The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition

Eva D. Poort[*], Jennifer M. Rodd

Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

A R T I C L E   I N F O

A B S T R A C T

Cognates share their form and meaning across languages: "winter" in English means the same as "winter" in Dutch. Research has shown that bilinguals process cognates more quickly than words that exist in one language only (e.g. "ant" in English). This finding is taken as strong evidence for the claim that bilinguals have one integrated lexicon and that lexical access is language non-selective. Two English lexical decision experiments with Dutch–English bilinguals investigated whether the cognate facilitation effect is influenced by stimulus list composition. In Experiment 1, the 'standard' version, which included only cognates, English control words and regular non-words, showed significant cognate facilitation (31 ms). In contrast, the 'mixed' version, which also included interlingual homographs, pseudohomophones (instead of regular non-words) and Dutch-only words, showed a significantly different profile: a non-significant disadvantage for the cognates (8 ms). Experiment 2 examined the specific impact of these three additional stimuli types and found that only the inclusion of Dutch words significantly reduced the cognate facilitation effect. Additional exploratory analyses revealed that, when the preceding trial was a Dutch word, cognates were recognised up to 50 ms more slowly than English controls. We suggest that when participants must respond 'no' to non-target language words, competition arises between the 'yes'- and 'no'-responses associated with the two interpretations of a cognate, which (partially) cancels out the facilitation that is a result of the cognate's shared form and meaning. We conclude that the cognate facilitation effect is a real effect that originates in the lexicon, but that cognates can be subject to competition effects outside the lexicon.

PsychINFO classification code

2340 (Cognitive Processes)

1. Introduction

One of the most researched phenomena within the field of bilingualism is the *cognate facilitation effect*. Cognates are words that exist in an identical (or near identical) form in more than one language and carry the same meaning, like "winter" in Dutch and English. Many studies have shown that bilinguals process these words more quickly than words that exist in one language only (i.e. that do not share their form with their translation), like "ant" in English and it's translation "mier" in Dutch. This effect is at the heart of the Bilingual Interactive Activation plus (BIA+) model (Dijkstra & Van Heuven, 2002), the most commonly used model of the bilingual mental lexicon, and is taken as strong evidence for the claim that all the languages a bilingual speaks are stored in a single, integrated lexicon and that access to this lexicon is language non-selective.

The cognate facilitation effect has most commonly been observed in visual lexical decision experiments when the target words are presented in isolation (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; Font, 2001; Lemhöfer et al., 2008; Lemhöfer & Dijkstra, 2004; Peeters, Dijkstra, & Grainger, 2013; Sánchez-Casas, García-Albea, & Davis, 1992; Van Hell & Dijkstra, 2002), but also when they are embedded in sentences, although in the latter case the effect is often smaller (Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009; Van Hell & De Groot, 2008). In addition, the cognate facilitation effect has been observed in word production: bilinguals are faster to name pictures of cognates (e.g. Costa, Caramazza, & Sebastian-Galles, 2000) and to read aloud cognate words (e.g. Schwartz, Kroll, & Diaz, 2007). It has been demonstrated most often in experiments in the bilinguals' second language, but it has also been observed in native-language only experiments (Van Hell & Dijkstra, 2002). Finally, the size of the effect is greater for

* Corresponding author at: 26 Bedford Way, London WC1H 0AP, United Kingdom.
E-mail addresses: eva.poort.12@ucl.ac.uk (E.D. Poort), j.rodd@ucl.ac.uk (J.M. Rodd).

cognates that are identical compared to non-identical cognates (e.g. "meloen" in Dutch and "melon" in English; Comesaña et al., 2015; Dijkstra et al., 2010; Duyck et al., 2007; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011) and for cognates that exist in three languages compared to cognates that exist in only two languages (Lemhöfer, Dijkstra, & Michel, 2004; Van Hell & Dijkstra, 2002). This wealth of research suggests that the cognate facilitation effect is very robust and universal.

Research with interlingual homographs paints a more nuanced picture. Interlingual homographs are words that, like cognates, share their form in more than one language, but carry a different meaning, such that "angel" means "insect's sting" in Dutch. Also like cognates, bilinguals process interlingual homographs differently than single-language control words. In contrast to cognates, however, interlingual homographs are often processed more *slowly* than control words. This *interlingual homograph inhibition effect* has been reported in experiments examining bilinguals' visual word recognition (Dijkstra et al., 1998; Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004; Van Heuven, Schriefers, Dijkstra, & Hagoort, 2008), auditory word recognition (Lagrou, Hartsuiker, & Duyck, 2011; Schulpen, Dijkstra, Schriefers, & Hasper, 2003) and word production (Jared & Szucs, 2002; Smits, Martensen, Dijkstra, & Sandra, 2006). As with the cognate effect, this effect forms an important part of the BIA+ and is usually interpreted as evidence that both of the languages a bilingual speaks are stored in one integrated lexicon and that lexical access is language non-selective.

Importantly, most experiments that have focused on the interlingual homograph inhibition effect used single-language visual lexical decision tasks, during which participants have to decide whether letter strings are words in a specific language (usually the bilingual's second language). Further research has shown that when using such tasks, interlingual homographs are more likely to be recognised more slowly than control words when the experiment also includes words from the bilingual's other language (the non-target language, usually the bilingual's first language) that require a 'no'-response (De Groot, Delmaar, & Lupker, 2000; Dijkstra et al., 1998; Dijkstra, De Bruijn, Schriefers, & Ten Brinke, 2000; Von Studnitz & Green, 2002). For example, in Experiment 1 of their study, Dijkstra et al. (1998) asked Dutch–English bilinguals to complete an English lexical decision task which included cognates, interlingual homographs, English controls and regular non-words, but no words from the bilinguals' native language, Dutch. In this experiment, they observed no significant difference in average reaction times for the interlingual homographs and the English controls (cf. Van Heuven et al., 2008, who did find evidence for an inhibition effect under the same conditions). In Experiment 2, the English lexical decision task also included a number of Dutch words which the participants were told required a 'no'-response. This time, the analysis did reveal a significant difference between the interlingual homographs and the English (but not the Dutch) control words: the participants were slower to respond to the interlingual homographs than the English controls.

This pattern of results is interpreted within the framework of the BIA+ model (Dijkstra & Van Heuven, 2002) by assuming that there are two points at which language conflict can arise for an interlingual homograph. According to this model, there are two components to the (bilingual) word recognition system: the word identification system and the task/decision system (inspired by Green's, 1998 Inhibitory Control model). In the word identification system, the visual input of a string of letters first activates letter features, which in turn activate the letters that contain these features and inhibit those that do not. The activated letters then activate words that contain those letters in both languages the bilingual speaks. These activated words inhibit each other through lateral inhibition, irrespective of the language to which they belong. The task/decision system continuously reads out the activation in the word identification system and weighs the different levels of activation to arrive at a response relevant to the task at hand. In this system,

stimulus-based conflict can arise in the lexicon due to competition (lateral inhibition) between the two (orthographic) representations of the interlingual homograph (Van Heuven et al., 2008). *Response-based* conflict arises outside the lexicon at the level of decision making (i.e. in the task/decision system) and is the result of one of those two lexical representations being linked to the 'yes'-response, while the other is linked to the 'no'-response (Van Heuven et al., 2008).

In short, in Experiment 1 of the Dijkstra et al. (1998) study, the interlingual homographs most likely only elicited stimulus-based language conflict, which it appears does not always translate to an observable effect in lexical decision reaction times. In contrast, in Experiment 2 the interlingual homographs elicited both stimulus-based and response-based conflict, as the participants linked the Dutch reading of the interlingual homographs to the 'no'-response, due to the presence of the Dutch words that required a 'no'-response. This response-based conflict resulted in a clear inhibition effect. In other words, in Experiment 1, the participants could base their decisions on a sense of familiarity with each stimulus (essentially reinterpreting the instructions as 'Is this a word in *general*?'), whereas in Experiment 2, they were forced to be very specific (adhering to the instructions 'Is this a word in *English*?').

Recent work indicates that the cognate facilitation effect may also be influenced by the composition of the experiment's stimulus list. Poort, Warren, and Rodd (2016) designed an experiment to investigate whether recent experience with a cognate or interlingual homograph in one's native language (e.g. Dutch) affects subsequent processing of those words in one's second language (e.g. English). They asked their participants to read sentences in Dutch that contained cognates or interlingual homographs. After an unrelated filler task that lasted approximately 16 minutes, the participants completed a lexical decision task in English. Some of the words included in the lexical decision task were the same cognates and interlingual homographs the participants had seen before in Dutch. The analysis revealed that their recent experience with these words in Dutch affected how quickly they were able to recognise them in English and, crucially, that this depended on whether the Dutch and English meaning were shared: recent experience with a cognate in Dutch was shown to speed up recognition in English (by 28 ms), while recent experience with an interlingual homograph slowed the participants down (by 49 ms). In contrast to the studies mentioned previously, however, they found that the (unprimed) cognates in their experiment were recognised 35 ms more *slowly* than the English controls (see panel A of Figure 1 of their article), although a subsequent re-analysis of their data revealed this difference to be non-significant.

