# The representation of cognates and interlingual homographs in the bilingual lexicon

Eva Denise Poort

2018

Experimental Psychology

University College London

A thesis submitted for the degree of Doctor of Philosophy

I, Eva Denise Poort, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

# ABSTRACT

Cognates and interlingual homographs are words that exist in multiple languages. Cognates, like "wolf" in Dutch and English, also carry the same meaning. Interlingual homographs do not: the word "angel" in English refers to a spiritual being, but in Dutch to the sting of a bee. The six experiments included in this thesis examined how these words are represented in the bilingual mental lexicon.

Experiment 1 and 2 investigated the issue of task effects on the processing of cognates. Bilinguals often process cognates more quickly than single-language control words (like "carrot", which exists in English but not Dutch). These experiments showed that the size of this cognate facilitation effect depends on the other types of stimuli included in the task. These task effects were most likely due to response competition, indicating that cognates are subject to processes of facilitation and inhibition both within the lexicon and at the level of decision making.

Experiment 3 and 4 examined whether seeing a cognate or interlingual homograph in one's native language affects subsequent processing in one's second language. This method was used to determine whether non-identical cognates share a form representation. These experiments were inconclusive: they revealed no effect of cross-lingual long-term priming. Most likely this was because a lexical decision task was used to probe an effect that is largely semantic in nature.

Given these caveats to using lexical decision tasks, two final experiments used a semantic relatedness task instead. Both experiments revealed evidence for an interlingual homograph inhibition effect but no cognate facilitation effect. Furthermore, the second experiment found evidence for a small effect of cross-lingual long-term priming. After comparing these findings to the monolingual literature on semantic ambiguity resolution, this thesis concludes that it is necessary to explore the viability of a distributed connectionist account of the bilingual mental lexicon.

# IMPACT STATEMENT

Approximately half of the world's population speaks at least one language, so one of the major issues in research on bilingualism is to determine how words are stored and accessed in the bilingual mental lexicon. The work presented in this thesis has added to the existing literature and proposes a new account of the representation of cognates (words that are spelled similarly or identically across languages and refer to the same meaning) and interlingual homographs (words that are spelled identically but refer to a different meaning). In addition, it underscores the need for researchers in the field of bilingual language processing to work closely together with their colleagues in the field of monolingual language processing, so that both fields may progress in their quest to determine how people resolve semantic ambiguity, both within a language and between languages. This work falls squarely in the category of 'basic' or 'theoretical' science. As such, there are few direct applications of this work to the world outside academia. Nevertheless, this kind of research is essential for the development of evidence-based second-language educational strategies and/or language interventions.

# ACKNOWLEDGEMENTS

I could not have written this thesis without the help of these incredibly important people. First, I would like to thank my supervisor, Jenni, for all of her help over the years, especially for always making me think of the big picture and not just the details. I am also grateful to my mum, Trix, for her emotional support, especially during the last few weeks of writing this thesis and my dad, René, for all of his help proofreading my stimuli and correcting my Anglicisms in Dutch. I am thankful to my brother, Jonah, for being my guinea pig for many of these experiments and for never asking me how things were going. I would also like to thank my friends and colleagues, especially Rachael, Victoria, Emilia, Ella, Lizzie, Naomi, Hannah and Becky, for laughing with me over the years and (when I didn't feel like laughing) being sounding boards. Finally, I am grateful to my grandmothers, oma Niesje and oma Eef, for being enormously proud of me. (Oma Niesje, ik weet dat je trots op me geweest zou zijn.)

In more practical terms, I would like to thank UCL for the demonstratorship that made this research possible and the EPS for funding many of the conferences that I have attended, as well as my study visit at Ghent University in the summer of 2017. I am also thankful to Prof Marc Brysbaert, Tessa Obbens, Dr Rosemary Orr and Dr Jocelyn Ballantyne, for helping me with recruitment for the web-based experiments. Although the work I did during my study visit did not make it into this thesis, I would like to thank Prof Marc Brysbaert for hosting me during my study visit, Nicolas Dirix, for all of his help with the eye-tracker, and Heleen Vander Beken, for all of the great chats in between testing participants.

Finally, I would like to thank my two examiners, Dr Wing Yee Chow and Dr Walter van Heuven. One of the best things about doing research is discussing it with others, so thank you for the excellent exchange of ideas we had during my *viva voce*.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1: General introduction

# 1 The organisation of the bilingual mental lexicon

It is estimated that half of the world's population speaks more than one language, so one of the key issues in research on bilingualism is to determine how words are stored and accessed in a bilingual's mental lexicon. Two overarching questions dominate the field of bilingualism: (1) are all of the words that a bilingual (or multilingual) knows stored in the same lexicon, regardless of which language they belong to? and (2) when a bilingual tries to retrieve a word from that lexicon, do words from both languages (at least initially) get accessed? Much of the research so far seems to agree that (1) all of the languages a bilingual or multilingual speaks are indeed stored in the same lexicon (this is also known as the 'shared-lexicon' account) and (2) words from all of those languages become active during lexical retrieval (also known as the 'non-selective access' account) (for a review, see Dijkstra, 2005; Dijkstra & Van Heuven, 2012).

Studies have shown, for example, that processing words in one language is affected by the orthographical or phonological neighbours those words have in another language (Bijeljac-Babic, Biardeau, & Grainger, 1997; Grainger & Dijkstra, 1992; Jared & Kroll, 2001; Marian & Spivey, 2003; Midgley, Holcomb, Van Heuven, & Grainger, 2008; Spivey & Marian, 1999; Van Heuven, Dijkstra, & Grainger, 1998). In the monolingual domain, early studies of word recognition showed that identification of a word is sensitive to the number of neighbours (orthographically or phonologically similar words) that word has, as well as the frequency of those neighbours (e.g. Andrews, 1989; Grainger, 1990). Van Heuven et al. (1998) found that, for Dutch–English bilinguals, recognition of the *English* word "lame" in a progressive demasking task was affected by the number of orthographically similar neighbours the word had in *Dutch* (e.g. "lade"). This effect was also present in the participants' first language: recognition of Dutch words was slowed down by the number of neighbours in English. If the bilingual participants had been capable of restricting lexical access to only the words from the task-relevant language, the number of cross-lingual neighbours a word has should not have affected word recognition in that target language. Consequently, these studies provide some of the most convincing evidence in favour of the idea that, during lexical access, word candidates from both of the languages a bilingual speaks are (at least initially) accessed.

Some of the strongest evidence in favour of the idea that all of the words a bilingual knows are stored in the same, integrated lexicon comes from research on translation priming and cross-lingual semantic priming. Such studies have shown, for example, that reaction times to an English word like "girl" for Dutch–English bilinguals are facilitated if the word "meisje" (the Dutch translation of "girl") is presented briefly before "girl" (e.g. Schoonbaert, Duyck, Brysbaert, & Hartsuiker, 2009). Many studies have also found that the reverse applies as well: priming "meisje" with "girl" facilitates response times to "meisje". Note, however, that priming from the second language to the native language is not always observed and is often weaker (for a review, see Wen & Van Heuven, 2017). Similarly, studies have also found that Dutch-English bilinguals respond more quickly to the Dutch word "meisje" if the English word "boy" is briefly presented first. As with translation priming, this cross-linguistic semantic priming from the second language to the native language is often weaker or even absent (Schoonbaert et al., 2009). Nevertheless, taken together these findings convincingly suggest that, in the bilingual mental lexicon, word forms and meanings are interconnected: a single meaning can be connected to both a Dutch word form and an English word form.

Further strong evidence in favour of the shared-lexicon account comes from research on the cognate facilitation and interlingual homograph inhibition effect. *Cognates* are words that exist in an identical (or near identical) form in more than one language and refer to the same meaning, like the word "wolf" in Dutch and English. *Interlingual homographs* are words that, like cognates, share their form in more than one language, but that refer to a different meaning in those languages: "angel" means "insect's sting" in Dutch. Briefly, these studies have shown that cognates are processed more quickly than words that exist only in one of the languages they speak, such as the word "carrot" in English (which translates to "wortel" in Dutch). This effect is called the *cognate facilitation effect*. Interlingual homographs, in contrast, are often processed more slowly than single-language control words. Researchers use the term *interlingual homograph inhibition effect* to describe this finding. Facilitation for cognates and inhibition for interlingual homographs would be unlikely to arise if the two language-specific readings of these words were stored in two separate lexicons. Therefore, the existence of the cognate facilitation effect and interlingual homograph inhibition effect provides more evidence in favour of the idea that the mental lexicon in bilinguals contains words from both languages that are fully interconnected.

# 2 Processing of cognates and interlingual homographs

More importantly, the existence of the cognate facilitation effect and the interlingual homograph inhibition effect suggests that cognates and interlingual homographs, compared to words that exist in only one language, have a special representation in this shared and interconnected bilingual mental lexicon. Another important question in the field of bilingualism, therefore, concerns the processing and representation of cognates and interlingual homographs. Most of the research to date has been concerned with the representation and processing of cognates. The cognate facilitation effect has most commonly been observed in visual lexical decision experiments (e.g. Caramazza & Brones, 1979; Cristoffanini, Kirsner, & Milech, 1986; De Groot & Nas, 1991; Dijkstra et al., 1999; Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Lemhöfer & Dijkstra, 2004; Peeters et al., 2013; Sánchez-Casas, García-Albea, & Davis, 1992; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009; Van Hell & Dijkstra, 2002). In addition, the cognate facilitation effect has been observed in word production: bilinguals are faster to name pictures of cognates (e.g. Costa, Caramazza, & Sebastián-Gallés, 2000; Kroll, Dijkstra, Janssen, & Schriefers, 1999) and to read aloud cognate words (e.g. Schwartz, Kroll, & Diaz, 2007). Furthermore, word association times to cognates are often faster than to control words (Van Hell & De Groot, 1998; Van Hell & Dijkstra, 2002).

The cognate facilitation effect has been demonstrated most often in experiments in the bilinguals' second language, but it has also been observed in experiments where only the native language of the participants was used (Van Hell & Dijkstra, 2002). Moreover, language proficiency appears to play a role in this effect: the size of the cognate facilitation effect is smaller (or in some cases the effect is not there) for participants with a low proficiency in their second (or third) language (Brenders, Van Hell, & Dijkstra, 2011; Bultena, Dijkstra, & Van Hell, 2014; Titone, Libben, Mercier, Whitford, & Pivneva, 2011; Van Hell & Dijkstra, 2002). Along the same lines, research has shown that the effect is bigger for cognates that exist in three languages compared to cognates that exist in only two languages (Lemhöfer, Dijkstra, & Michel, 2004; Van Hell & Dijkstra, 2002). Evidence for a cognate facilitation effect has even been found when cognates are embedded in sentences (Duyck et al., 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche et al., 2009; Van Hell & De Groot, 2008), although the effect is often smaller in these experiments than in experiments that have presented the words in isolation. This wealth of research suggests that the cognate facilitation effect is very robust and universal and that cognates have a special status in the bilingual mental lexicon.

There is one final important thing to note regarding the cognate facilitation effect and that is that the size of the cognate facilitation effect is greater for cognates that are identical compared to non-identical cognates (e.g. "kat" in Dutch and "cat" in English; Comesaña, Ferré, Romero, Guasch, Soares, & García-Chico, 2015; Dijkstra et al., 2010; Duyck et al., 2007; Font, 2001; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). It is unclear, however, whether the cognate facilitation effect is linearly related to orthographic overlap or whether there is an additional benefit for identical cognates. For example, Dijkstra et al. (2010) observed a linear relationship between orthographic similarity and processing time but also found that identical cognates like "tennis"–"tennis" were recognised much more quickly in a lexical decision task (and much more slowly in a language decision task) than non-identical cognates like "rijk"–"rich" and "metaal"–"metal". In other words, the facilitation effect for identical cognates was much greater than for non-identical cognates, even for those non-identical cognates that differed only little in form like "metaal"–"metal". In contrast, Duyck et al. (2007) found that the size of the cognate facilitation effect was positively and linearly related to the degree of orthographic overlap, both in lexical decision reaction times and first fixation durations during sentence reading. Similarly, Van Assche et al. (2011) found that reaction times in lexical decision and first fixation durations, gaze durations and go-past times during sentence reading were facilitated by orthographic (and phonological) overlap. The issue of whether there is an additional benefit for identical cognates compared to non-identical cognates is relevant to models of the bilingual lexicon and will be discussed in more detail later on.

Much research has also focused on the representation and processing of interlingual homographs. Like the cognate facilitation effect, this effect has been found most frequently in experiments investigating visual word recognition (De Bruijn, Dijkstra, Chwilla, & Schriefers, 2001; Dijkstra et al., 1999; Dijkstra et al., 1998; Kerkhofs et al., 2006; Lemhöfer & Dijkstra, 2004; Van Heuven, Schriefers, Dijkstra, & Hagoort, 2008), but it has also been observed in auditory word recognition (using interlingual homophones; Lagrou, Hartsuiker, & Duyck, 2011; Schulpen et al., 2003) and word production (Jared & Szucs, 2002; Smits et al., 2006). Finally, the effect has also been observed during sentence reading (Libben & Titone, 2009; Titone et al., 2011).

Importantly, however, most experiments that have researched the interlingual homograph inhibition effect have used single-language visual lexical decision tasks. In a single-language lexical decision task participants have to decide whether letter strings are words in a *specific* language (usually the bilingual's second language). In such tasks, a disadvantage for interlingual homographs compared to control words is more likely to be observed when the experiment also includes words from the bilingual's other language (i.e. the non-target 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 single-language lexical decision tasks that do not include words from the non-target language, interlingual homographs are often processed as quickly as control words. Furthermore, in general-language lexical decision tasks, during which participants have to decide whether the stimuli are words in either of the languages they know, interlingual homographs are often processed more quickly than control words (Lemhöfer & Dijkstra, 2004). The finding that the interlingual homograph inhibition effect is to some degree task-dependent puts constraints on models of the bilingual lexicon and will also be discussed in more detail later on.

