} = 0.018$, $\Delta = 0.05\%$; identical cognates vs translation equivalents: $\chi^2(1) < 0.001$, $p = .993$, $BF_{10} = 0.017$, $\Delta = 0.003\%$; non-identical cognates vs translation equivalents: $\chi^2(1) = 0.014$, $p = .905$, $BF_{10} = 0.018$, $\Delta = -0.05\%$]. Again, the Bayes factors provided strong or even very strong evidence for the null hypothesis.

4 GENERAL DISCUSSION

The aim of these two experiments was to determine whether non-identical cognates share an orthographic representation in the bilingual mental lexicon, using the same cross-lingual long-term priming paradigm employed by Poort et al. (2016). Unfortunately, we did not observe the expected priming effect in either of the two experiments, so we are unable to provide a definitive answer to our research question. In Experiment 1, although the main effect of priming was marginally significant, it was considerably smaller than expected based on Poort et al.'s (2016) findings. In contrast to our predictions, Experiment 2 also did not show the predicted priming effect in either the –IH version or the +IH version for either the identical cognates or the interlingual homographs, nor for the non-identical cognates. Indeed, in all cases the Bayes Factors indicated that our data provided strong evidence for the null hypothesis that there was no effect of cross-lingual long-term word-meaning priming.

The most likely explanation for the fact that we did not find evidence for a cross-lingual long-term word-meaning priming effect is that the effect is largely semantic in nature. Based on a series of three experiments, Rodd et al. (2013) concluded that the most likely locus for the long-term word-meaning priming effect is in the connection between a word's form and the meaning that is accessed. When this meaning is accessed, this connection is strengthened so that it becomes easier in future to access that meaning again (at the cost of any other meanings associated with that word form). Rodd et al. (2016) further note that it may also be the case that priming results in changes to the connections between the semantic units that represent the primed meaning, so that this meaning becomes a more stable attractor basin relative to the unprimed meaning. The lexical decision tasks we used, however, did not necessarily require the participants to access the meaning of our stimuli, as lexical decisions can be made on the basis of a general sense of familiarity with a stimulus. If our participants did not wait to make a decision until they had fully accessed the meanings of the words they saw, but instead made their decisions on the basis of a general sense of word-likeness, this could explain why we did not observe priming.

Of course, Poort et al. (2016) also used a lexical decision task and did observe an effect of cross-lingual long-term word-meaning priming. This may be explained by the fact that the participants in their experiment took approximately 200 ms longer to respond than the participants in the current experiments, most likely because Poort et al. (2016) did not limit the time participants could take to respond. Research has shown that the meaning of a word is more likely to play a role in a lexical decision when the decision takes longer, as shown in experiments that varied the characteristics of the non-words. Azuma and Van Orden (1997), for example, found that the number of meanings of a word (and how related those meanings are to each other) affects lexical decision reaction times only when the non-words were word-like and so the lexical decisions were harder to make and took longer. Armstrong and Plaut (2008) further found evidence for a homonymy disadvantage only when the non-words included in their lexical decision task were of medium or high difficulty. It may have been the case that the participants in Poort et al.'s (2016) experiment accessed the meanings of the words more often than the participants in the current experiments and so they observed an effect of priming.

Although we did not find evidence for priming in the current experiments, recent research in the monolingual domain indicates that the long-term word-meaning priming effect is a real effect. Rodd et al. (2013) initially showed that a single encounter with the subordinate meaning of an ambiguous word like "ball" (i.e. using the "dance" meaning as opposed to the "toy" meaning) is sufficient to bias participants' future interpretation of the word "ball" towards that lesser-used meaning. Further research by Rodd et al. (2016) has since shown that participants can remain biased towards that primed meaning for up to 40 minutes. They also found that the priming effect was bigger for younger participants than older participants and for shorter priming delays than for longer delays (Rodd et al., 2016; Exp. 1 and 2). Finally, they demonstrated that repeated exposures to the subordinate meaning over longer time periods more permanently altered participants' meaning preferences: rowers with a higher number of years of experience with rowing were more likely to provide rowing-related meanings for a set of ambiguous words than rowers with less experience (Exp. 3 & 4).

With regards to our predictions for the unprimed trials, the data from Experiment 2 did confirm our hypotheses. First, we did not find evidence for a cognate facilitation effect for either the identical cognates or the non-identical cognates in either version of the experiment, neither based on our traditional frequentist *p*-values nor based on the Bayes Factors. Although a great number of studies has found evidence for a cognate facilitation effect in lexical decision (Cristoffanini et al., 1986; De Groot & Nas, 1991; Dijkstra et al., 1999; Dijkstra et al., 2010; Dijkstra et al., 1998; Font, 2001; Lemhöfer & Dijkstra, 2004; Lemhöfer et al., 2008; Peeters et al., 2013; Sánchez-Casas et al., 1992; Van Hell & Dijkstra, 2002), recent research conducted by Poort and Rodd (2017, May 9) indicates that this effect can disappear when the experiment includes non-target language words like our Dutch words that the participants are required to respond 'no' to. Our data are fully consistent with their findings and further show that the same appears to be true for non-identical cognates.