Notably, in contrast to those previous lexical decision experiments, Poort et al. (2016) also included some non-target language (Dutch) words (e.g. "vijand", meaning "enemy") in their English lexical decision task as non-English words which required a 'no'-response. They furthermore included both cognates and interlingual homographs in the same experiment and used pseudohomophones—non-words designed to sound like existing words, like "mistaik"—instead of 'regular' non-words—non-words derived from existing words by changing one or two letters, like "grousp". As far as we are aware, no research has systematically investigated whether the cognate facilitation effect, like the interlingual homograph inhibition effect, could be affected by the composition of the stimulus list. However, given the significance of the cognate facilitation effect to theories of the bilingual lexicon, it is important to determine whether the unusual composition of Poort et al.'s (2016) stimulus list is the reason behind this apparent inconsistency with the studies mentioned previously.

Indeed, there are good reasons to suspect that any (or all) of the 'extra' stimuli types Poort et al. (2016) included—the interlingual homographs, pseudohomophones and Dutch words—might have affected the size and/or direction of the cognate effect. As discussed previously, the presence of non-target language words in a single-language lexical decision has notable consequences for how bilinguals

process interlingual homographs. As Poort et al. (2016) also included such items in their experiment, participants in their study may have adopted a different response strategy (i.e. constructed a different task schema) compared with participants in the 'standard' experiments, which did not include non-target language words (e.g. Dijkstra et al., 1999). Although according to the BIA+ model cognates are not subject to stimulus-based competition, they are, like interlingual homographs, ambiguous with respect to their language membership. As such, in a task that includes non-target language words, participants will have to determine whether the cognates are words in English specifically, instead of in general. The BIA+ does not exclude the possibility that including non-target language words could result in competition between the 'yes'-response linked to one interpretation of the cognate and the 'no'-response linked to the other.

Previous research with young second-language learners has also found that including interlingual homographs in a single-language lexical decision task can result in a disadvantage for cognates compared to control words. Brenders, Van Hell, and Dijkstra (2011) found that 10-year-olds, 12-year-olds and 14-year-olds who spoke Dutch as their native language and had 5 months, 3 years and 5 years of experience with English, respectively, already showed a cognate facilitation effect in an English lexical decision task (Exp. 1), though not in a Dutch lexical decision task (Exp. 2). In an English lexical decision task that included both cognates and interlingual homographs (Exp. 3), however, the participants responded more slowly to the cognates than to the English controls. (Indeed, the disadvantage for the cognates was of about the same size as the disadvantage for the interlingual homographs.) As Brenders et al. (2011) suggest, it is possible that the interlingual homographs drew the children's attention to the fact that the cognates were also ambiguous with respect to their language membership and may have prompted them to link the Dutch interpretation of the cognates to the 'no'-response, resulting in response competition. As such, it could also have been the presence of the interlingual homographs in Poort et al.'s (2016) experiment that was responsible for the non-significant cognate disadvantage they observed.

Finally, in the monolingual domain, research has shown that semantically ambiguous words with many senses like "twist"—which are, essentially, the monolingual equivalent of cognates—are recognised more quickly than semantically unambiguous words like "dance" (e.g. Rodd, Gaskell, & Marslen-Wilson, 2002). Rodd, Gaskell, and Marslen-Wilson (2004) used a distributed connectionist network to model these effects of semantic ambiguity on word recognition and found that their network was indeed more stable for words with many senses, but only early in the process of word recognition. This many-senses benefit reversed during the later stages of word recognition and became a benefit for words with few senses. It could have been the case that Poort et al.'s (2016) decision to use pseudohomophones—which tend to slow participants down—instead of 'regular' non-words similarly affected the processing of their cognates.

To determine whether the cognate facilitation effect is indeed influenced by stimulus list composition, we set up two online English lexical decision experiments. The aim of Experiment 1 was to determine whether Poort et al.'s (2016) unexpected findings were indeed due to differences in the composition of their stimulus list (and not some other factor, such as the priming manipulation or differences in the demographics of their participants or the characteristics of their stimuli). Having confirmed, based on the results of Experiment 1, that stimulus list composition does influence the cognate facilitation effect, Experiment 2 investigated which of the three additional types of stimuli included by Poort et al. (2016) can significantly influence the direction and/or magnitude of the cognate effect. The experiments were conducted online, in order to recruit highly proficient bilinguals immersed in a native-language environment, which is a similar population as the populations sampled in previous studies.

In Experiment 1, one version of the experiment was designed to replicate the experimental conditions of a 'standard' cognate effect experiment (e.g. Dijkstra et al., 1999) and included identical cognates, English controls and 'regular' non-words. The other version was designed to replicate the experimental conditions of Poort et al.'s (2016) experiment, but without the priming manipulation. It included the same cognates and English controls, but also identical interlingual homographs. The regular non-words were replaced with English-sounding pseudohomophones and some Dutch-only words. We use the term 'standard version' to refer to the first version and 'mixed version' to refer to the second. If the differences between Poort et al.'s (2016) findings and the findings reported in the literature do indeed reflect a difference in stimulus list composition, we would expect to see a different pattern of reaction times for the cognates and English controls in the two versions. In accordance with the literature, we predict to find a significant cognate facilitation effect in the standard version, but, based on Poort et al.'s (2016) findings, we expect to find no advantage (or even a disadvantage) for the cognates in the mixed version.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Forty-one Dutch–English bilinguals were recruited through Prolific Academic and social media and personal contacts resident in the Netherlands and Belgium. The participants gave informed consent and were paid for their participation in the experiment. They were all living in the Netherlands or Belgium at the time of the experiment and were native speakers of Dutch (or Flemish) and fluent speakers of English. The data from one participant who completed the mixed version were excluded from the analysis, as this participant's overall accuracy (83.0%) for the target items (cognates and English controls) was more than three standard deviations below the version mean ($M = 95.7\%$, $SD = 3.8\%$). The remaining 40 participants, 20 in each version (26 male; $M_{age} = 26.23$ years, $SD_{age} = 6.7$ years) had an average of 18.8 years of experience with English ($SD = 6.9$). The participants rated their proficiency as 9.8 out of 10 in Dutch and 8.8 in English. These ratings were confirmed by their high LexTALE scores in both languages, which a paired t-test showed were slightly higher in Dutch (Dutch: $M = 91.2\%$, $SD = 6.2\%$; English: $M = 86.1\%$, $SD = 8.6\%$; $p < .001$). The LexTALE (Lemhöfer & Broersma, 2012) is a simple test of vocabulary knowledge that provides a fair indication of a participant's general language proficiency. There were no significant differences between the versions on any of the variables reported here (as shown by chi-square tests and independent-samples Welch's t-tests; all $ps > .09$).

2.1.2. Materials

Table 1 lists the number of items of each word type included in the two versions of the experiment. The full set of stimuli used can be found in Supplementary materials 1.

2.1.2.1. Words. A large number of cognates, English controls and interlingual homographs were selected from Dijkstra et al. (2010), Poort et al. (2016) and Tokowicz, Kroll, De Groot, and Van Hell (2002). Some additional interlingual homographs were identified by selecting orthographically identical entries in the SUBTLEX-US and SUBTLEX-NL databases (Brysbaert & New, 2009; Keuleers, Brysbaert, & New, 2010, respectively) that had dissimilar meanings. 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. The interlingual homographs were more difficult to find, so six interlingual homographs with frequencies below 2 in Dutch (e.g. "gulp") and three with frequencies below 2 in English (e.g. "slang") were included, as they were considered to be well-known to the participants despite their low frequency.