# 3 The representation of cognates and interlingual homographs in the bilingual mental lexicon

As discussed previously, the large body of research on the cognate facilitation effect and the interlingual homograph inhibition effect suggests that cognates and interlingual homographs have a special status in the bilingual lexicon, compared to words that exist in only one of the languages that a bilingual knows. Regarding cognates, Dijkstra et al. (2010) note that there are at least four different viewpoints that offer an explanation of how these words are stored in the bilingual mental lexicon and how this special representation causes the cognate facilitation effect. Because the experiments in this thesis examined visual word processing, the discussion in this section will focus on the orthographic and semantic representations of cognates and interlingual homographs.

All in all, the most cited and detailed view on cognates has been proposed within the localist connectionist framework of the Bilingual Interactive Activation plus (BIA+) model (Dijkstra & Van Heuven, 2002). The BIA+ model is the successor to the Bilingual Interactive Activation (BIA) model (Dijkstra & Van Heuven, 1998) and is a bilingual version of a well-known monolingual localist connectionist model of word recognition, the Interactive Activation model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). According to the BIA+ model (depicted in Figure 1-1), the bilingual word recognition system consists of two components: a word identification system and a task system (which was inspired by Green's (1998) Inhibitory Control model). In the word identification system, the visual input of a string of letters 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 in both of the languages the bilingual speaks. These words inhibit each other through a process called lateral inhibition, irrespective of the language to which they belong. Eventually, after a period of competition between these activated words, one will reach the recognition threshold. The task system continuously reads out the activation in the word identification

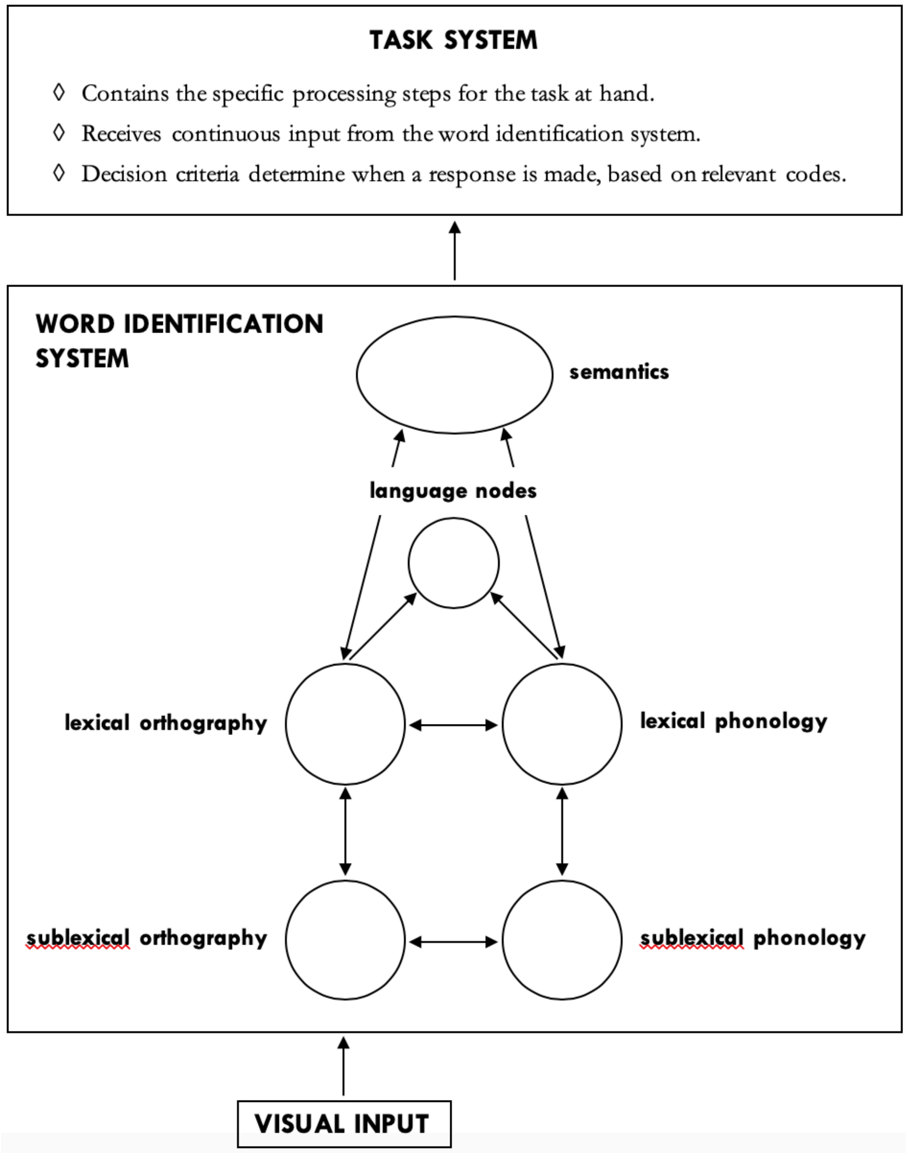


**Figure 1-1:** The BIA+ model of bilingual visual word recognition. The model is described in more detail in the text. Arrows in the figure indicate how activation flows between levels of representation. Inhibitory connections within levels are not depicted. Adapted from Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175–197. doi: 10.1017/S1366728902003012.

system and weighs the different levels of activation to arrive at a response relevant to the task at hand.

In the BIA+ model, identical cognates are represented in a qualitatively different way than non-identical cognates. According to Peeters et al. (2013), the Dutch and English readings of identical cognates share one orthographic representation (i.e. one localist node)

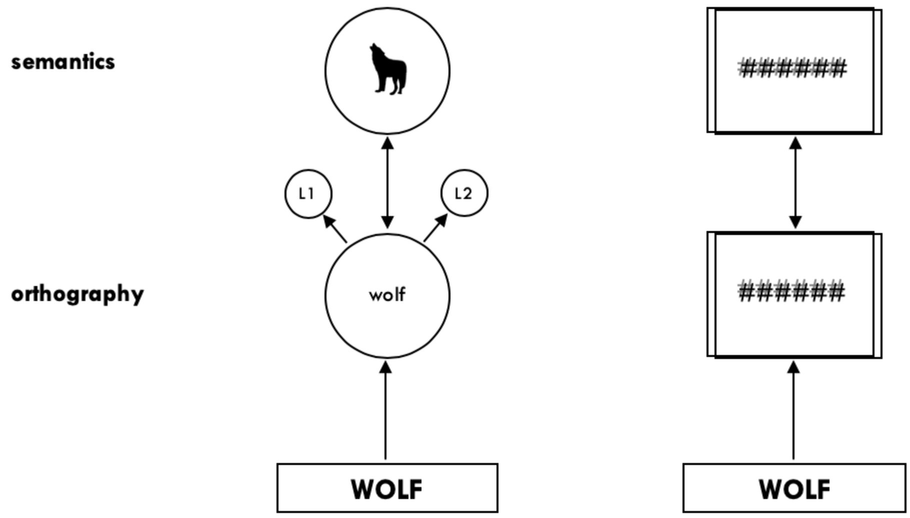


**Figure 1-2:** Two viewpoints on the representation of identical cognates. The localist connectionist viewpoint (as proposed by Peeters et al. (2013) within the BIA+ framework) is depicted on the left. The distributed connectionist viewpoint (as proposed by Van Hell and De Groot (1998) within the DFM framework) is depicted on the right.

and one semantic representation. Two language-specific morphemes store information about the cognate's frequency and morpho-syntactic characteristics in each language separately (Peeters et al., 2013). In this model, the cognate facilitation effect is assumed to be a consequence of greater activation at the semantic level (due to the shared meaning of a cognate; Dijkstra et al., 2010) as well as resonance between the semantic representation and the orthographic and phonological representations (if also shared; Peeters et al., 2013). With regards to non-identical cognates, Dijkstra et al. (2010) argue that these words must consist of two separate orthographic nodes connected to a single shared semantic node. In their view, this explains why they found that the cognate facilitation effect was disproportionally smaller for non-identical cognates than identical cognates: the lateral inhibition between these two highly similar but not identical orthographic nodes for a large part cancels out the facilitation effect.

Another viewpoint was proposed within the framework of a distributed connectionist model, like that of the Distributed Feature Model (DFM; De Groot, 1992, 1993, 1995; De Groot, Dannenburg, & Van Hell, 1994; Van Hell, 1998; Van Hell & De Groot, 1998). In such a model, identical cognates are represented as consisting of two identical and therefore entirely overlapping orthographic patterns (and two identical or highly overlapping phonological patterns) connected to two highly overlapping semantic patterns. Because the orthographic and semantic representations of a cognate are so similar across languages, they form stable representations in lexico-semantic space that are easy to access. In the distributed connectionist framework, this explains the existence of the cognate facilitation effect.

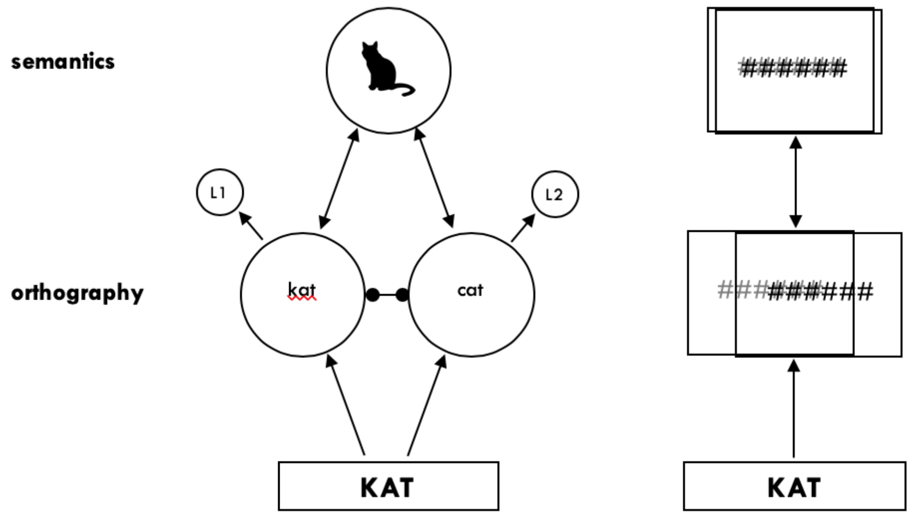


**Figure 1-3:** Two viewpoints on the representation of non-identical cognates. The localist connectionist viewpoint (as proposed by Dijkstra et al. (2010) within the BIA+ framework) is depicted on the left. The distributed connectionist viewpoint (as proposed by Van Hell and De Groot (1998) within the DFM framework) is depicted on the right.

Furthermore, according to this account, non-identical cognates are different only to identical cognates in that their orthographic patterns are less similar. This is in line with the observations of Duyck et al. (2007) and Van Assche et al. (2011) who demonstrated a linear relationship between form overlap and the cognate facilitation effect. The localist and distributed accounts of identical and non-identical cognates are depicted in Figure 1-2 and Figure 1-3.

Dijkstra et al. (2010) also discuss two additional accounts of how cognates may be stored in the bilingual lexicon. In the first of these, proposed by Kirsner and colleagues (Cristoffanini et al., 1986; Kirsner, Lalor, & Hird, 1993; Lalor & Kirsner, 2000) and Sanchéz-Casas and colleagues (Sánchez-Casas & García-Albea, 2005; Sánchez-Casas et al., 1992), cognates are thought to have developed a single shared morphological representation. The second account is based on the framework provided by the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994) and holds that cognates are special because of strong associative links between language-specific form representations. In particular, these links are stronger for cognates than for control words because of the high degree of form similarity. However, it is unclear exactly how the cognate facilitation effect arises in these accounts and, as Dijkstra et al. (2010) note, these accounts also do not offer an explanation for why the cognate facilitation effect is smaller for non-identical cognates. For this reason, these accounts are not discussed any further.

Much less attention has been given to the representation of interlingual homographs in the bilingual mental lexicon. In fact, it appears that the BIA+ model offers the only detailed

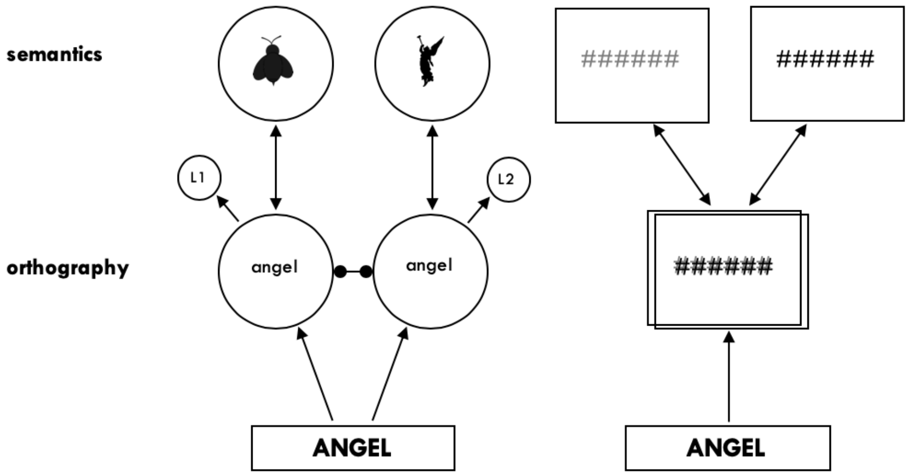


**Figure 1-4:** Two viewpoints on the representation of (identical) interlingual homographs. The localist connectionist viewpoint (as proposed by Dijkstra and Van Heuven (2002) within the BIA+ framework) is depicted on the left. The distributed connectionist viewpoint (as would fit within the DFM framework) is depicted on the right.

account. In the BIA+ model, the Dutch and English readings of an identical interlingual homograph share none of their representations: both readings have their own orthographic and semantic nodes (Kerkhofs et al., 2006). The interlingual homograph inhibition effect is explained by this model as a disproportionately strong effect of lateral inhibition. As with any two words, the two orthographic nodes of the interlingual homograph laterally inhibit each other. However, because these representations are identical, this competition is stronger than that between two regular words. The finding that the interlingual homograph inhibition effect is stronger in single-language lexical decision tasks that include non-target language words is interpreted within the framework of the BIA+ model by assuming that there are two points at which language conflict can arise for an interlingual homograph (Van Heuven et al., 2008). The lateral inhibition between the two identical orthographic representations of an interlingual homograph represents *stimulus-based* conflict. *Response-based* conflict arises outside the lexicon in the task 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.