Second, as predicted, we did find evidence for an interlingual homograph inhibition effect in both the reaction time data and the accuracy data of the +IH version, with participants responding 50-60 ms more slowly and 4% less accurately to the interlingual homographs than to any of the three other word types. This pattern of results was confirmed by both our traditional frequentist *p*-values and our Bayes factors. These findings are in line with research indicating that the interlingual homograph inhibition effect in lexical decision tasks depends on the presence of non-target language words in the stimulus list. These studies have shown that when bilinguals complete a lexical decision task in one of their languages (usually their second language), interlingual homographs are more likely to elicit inhibition compared to control words when the experiment also includes words from the bilingual's other language (usually the bilingual's first language) that require a 'no'-response (De Groot, Delmaar & Lupker, 2000; Dijkstra, De Bruijn, Schriefers & Ten Brinke, 2000; Dijkstra et al., 1998; Von Studnitz & Green, 2002).

In summary, we did not find evidence for a cross-lingual long-term word-meaning priming effect for any of the word types we examined, so we cannot answer our primary research question. The fact that we did find evidence for an interlingual homograph inhibition effect, but no evidence for a cognate facilitation effect, is in line with previous research and confirms the quality of our data. Therefore, we think the most likely reason for the absence of a long-term word-meaning priming effect is that a lexical decision task is not the appropriate task to investigate an effect that is largely semantic in nature. Based on research with monolinguals, we argue that the long-term word-meaning priming effect is real, but future research should investigate whether evidence for cross-lingual long-term word-meaning priming can be found when using a task that is more semantic in nature, or when conditions in the lexical decision task encourage slower responding.

ACKNOWLEDGEMENTS

EDP was supported by a demonstratorship from the Division of Psychology and Language Sciences, University College London. JMR was supported by a grant from the Economic and Social Research Council (grant number ES/K013351/1). The authors would like to thank René Poort for proofreading the experimental materials and Rachael Hulme for providing valuable feedback on this manuscript.

5 REFERENCES

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. *Proceedings of the Cognitive Science Society, 30*, http://www.escholarship.org/uc/item/6rv2c5hh.

Azuma, T., & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36*, 484–504.

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

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

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch lexicon project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 441-458.

Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 614–635.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1283–1296.

Cristoffanini, P., Kirsner, K., & Milech, D. (1986). Bilingual lexical representation: The status of Spanish–English cognates. *The Quarterly Journal of Experimental Psychology, 38*, 367–393.

Damer, E., & Bradley, P. (2014). Prolific Academic. Available at www.prolific.ac.

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: Section A, 53*, 397–428.

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

Dijkstra, T., De Bruijn, E., Schriefers, H., & Ten Brinke, S. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition, 3*, 69–78.

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*, 496–518.

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

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

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*, 51–66.

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 663-679.

Evershed, J., & Hodges, N. (2016). Gorilla: Cauldron (cauldron.sc). Available at www.gorilla.sc.

Font, N. (2001). *Rôle de la langue dans l'accès au lexique chez les bilingues: Influence de la proximité orthographique et sémantique interlangue sur la reconnaissance visuelle de mots.* Unpublished Ph.D. dissertation, Université Paul Valery, Montpellier, France.

Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition, 5*, 225–239.

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford, England: Oxford University Press.

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & De Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research, 1068*, 170–183.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*, 643–650.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology, 68*, 1665–1692.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 952.

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

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition, 32*, 533–550.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 12–31.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 381–390.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language, 68*, 315–332.

Poort, E. D., & Rodd, J. M. (2017, January 24). Recent experience with identical and non-identical cognates and identical interlingual homographs on visual word recognition. Available at http://osf.io/33r86.

Poort, E. D., & Rodd, J. M. (2017, May 9). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition. *OSF Preprints*, http://osf.io/rvmz3.

Poort, E. D., Warren, J. E., & Rodd, J. M. (2016). Recent experience with cognates and interlingual homographs in one language affects subsequent processing in another language. *Bilingualism: Language and Cognition, 19*, 206–212.

R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2). Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org/.

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

Rodd, J. M. (2000). *Semantic representation and lexical competition: Evidence from ambiguity.* Unpublished Ph.D. dissertation, Cambridge University, Cambridge, UK.

Rodd, J. M., Cai, Z. G., Betts, H. N., Hanby, B., Hutchinson, C., & Adler, A. (2016). The impact of recent and long-term experience on access to word meanings: Evidence from large-scale internet-based experiments. *Journal of Memory and Language, 87*, 16-37.

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68*, 180–198.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*, 245–266.

Sánchez-Casas, R. M., García-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology, 4*, 293–310.

Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1155–1178.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*, 197–212.

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes, 22*, 106–129.

The Mathworks Inc. (2012). Matlab (Version R2012a). Natick, Massachusetts, United States.
Tokowicz, N., Kroll, J. F., De Groot, A. M. B., & 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*, 435–451.
Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64*, 88–107.
Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science, 20*, 923–927.
Van Casteren, M., & Davis, M. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods, 39*, 973–978.
Van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica, 128*, 431–451.
Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*, 780–789.
Van Heuven, W. J. B., Schriefers, H. J., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*, 2706–2716.
Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*, 181-198.
Von Studnitz, R. E., & Green, D. W. (2002). Interlingual homograph interference in German–English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition, 5*, 1–23.
Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review, 14*, 779–804.
Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979.