Of this initial set, we obtained spelling, pronunciation and meaning

Table 1
Overview of the types and numbers of stimuli included in each version of Experiment 1 and 2, as well as durations of the different tasks in mm:ss. *N* is the number of participants included in the analysis for that version.

| | *N* | Number of items per stimulus type | | | | | | Task duration | | |
		Cognates	English controls	Interlingual homographs	Regular non-words	Pseudohomophones	Dutch words	Main task	English LexTALE	Dutch LexTALE
Experiment 1										
Standard	20	56	56	0	112	0	0	12:29	2:40	2:28
Mixed	20	56	56	56	0	140	28	18:45	2:48	2:29
Experiment 2										
Standard	21	56	56	0	112	0	0	13:04	2:53	2:24
Mixed	20	56	56	56	0	140	28	18:45	2:45	2:35
+Dutch words	20	56	56	0	94	0	18	13:03	2:33	2:23
+Interlingual homographs	20	56	56	56	168	0	0	18:57	3:10	2:29
+Pseudohomophones	19	56	56	0	0	112	0	12:48	2:46	2:28

Table 2
Means (and standard deviations) for all key matching variables, similarity ratings and raw word frequency. SUBTLEX-WF refers to the SUBTLEX raw word frequency in occurrences per million (see Keuleers et al., 2010 for Dutch and Brysbaert & New, 2009 for English); LG10-WF refers to the SUBTLEX log-transformed word frequency (log10[raw frequency + 1]); length refers to the number of letters in a word; OLD20 refers to Yarkoni et al.'s (2008) measure of orthographic complexity of a word expressed by its mean orthographic Levenshtein distance to its 20 closest neighbours. The Dutch characteristics are listed for completeness only; the items were not matched on these characteristics. Meaning, spelling and pronunciation similarity ratings were obtained through pre-tests and were given on a scale of 1 (not at all similar) to 7 (almost identical).

| | Dutch characteristics | | | | English characteristics | | | | Similarity ratings | | |
	SUBTLEX-WF	LG10-WF	Length	OLD20	SUBTLEX-WF	LG10-WF	Length	OLD20	Meaning	Spelling	Pronunciation
Cognates	36.0 (56.9)	2.89 (0.49)	4.54 (1.08)	1.61 (0.41)	41.5 (54.8)	3.6 (0.48)	4.54 (1.08)	1.63 (0.35)	6.83 (0.22)	7.00 (0.01)	5.89 (0.67)
Interlingual homographs	52.4 (115)	2.74 (0.71)	4.23 (0.93)	1.31 (0.34)	52.2 (102)	2.96 (0.65)	4.23 (0.93)	1.43 (0.32)	1.16 (0.30)	7.00 (0.01)	5.45 (0.80)
English controls	–	–	–	–	29.4 (25.8)	3.01 (0.40)	4.46 (0.93)	1.59 (0.31)	6.86 (0.18)	1.12 (0.23)	1.10 (0.20)

similarity ratings for 65 cognates, 80 interlingual homographs and 80 English controls (and their Dutch translations) across two different pre-tests using a total of 90 Dutch–English bilinguals who did not take part in the main experiment. Each item received ratings from at least 11 participants. Cognates and English controls with meaning similarity ratings below 6 on our 7-point scale were discarded, as were interlingual homographs with ratings above 2.5. English controls with spelling similarity ratings higher than 2 were also discarded. The software package Match (Van Casteren & Davis, 2007) was then used to select the 56 best-matching cognates, interlingual homographs and English controls, where matching was based on log-transformed English word frequency (weight: 1.5), the number of letters of the English word (weight: 1.0) and orthographic complexity of the English word 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 the spelling, pronunciation and meaning similarity ratings obtained from the pre-tests. Only the cognates and English controls were included in the standard version for a total of 112 words; the mixed version also included the 56 interlingual homographs for a total of 168 words in this version.

Independent-samples Welch's *t*-tests showed that he differences between the cognates and English controls on the matching criteria were not significant (all *p*s > .5). The cognates and English controls were significantly more orthographically complex than the interlingual homographs as evidenced by their higher average OLD20 (*p* = .002, *p* = .008, respectively). The cognates and English controls did not significantly differ from the interlingual homographs on any of the other measures (all *p*s > .1). An analysis of the meaning similarity ratings confirmed that the cognates and English controls both differed significantly from the interlingual homographs, as intended (both *p*s < .001), but not from each other (*p* > .4). The cognates and interlingual homographs were significantly different from the English controls in terms of spelling similarity ratings (both *p*s < .001), but not from each other (*p* > .7). In terms of pronunciation similarity, all three word types were significantly different to each other (*p* < .005).

2.1.2.2. Non-words. Each version included the same number of non-words as words. In the mixed version, the 168 non-words comprised 140 English-sounding pseudohomophones selected from Rodd (2000) and the ARC non-word and pseudohomophone database (Rastle, Harrington, & Coltheart, 2002), as well as 28 Dutch words (e.g. "vijand") of a similar frequency to the target items, selected pseudo-randomly from the SUBTLEX-NL database. In the standard version, the 112 non-words were pronounceable nonsense letter strings generated using the software package Wuggy (Keuleers & Brysbaert, 2010), which creates non-words from words while respecting their subsyllabic structure and the phonotactic contraints of the target language. The 112 words given to Wuggy were of a similar frequency as the target items and had been pseudo-randomly selected from the SUBTLEX-US database. In both versions, the non-words were matched word-for-word to a target in terms of number of letters.

2.1.3. Design and procedure

The experiment comprised three separate tasks: (1) the English lexical decision task, (2) the English version of the LexTALE and (3) the Dutch version of the LexTALE. At the start of the experiment, the participants completed a self-report language background survey in Dutch. Participants were randomly assigned to one of the two versions of the experiment. The experiment was created using version 15 of the Qualtrics Reaction Time Engine (QRTE; Barnhoorn, Haasnoot,

---

[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).

Bocanegra, & Van Steenbergen, 2014). Due to Qualtrics updating their Survey Engine, QRTE version 15 stopped working after only 18 participants had been tested (8 in the standard version and 10 in the mixed version). The remaining 23 participants were tested using QRTE version 16 (12 in the standard version, 11 in the mixed version).

During the English lexical decision task, the participants saw all 224 (standard version) or 336 (mixed version) stimuli and were asked to indicate, by means of a button press, as quickly and accurately as possible, whether the letter string they saw was a *real* English word or not (emphasis was also present in the instructions). Participants in the mixed version were explicitly instructed to respond 'no' to items that were words in another language (i.e. the Dutch words). A practice block of 16 or 24 letter strings was followed by 8 blocks of 28 or 42 experimental stimuli for the standard and mixed versions, respectively. The order of the items within blocks was randomised for each participant, as was the order of the blocks. Four or six fillers were presented at the beginning of each block, with a 10-second break after each block. All items remained on screen until the participant responded, or until 2000 ms passed. The inter-trial interval was 500 ms.

2.2. Results

Only the cognates and English controls were initially analysed, as the other stimuli (i.e. the interlingual homographs, regular non-words, pseudohomophones and Dutch words) differed between the two versions and were considered fillers. Two items (the English controls "flu" and "treaty") were excluded from the analysis, as the overall percentages correct for those items (70.0%, 80.0%) were more than three standard deviations below the mean of all experimental items ($M = 96.6\%$, $SD = 4.9\%$).

All analyses were carried out in R (version 3.2.1; R Core Team, 2015) using the lme4 package (version 1-1.10; Bates, Maechler, Bolker, & Walker, 2015), following Barr, Levy, Scheepers, and Tily's (2013) guidelines for confirmatory hypothesis testing and using likelihood ratio tests to determine significance (comparing against an α of .05 unless otherwise stated). Two fixed factors were included in the main 2×2 analysis: word type (2 within-participant/between-items levels: cognate, English control) and version (2 between-participants/within-items levels: standard, mixed). Effect coding was used to specify contrasts for both factors. The simple effects analyses, looking at the effect of word type within each version, included only one fixed factor, word type (2 within-participant/between-items levels: cognate, English control). Detailed results of all analyses for Experiment 1 can be found in Supplementary materials 2.

2.2.1. Reaction times

Lexical decision reaction times are shown in panel A of Fig. 1. Reaction times (RTs) for incorrect trials were discarded (3.0% of the data), as were RTs more than three standard deviations above or below a participant's mean RT for all experimental items (2.3% of the remaining data). All remaining RTs were greater than 300 ms. The maximal model converged for the 2×2 and included a correlated random intercept and slope for word type by participants and a correlated random intercept and slope for version by items. An inspection of a histogram of the residuals and a predicted-vs-residuals plot showed that the assumptions of homoscedasticity and normality were violated. To remedy this, the RTs were inverse transformed and the maximal model refitted to the inverse-transformed RTs (inverse-transformed RT = 1000/raw RT; the inverse-transform achieved a better distribution of the residuals than the log-transform). Finally, it should be noted that the graph in panel A of Fig. 1 displays the harmonic participant means, while the effects (and means) reported in the text were derived from the estimates of the fixed effects provided by the model.

The main effect of word type was marginally significant [$\chi^2(1) = 2.789$, $p = .095$], with cognates being recognised on average 12 ms more quickly than English controls. The main effect of version was also

marginally significant [$\chi^2(1) = 3.347$, $p = .067$], with participants in the mixed version responding on average 38 ms more slowly than participants in the standard version. Crucially, the interaction between word type and version was significant [$\chi^2(1) = 15.10$, $p < .001$]. The simple effects analyses revealed that the 31 ms cognate *facilitation* effect observed in the standard version was significant [$\chi^2(1) = 13.52$, $p < .001$], though the 8 ms *disadvantage* for cognates in the mixed version was not [$\chi^2(1) = 0.744$, $p = .388$]. For the simple effects analyses, the maximal model also converged and included a random intercept and random slope for word type by participants and a random intercept by items.