The representation of interlingual homographs is not considered explicitly in the DFM. Nevertheless, Van Hell and De Groot (1998) do suggest that homophones such as "dear" and "deer" would share orthographic features but not semantic features. It seems reasonable to assume, then, that in a distributed connectionist model like the DFM, interlingual homographs, like cognates, would consist of two entirely overlapping orthographic patterns (and two identical or highly similar phonological patterns). In contrast to cognates, however, these representations would be connected to two distinct semantic patterns that are far away from each other in lexico-semantic space. In this account, interlingual homographs would elicit inhibition because of the difficulty of accessing only one of these semantic patterns instead of a nonsensical blend state comprising of features that are part of both meanings. The localist and distributed accounts of interlingual homographs are depicted in Figure 1-4.

In sum, both the localist and distributed connectionist accounts of identical cognates (see Figure 1-2) assume that these words share both their orthographic and semantic representation across languages (although the distributed account leaves open the possibility that the two meanings of a cognate are not exactly the same). However, these models differ in their viewpoints on non-identical cognates (see Figure 1-3): while the localist BIA+ model argues that non-identical cognates have completely separate orthographic representations, the distributed DFM claims that the two patterns of orthographic features for non-identical cognates are highly overlapping. Similarly, the two accounts disagree when it comes to the orthographic representation of (identical) interlingual homographs (see Figure 1-4). Although the DFM has not explicitly offered an account of these words, it is likely that it would assume that the form representation of an interlingual homograph is similar to that of an identical cognate: completely overlapping. In contrast, in the BIA+ an interlingual homograph is associated with two identical orthographic representations.

# 4 The present experiments

Despite the great abundance of research concerning cognates and interlingual homographs, as the previous section showed, some important questions regarding the representation of these words in the bilingual mental lexicon remain unanswered. This thesis presents two rating experiments and six empirical experiments that attempt to shed more light on the matter of the representation of cognates and interlingual homographs. The purpose of the rating experiments included in Chapter 2 was to develop a database of identical cognates, non-identical cognates, identical interlingual homographs and translation equivalents as control items (i.e. English words that only minimally shared their form with their Dutch translations), for use in the experiments included in the other chapters. The six empirical experiments then each addressed one or more of the questions highlighted in this section that remain unanswered. Across all of these experiments, 537 participants were tested. The significance level for all analyses was set at .05 unless otherwise noted.

The first open question that is addressed concerns the issue of response competition. As discussed previously, research has shown that interlingual homographs are subject to response competition, when one reading of the interlingual homograph is linked to the “yes”-

response and competes with the other reading that is linked to the "no"-response. Furthermore, this response competition can be elicited both by task demands, as Lemhöfer and Dijkstra (2004) demonstrated, and stimulus list competition, as Dijkstra et al. (1998) showed. The aim of the experiments presented in Chapter 3 (Experiment 1 and 2) was to determine whether stimulus list composition effects can also cause response competition for cognates and whether this has implications for the interpretation of the cognate facilitation effect in the literature.

The experiments in Chapter 4 (Experiment 3 and 4) attempted to answer a second open question, that concerning the representation of non-identical cognates. As discussed, both the localist and distributed account agree that identical cognates consist of a single, shared orthographic representation. There is less agreement amongst these accounts, as well as in the empirical literature, regarding the form representation of non-identical cognates. While localist models like the BIA+ argue that they must consist of two separate (but highly similar) orthographic representations, distributed connectionist models like the DFM claim that non-identical cognates consist of two greatly overlapping representations (i.e. patterns of orthographic and phonological features).

The method that Experiment 3 and 4 use to address this question involves cross-lingual long-term priming. Poort et al. (2016) initially used this method to show that recent experience with an identical cognate or interlingual homograph in a bilingual's native language affects how they subsequently process these words in their second language. Poort et al. (2016), following Rodd, Cutrin, Kirsch, Millar, and Davis's (2013) example, designed a three-phase experiment to demonstrate this. In the first phase, Dutch–English bilinguals read *Dutch* sentences that contained either a cognate (e.g. "Hij nam elke dag de **bus** naar school", i.e. "He took the **bus** to school every day"), an interlingual homograph (e.g. "Alleen vrouwelijke bijen en wespen hebben een **angel**", i.e. "Only female bees and wasps have a **sting**") or the Dutch translation of an English control word (e.g. "De schrijver zat achter zijn **bureau** te schrijven", i.e. "The writer was writing at his **desk**"). To ensure the participants read the sentences carefully, they indicated for each sentence whether a subsequent probe was semantically related to it or not. During the second phase of the experiment, the participants completed a digit span task (in English), to create a delay before the third phase, during which they completed a lexical decision task in *English*. Half of (the translations of) the cognates, interlingual homographs and English controls presented during this lexical decision task had been presented during the first phase of the experiment.

The data revealed that the effect of this cross-lingual long-term priming manipulation 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 encountered them in Dutch. According to Rodd et al. (2013) and Rodd, Cai, Betts, Hanby, Hutchinson and Adler (2016) this long-term word-meaning priming is caused by a strengthening of the connection between a word's form and meaning (and possibly also the recurrent connections within the semantic layer itself; Rodd et al., 2016) at the time of priming. Consequently, when the word is seen again, the meaning that has been encountered most recently is more readily accessible (potentially at the cost of any alternative meanings). Because cognates share their form and their meaning across languages, cross-lingual priming was facilitative. In contrast, because interlingual homographs share only their form but not their meaning, priming was disruptive. The experiments in Chapter 4 used this method to determine whether non-identical cognates share a form representation.

Finally, based on the results of Chapter 3 and Chapter 4, the experiments presented in Chapter 5 (Experiment 5 and 6) examined the processing of cognates and interlingual homographs using a semantic relatedness judgements task and contrasted the findings obtained with this task to the findings from the lexical decision tasks used in Chapter 3 and Chapter 4. This chapter attempts to address the following questions: how much of the cognate facilitation effect and the interlingual homograph inhibition effect is due to task artefacts? What can the different patterns of results in these types of tasks tell us about how cognates and interlingual homographs are stored in the bilingual mental lexicon? Finally, this chapter also (again) seeks to answer the question of how cross-lingual experience with cognates and interlingual homographs affects how these words are processed and what this means for how they must be represented in the lexicon.

# CHAPTER 2: Developing a database of stimuli for future experiments

# 1 Introduction

The aim of the experiments included in this chapter was to develop a database of stimuli — identical cognates, non-identical cognates, identical interlingual homographs and translation equivalents — for use in the experiments included in the later chapters of this thesis. The purpose of the first rating experiment was to collect a database of identical cognates, non-identical cognates and translation equivalents. The purpose of the second rating experiment was to collect a database of identical interlingual homographs.[1] Although most of the stimuli (specifically the non-identical cognates and translation equivalents) that were pre-tested in these experiments were selected from published research articles, many of the identical cognates and interlingual homographs had never previously been pre-tested. The two rating experiments were conducted to ensure the quality of these stimuli for future experiments.

Identical cognates were defined as words that had an identical written form in both Dutch and English and highly similar meanings in both languages (e.g. "wolf"–"wolf"). Non-identical cognates were defined as having very similar but not identical forms in Dutch and English and highly similar meanings in both languages (e.g. "kat"–"cat"). The definition of identical interlingual homographs, like the identical cognates, included having identical forms in Dutch and English, but different and unrelated meanings (e.g. "angel" –"angel", where "angel" means "insect's sting" in Dutch). Finally, the translation equivalents were defined as English words with a clear Dutch translation whose written form was not or only minimally similar to the English word (e.g. "wortel"–"carrot").[2] For Experiment 1, 2 and 5, in which participants would see these items only in English, these stimuli further needed to be matched on characteristics like word length, log-transformed frequency and orthographic complexity in English. For Experiment 3, 4 and 6, in which participants would see these items in both Dutch and English, these items also needed to be matched on these variables in Dutch.

Moreover, Experiment 3, 4 and 6 used the cross-lingual priming paradigm discussed in Chapter 1. It was, therefore, important that the sentences that would be used to prime the stimuli in Dutch provided a context in which the Dutch word forms were — in the case of the two types of cognates and the translation equivalents — clearly related in meaning to the

[1] The two rating experiments were conducted separately because, for the experiment that was conducted first chronologically (Experiment 3 from Chapter 4), initially only a set of identical cognates, non-identical cognates and translation equivalents was needed. A set of interlingual homographs only became necessary for the second experiment chronologically (Experiment 1 from Chapter 3). As the Additional analyses reported for the second rating experiment show, the participants in the two experiments did not use the scales in different ways. This indicates that the fact that the ratings were obtained in separate experiments did not affect their validity.

[2] Note that, in the experiments that did not involve priming or did not prime the translation equivalents, the term "English control" is used to refer to the translation equivalents.

items in English or — in the case of the interlingual homographs — not at all related. As mentioned previously, most of the stimuli that were pre-tested in these experiments were selected from published research articles. Many of these had already been pre-tested, but never in a sentence context (in either Dutch or English). For this reason, the participants in these rating experiments were asked to read the Dutch sentences that would later be used in the priming experiments and rate the similarity in meaning, spelling and pronunciation between the Dutch word within that sentence context and its English counterpart. Ratings were obtained for meaning and spelling similarity as these variables critically affect word processing in bilinguals (e.g. Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004). Similarly, pronunciation similarity has also been shown to affect word processing, but phonological similarity is not usually considered a core feature of the definitions of the word types. For the sake of completeness, pronunciation similarity ratings were also obtained, but these were not used to discard any items from the database.

# 2 Rating experiment 1: Identical cognates, non-identical cognates and translation equivalents

The first rating experiment was conducted to gather a database of potential identical cognates, non-identical cognates and translation equivalents initially for use in Experiment 3 and later also in the other experiments included in this thesis. (The non-identical cognates were only included in Experiment 3 and 4.)

## 2.1 Methods

### 2.1.1 Materials

An initial set of 96 identical cognates, 134 non-identical cognates and 444 translation equivalents was selected, most from the following two sources: Dijkstra et al. (2010) and Tokowicz, Kroll, De Groot, and Van Hell (2002). Several criteria were used to guide the selection of these items. First, the items from Dijkstra et al. (2010) and Tokowicz et al. (2002) had been previously pre-tested in terms of their meaning and form similarity and these ratings constituted the main source of information on which decisions were based. (Note that, when judging spelling and pronunciation similarity, Tokowicz et al. (2002) had asked their participants to take both spelling and pronunciation into account for a single 'form similarity' rating, on a scale from 1 to 7. In contrast, Dijkstra et al. (2010) had asked their participants to rate the pairs' spelling and pronunciation similarity separately, also on scales from 1 to 7. An average of these spelling and pronunciation similarity ratings was calculated for each item in Dijkstra et al.'s (2010) materials, to be more comparable to Tokowicz et al.'s (2002) form similarity rating.) Second, an objective measure of orthographic similarity was used in addition to these ratings. This measure was calculated by dividing the Levenshtein distance between the Dutch and English forms of the word by the number of letters of the longest form of the word, which yielded a score between 0 and 1. Finally, the items were selected to be roughly matched on word length, frequency[3] and orthographic complexity in both Dutch and English. Frequency information was obtained from the SUBTLEX-US[4] (Brysbaert & New, 2009) and SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010) databases. For orthographic complexity, the words' OLD20 values were used. A word's OLD20 value is calculated as its mean orthographic Levenshtein distance to a its 20 closest neighbours (Yarkoni, Balota, & Yap, 2008).

Thirty-two identical cognates were selected from Dijkstra et al. (2010) and Tokowicz et al. (2002) based on the following criteria: (1) the item's score on the objective orthographic similarity measure was exactly 1 (which meant that the Dutch and English word form were identical) and (2) the item's meaning similarity rating was greater than 6. An additional 64 identical cognate pairs were selected from Poort et al. (2016) and the other sources listed in

| sources of identical cognates | sources of identical interlingual homographs |
|---|---|
| Dijkstra, Grainger, and Van Heuven (1999) | Dijkstra, Grainger, and Van Heuven (1999) |
| Dijkstra, Van Jaarsveld, and Ten Brinke (1998) | Dijkstra, Timmermans, and Schriefers (2000) |
| Lemhöfer and Dijkstra (2004) | Dijkstra, Van Jaarsveld, and Ten Brinke (1998) |
| Peeters, Dijkstra, and Grainger (2013) | Kerkhofs, Dijkstra, Chwilla, and De Bruijn (2006) |
| Poort, Warren, and Rodd (2016) | Schulpen, Dijkstra, Schriefers, and Hasper (2003) |
| Van Hell and De Groot (1998) | Smits, Martensen, Dijkstra, and Sandra (2006) |
| Van Hell and Dijkstra (2002) | |

**Table 2-1: Rating experiment 1 & 2. Published articles from which were selected many of the identical cognates and interlingual homographs that were pretested in the two rating experiments. The first column lists the sources of identical cognates for the first rating experiment. The second column lists the sources of identical interlingual homographs for the second rating experiment.**

---

[3] When selecting items, only information about their raw frequency (in occurrences per million) was used, as this is more straightforwardly interpretable. When matching the word types, only the items' log-transformed frequency was considered, as the relationship between reaction times and word frequency approximates a logarithmic function (see e.g. Howes & Solomon, 1951; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014).
[4] The SUBTLEX-US database was used instead of the SUBTLEX-UK database because a regression analysis of the data collected by Poort et al. (2016) revealed that the US frequencies were a better predictor of these Dutch–English bilinguals' lexical decision times than the UK frequencies.