In addition, although it was not the primary focus of our experiment, for the mixed version we also compared the English controls to the interlingual homographs. The participant who had been excluded for the main analysis was included in this analysis, as their overall percentage correct (81.3%) for the target items included in this analysis (interlingual homographs and English controls) was within three standard deviations of the mean ($M = 92.2\%$, $SD = 5.1\%$). Three items with an average accuracy more than three standard deviations below their word type's mean were excluded. These were the interlingual homograph "hoop" (33.3%; $M = 88.7\%$, $SD = 14.8\%$) and the English controls "flu" and "treaty" (71.4%, 66.7%; $M = 95.7\%$, $SD = 7.4\%$). Since there was a significant difference with respect to the English OLD20 measure between the English controls and the interlingual homographs, we included this variable in the analysis as a covariate (though it was not significant, $p = .790$). The maximal model with a random intercept and random slope for word type by participants and a random intercept by items converged and revealed a significant inhibition effect of 43 ms for the interlingual homographs ($M = 724$ ms) compared to the English controls ($M = 681$ ms) [$\chi^2(1) = 14.05$, $p < .001$].
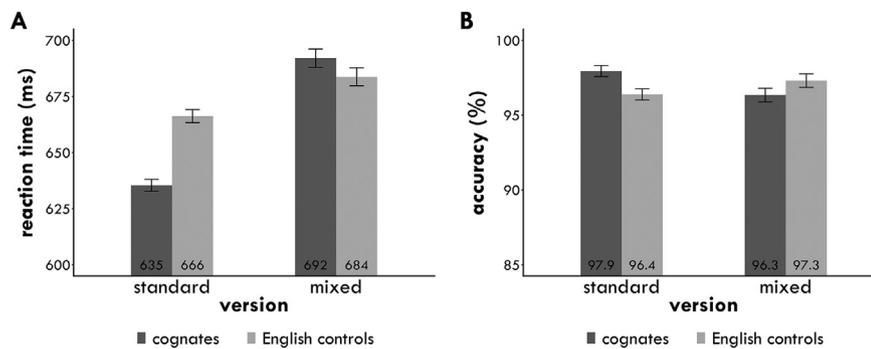
2.2.2. Accuracy

Task accuracy is shown in panel B of Fig. 1. The maximal model with a random intercept and slope for word type by participants and a random intercept and slope for version by items converged when the bobyqa optimiser was used. This model revealed that the main effect of word type was not significant [$\chi^2(1) = 0.157$, $p = .692$], nor was the main effect of version [$\chi^2(1) = 0.088$, $p = .767$]. The interaction between word type and version was marginally significant [$\chi^2(1) = 3.231$, $p = .072$]. The simple effects analyses showed that the small cognate advantage in the standard version was not significant [$\chi^2(1) = 1.415$, $p = .234$], nor was the slight cognate disadvantage in the mixed version [$\chi^2(1) = 0.651$, $p = .420$].

2.2.3. Exploratory analysis: Effect of the preceding trial

For the reaction time data of the mixed version, we also investigated whether the stimulus type of the preceding trial (cognate, English control, interlingual homograph, pseudohomophone or Dutch word) interacted with the word type of the current trial (cognate or English control). From the total number of trials included for that version in the confirmatory analysis, we selected only current trials for which the preceding trial had received a correct response (93.1%). Note that this was a post-hoc exploratory analysis that was carried out in response to effects observed in Experiment 2 and according to the analysis plan for the confirmatory analyses for that experiment.

We first conducted five simple effects analyses, to determine whether or not there was evidence for a cognate facilitation effect for each of the five preceding trial stimulus types. These models included only one (effect-coded) fixed factor: word type of the current trial (2 within-participant/between-items levels: cognate, English control). The maximal random effects structure included a random intercept and random slope for word type by participants and a random intercept by items. The *p*-values for these five analyses were compared against a Bonferroni-corrected α of .01. We also conducted ten 2×2 analyses, focusing on two of the five preceding trial stimulus types at a time, to

Fig. 1. Experiment 1. **A** Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) and **B** participant means of task accuracy (percentages correct). Both panels display the data by version (standard, mixed; x-axis) and word type (cognates, dark grey; English controls, light grey). Error bars represent the standard error of the mean adjusted for a within-participants design, using version means to calculate the adjustment factor (Cousineau, 2005).

determine whether the influence of each of the five types on the cognate facilitation effect was significantly different to that of the others. These models included two (effect-coded) fixed factors: word type of the current trial (2 within-participants/between-items levels: cognate, English control) and stimulus type of the preceding trial (using only 2 of the 5 within-participants/within-items levels: cognate, English control, interlingual homograph, pseudohomophone or Dutch word). The maximal random effects structure included a random intercept and random slopes for all fixed effects by participants and a random intercept only by items. Although stimulus type of the preceding trial was a within-items factor, we did not include a by-items random slope for this factor as across participants not every item was necessarily preceded by each of the five stimulus types. Correlations between the by-participants random effects were removed, as the models did not converge when the random effects were allowed to correlate. Finally, for these analyses, we were only interested in the interactions, so only those are reported and the p-values were compared against a Bonferroni-corrected α of .005.

The simple effects analyses revealed that having seen a Dutch word on the preceding trial resulted in a numerically large disadvantage of 50 ms for the cognates that was only significant at an uncorrected α of .05 [$\chi^2(1) = 4.864$, $p = .027$]. Importantly, as can be seen in Fig. 2, this effect appears to be due to the participants responding more slowly to the cognates and not more quickly to the English controls. The interlingual homographs and pseudohomophones elicited small, but non-significant cognate disadvantages of 12 ms and 4 ms, respectively [interlingual homographs: $\chi^2(1) = 0.529$, $p = .467$; pseudohomophones: $\chi^2(1) = 0.144$, $p = .705$], while the cognates and English controls
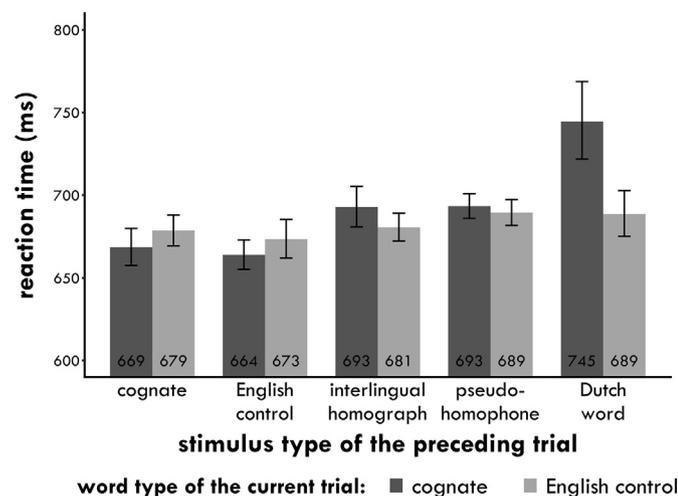
elicited small, but non-significant facilitation effects of 7 ms and 9 ms, respectively [cognates: $\chi^2(1) = 0.174$, $p = .677$; English controls: $\chi^2(1) = 0.307$, $p = .580$; respectively]. The 2×2 analyses further showed that the effect the Dutch words had on the size and direction of the cognate effect was significantly or marginally significantly different to that of all of the other stimulus types, but again only at an uncorrected α of .05 [English controls: $\chi^2(1) = 5.516$, $p = .019$; cognates: $\chi^2(1) = 6.427$, $p = .011$; interlingual homographs: $\chi^2(1) = 2.850$, $p = .091$; pseudohomophones: $\chi^2(1) = 4.005$, $p = .045$]. None of the other interactions were significant (all $ps > .3$).

2.3. Discussion

The results of Experiment 1 demonstrate that the cognate facilitation effect is indeed influenced by stimulus list composition. In the standard version of Experiment 1, we found a significant cognate facilitation effect of 31 ms, while cognates in the mixed version were recognised 8 ms more slowly than the English controls. Although this latter effect was not significant, the interaction between word type and version was highly significant, suggesting that the types of other stimuli included in the experiment had a reliable effect on the direction of the cognate effect. Before we discuss these findings in detail, it should be noted that our participants completed a language background questionnaire in Dutch at the start of the experiment, which may have increased the activation of their Dutch lexicon and made them operate in a more bilingual mode. This could have increased the salience of the Dutch items in the mixed version, but may also have increased the size of the cognate effect in general. As this factor was kept constant across the different versions of the experiment, we think it unlikely that this could have affected our results.

Notably, the cognate facilitation effect we observed in the standard version mirrors the effect described in the literature (e.g. Cristoffanini et al., 1986; De Groot & Nas, 1991; Dijkstra et al., 1998; Dijkstra et al., 1999; Dijkstra et al., 2010; Font, 2001; Lemhöfer et al., 2008; Lemhöfer & Dijkstra, 2004; Peeters et al., 2013; Sánchez-Casas et al., 1992; Van Hell & Dijkstra, 2002), while the absence of a cognate advantage in the mixed version replicates Poort et al.'s (2016) findings. Also in agreement with previous findings demonstrating that an interlingual homograph inhibition effect should be observed in single-language lexical decision tasks when those include non-target language words that require a 'no'-response (e.g. Dijkstra et al., 1998; Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004; Van Heuven et al., 2008), the interlingual homographs in the mixed version were recognised on average 43 ms more slowly than English controls.