Table 2-1. The additional identical cognates selected from the sources mentioned in Table 2-1 had not been previously pre-tested. Non-identical cognates were selected solely from Dijkstra et al. (2010) and Tokowicz et al. (2002) based on the following criteria: (1) the item's score on the objective orthographic similarity measure was more than 0.5 but less than 1, (2) the item's form (or average orthography-phonology) similarity rating was above 5 (on a scale from 1 to 7) and (3) the item's meaning similarity rating was greater than 6. A total of 134 non-identical cognates met these criteria. The selection criteria for the translation equivalents were as follows: (1) the item's score on the objective orthographic similarity measure was less than 0.5, (2) the item's form (or average orthography-phonology) similarity rating was below 3 and (3) the item's meaning similarity rating was greater than 6. The number of translation equivalents that met these criteria was 444 and all came from the Dijkstra et al. (2010) and Tokowicz et al. (2002) materials. All selected words were nouns or adjectives between 3 and 8 letters long.

From this initial selection, any words that had a frequency of less than 2 occurrences per million in either Dutch or English were discarded. As the identical cognates were less frequent than the other two word types, very frequent non-identical cognates and translation equivalents were also discarded. Similarly, to make it easier later to match the non-identical cognates and translation equivalents to the identical cognates on orthographic complexity, items with a high OLD20 in either Dutch or English were excluded. Next, any items that had a mean lexical decision accuracy in the English Lexicon Project (Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson, & Treiman, 2007) of less than 85% were discarded. Offensive words and words within the set of non-identical cognates and translation equivalents that could also be part of one of the other sets, or that could be considered a (non-identical) interlingual homograph, were excluded. (For example, the Dutch word "brood" is a non-identical cognate with the English word "bread", but also an identical interlingual homograph with the English word "brood".) After this second step in the selection procedure, a total of 65 identical cognates, 102 non-identical cognates and 315 translation equivalents remained. The software package Match (Van Casteren & Davis, 2007) was then used to select the 80 non-identical cognates and 80 translation equivalents that best matched the 65 identical cognates. Matching was based on log-transformed word frequency, word length and OLD20 in both Dutch and English. Table 2-2 lists means, minimums, maximums and standard deviations per word type for each of these measures (and raw word frequency) for both English and Dutch.

| | **Dutch characteristics** | | | | **English characteristics** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | frequency | log10(fre-quency) | word length | OLD20 | frequency | log10(fre-quency) | word length | OLD20 | orthographic similarity |
| identical | 41.5 (61.2) | 2.94 (0.51) | 4.52 (1.08) | 1.58 (0.42) | 44.9 (61.5) | 3.08 (0.49) | 4.52 (1.08) | 1.60 (0.36) | 1.00 (0.00) |
| cognates | *min:* 2.17 | *min*: 1.98 | *min:* 3 | *min:* 1.00 | *min:* 2.35 | *min:* 2.08 | *min:* 3 | *min:* 1.00 | *min:* 1.00 |
| | *max:* 254 | *max:* 4.05 | *max:* 8 | *max:* 2.50 | *max:* 308 | *max:* 4.20 | *max:* 8 | *max:* 2.60 | *max:* 1.00 |
| non-identical | 37.7 (44.7) | 2.95 (0.50) | 4.95 (1.05) | 1.55 (0.35) | 47.9 (57.0) | 3.15 (0.46) | 4.96 (1.00) | 1.69 (0.39) | 0.69 (0.12) |
| cognates | *min:* 2.26 | *min*: 2.00 | *min:* 3 | *min:* 1.00 | *min:* 2.59 | *min:* 2.12 | *min:* 3 | *min:* 1.00 | *min:* 0.50 |
| | *max:* 244 | *max:* 4.03 | *max:* 8 | *max:* 2.45 | *max:* 266 | *max:* 4.13 | *max:* 8 | *max:* 2.60 | *max:* 0.83 |
| interlingual | 39.2 (95.1) | 2.57 (0.77) | 4.22 (1.13) | 1.32 (0.37) | 65.8 (153) | 2.81 (0.81) | 4.22 (1.13) | 1.43 (0.36) | 1.00 (0.00) |
| homographs | *min:* 0.09 | *min*: 0.70 | *min:* 3 | *min:* 1.00 | *min:* 0.22 | *min:* 1.08 | *min:* 3 | *min:* 1.00 | *min:* 1.00 |
| | *max:* 580 | *max:* 4.40 | *max:* 7 | *max:* 2.70 | *max:* 828 | *max:* 4.63 | *max:* 7 | *max:* 2.80 | *max:* 1.00 |
| translation | 34.1 (35.6) | 2.96 (0.45) | 4.90 (1.00) | 1.49 (0.31) | 37.5 (38.4) | 3.10 (0.41) | 4.64 (1.02) | 1.63 (0.34) | 0.11 (0.14) |
| equivalents | *min:* 2.15 | *min*: 1.98 | *min:* 3 | *min:* 1.00 | *min:* 3.63 | *min:* 2.27 | *min:* 3 | *min:* 1.00 | *min:* 0.00 |
| | *max:* 179 | *max:* 3.89 | *max:* 7 | *max:* 2.25 | *max:* 215 | *max:* 4.04 | *max:* 8 | *max:* 2.50 | *max:* 0.50 |

**Table 2-2:** Rating experiment 1 & 2. Means (and standard deviations) and minimum and maximum values for the Dutch and English characteristics and orthographic similarity measure for the 65 identical cognates, 80 non-identical cognates, 87 identical interlingual homographs and 80 translation equivalents pre-tested across both rating experiments. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; orthographic similarity refers to the measure of objective orthographic similarity discussed in the text (measured on a scale from 0 to 1), which was calculated as the Levenshtein distance between the Dutch and English forms of the words divided by the length of the longest of the two forms.

| | Dutch word form | English word form | prime sentence (Dutch original) | prime sentence (English translation) |
|---|---|---|---|---|
| identical cognate | wolf | wolf | De hond is een gedomesticeerde ondersoort van de **wolf**. | The dog is a domesticated subspecies of the **wolf**. |
| non-identical cognate | kat | cat | Haar ouders hebben een dikke, grijze **kat**. | Her parents have a fat, grey **cat**. |
| translation equivalent | wortel | carrot | Een ezel kun je altijd blij maken met een **wortel**. | You can always make a donkey happy with a **carrot**. |
| identical interlingual homograph | angel | angel | Alleen vrouwelijke bijen hebben een **angel**. | Only female bees have a **sting**. |
| non-identical interlingual homograph | brutaal | brutal | Als klein meisje was ze behoorlijk **brutaal**. | She was quite **cheeky** when she was a little girl. |
| trick item | vorst | frost | Een andere aanduiding voor monarch is **vorst**. | A different term for monarch is **sovereign**. |

**Table 2-3:** Rating experiment 1 & 2. Examples of items for each of the word types, along with their Dutch prime sentences (with English translations). The non-identical interlingual homographs only served as fillers in these rating experiments. The trick items were included to determine whether the participants were carefully reading the sentences. During the experiments, the participants were only shown the Dutch sentence (with the Dutch word forms, as here, marked in bold) and the English word form.

As mentioned in the Introduction, it was important that the Dutch sentences that would be used to prime the stimuli in Experiment 3, 4 and 6 provided a clear context for the Dutch word forms to prime the English word forms. The next step involved writing the Dutch sentences for these items that would be used in the priming experiments (see Table 2-3 for example sentences). The sentences were between 6 and 12 words long and were written so that the target word (in its stem form) was placed as far towards the end of the sentence as possible, as this minimises ambiguity. Each target word appeared only in its own sentence and not in any other sentence. For nine of the non-identical cognates and 21 of the translation equivalents that Match selected, it was difficult to write a clear and concise sentence that complied with all of the criteria. These pairs were manually replaced with more suitable pairs of a similar frequency, length and OLD20. Finally, to ensure participants made full use of the rating scale for all three aspects — meaning, spelling and pronunciation similarity — across all items they rated, 40 identical interlingual homographs and 21 non-identical interlingual homographs were selected from Poort et al. (2016) and lists of Dutch–English false friends found on Wikipedia and other web pages. Any items for which either the Dutch or English word had a frequency of less than 2 occurrences per million or more than 700 were discarded, as well as all items that could also be considered a (non-)identical cognate[5] and items for which it was difficult to write a clear and concise sentence in Dutch. This left 31 identical and 14 non-identical interlingual homograph pairs to serve as fillers in the first rating experiment. These words were between 3 and 7 letters long. In terms of their orthographic complexity, they ranged between 1 and 2.8 on the OLD20 measure. The sentences for these items were written according to the same criteria as for the identical and non-identical cognates and the translation equivalents. A native speaker of Dutch then proofread all 270 sentences and suggested corrections and clarifications where necessary.

### 2.1.2 Design and procedure

The experiment was set up in Qualtrics (Qualtrics, 2015). Participants saw the English word (in bold) on the left and the Dutch sentence with the Dutch target word in bold on the right and were asked to rate, on a scale from 1 to 7, how similar the two words in bold were in terms of their meaning, spelling and pronunciation. As there were 225 items to obtain ratings for, five versions of the experiment were created, each containing 45 target items plus the 45 identical and non-identical interlingual homograph fillers. Each version also included an additional five items for which the Dutch and English words could be translations of each other (varying in their degree of orthographic similarity), but in the context of the sentence the Dutch word required a different English translation. For example, the word "vorst" in Dutch can be translated as "frost" in English, but also means "monarch". The word "vorst" was then used in a Dutch sentence to mean "monarch", but the participants were asked to rate the similarity in meaning (and spelling and pronunciation) between "vorst" and "frost". These five 'trick' items were included to check whether the participants were carefully reading the sentences. Participants were randomly assigned to one of the five versions of the experiment and the order of items was randomised. Only five items were presented per screen, for a total of 19 screens. The participants were shown six examples (including an example of a trick item) with suggested ratings at the start of the rating experiment and filled in a language background questionnaire at the end.

### 2.1.3 Participants

The aim was to recruit between 10 and 15 participants for each of the five versions of the experiment. Participants were eligible to participate in the experiment if they were a native speaker of Dutch and fluent speaker of English, with no diagnosis of a language disorder. They also had to be between the ages of 18 and 50, of Dutch nationality and resident in the

[5] Unfortunately, a few of these words, like the identical interlingual homograph "beer"–"beer", slipped the net. (The word "beer" in Dutch translates to "bear" in English; the word "beer" in English means "bier" in Dutch.)

Netherlands at the time of the experiment. In the end, a total of 77 participants were recruited through personal contacts resident in the Netherlands and word-of-mouth. The participants gave informed consent and participated for a chance to win an electronic gift card worth €100 (then roughly £75).

The data from one participant were excluded because this participant regularly rated the spelling and pronunciation similarity of the identical and non-identical cognates a 1 or 2. The data from an additional 9 participants were excluded because these participants made more than 3 mistakes on the 5 trick items.

The remaining 67 participants (14 males; $M_{age}$ = 23.5 years, $SD_{age}$ = 5.4 years) had started learning English from an average age of 7.7 ($SD$ = 3.3 years) and so had an average of 15.8 years of experience with English ($SD$ = 5.8 years). The participants rated their proficiency as 9.7 out of 10 in Dutch ($SD$ = 0.6) and 9.2 in English ($SD$ = 0.7). A two-sided paired *t*-test showed this difference to be significant [$t(66) = 4.729$, $p < .001$]. Thirteen participants completed version 1, 14 completed version 2, 12 completed version 3, 15 completed version 4 and 13 completed version 5. There were no differences between the versions with respect to the demographic variables reported here (as shown by ANOVAs and chi-square tests where appropriate; all $p$s > .125).

## 2.2 Findings

Mean ratings for the three word types (identical cognates, non-identical cognates and translation equivalents) for all three aspects (meaning, spelling and pronunciation similarity) can be found in Table 2-4 at the end of this chapter. Overall, most items had received high (or low) ratings for the three aspects as expected for their word type. All translation equivalents had received meaning similarity ratings of 6 or greater. Seven identical and 3 non-identical cognates with meaning similarity ratings below 6 on the 7-point scale were discarded from the database of potential stimuli for future experiments. Strangely, all but two identical cognates had received spelling similarity ratings of 7. Since these two items were truly identical, they were not discarded. Six translation equivalents with spelling similarity ratings higher than 2 were discarded. The intention was to discard non-identical cognates with spelling similarity ratings below 5, in line with the selection criteria, but 21 non-identical cognates met this criterion. In order not to limit the number of stimuli for future experiments too much, only the one non-identical cognate with a spelling similarity rating of less than 4 was discarded. In sum, the first rating experiment produced a database of potential stimuli that included 58 identical cognates, 76 non-identical cognates and 74 translation equivalents.

# 3 Rating experiment 2: Interlingual homographs

A second rating experiment was conducted to gather a database of potential identical interlingual homographs, for use initially in Experiment 1 and 2 and later also in Experiment 4, 5 and 6. This second rating experiment was designed in an identical manner as the first experiment.

## 3.1 Methods

### 3.1.1 Materials

Seventy additional identical interlingual homographs were selected from Poort et al. (2016), the other research articles listed in Table 2-1 and the SUBTLEX-US and SUBTLEX-NL databases (Brysbaert & New, 2009; Keuleers et al., 2010, respectively). In the latter case, noun, adjective and verb entries of similar length as the cognates and translation equivalents were extracted from the SUBTLEX-US and SUBTLEX-NL databases. Those with identical forms but different meanings in Dutch and English were then manually selected.

From this initial selection, as previously, any items for which the English word had a mean lexical decision accuracy in the English Lexicon Project (Balota et al., 2007) of less than 85% were discarded. Since it was difficult to find enough identical interlingual homographs, words with frequencies of less than 2 occurrences per million that were considered to be well-known words were retained, as well as words with a very high frequency or high OLD20. Items that could also be considered a (non-)identical cognate were discarded (but see Footnote 5). Lastly, items for which it was difficult to write a clear and concise sentence in Dutch were excluded, as well as items that were only identical when inflected (such as "arts"–"arts", where "arts" is the uninflected, singular form of the word in Dutch but the plural of "art" in English).

A total of 56 items met these criteria. These newly selected items ranged in frequency from 0.09 to 828.45. All words were between 3 and 7 letters long. In terms of orthographic complexity, the items ranged from 1 to 2.8 on the OLD20 measure. Table 2-2 lists means, minimums, maximums and standard deviations for each of these measures (and raw word frequency) for both English and Dutch. The sentences for these items were written according to the same criteria as for the first rating experiment and they were proofread by the same native speaker of Dutch who proofread those sentences.

Finally, to ensure again that the participants would make full use of the entire rating scale across all items for all three aspects they were asked to judge, seven identical cognates, seven non-identical cognates, seven non-identical interlingual homographs and 14 translation equivalents were selected from the first rating experiment to serve as fillers in the second rating experiment.

### 3.1.2 Design and procedure

The experimental design and procedure of the second rating experiment was the same as that of the first, except that participants were now also able to indicate if they were not familiar with a word, as not all words met the initial frequency criteria. Two versions of the experiment were created, each containing 28 targets plus the 35 identical and non-identical cognate and translation equivalent fillers and the five trick items from the first rating experiment.

### 3.1.3 Participants

Again, the aim was to recruit between 10 and 15 participants for each of the two versions of the experiment. In the end, a total of 24 participants were recruited using the same eligibility criteria and recruitment procedure as for the first rating experiment. Due to the smaller number of stimuli, participants in this experiment had the chance to win an electronic gift card worth £50.

The data from one participant were excluded because this participant regularly rated the spelling and pronunciation similarity of the identical and non-identical cognates a 1 or 2. No participants made more than 3 mistakes on the 5 trick items.

The remaining 23 participants (8 males; $M_{age}$ = 24.5 years, $SD_{age}$ = 5.9 years) had started learning English from an average age of 6.3 (*SD* = 4.0 years) and so had an average of 18.2 years of experience with English (*SD* = 5.0 years). The participants rated their proficiency as 9.5 out of 10 in Dutch (*SD* = 0.7) and 9.2 in English (*SD* = 0.7). A two-sided paired *t*-test showed this difference to be non-significant [$t(66) = 1.628$, $p = .118$]. Eleven participants completed version 1 and 12 completed version 2. A two-sided independent-samples Welch's *t*-test showed that there was a significant difference in age between the two versions [version 1: *M* = 22.4 years, *SD* = 1.9 years; version 2: *M* = 26.5 years, *SD* = 6.0 years; $t(13.4) = -2.264$, $p = .041$]. There were no significant differences between the versions with respect to the other demographic variables reported here (as shown by additional independent-samples Welch's *t*-tests and chi-square tests where appropriate; all *p*s > .1).

## 3.2 Findings

Mean ratings for the identical interlingual homographs for all three aspects (meaning, spelling and pronunciation similarity) can be found in Table 2-4 at the end of this chapter. Of the 87 interlingual homographs that had received meaning, spelling and pronunciation similarity ratings across both rating experiments, most had received high (or low) ratings as expected for the three aspects. Only a total of 15 items was discarded from the database. One item was excluded because it turned out the English word had a mean accuracy in the English Lexicon Project (Balota et al., 2007) of less than 85%. Nine items were discarded because they had received an average meaning similarity rating of more than 2. Three other items had received an average meaning similarity rating of more than 2, but for these items, one or two participants had given them a high rating of 7 while all other participants had given them a rating of 1 or 2. As the majority of participants agreed that these items' meanings were highly dissimilar, these items were retained. Finally, four of the items that had been included in the second rating experiment were discarded because they had received ratings from fewer than ten participants, as some participants had indicated that they did not know those items. At the end of the second rating experiment, the database of potential stimuli had been extended to include a set of 72 identical interlingual homographs.

### 3.2.1 Additional analyses

The 28 identical cognate, non-identical cognate and translation equivalent fillers included in the second rating experiment had also been included (as target items) in the first rating experiment. As such, it was possible to test the similarity of the ratings given in the two rating experiments by comparing the differences between the ratings from the two experiments. The aim was to determine whether the participants in the second experiment had used the rating scales in a meaningfully different way than the participants in the first experiment had. Overall, the differences between the ratings from the two experiments for the three aspects were small. For meaning similarity, the average difference was 0.04 ($SD = 0.16$, *range* = -0.36-0.43). For spelling similarity, it was 0.04 ($SD = 0.17$, *range* = -0.61-0.34). Finally, for pronunciation similarity, the average difference was 0.01 ($SD = 0.16$, *range* = -0.50-0.25). Two-tailed paired-samples *t*-tests indicated that these differences between the two experiments were not significant for any of the three aspects [for meaning similarity: $t(27) = -1.495$, $p = .147$; for spelling similarity: $t(27) = -1.379$, $p = .179$; for pronunciation similarity: $t(27) = -0.489$, $p = .629$]. This indicates that the ratings across both rating experiments were very similar.

# 4 Conclusion

The two rating experiments produced a database of high quality potential stimuli for the other experiments included in this thesis. This database included 58 potential identical cognates, 76 non-identical cognates, 74 translation equivalents and 72 identical interlingual homographs. These items can be found in Appendix A.

| | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | meaning | spelling | pronunciation |
| identical cognates | 37.0 (56.3)<br>*min:* 2.17<br>*max:* 254 | 2.90 (0.49)<br>*min:* 1.98<br>*max:* 4.05 | 4.57 (1.11)<br>*min:* 3<br>*max:* 8 | 1.61 (0.42)<br>*min:* 1.00<br>*max:* 2.50 | 41.5 (54.0)<br>*min:* 2.35<br>*max:* 280 | 3.07 (0.47)<br>*min:* 2.08<br>*max:* 4.15 | 4.57 (1.11)<br>*min:* 3<br>*max:* 8 | 1.63 (0.35)<br>*min:* 1.00<br>*max:* 2.60 | 6.83 (0.22)<br>*min:* 6.20<br>*max:* 7.00 | 7.00 (0.02)<br>*min:* 6.92<br>*max:* 7.00 | 5.91 (0.67)<br>*min:* 4.21<br>*max:* 7.00 |
| non-identical cognates | 38.3 (45.6)<br>*min:* 2.26<br>*max:* 244 | 2.96 (0.50)<br>*min:* 2.00<br>*max:* 4.03 | 5.00 (1.06)<br>*min:* 3<br>*max:* 8 | 1.57 (0.35)<br>*min:* 1.00<br>*max:* 2.45 | 48.8 (58.1)<br>*min:* 2.59<br>*max:* 266 | 3.16 (0.46)<br>*min:* 2.12<br>*max:* 4.13 | 4.99 (1.01)<br>*min:* 3<br>*max:* 8 | 1.69 (0.39)<br>*min:* 1.00<br>*max:* 2.55 | 6.86 (0.21)<br>*min:* 6.00<br>*max:* 7.00 | 5.35 (0.53)<br>*min:* 4.00<br>*max:* 6.08 | 5.06 (0.72)<br>*min:* 3.62<br>*max:* 6.80 |
| interlingual homographs | 46.2 (103)<br>*min:* 0.57<br>*max:* 580 | 2.71 (0.70)<br>*min:* 1.42<br>*max:* 4.40 | 4.01 (0.94)<br>*min:* 3<br>*max:* 7 | 1.26 (0.32)<br>*min:* 1.00<br>*max:* 2.70 | 70.9 (163)<br>*min:* 0.29<br>*max:* 828 | 2.91 (0.73)<br>*min:* 1.20<br>*max:* 4.63 | 4.01 (0.94)<br>*min:* 3<br>*max:* 7 | 1.37 (0.32)<br>*min:* 1.00<br>*max:* 2.80 | 1.16 (0.28)<br>*min:* 1.00<br>*max:* 2.20 | 7.00 (0.01)<br>*min:* 6.91<br>*max:* 7.00 | 5.49 (0.79)<br>*min:* 3.83<br>*max:* 7.00 |
| translation equivalents | 33.6 (35.1)<br>*min:* 2.15<br>*max:* 179 | 2.96 (0.45)<br>*min:* 1.98<br>*max:* 3.89 | 4.95 (0.99)<br>*min:* 3<br>*max:* 7 | 1.50 (0.31)<br>*min:* 1.00<br>*max:* 2.25 | 35.5 (33.2)<br>*min:* 3.63<br>*max:* 175 | 3.08 (0.41)<br>*min:* 2.27<br>*max:* 3.95 | 4.64 (1.03)<br>*min:* 3<br>*max:* 8 | 1.64 (0.34)<br>*min:* 1.00<br>*max:* 2.50 | 6.88 (0.17)<br>*min:* 6.23<br>*max:* 7.00 | 1.12 (0.23)<br>*min:* 1.00<br>*max:* 2.00 | 1.10 (0.19)<br>*min:* 1.00<br>*max:* 1.92 |

**Table 2-4:** Rating experiment 1 & 2. Means (and standard deviations) and minimum and maximum values for the Dutch and English characteristics and similarity ratings for the set 58 identical cognates, 76 non-identical cognates, 72 identical interlingual homographs and 74 translation equivalents selected for use in the experiments included in this thesis based on the two rating experiments. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours. The similarity ratings were provided on a scale from 1 (not at all similar) to 7 ((almost) identical). For the 28 items (7 identical cognates, 7 non-identical cognates and 14 translation equivalents) that were included in both rating experiments, only the average ratings from the first experiment were used.

# CHAPTER 3: The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition

An earlier version of this chapter was published in *Acta Psychologica*:

Poort, E. D., & Rodd, J. M. (2017). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition. *Acta Psychologica, 180*, 52–63. doi: 10.1016/j.actpsy.2017.08.008.

The Methods and Results sections as well as all figures were modified to be more in line with those of the other chapters in this thesis.

The *Acta Psychologica* article won a Center for Open Science Preregistration Challenge Award in January 2018 (see https://cos.io/about/news/Jan-2018-Prereg-Winners).

# 1 Introduction

As Discussed in Chapter 1, one of the most researched phenomena within the field of bilingualism is the *cognate facilitation effect*, that is the finding that bilinguals process cognates more quickly than words that exist in one language only (i.e. that do not share their form with their translation, like “carrot” and its Dutch translation “wortel”). The cognate facilitation effect has most commonly been observed in visual lexical decision experiments when the target words are presented in isolation (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, Dijkstra, Schriefers, Baayen, Grainger, & Zwitserlood, 2008; Peeters et al., 2013; Sánchez-Casas et al., 1992; Van Hell & Dijkstra, 2002), but also in many other tasks like word association (Van Hell & De Groot, 1998; Van Hell & Dijkstra, 2002), picture naming (e.g. Costa et al., 2000), single-word reading aloud (e.g. Schwartz et al., 2007) and sentence reading (Duyck et al., 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche et al., 2009; Van Hell & De Groot, 2008). The cognate facilitation effect 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. Furthermore, this wealth of research suggests that the cognate facilitation effect is a very robust and universal effect.

Research with interlingual homographs paints a more nuanced picture: the *interlingual homograph inhibition effect* appears to depend on the other stimuli that are included in the experiment. As mentioned, the interlingual homograph inhibition effect has been reported in experiments examining bilinguals’ visual word recognition (Dijkstra et al., 1999; Dijkstra et al., 1998; Lemhöfer & Dijkstra, 2004; Van Heuven et al., 2008), during sentence reading (Libben & Titone, 2009; Titone et al., 2011), in auditory word recognition (Lagrou et al., 2011; Schulpen et al., 2003) and during word production (Jared & Szucs, 2002; Smits et al., 2006). Crucially, however, this effect is much larger (or in some cases only present) in single-language visual lexical decision tasks (e.g. an *English* lexical decision task) if the task also includes words from the bilingual’s other language (e.g. Dutch) that participants must say “no” to (De Groot et al., 2000; Dijkstra et al., 2000; Dijkstra et al., 1998; 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 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.