In sum, our data suggest that the (non-significant) disadvantage for the cognates compared to the English control in Poort et al.'s (2016) study was most likely due to the composition of their stimulus list. The most plausible explanation for this pattern of results is that the participants in the standard version responded on the basis of qualitatively different information compared to the participants in the mixed version. In other words, the composition of the stimulus list (for both versions) prompted the participants to adapt their response strategy

Fig. 2. Experiment 1. Harmonic participant means of the inverse-transformed reaction times (in milliseconds) by stimulus type of the preceding trial (cognate, English control, interlingual homograph, pseudohomophone, Dutch word; x-axis) and word type of the current trial (cognate, dark grey; English control, light grey). Error bars represent the standard error of the mean adjusted for a within-participants design (Cousineau, 2005).

(task schema) to the specific stimuli they encountered, presumably to allow them to execute the task as efficiently as possible. Of the three extra stimuli types Poort et al. (2016) included in their experiment, the most likely stimuli to elicit such a change in the participants' behaviour are the Dutch words.

By way of requiring a 'no'-response, the Dutch words probably prompted the participants to link the Dutch reading of the cognates to the 'no'-response, resulting in competition with the 'yes'-response linked to the English reading. Indeed, the exploratory analysis examining the direct effects of the different types of stimuli on the processing of the cognates and English controls in the mixed version suggests that the Dutch words directly and adversely affected the processing of the cognates. Cognates immediately following a Dutch word were recognised 50 ms more slowly than English controls following a Dutch word, although this effect was not significant when correcting for multiple comparisons. In contrast to the Dutch words, neither the pseudohomophones nor the interlingual homographs seemed to have a strong direct effect on how the cognates were processed, although notably both stimuli types seemed to negatively affect the cognates.

An alternative explanation for why we did not observe facilitation for the cognates in the mixed version of Experiment 1 is that this version tapped into a later stage of processing than the standard version due to the increased difficulty of this task. Indeed, the main effect of version on the reaction time data was marginally significant, indicating that the participants in the mixed version on average seemed to take longer to make a decision than the participants in the standard version. As discussed in the Introduction, in the monolingual domain, using a computational model to simulate the time course of semantic ambiguity resolution, Rodd et al. (2004) found that in the later cycles of processing, the 'sense benefit' that is usually observed in lexical decision tasks reversed and became a 'sense disadvantage'. If the settling process for cognates has a similar profile, then it is possible that by slowing participants down, the mixed version may have tapped into a later stage of processing, when cognates are no longer at an advantage compared to single-language control words.

In sum, the results of Experiment 1 demonstrate that the cognate facilitation effect is influenced by stimulus list composition. It seems most likely that the participants adapted their response strategy to the types of stimuli they encountered during the experiment, although we cannot draw any firm conclusions as to which of the three additional stimuli types included in the mixed version had the biggest influence. In addition, it is also possible that the participants were slower to respond to the cognates in the mixed version because that version of the experiment was sensitive to a later stage of processing, when perhaps the cognate advantage no longer exists. Experiment 2 was designed to investigate further.

3. Experiment 2

Experiment 2 was preregistered as part of the Center for Open Science's Preregistration Challenge (cos.io/prereg/). All of the experimental materials, processing and analysis scripts and data can be found in our project on the Open Science Framework (osf.io/zadys). The preregistration can be retrieved from osf.io/9b4a7 (Poort & Rodd, 2016, February 8). Where applicable, deviations from the pre-registration will be noted.

The primary aim of Experiment 2 was to examine separately the influence of each of the three additional filler types on the cognate effect. In addition to the two experimental versions used in Experiment 1, three more versions of the experiment were created that were all based on the standard version. Consequently, Experiment 2 consisted of five versions: (1) the standard version of Experiment 1, (2) the mixed version of Experiment 1, (3) a version in which we replaced some regular non-words with Dutch words (the +DW version), (4) a version that included interlingual homographs (the +IH version) and, finally,

(5) a version in which we replaced all of the regular non-words with pseudohomophones (the +P version).

On the basis of the two explanations outlined above, if we find that the cognate facilitation effect is specifically reduced (or potentially reversed) in the experimental versions that contain Dutch words then this would be consistent with the view that the cognate effect in the mixed version was reversed because of response competition between the 'yes'- and 'no'-responses linked to the two interpretations of a cognate. Similarly, if we find that the cognate effect is reduced or reversed in the versions of the experiment that include interlingual homographs, this would suggest that the interlingual homographs drew attention to the cognates' double language membership and this also resulted in response competition. In contrast, if the effect is reduced or reversed when the task is made more difficult by the presence of pseudohomophones then this would imply that the cognates in the mixed version of Experiment 1 (and in Poort et al.'s, 2016 experiment) were at a disadvantage to the English controls because the task tapped into a later stage of processing.

3.1. Methods

3.1.1. Participants

Given the uncertainty surrounding the size of the cognate facilitation effect in any but the standard version, we decided to recruit (at least) 20 participants per version, consistent with Experiment 1. In the end, a total of 107 participants were recruited using the same recruitment methods as for Experiment 1. Excluding participants happened in two stages. First, while testing was still on-going, five participants who scored less than 80% correct on the lexical decision task were excluded and five new participants tested in their stead. Second, after testing had finished and a total of 102 useable datasets had been gathered, the data from a further two participants were excluded, as their overall accuracy for the cognates and English controls (84.8%; 85.7%) was more than three standard deviations below the mean for their version (mixed version: $M = 95.6\%$, $SD = 3.6\%$; +P version: $M = 96.8\%$, $SD = 3.4\%$). The remaining 100 participants (see Table 1 for numbers per version; 44 males; $M_{age} = 25.1$ years, $SD_{age} = 7.1$ years) had an average of 17.0 years of experience with English ($SD = 7.2$ years). The participants rated their proficiency in Dutch a 9.6 out of 10 and in English an 8.7. These ratings were confirmed by their high LexTALE scores in both languages, which a paired t-test showed were slightly higher in Dutch (Dutch: $M = 88.4\%$, $SD = 8.3\%$; English: $M = 84.4\%$, $SD = 11.0\%$; $p < .001$). Again, there were no differences between the versions with respect to the variables reported here (as shown by chi-square tests and independent-samples Welch's t-tests; all $ps > .2$).

3.1.2. Materials

See Table 1 for an overview of the types of stimuli included in each version. We used the same materials as for Experiment 1. Where necessary, additional regular non-words, pseudohomophones and Dutch words were selected from the same sources or created to ensure that, in all versions, each word was matched in terms of length to a non-word, as in Experiment 1.

3.1.3. Design and procedure

The experimental design and procedure was identical to that of Experiment 1. For any versions of the experiment that included Dutch words, the participants were explicitly instructed to respond 'no' to these.

3.2. Results

3.2.1. Confirmatory analyses

Five items with overall accuracy more than three standard deviations below the average of all experimental items ($M = 97.0\%$, $SD = 4.5\%$) were excluded. These were the cognate "lens"
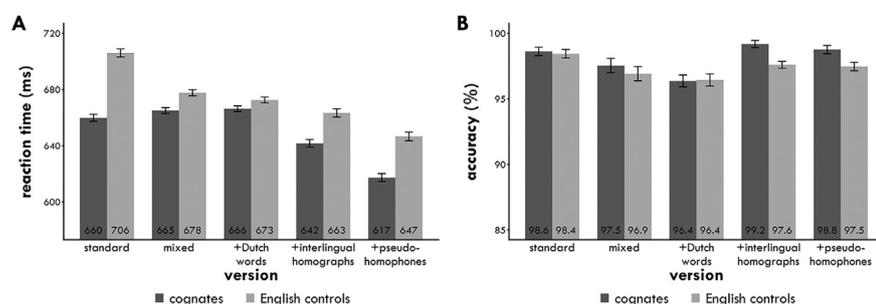
Fig. 3. Experiment 2. **A** Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) and **B** Participant means of task accuracy (percentages correct). Both panels display the data by version (standard, mixed, +Dutch words, +interlingual homographs, +pseudohomophones; x-axis) and word type (cognates, dark grey; English controls, light grey). Error bars represent the standard error of the mean adjusted for a within-participants design, using version means to calculate the adjustment factor (Cousineau, 2005).

($M = 83.0\%$) and the English controls "flu" ($M = 73.0\%$), "poem" ($M = 83.0\%$), "rifle" ($M = 81.0\%$) and "treaty" ($M = 82.0\%$).