As discussed in Chapter 1, 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. *Stimulus-based* conflict can arise in the lexicon (or word identification system) and is due to competition (i.e. lateral inhibition) between the two (orthographic) representations of the interlingual homograph (Van Heuven et al., 2008). *Response-based* conflict takes place outside the lexicon at the level of decision making (i.e. in the task system) and happens when one of those two representations is 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: the participants had linked the Dutch readings 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 disadvantage for the interlingual homographs compared to the English control words. 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. As mentioned in Chapter 1, Poort et al. (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). While this cross-language long-term priming manipulation had the expected effect of making the cognates easier to recognise and the interlingual homographs more difficult, the data of the unprimed trials was surprising. In contrast to the studies mentioned previously, 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 most of the lexical decision experiments that have found evidence for a cognate facilitation effect, 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 “vasui”. It appears that 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. Cognates, like interlingual homographs, are 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. 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). Therefore, including non-target language words could have resulted in competition between the “yes”-response linked to one interpretation of the cognate and the “no”-response linked to the other.

This is consistent with the findings of Vanlangendonck (2012). She asked Dutch–English bilingual participants to perform two English lexical decision tasks. One task included identical cognates, non-identical cognates, interlingual homographs, English control words and a set of non-words. The second task also included a set of Dutch words. She found that in the task that did not include the Dutch words, participants responded more quickly to the identical cognates than to the English controls, but there was no evidence for an interlingual homograph effect. In the task that did include the Dutch words, however, participants did respond more slowly to the interlingual homographs. They also responded more slowly to the identical cognates (though not as slowly as to the interlingual homographs). Furthermore, in a second experiment where participants completed the lexical decision tasks in an fMRI scanner, she found increased activity in the left inferior frontal and medial frontal cortex for identical cognates in the task that included the Dutch words, which Van Heuven et al. (2008) claim reflects response-based conflict.

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 et al. (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 (Experiment 1), though not in a Dutch lexical decision task (Experiment 2). In an English lexical decision task that included both cognates and interlingual homographs (Experiment 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 presence of 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, therefore, also 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, two online English lexical decision experiments were conducted. 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.

# 2 Experiment 1: Stimulus list composition effects on the cognate facilitation effect in lexical decision

## 2.1 Introduction

One version of Experiment 1 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. The term 'standard version' is used 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 are indeed due to a difference in stimulus list composition, a different pattern of reaction times for the cognates and English controls would be expected in the two versions of this experiment. In accordance with the literature, a cognate facilitation effect was predicted for the standard version, but, based on Poort et al.'s (2016) findings, no advantage (or even a disadvantage) was predicted for the cognates in the mixed version.

## 2.2 Methods

### 2.2.1 Participants

Previous experiments that examined the cognate facilitation effect have usually included approximately 30 participants and 30 items per word type. Given that the set of stimuli used in this experiment was almost double the size, the aim was to recruit at least 20 participants for each of the two versions of the experiment. As for the rating experiments, participants were eligible to participate in this experiment if they were a native speaker of Dutch and fluent speaker of English, with no diagnosis of a language disorder. They also had to be between the ages of 18 and 50, of Dutch or Belgian nationality and resident in the Netherlands or Belgium at the time of the experiment. In the end, a total of 41 participants was recruited through Prolific Academic (Damer & Bradley, 2014), social media and personal contacts resident in the Netherlands and Belgium. The participants gave informed consent and received a gift card worth €5 (then roughly £4) in return for their participation in the experiment.

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; $M_{age}$ = 26.2 years, $SD_{\text{age}}$ = 6.7 years) had started learning English from an average age of 7.4 ($SD$ = 2.5 years) and so had an average of 18.8 years of experience with English ($SD$ = 6.9 years). The participants rated their proficiency as 9.8 out of 10 in Dutch ($SD$ = 0.5) and 8.8 in English ($SD$ = 0.8). A two-sided paired $t$-test showed this difference to be significant [$t(39) = 7.110$, $p < .001$]. These self-ratings were confirmed by their high LexTALE scores in both languages, which a two-sided paired $t$-test showed were also higher in Dutch [Dutch: $M$ = 91.2%, $SD$ = 6.2%; English: $M$ = 86.1%, $SD$ = 8.6%; $t(39) = 5.008$, $p < .001$]. 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 demographic variables reported here (as shown by independent-samples Welch's $t$-tests and chi-square tests where appropriate; all $p$s > .09).

### 2.2.2 Materials

Table 3-1 lists the number of items of each stimulus type included in the two versions of the experiment. The tables in Appendix A indicate which cognates, interlingual homographs and English controls were included in Experiment 1.

| | | NUMBER OF ITEMS PER STIMULUS TYPE | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | *N* | cognates | English controls | interlingual homographs | regular non-words | pseudo-homophones | Dutch words | task duration |
| **Experiment 1** | | | | | | | | |
| standard | 20 | 56 | 56 | 0 | 112 | 0 | 0 | 12:29 |
| mixed | 20 | 56 | 56 | 56 | 0 | 140 | 28 | 18:45 |
| **Experiment 2** | | | | | | | | |
| standard | 21 | 56 | 56 | 0 | 112 | 0 | 0 | 13:04 |
| mixed | 20 | 56 | 56 | 56 | 0 | 140 | 28 | 18:45 |
| +Dutch words | 20 | 56 | 56 | 0 | 94 | 0 | 18 | 13:03 |
| +interlingual homographs | 20 | 56 | 56 | 56 | 168 | 0 | 0 | 18:57 |
| +pseudo-homophones | 19 | 56 | 56 | 0 | 0 | 112 | 0 | 12:48 |

**Table 3-1:** Experiment 1 & 2. Overview of the types and numbers of stimuli included in each version of Experiment 1 and 2, as well as durations of the lexical decision task. *N* is the number of participants included in the analysis for that version.

| | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word 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) | 3.60 (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) |

**Table 3-2:** Experiment 1 & 2. Means (and standard deviations) for the Dutch and English characteristics and similarity ratings for the cognates, interlingual homographs and English controls. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours. *Note*. The Dutch characteristics are listed for completeness only; the items were not matched on these variables.

#### 2.2.2.1 Words

Using the software package Match (Van Casteren & Davis, 2007) again, 56 identical cognates, 56 identical interlingual homographs and 56 English controls were selected from the pre-tested materials from Chapter 2.[6] Matching was based on English log-transformed 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 et al., 2008; weight: 0.5)[7]. Table 3-2 lists means and standard deviations per word type for each of these measures (and raw word frequency) for both English and Dutch, as well as the meaning, spelling and pronunciation 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 included in that version; the mixed version also included the 56 interlingual homographs for a total of 168 words.

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 [$t(109.6) = 3.117$, $p = .002$, $t(109.9) = 2.698$, $p = .008$, 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 (all *p*s < .002). This confirmed the word types' intended status.

#### 2.2.2.2 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, &

[6] All of the cognates and most of the English controls had also been used in Experiment 3, which was chronologically the first experiment.

[7] The matching variables were weighted 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).

Coltheart, 2002), as well as 28 Dutch words (e.g. "vijand") of a similar frequency as the target items, selected pseudo-randomly from the SUBTLEX-NL database (Keuleers et al., 2010). In the standard version, the 112 regular non-words were pronounceable nonsense letter strings generated using the software package Wuggy (Keuleers & Brysbaert, 2010), which creates non-words from existing 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 (Brysbaert & New, 2009). In both versions, the non-words were matched word-for-word to a target in terms of number of letters.

### 2.2.3 Design and procedure

The experiment employed a mixed design. Word type was a within-participants/between-items factor: participants saw all words of each word type, but each word of course belonged to only one word type. In contrast, version was a between-participants/within-items factor: participants were randomly assigned to one of the two versions of the experiment, but each cognate and English control was included in both versions.

The experiment comprised three separate tasks: (1) the English lexical decision task, (2) the English version of the LexTALE (Lemhöfer & Broersma, 2012) and (3) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012). At the start of the experiment, the participants completed a self-report language background survey in Dutch to verify their eligibility to take part in the experiment. The experiment was created using version 15 of the Qualtrics Reaction Time Engine (QRTE; Barnhoorn, Haasnoot, 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). The differences between the two QRTE versions were minimal.

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. Each item remained on screen until the participant responded or until 2000 ms passed. The inter-trial interval was 500 ms.

## 2.3 Results

Although Experiment 1 was not pre-registered, the data were analysed following same procedures as for the confirmatory (and exploratory) analyses conducted for Experiment 2, which was pre-registered. 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 guidelines for confirmatory hypothesis testing proposed by Barr, Levy, Scheepers, and Tily (2013) and using Type III Sums of Squares likelihood ratio tests to determine significance. Reaction times were analysed using the lmer() function with the default optimiser; accuracy data were analysed using the glmer() function with the bobyqa optimiser[8]. Detailed results of all of the analyses for Experiment 1 can be found in Appendix B, section 1. Furthermore, it should be noted that the graphs in the figures display the (harmonic) participant means, while the effects reported in the text were derived from the estimates of the fixed effects provided by the model summary.

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 "griep"–"flu" and "verdrag"–"treaty") were excluded from the analyses, as the percentages correct for these items (70.0%, 80.0%) were more than three standard deviations below the mean of all experimental items ($M$ = 96.6%, $SD$ = 4.9%). Excluding these items did not affect the matching of the word types.

### 2.3.1 'Confirmatory' analyses

#### 2.3.1.1 Analysis procedure

The same analysis procedure was employed for the reaction times and accuracy data. In all cases, positive effects of word type indicate an advantage for the cognates over the English controls (i.e. faster reaction times and higher accuracy), while negative effects indicate a disadvantage (i.e. slower reaction times and lower accuracy). Positive (negative) effects of version indicate an advantage (disadvantage) for the standard version over the mixed version.

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-

[8] The bobyqa optimiser was used instead of the default Nelder-Mead optimiser as more complex models converged when this optimiser was used.

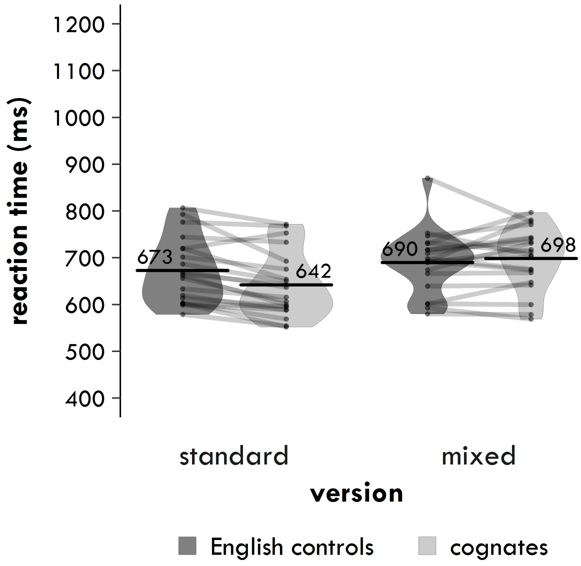


**Figure 3-1:** Experiment 1. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by version (standard, mixed; x-axis) and word type (English controls, dark grey; cognates, light grey). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

participants/within-items levels: standard, mixed). The maximal random effects structure included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for version by items. This maximal model converged for both the reaction times and accuracy analyses.

In addition, two pairwise comparisons were conducted, comparing the cognates and English controls separately for each of the two versions. The maximal random effects structure for these analyses included a correlated random intercept and random slope for word type by participants and a random intercept by items. Again, the maximal model converged for both the reaction time and accuracy analysis.

#### 2.3.1.2 Reaction times

Reaction times are shown in Figure 3-1. Reaction times (RTs) for incorrect trials and trials that participants had not responded to were discarded (3.0% of the data), as were RTs more than three standard deviations above or below a participant's mean RT (2.3% of the remaining data). All remaining RTs were greater than 300 ms. The RTs were 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 analysis showed that the assumptions of

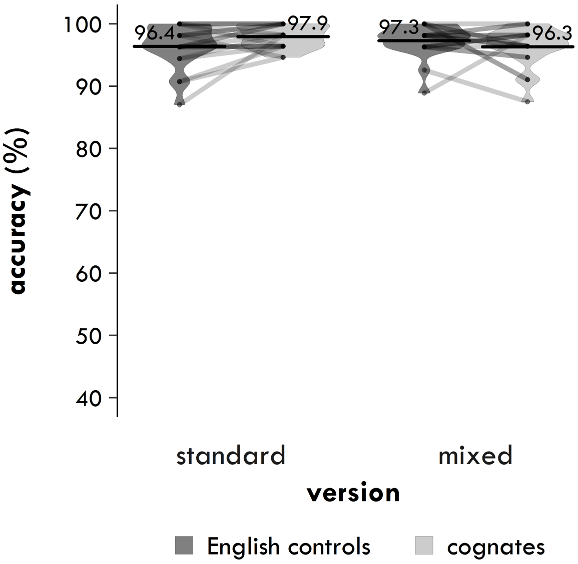


**Figure 3-2:** Experiment 1. Participant means of lexical decision accuracy (percentages correct) by version (standard, mixed; x-axis) and word type (English controls, dark grey; cognates, light grey). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

normality and homoscedasticity were violated. (The inverse-transform achieved a better distribution of the residuals than the log-transform.)

***2×2***

In the 2×2 analysis, 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 [$\chi^2(1) = 15.10$, $p < .001$].

***Pairwise comparisons***

The pairwise comparisons further revealed that the cognate *facilitation* effect of 31 ms in the standard version was significant [$\chi^2(1) = 13.52$, $p < .001$], while the *disadvantage* for cognates of 8 ms in the mixed version was not [$\chi^2(1) = 0.744$, $p = .388$].

#### 2.3.1.3 Accuracy

Accuracy is shown in Figure 3-2. Broadly speaking, the analyses on the accuracy data revealed a similar pattern of results as the analyses on the reaction time data, except that the effects appeared weaker.