As for Experiment 1, all analyses were carried out in R using the lme4 package, following Barr et al.'s (2013) guidelines and using likelihood ratio tests to determine significance of main and interaction effects. Two factors were included in the main 5×2 analysis: word type (2 within-participant/between-items levels: cognate, English control) and version (5 between-participant/within-items levels: standard, mixed, +DW, +IH, +P). Helmert coding (using fractions instead of integers) was used to specify contrasts for the effect of version, whereas effect coding was used to specify a contrast for word type. The *p*-values were compared against an α of .05. To examine more closely which versions of the experiment differed in size and/or direction of the cognate facilitation effect, we also conducted ten 2×2 analyses which included the same factors as the 5×2 analysis, but focused on only two versions at a time. For these analyses, we were only interested in the interaction between word type and version, so we compared the resulting *p*-values against a Bonferroni-corrected α of .005. Finally, we carried out five simple effects analyses, to determine whether the effect of word type was significant in each version. The *p*-values for these analyses were compared against a Bonferroni-corrected α of .01. Detailed results of all analyses for Experiment 2 can be found in Supplementary materials 3.

3.2.1.1. Reaction times. Reaction times (RTs) for incorrect trials were discarded (2.0% of the data), as were RTs less than 300 ms, more than three standard deviations below a participant's mean or more than three standard deviations above a participant's mean RT for all experimental items (2.1% of the remaining data). It should be noted that the 300 ms criterion was not mentioned in our pre-registration. After trimming the data according to our pre-registered exclusion criteria, we discovered two of the remaining data points were below 300 ms. We decided to exclude these, as they were likely accidental key-presses. These exclusions did not affect the significance level of any of the confirmatory or exploratory analyses, but for transparency Table S3.2 of Supplementary materials 3 lists the results of the analyses using the original trimming criteria.

The maximal model, which included a correlated random intercept and effect for word type by participants and a correlated random intercept and effect for version by items, did not converge, nor did a model without correlations between the random effects or a model without random intercepts. Therefore, the final model for the 5×2 analysis included only a random intercept by participants and by items. Because the assumptions of homoscedasticity and normality were violated, the RTs were inverse transformed (inverse-transformed RT = 1000/raw RT) and the model refitted to the inverse-transformed RTs. (The intercepts-only model was also the most complex model that would converge for the inverse-transformed RTs.) Again, it should be noted that panel A of Fig. 3 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.

The main effect of word type was significant [$\chi^2(1) = 18.13$, $p < .001$], with cognates being recognised on average 23 ms more

quickly than English controls. The main effect of version was not significant [$\chi^2(4) = 5.305$, $p = .257$]. As can be seen in Fig. 3, the interaction between word type and version was highly significant [$\chi^2(4) = 45.65$, $p < .001$].

The maximal model for the 10 2×2 analyses, which included a correlated random intercept and slope for word type by participants and a correlated random intercept and slope for version by items, converged for all analyses, so despite the fact that this model did not converge for the 5×2, we decided to take advantage of this extra complexity for the 2×2 analyses. As in Experiment 1, the interaction between word type and version for the standard and mixed versions was significant [$\chi^2(1) = 16.23$, $p < .001$]. The interaction was also significant in the analysis of the standard and +DW versions [$\chi^2(1) = 23.83$, $p < .001$], but not in the analysis of the mixed and +DW versions [$\chi^2(1) = 0.878$, $p = .349$]. It was also not significant in the analyses of the standard and +IH versions [$\chi^2(1) = 6.657$, $p = .010$], the standard and +P versions [$\chi^2(1) = 1.678$, $p = .195$] and the +IH and +P versions [$\chi^2(1) = 1.263$, $p = .261$]. Finally, it was significant in the analysis of the +DW and +P versions [$\chi^2(1) = 10.31$, $p = .001$], but not in any of the remaining 2×2 analyses (all *p*s > .01).

The simple effects analyses revealed significant facilitation for cognates compared to English controls in the standard version [$\chi^2(1) = 27.99$, $p < .001$, $\Delta = 46$ ms], +IH version [$\chi^2(1) = 7.490$, $p = .006$, $\Delta = 22$ ms] and +P version [$\chi^2(1) = 12.11$, $p < .001$, $\Delta = 30$ ms]. The cognate facilitation effects in the mixed and +DW versions were not significant [mixed version: $\chi^2(1) = 3.357$, $p = .067$, $\Delta = 13$ ms; +DW version: $\chi^2(1) = 0.778$, $p = .378$, $\Delta = 6$ ms]. For the simple effects analyses, the maximal model also converged and included a random intercept and random slope for word type by participants and a random intercept by items.

3.2.1.2. Accuracy. Task accuracy is shown in panel B of Fig. 3. The maximal model with a random intercept and random effect for word type by participants and a random intercept and random effect for version by items converged when the bobyqa optimiser was used. This model revealed that the main effect of word type was not significant [$\chi^2(1) = 1.243$, $p = .165$]. The main effect of version was significant [$\chi^2(1) = 9.575$, $p = .048$].[2] The interaction was not significant [$\chi^2(1) = 6.885$, $p = .142$], nor were any of the 10 2×2 analyses (all *p*s > .01; the Bonferroni-corrected α was .005). None of the simple effects analyses were significant either (all *p*s > .06; the Bonferroni-corrected α was .01).

3.2.2. Exploratory analyses

In addition to the confirmatory analyses listed in our

---

[2] To investigate further, we conducted ten exploratory pairwise comparisons for all the five versions. At a Bonferroni-corrected α of .005, none of these pairwise comparisons were significant. At an uncorrected α of .05, this analysis revealed participants performed marginally significantly better in the standard version on the one hand than in the +DW and mixed versions on the other hand [$\chi^2(1) = 3.159$, $p = .076$; $\chi^2(1) = 3.440$, $p = .064$]. Similarly, participants in the +IH version performed marginally significantly better than participants in the mixed version [$\chi^2(1) = 2.745$, $p = .098$].

preregistration, we conducted a number of exploratory analyses on the reaction time data of Experiment 2. Detailed results of all exploratory analyses can be found in Supplementary materials 3.

3.2.2.1. Comparing interlingual homographs and English controls. Although it was not the primary focus of the experiment, our design allowed us to test whether the interlingual homograph inhibition effect does indeed depend on the presence of non-target language words, since the mixed version included interlingual homographs and English controls and some Dutch words, while the + IH version included interlingual homographs and English controls and no Dutch words. We conducted a 2 × 2 analysis with factors word type (2 within-participant/between-items levels: interlingual homograph, English control) and version (2 between-participant/within-items levels: mixed, + IH) and OLD20 as a covariate. We also conducted two simple effects analyses for word type within each version. The design of these analyses was identical to the analogous confirmatory analyses that compared the cognates and English controls.

For the mixed version, we re-included the participant who had been excluded for the confirmatory analysis, while for the + IH version we excluded one participant whose overall percentage correct for the target items included in this analysis (81.3%) was more than three standard deviations below the mean ($M = 95.0\%$, $SD = 4.5\%$). We excluded two interlingual homographs ("hoop", 52.5%, and "lever", 65.0%; $M = 92.4\%$, $SD = 8.8\%$) and one English control ("flu"; 72.5%; $M = 95.8\%$, $SD = 5.8\%$).

The interaction between word type and version in the 2 × 2 analysis was marginally significant [$\chi^2(1) = 2.889$, $p = .089$]. The effect of word type was significant in the mixed version: there was an inhibition effect of 24 ms for the interlingual homographs ($M = 707$ ms) compared to the English controls ($M = 684$ ms) [$\chi^2(1) = 6.9871$, $p = .008$]. In contrast, the effect of word type was not significant in the + IH version, although the interlingual homographs ($M = 658$ ms) were recognised on average 8 ms more slowly than the English controls ($M = 651$ ms) [$\chi^2(1) = 0.693$, $p = .405$]. The effect of OLD20 was not significant in any of these analyses ($p > .3$). In summary, these results are consistent with the literature that has demonstrated that the interlingual homograph inhibition effect depends on or is increased by the presence of non-target language words.

3.2.2.2. Effect of the preceding trial. As for Experiment 1, we investigated whether the stimulus type of the preceding trial interacted with the word type of the current trial in the mixed version. The simple effects analyses showed that having seen a Dutch word on the preceding trial resulted in a strong and significant cognate disadvantage of 49 ms [$\chi^2(1) = 6.722$, $p = .0095$] and as can be seen in Fig. 4, again, this effect was due to the participants taking more time to respond to the cognates and not less time to respond to the English controls. Having seen a cognate, English control or pseudohomophone on the preceding trial resulted in small to moderate but non-significant facilitation effects of 25 ms, 11 ms and 25 ms, respectively [cognates: $\chi^2(1) = 3.237$, $p = .072$; English controls: $\chi^2(1) = 0.635$, $p = .426$; pseudohomophones: $\chi^2(1) = 6.011$, $p = .014$]. In contrast but in line with the findings from Experiment 1, having seen an interlingual homograph resulted in a non-significant cognate disadvantage of 10 ms [$\chi^2(1) = 0.541$, $p = .462$].

The 2 × 2 analyses further showed that the effect the Dutch words had on the size and direction of the cognate effect was significantly different compared to that of the cognates and pseudohomophones [cognates: $\chi^2(1) = 10.70$, $p = .001$; pseudohomophones: $\chi^2(1) = 10.65$, $p = .001$], but compared to the English controls and interlingual homographs only at an uncorrected α of .05 [English controls: $\chi^2(1) = 5.572$, $p = .018$; interlingual homographs: $\chi^2(1) = 4.037$, $p = .045$]. Also at an uncorrected α of .05, the cognate effect was significantly different in cases when the preceding trial was an interlingual homograph compared to when it was a cognate or pseudohomophone
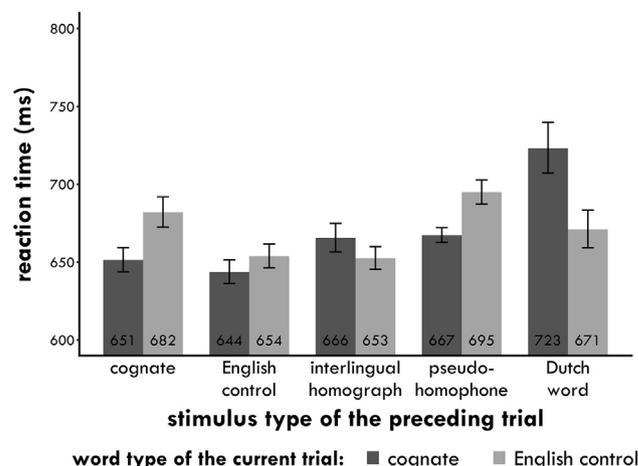

Fig. 4. Experiment 2. Harmonic participant means of the inverse-transformed reaction times (in milliseconds) by stimulus type of the preceding trial (cognate, English control, interlingual homograph, pseudohomophone, Dutch word; x-axis) and word type of the current trial (cognate, dark grey; English control, light grey). Error bars represent the standard error of the mean adjusted for a within-participants design (Cousineau, 2005).

[cognates: $\chi^2(1) = 4.971$, $p = .026$; pseudohomophones: $\chi^2(1) = 4.360$, $p = .037$]. None of the other interactions were significant (all ps > .2).

4. General discussion

We set out to determine whether the cognate facilitation effect in bilingual lexical decision is affected by on the other types of stimuli included in the experiment. In Experiment 1, cognates in the standard version of our English lexical decision task—which included only cognates, English controls and 'regular' non-words—were recognised 31 ms more quickly than English controls, consistent with previous findings (e.g. Cristoffanini et al., 1986; De Groot & Nas, 1991; Dijkstra et al., 1998; Dijkstra et al., 1999; Dijkstra et al., 2010; Font, 2001; Lemhöfer et al., 2008; Lemhöfer & Dijkstra, 2004; Peeters et al., 2013; Sánchez-Casas et al., 1992; Van Hell & Dijkstra, 2002). In contrast, cognates in the mixed version—which included, in addition to the same cognates and English controls, interlingual homographs, pseudohomophones and Dutch words—were recognised 8 ms more slowly, although this difference was not significant. This pattern of results confirms the idea that the difference between Poort et al.'s (2016) findings and the 'standard' experiments reported in the literature were due to their stimulus list composition and not to any other differences between these experiments. Experiment 2 replicated this effect of list composition: there was a significant cognate facilitation effect of 46 ms in the standard version, while the facilitation effect of 13 ms in the mixed version was not significant. Crucially, as in Experiment 1, the effect in the mixed version was significantly smaller than the effect in the standard version. These findings suggest that it is indeed the case that the size and direction of the cognate effect can be influenced by stimulus list composition.

Specifically, it appears that it was the presence or absence of the Dutch words that was critical in determining whether a cognate advantage was observed. In both versions of Experiment 2 that included Dutch words, the cognate facilitation effect was significantly reduced compared to the standard version. Furthermore, the cognate facilitation effects in these versions—13 ms in the mixed version and 6 ms in the + DW version—were not significantly different from zero. Notably, in the mixed versions of both Experiment 1 and 2, we also found that the Dutch words affected the cognates more directly on a trial-by-trial basis: when the preceding trial had been a Dutch word, we found that cognates were recognised more slowly than the English controls, by 49 ms and 50 ms, respectively. (After correcting for multiple

comparisons this effect was only significant in Experiment 2.) Such strong negative effects were not found for any of the other word types.

In contrast to this clear influence of the Dutch words on the magnitude of the cognate advantage, we found no evidence that introducing pseudohomophones had a similar impact on performance. Although the significant cognate facilitation effect of 30 ms in the +P version was numerically smaller than in the standard version, it was not significantly so. Furthermore, the cognate effect in the version with the pseudohomophones was significantly larger compared to the version that included Dutch words, confirming that the pseudohomophones were less effective than the Dutch words in reducing the size of the cognate effect.

The picture remains unclear for the interlingual homographs, however. As for the pseudohomophones, the significant cognate facilitation effect of 22 ms in the +IH version was numerically but not significantly smaller than in the standard version. Unlike for the pseudohomophones, the cognate effect in the +IH version was not significantly bigger than that in the +DW version. As Brenders et al. (2011) note for their younger participants, it may have been the case that the interlingual homographs drew attention to the fact that cognates are words in both English and Dutch. However, it should also be noted that Dijkstra et al. (1998, Exp. 1) and Dijkstra et al. (1999, Exp. 2) also included both cognates and interlingual homographs in the same experiment and did not observe a disadvantage for the cognates. Further research is required, therefore, to determine whether the interlingual homographs may have mimicked, to a lesser extent, the effect of the Dutch-only words.

Taken together, our findings are fully consistent with the idea that the participants constructed a task schema specifically to account for and respond accurately to the stimuli they encountered during the experiment. In single-language lexical decision tasks that do not include non-target language words (such as our standard, +IH and +P versions), the cognate facilitation effect is a consequence of the cognates' overlap in form and meaning in the two languages the bilingual speaks (see Peeters et al., 2013, for a proposal of how this is instantiated in the BIA+ model, Dijkstra & Van Heuven, 2002). Importantly, this is only possible because in such tasks, the participants really only need to decide whether the stimuli they see are word-like or familiar. In terms of the BIA+ model, in such tasks participants can construct a task schema to check merely whether the activation in their word identification system meets a certain threshold of 'word-likeness'.

In contrast, when single-language lexical decision tasks *do* include non-target language words to which the participants should respond 'no' (such as in our mixed and +DW versions), bilinguals can only perform the task accurately if they respond 'yes' solely to stimuli that are words in a *specific* language and 'no' to anything else, including words from the non-target language. In terms of the BIA+ model, this means that they must construct a task schema that checks not only whether the current stimulus meets the threshold for 'word-likeness', but also whether it is of the correct language. Because the Dutch words in the English lexical decision task required a 'no'-response, the participants in the mixed and +DW versions likely linked the Dutch reading of the cognates to the 'no'-response in their task schema, while the English reading was linked to the 'yes'-response. Indeed, the fact that the Dutch words appeared to directly and negatively affect the cognates suggests that the cognates suffered from response competition as a result of this. We suggest that this response competition then (partially) cancelled out the facilitation that is a result of the cognates' overlap in form and meaning.

Further support for the idea that cognates suffer from response competition in single-language lexical decision tasks when those tasks include non-target language words comes from experiments conducted by Lemhöfer and Dijkstra (2004). For Experiment 4, they designed a generalised lexical decision task in which their Dutch–English bilingual participants were asked to decide whether the stimuli they saw were words in either of the two languages they spoke fluently. The stimuli

included cognates, English controls and Dutch words, as well as English-like, Dutch-like and neutral non-words. In this experiment, the participants would have connected both the English and the Dutch interpretation of the cognates to the 'yes'-response, so the presence of the Dutch words should not have elicited response competition. Indeed, Lemhöfer and Dijkstra (2004) found that the participants responded more quickly to the cognates compared to both the English controls and the Dutch words.

Our findings nicely complement research carried out by Dijkstra et al. (1998) (and replicated by De Groot et al., 2000; Dijkstra et al., 2000; Von Studnitz & Green, 2002), who demonstrated that the interlingual homograph inhibition effect in single-language lexical decision tasks depends on the presence of non-target language words. As mentioned in the Introduction, Dijkstra et al. (1998) found no evidence for an inhibition effect for interlingual homographs when their stimulus list only included interlingual homographs, cognates, English controls and regular non-words (Exp. 1), but they did observe significant inhibition for the interlingual homographs compared to the English controls when they also included some Dutch words that the participants were told to respond 'no' to (Exp. 2). Indeed, one of our exploratory analyses replicates this finding. In the +IH version of Experiment 2, which did not include any Dutch words, we did not observe a significant difference between the interlingual homographs and the English controls (although there was an 8 ms trend towards inhibition). In contrast, we did find a significant interlingual homograph inhibition effect of 44 ms and 23 ms in the mixed versions of both Experiment 1 and 2, respectively, which did include Dutch words. The interaction between word type and version in +IH and mixed versions of Experiment 2 was marginally significant.