***2×2***

In the 2×2 analysis, the main effect of word type was not significant [$\chi^2(1) = 0.157$, $p = .692$, $\Delta = 0.2\%$], nor was the main effect of version [$\chi^2(1) = 0.088$, $p = .767$, $\Delta = -0.2\%$]. The interaction between word type and version was marginally significant [$\chi^2(1) = 3.231$, $p = .072$].

***Pairwise comparisons***

The pairwise comparisons revealed that the small cognate advantage of 0.8% in the standard version was not significant [$\chi^2(1) = 1.415$, $p = .234$], nor was the slight cognate disadvantage of 0.5% in the mixed version [$\chi^2(1) = 0.651$, $p = .420$].

## 2.3.2 'Exploratory' analyses

Two 'exploratory' analyses were conducted on the reaction time data from the mixed version. Note that these were post-hoc exploratory analyses that were carried out in response to effects observed in Experiment 2 and according to the analysis plan for the confirmatory analyses for that experiment.

### 2.3.2.1 Comparing the interlingual homographs and English controls

Although it was not the primary focus of this experiment, the interlingual homographs included in the mixed version were also compared against the English controls. 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 (the interlingual homographs and English controls) was within three standard deviations of the version mean ($M = 92.2\%$, $SD = 5.1\%$). Three items with percentages correct more than three standard deviations below their word type's mean were excluded. These were the interlingual homograph "hoop"–"hoop" (33.3%; $M = 88.7\%$, $SD = 14.8\%$) and the English controls "griep"–"flu" (71.4%) and "verdrag"–"treaty" (66.7%, respectively; $M = 95.7\%$, $SD = 7.4\%$). Since there was a significant difference with respect to the English OLD20 measure between the English controls and the interlingual homographs, this variable was included in the analysis as a covariate[9] (though it was not significant, $\chi^2(1) = 0.071$, $p = .790$). The maximal model with a correlated random intercept and random slope for word type by participants and a random intercept by items converged and revealed a significant interlingual homograph inhibition effect of 43 ms [$\chi^2(1) = 14.05$, $p < .001$]. Reaction times for the interlingual homographs and English controls are shown in Figure 3-6 on page 82.

[9] A random slope for this covariate was not included in the model.

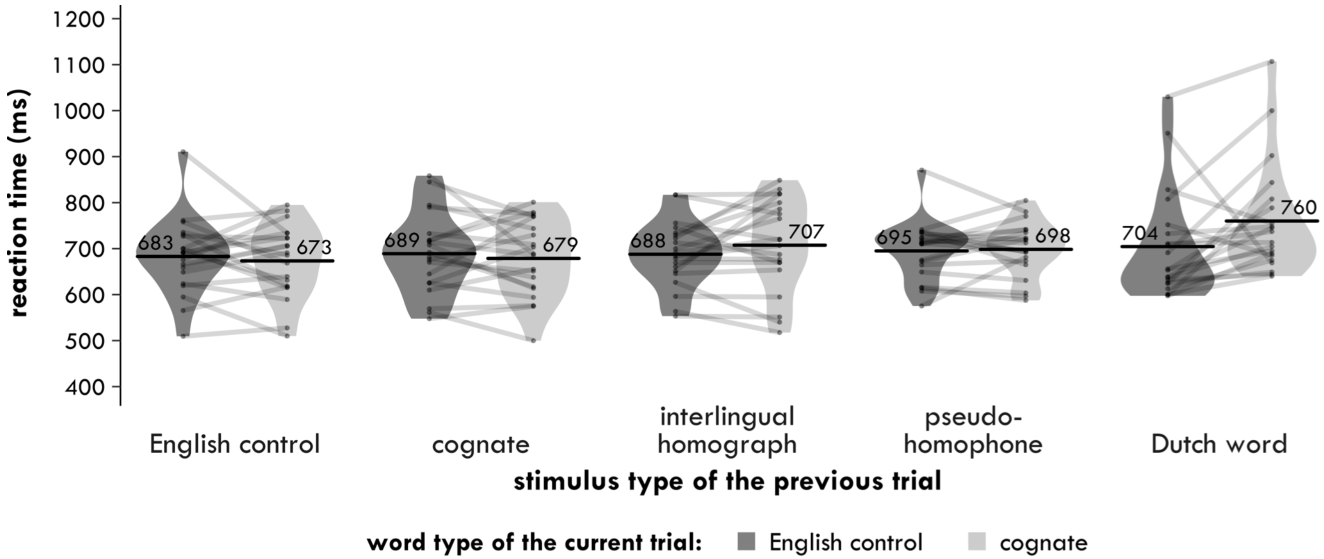


**Figure 3-3:** Experiment 1. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by stimulus type of the preceding trial (cognate, English control, interlingual homograph, pseudohomophone, Dutch word; x-axis) and word type of the current trial (English control, dark grey; cognate, light grey). Each point represents a condition mean for a participant with lines connecting word type means of the current trial from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

#### 2.3.2.2 Examining the effect of the preceding trial on the current trial

An analysis was also conducted on the reaction time data from the mixed version that 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 the mixed version in the confirmatory analysis, only current trials for which the preceding trial had received a correct response were selected (93.1%).

Five pairwise comparisons were first conducted, to determine whether or not there was evidence for a cognate facilitation effect for each of the five preceding trial stimulus types. The maximal random effects structure for these analyses converged and included a correlated 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 $\alpha$ of .01.

Ten 2×2 analyses were also conducted, which focused on two of the five preceding trial stimulus types at a time, to determine whether the influence of each of the five types on the cognate facilitation effect was significantly different to that of the others. Two fixed factors were included in these analyses: word type of the current trial (2 within-participants/between-items levels: cognate, English control) and stimulus type of the preceding trial (restricted to only 2 of the 5 within-participants/within-items levels for each analysis: cognate, English control, interlingual homograph, pseudohomophone or Dutch word). The maximal random effects structure included a correlated random intercept and random slopes for word type of the current trial, stimulus type of the preceding trial and their interaction by participants and a random intercept only by items. Although stimulus type of the preceding trial was a within-items factor, a by-items random slope was not included for this factor as across participants not every item was necessarily preceded by each of the five stimulus types. The maximal model did not converge, so the correlations between the by-participants random effects were removed. Finally, for these analyses, only the interactions were of interest, so only those are reported. The *p*-values for these interactions were compared against a Bonferroni-corrected $\alpha$ of .005.

The five pairwise comparisons 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 $\alpha$ of .05 [$\chi^2(1) = 4.864, p = .027$]. Importantly, as can be seen in Figure 3-3, 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 [for the interlingual homographs: $\chi^2(1) = 0.529$, $p = .467$; for the 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 [for the cognates: $\chi^2(1) = 0.174$, $p = .677$; for the 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 $\alpha$ of .05 [compared to the English controls: $\chi^2(1) = 5.516$, $p = .019$; compared to the cognates: $\chi^2(1) = 6.427$, $p = .011$; compared to the interlingual homographs: $\chi^2(1) = 2.850$, $p = .091$; compared to the pseudohomophones: $\chi^2(1) = 4.005$, $p = .045$]. None of the other interactions were significant (all $p$s > .3).

## 2.4 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, a significant cognate facilitation effect of 31 ms was found, 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 discussing these findings in detail, it should be noted that the 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, it seems unlikely that this could have affected the results.

Notably, the cognate facilitation effect in the standard version mirrors the effect described in the literature (e.g. 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), while the absence of a cognate advantage in the mixed version replicates Poort et al.'s (2016) findings and is in line with Vanlangendonck's (2012) results. 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., 1999; Dijkstra et al., 1998; 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, the 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 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 after 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 a facilitation effect for the cognates was not observed 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 to this chapter, 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 it is impossible to 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: Stimulus list composition effects on the cognate facilitation and interlingual homograph inhibition effect in lexical decision

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

## 3.1 Introduction

The primary aim of Experiment 2 was to examine separately the influence of each of the three additional filler types included in the mixed version of Experiment 1 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 some regular non-words were replaced with Dutch words (the +DW version), (4) a version that included interlingual homographs (the +IH version) and, finally, (5) a version in which all of the regular non-words were replaced with pseudohomophones (the +P version).

On the basis of the two explanations outlined above, if the cognate facilitation effect is specifically reduced (or potentially reversed) in the experimental versions that contain Dutch words (the mixed version and the +DW version) then this would be consistent with the view that the cognate effect in the mixed version of Experiment 1 was reversed because of response competition between the "yes"- and "no"-responses linked to the two interpretations of a cognate. Similarly, if the cognate effect is reduced or reversed in the versions of the experiment that include interlingual homographs (the mixed version and the +IH version), 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 (in the +P version) 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.

## 3.2 Methods

### 3.2.1 Participants

Given the uncertainty surrounding the size of the cognate facilitation effect in any but the standard version, the aim was to recruit (at least) 20 participants per version, consistent with Experiment 1. In the end, a total of 107 participants was recruited using the same eligibility criteria and recruitment procedure as for Experiment 1.

Excluding participants happened in two stages. First, while testing was still on-going, five participants (two in the standard version, two in the +P version and one in the +DW version) 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, each participant's performance on the target items (cognates and English controls) included in the lexical decision task was compared to the mean of all participants who completed the same version to determine whether any more participants needed to be excluded. Two participants had performed worse than three standard deviations below their group's mean (one in the mixed version: 84.8%, $M$ = 95.6%, $SD$ = 3.6%; one in the +P version: 85.7%, $M$ = 96.8%, $SD$ = 3.4%) and were excluded.

The remaining 100 participants (see Table 3-1 for numbers per version; 44 males; $M_{age}$ = 25.1 years, $SD_{age}$ = 7.1 years) had started learning English from an average age of 8.1 ($SD$ = 2.7 years) and so had an average of 17.0 years of experience with English ($SD$ = 7.2 years). The participants rated their proficiency as 9.6 out of 10 in Dutch ($SD$ = 0.6) and 8.7 in English ($SD$ = 0.8). A two-sided paired $t$-test showed this difference to be significant [$t(99) = 9.501$, $p < .001$]. These self-ratings were confirmed by their high LexTALE scores in both languages, which a two-sided paired $t$-test showed were also higher in Dutch [Dutch:

*M* = 88.4%, *SD* = 8.3%; English: *M* = 84.4%, *SD* = 11.0%; $t(99) = 4.198$, $p < .001$]. There were no differences between the versions with respect to the demographic variables reported here (as shown by ANOVAs and chi-square tests where appropriate; all *p*s > .2).

### 3.2.2 Materials

See Table 3-1 for an overview of the types of stimuli included in each version. The same materials were used as for Experiment 1. Where necessary, additional regular non-words, pseudo-homophones and Dutch words were selected from the same sources or created according to the same criteria to ensure that, in all versions, each word was matched in terms of length to a non-word, as in Experiment 1.

### 3.2.3 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.

## 3.3 Results

As for Experiment 1, all analyses were carried out in R using the lme4 package, following Barr et al.'s (2013) guidelines and using Type III Sums of Squares likelihood ratio tests to determine significance. Reaction times were analysed using the lmer() function with the default optimiser; accuracy data were analysed using the glmer() function with the bobyqa optimiser. Detailed results of all of the analyses for Experiment 2 can be found in Appendix B, section 2. The graphs in the figures again display the (harmonic) participant means, while the effects reported in the text were derived from the estimates of the fixed effects provided by the model summary.

Five items (the cognate "lens"–"lens" and the English controls "gedicht"–"poem", "geweer"–"rifle", "griep"–"flu" and "verdrag"–"treaty") were excluded from the analyses, as the percentages correct (83.0%, 73.0%, 83.0%, 81.0% and 82.0%, respectively) for these items were more than three standard deviations below the mean of all experimental items (*M* = 97.0%, *SD* = 4.5%). Excluding these items did not affect the matching of the word types.

### 3.3.1 Confirmatory analyses

#### 3.3.1.1 Analysis procedure

The same analysis procedure was again employed for the reaction times and accuracy data. Again, positive effects of word type indicate an advantage for the cognates over the English controls (i.e. faster reaction times and higher accuracy), while negative effects indicate a disadvantage (i.e. slower reaction times and lower accuracy). Positive (negative) effects of version indicate an advantage (disadvantage) for the first-named version of the second.

Two fixed factors were included in the main 2×5 analysis: word type (2 within-participant/between-items levels: cognate, English control) and version (5 between-participants/within-items levels: standard, mixed, +DW, +IH, +P). The maximal random effects structure included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for version by items. This maximal model converged for the accuracy analysis. It did not converge for the reaction times analysis, nor did a model without correlations between the random effects or a model without random intercepts, but a model with only the random intercepts did.

To examine more closely which versions of the experiment differed in the size and/or direction of the cognate facilitation effect, 10 2×2 analyses were also conducted. These included the same factors as the 2×5 analysis, but focused on only two versions at a time. The maximal random effects structure again included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for version by items and converged for both the reaction times and accuracy analysis. For these 10 2×2 analyses, only the interactions between word type and version were of interest, so significance was only determined for these effects and the resulting *p*-values were compared against a Bonferroni-corrected $\alpha$ of .005.

Finally, five pairwise comparisons were conducted, comparing the cognates and English controls separately for each of the five versions. The maximal random effects structure for these analyses included a correlated random intercept and random slope for word type by participants and a random intercept by items. Again, the maximal model converged for both the reaction time and accuracy analysis. The *p*-values for these analyses were compared against a Bonferroni-corrected $\alpha$ of .01.

#### 3.3.1.2 Reaction times

Reaction times are shown in Figure 3-4. Reaction times (RTs) for incorrect trials and trials that participants had not responded to were discarded (2.0% of the data), as were RTs less than 300 ms, more than three standard deviations below a participant's mean RT for all experimental items or more than three standard deviations above a this mean (2.1% of the remaining data). It should be noted that the 300 ms criterion was not mentioned in the preregistration. After trimming the data according to the pre-registered exclusion criteria, two remaining data points were discovered to be below 300 ms. These were excluded, 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 B-7 of Appendix B lists the results of the analyses using the original trimming criteria. 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 2×5 analysis showed that the assumptions of normality and homoscedasticity were violated.