Dijkstra et al. (2000) further found that it was specifically the presence of the Dutch words in Dijkstra et al.'s (1998) experiment that caused this inhibition effect and not the nature of the instructions. They designed an English lexical decision task that included interlingual homographs, English controls and non-words only in the first half of the task, but also included Dutch control words during the second half of the task. From the beginning, the participants were told to respond 'no' to the Dutch control words. Overall, they observed an inhibition effect for the interlingual homographs compared to the English controls only in the second half of the experiment. And as we did for our cognates, they also found that the Dutch words directly affected the processing of the interlingual homographs: the average reaction time for the first interlingual homograph their participants encountered after the first Dutch item was much longer than for the last interlingual homograph before the introduction of the Dutch words. The English controls in their task did not suffer from the introduction of the Dutch words. This suggests that it was the response competition elicited by the presence of the Dutch words that resulted in the interlingual homograph inhibition effect in their experiment and the observed reduction in the size of the cognate facilitation effect in our experiments.

In contrast, our results are not consistent with the view that the lack of a significant cognate facilitation effect in the mixed and +DW versions (and in Poort et al.'s, 2016 experiment) was a consequence of the task tapping into a later stage of processing when cognates are no longer at an advantage compared to single-language control words. This explanation assumes that, by including stimuli that make the task more difficult (like the pseudohomophones), participants will need more time to accumulate the pieces of information they require to make a decision. Accordingly, this account would have predicted that the cognate facilitation effect would be reduced by the presence of the pseudohomophones as well as by the Dutch words, for which we have no strong evidence. (Note that overall differences in response times between the different experimental versions should be interpreted with caution as it is not possible with this design to remove the (often large) individual differences in reaction times.)

In sum, it appears that when a single-language lexical decision task includes non-target language words, the cognate facilitation effect is

significantly reduced compared to when the task does not. By including such stimuli, the participants must rely on qualitatively different information to perform the task accurately (i.e. for each stimulus determining 'Is this a word in *English*?'), as opposed to when the task can be completed by relying on a sense of word-likeness (i.e. determining 'Is this a word in *general*?'). Analogous to explanations of similar effects for interlingual homographs (e.g. Van Heuven et al., 2008), we suggest that competition between the 'no'-response that becomes linked to the non-target language reading and the 'yes'-response that is linked to the target language reading of the cognate (partially) cancels out the facilitation that is a result of the cognate's overlap in form and meaning. This response-based conflict is in line with the tenets of the BIA+ model and is a direct result of the presence of the non-target language words, which require a 'no'-response. In other words, it seems that cognates, like interlingual homographs, are subject to processes of facilitation and competition both within the lexicon and outside it (at the level of decision making).

These findings highlight the difficulty that researchers face when trying to determine whether effects seen in lexical decision tasks have their origin in the lexicon or at the level of decision making. Based solely on the evidence gathered using lexical decision tasks, one could argue that the cognate facilitation effect in single-language lexical decision tasks without non-target language words is a consequence of facilitation at the decision stage of processing, as the task allows both readings of the cognate to be linked to the 'yes'-response. We therefore suggest that, taken in isolation, evidence for a cognate facilitation effect in lexical decision cannot provide strong evidence that the two languages of a bilingual are stored in a single lexicon and that access to this lexicon is non-selective with respect to language. However, these claims are strongly supported by evidence from other methods, where response facilitation or competition effects are likely to be less salient. For example, experiments using eye-tracking methods show that the cognate facilitation effect can be observed even when the task does not involve any decision component (e.g. Duyck et al., 2007; Libben & Titone, 2009; Van Assche et al., 2011). On the whole, it appears that the cognate facilitation effect is a true effect that is a consequence of how cognates are stored in the bilingual lexicon, but that this effect can be influenced by stimulus list composition and task demands.

Acknowledgements

Supplementary materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.actpsy.2017.08.008.

References

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & Van Steenbergen, H. (2014). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods, 47*(4), 918–929. http://dx.doi.org/10.3758/s13428-014-0530-7.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.
Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Brenders, P., Van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology, 109*(4), 383–396. http://dx.doi.org/10.1016/j.jecp.2011.03.012.
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*(4), 977–990. http://dx.doi.org/10.3758/BRM.41.4.977.
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*(3), 441–458. http://dx.doi.org/10.1037/xhp0000159.
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*(3), 614–635. http://dx.doi.org/10.1037/xlm0000065.
Core Team, R. (2015). *R: A language and environment for statistical computing (version 3.2.1).* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.
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*(5), 1283–1296. http://dx.doi.org/10.1037/0278-7393.26.5.1283.
Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1*(1), 42–45. http://dx.doi.org/10.20982/tqmp.01.1.p042.
Cristoffanini, P., Kirsner, K., & Milech, D. (1986). Bilingual lexical representation: The status of Spanish–English cognates. *The Quarterly Journal of Experimental Psychology, 38*(3), 367–393. http://dx.doi.org/10.1080/14640748608401604.
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*(2), 397–428. http://dx.doi.org/10.1080/713755891.
De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and non-cognates in compound bilinguals. *Journal of Memory and Language, 30*(1), 90–123. http://dx.doi.org/10.1016/0749-596X(91)90012-9.
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*(1), 69–78. http://dx.doi.org/10.1017/S1366728900000146.
Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41*(4), 496–518. http://dx.doi.org/10.1006/jmla.1999.2654.
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*(3), 284–301. http://dx.doi.org/10.1016/j.jml.2009.12.003.
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*(3), 175–197. http://dx.doi.org/10.1017/S1366728902003012.
Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*(1), 51–66. http://dx.doi.org/10.1017/S1366728998000121.
Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663–679. http://dx.doi.org/10.1037/0278-7393.33.4.663.
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)Montpellier, France: Université Paul Valery.
Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*, 67–81. http://dx.doi.org/10.1017/S1366728998000133.
Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition, 5*(3), 225–239. http://dx.doi.org/10.1017/S1366728902003024.
Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627–633. http://dx.doi.org/10.3758/BRM.42.3.627.
Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643–650. http://dx.doi.org/10.3758/BRM.42.3.643.
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*(8), 1665–1692. http://dx.doi.org/10.1080/17470218.2015.1022560.
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*(4), 952. http://dx.doi.org/10.1037/a0023217.
Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325–343. http://dx.doi.org/10.3758/s13428-011-0146-0.
Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition, 32*(4), 533–550. http://dx.doi.org/10.3758/BF03195845.
Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language & Cognitive Processes, 19*(5), 585–611. http://dx.doi.org/10.1080/01690960444000007.

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*(1), 12–31. http://dx.doi.org/10.1037/0278-7393.34.1.12.

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*(2), 381–390. http://dx.doi.org/10.1037/a0014875.

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*(4), 315–332. http://dx.doi.org/10.1016/j.jml.2012.12.003.

Poort, E. D., & Rodd, J. M. (2016, February 8). *Does the cognate facilitation effect depend on task demands?* Retrieved from http://www.osf.io/9b4a7 http://dx.doi.org/10.17605/osf.io/9b4a7.

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*(1), 206–212. http://dx.doi.org/10.1017/S1366728915000395.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *The Quarterly Journal of Experimental Psychology: Section A, 55*(4), 1339–1362. http://dx.doi.org/10.1080/02724980244000099.

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

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*(2), 245–266. http://dx.doi.org/10.1006/jmla.2001.2810.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89–104. http://dx.doi.org/10.1016/j.cogsci.2003.08.002.

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*(4), 293–310. http://dx.doi.org/10.1080/09541449208406189.

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*(6), 1155–1178. http://dx.doi.org/10.1037/0096-1523.29.6.1155.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*(2), 197–212. http://dx.doi.org/10.1016/j.jml.2006.03.004.

Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language & Cognitive Processes, 22*(1), 106–129. http://dx.doi.org/10.1080/01690960500463920.

Smits, E., Martensen, H., Dijkstra, T., & Sandra, D. (2006). Naming interlingual homographs: Variable competition and the role of the decision system. *Bilingualism: Language and Cognition, 9*, 281–297. http://dx.doi.org/10.1017/S136672890600263X.

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*(3), 435–451. http://dx.doi.org/10.3758/BF03195472.

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*(1), 88–107. http://dx.doi.org/10.1016/j.jml.2010.08.006.

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*(8), 923–927. http://dx.doi.org/10.1111/j.1467-9280.2009.02389.x.

Van Casteren, M., & Davis, M. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods, 39*(4), 973–978. http://dx.doi.org/10.3758/BF03192992.

Van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica, 128*(3), 431–451. http://dx.doi.org/10.1016/j.actpsy.2008.03.010.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*(4), 780–789. http://dx.doi.org/10.3758/BF03196335.

Van Heuven, W. J. B., Schriefers, H. J., Dijkstra, T., & Hagoort, P. (2008). Language conflict in the bilingual brain. *Cerebral Cortex, 18*(11), 2706–2716. http://dx.doi.org/10.1093/cercor/bhn030.

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), 1–23. http://dx.doi.org/10.1017/S1366728902000111.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979. http://dx.doi.org/10.3758/PBR.15.5.971.