#### *2×5*

In the 2×5 analysis, the main effect of word type was significant [$\chi^2(1) = 18.13$, $p < .001$], with the cognates being recognised on average 23 ms more quickly than the English controls. The main effect of version was not significant [$\chi^2(4) = 5.305$, $p = .257$]. The interaction between word type and version was highly significant [$\chi^2(4) = 45.65$, $p < .001$].

#### *2×2 interactions*

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).

#### *Pairwise comparisons*

The pairwise comparisons further revealed a significant facilitation effect for the cognates compared to the English controls of 46 ms in the standard version [$\chi^2(1) = 27.99$, $p < .001$], of 22 ms in the +IH version [$\chi^2(1) = 7.490$, $p = .006$] and of 30 ms in the +P version [$\chi^2(1) = 12.11$, $p < .001$]. The much smaller cognate facilitation effects of 13 ms in the mixed and 6 ms in the +DW versions were not significant [mixed version: $\chi^2(1) = 3.357$, $p = .067$; +DW version: $\chi^2(1) = 0.778$, $p = .378$].

### 3.3.1.3 Accuracy

Accuracy is shown in Figure 3-5.

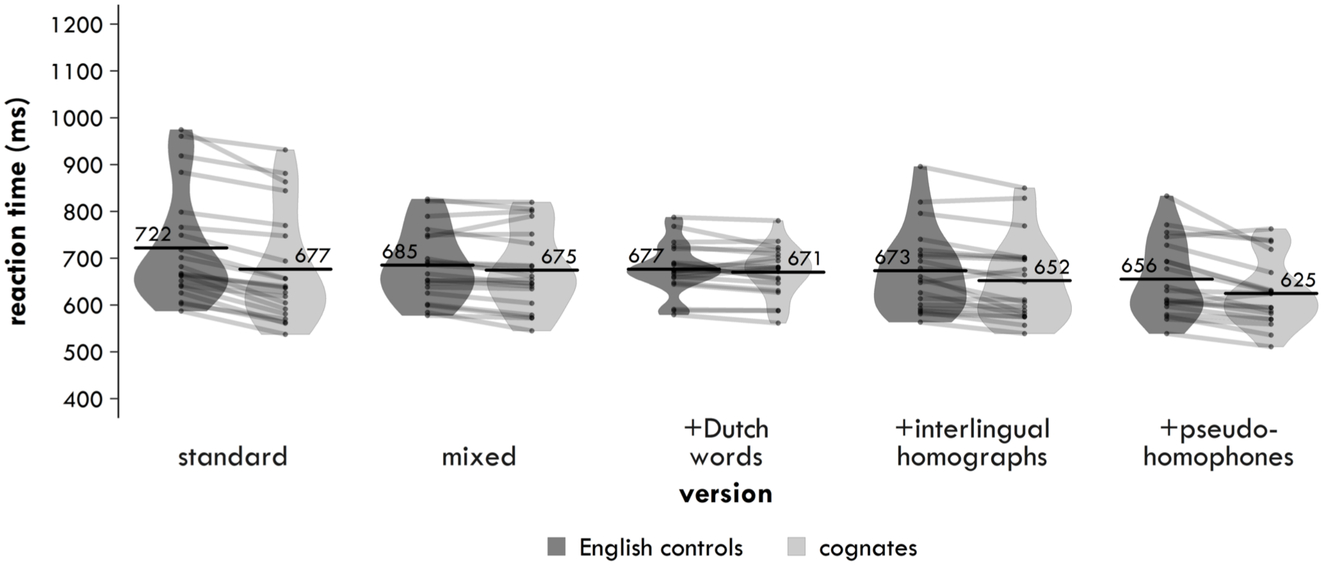


**Figure 3-4:** Experiment 2. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by version (standard, mixed, +Dutch words, +interlingual homographs, +pseudohomophones; x-axis) and word type (English controls, dark grey; cognates, light grey). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

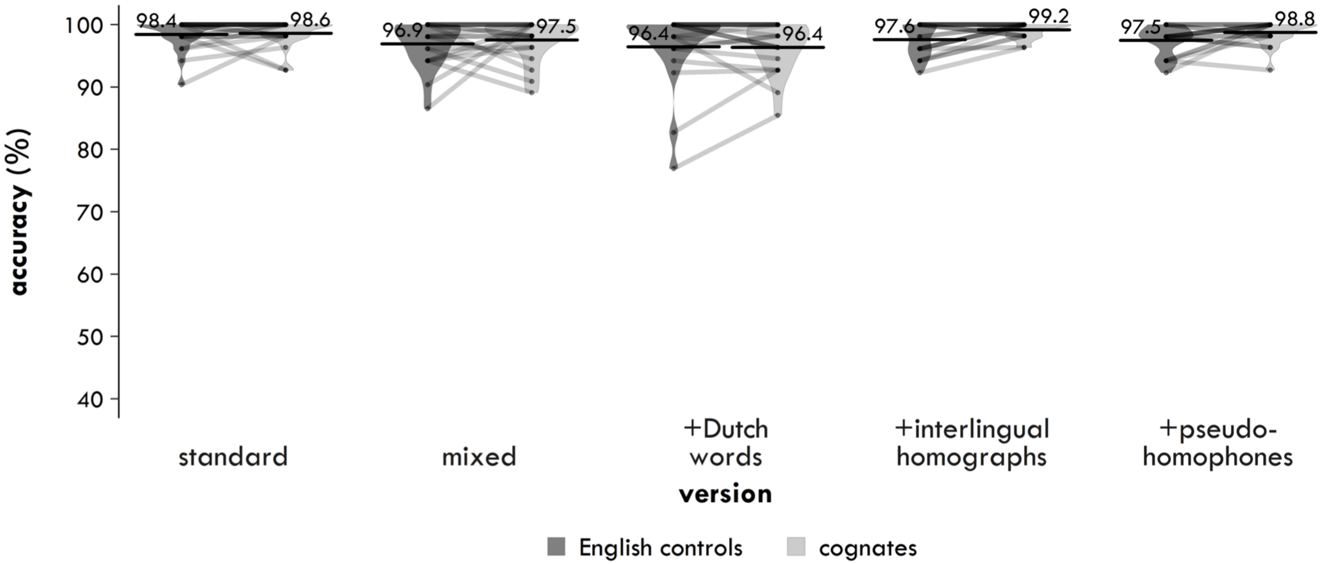


**Figure 3-5:** Experiment 2. Participant means of lexical decision accuracy (percentages correct) by version (standard, mixed, +Dutch words, +interlingual homographs, +pseudohomophones; x-axis) and word type (English controls, dark grey; cognates, light grey). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

### *2×5*

In the 2×5 analysis, the main effect of word type was not significant [$\chi^2(1) = 1.243$, $p = .165$, $\Delta = 0.3\%$], nor was the interaction between word type and version [$\chi^2(1) = 6.885$, $p = .142$]. The main effect of version was significant [$\chi^2(1) = 9.575$, $p = .048$][10].

### *2×2 interactions*

None of the interactions in the 10 2×2 analyses were significant (all *p*s > .01; $\alpha = .005$).

### *Pairwise comparisons*

None of the five pairwise comparisons were significant (all *p*s > .06; $\alpha = .01$).

## 3.3.2 Exploratory analyses

Two exploratory analyses were conducted on the reaction time data of Experiment 2.

### 3.3.2.1 Comparing the interlingual homographs and English controls

Although it was not the primary focus of the experiment, the design of Experiment 2 made it possible to test whether the interlingual homograph inhibition effect does indeed depend on the presence of non-target language words, since the mixed version included some Dutch words in addition to the sets of interlingual homographs and English controls, while the +IH version included interlingual homographs and English controls but no Dutch words. For the mixed version, the participant who had been excluded for the confirmatory analysis was re-included, while for the +IH version one participant was excluded whose percentage correct for the target items included in this analysis (81.3%) was more than three standard deviations below the mean of all experimental items ($M = 95.0\%$, $SD = 4.5\%$). Furthermore, three items with percentages correct more than three standard deviations below their word type's mean were excluded. These were the interlingual homographs "hoop"–"hoop" (52.5%) and "lever"–"lever" (65.0%; $M = 92.4\%$, $SD = 8.8\%$) and the English control "griep"–"flu" (72.5%; $M = 95.8\%$, $SD = 5.8\%$) were excluded.

[10] To investigate further, ten exploratory pairwise comparisons were conducted to compare all five versions to each other. At a Bonferroni-corrected $\alpha$ of .005, none of these pairwise comparisons were significant. At an uncorrected $\alpha$ of .05, this analysis revealed participants performed marginally significantly better in the standard version on the one hand than in the mixed and +DW versions on the other hand [compared to mixed version: $\chi^2(1) = 3.440$, $p = .064$, $\Delta = 0.9\%$; compared to +DW version: $\chi^2(1) = 3.159$, $p = .076$, $\Delta = 1.2\%$]. Similarly, participants in the mixed version performed marginally significantly worse than participants in the +IH version [$\chi^2(1) = 2.745$, $p = .098$, $\Delta = 0.7\%$].

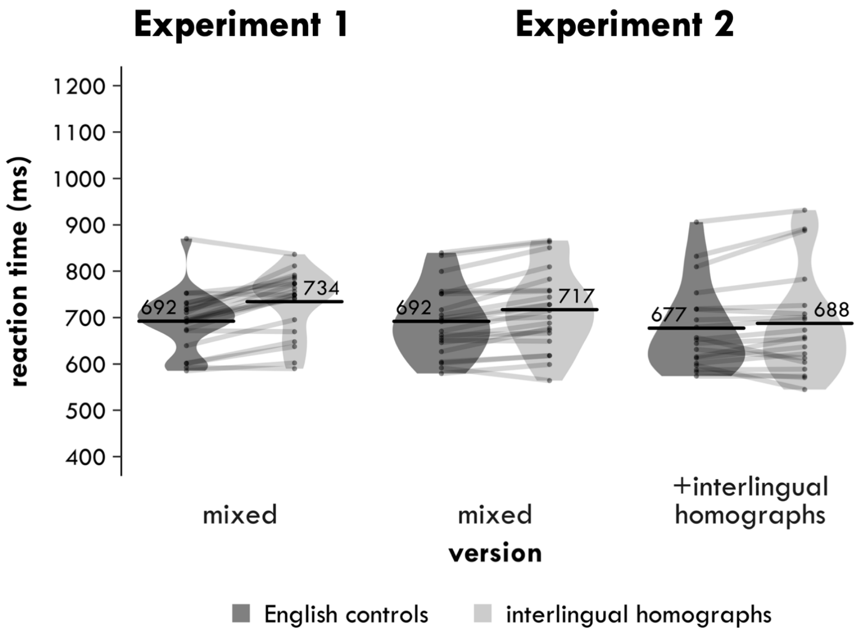


**Figure 3-6:** Experiment 1 & 2. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by version (mixed, +interlingual homographs; x-axis), word type (English controls, dark grey; interlingual homographs, light grey) and experiment (Experiment 1, Experiment 2). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

The design of these analyses was identical to the analogous confirmatory analyses that compared the cognates and English controls. A 2×2 analysis was conducted with the fixed factors word type (2 within-participant/between-items levels: interlingual homograph, English control) and version (2 between-participant/within-items levels: mixed, +IH). The maximal random effects structure included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for version by items and converged. Again, only the interaction between word type and version was of interest, so significance was only determined for this effect and the resulting *p*-value was compared against an $\alpha$ of .05. Two pairwise comparisons were also conducted, to compare the two word types within each version. The maximal random effects structure for these analyses included a correlated random intercept and random slope for word type by participants and a random intercept by items and also converged. The *p*-values for these analyses were compared against an $\alpha$ of .025. OLD20 was again included as a covariate in all of these analyses.

Reaction times for the interlingual homographs and English controls are shown in Figure 3-6. In the 2×2 analysis, the interaction between word type and version was marginally significant [$\chi^2(1) = 2.889$, $p = .089$]. The pairwise comparisons further revealed that, in the mixed version, there was an inhibition effect of 24 ms for the interlingual homographs compared to the English controls [$\chi^2(1) = 6.9871$, $p = .008$]. In contrast, the effect of word type was not significant in the +IH version, although the interlingual homographs were recognised on average 8 ms more slowly than the English controls [$\chi^2(1) = 0.693$, $p = .405$]. The effect of OLD20 was not significant in any of these analyses (both *p*s > .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.

#### 3.3.2.2 Examining the effect of the preceding trial on the current trial

As for Experiment 1, an analysis was conducted that examined whether the stimulus type of the preceding trial interacted with the word type of the current trial in the mixed version. The pairwise comparisons showed that having seen a Dutch word on the preceding trial had again resulted in a strong and significant cognate disadvantage of 49 ms [$\chi^2(1) = 6.722$, $p = .0095$] and as can be seen in Figure 3-7, 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 [for the cognates: $\chi^2(1) = 3.237$, $p = .072$; for the English controls: $\chi^2(1) = 0.635$, $p = .426$; for the 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 pseudo-homophones [compared to the cognates: $\chi^2(1) = 10.70$, $p = .001$; compared to the pseudohomophones: $\chi^2(1) = 10.65$, $p = .001$], but compared to the English controls and interlingual homographs only at an uncorrected $\alpha$ of .05 [compared to the English controls: $\chi^2(1) = 5.572$, $p = .018$; compared to the interlingual homographs: $\chi^2(1) = 4.037$, $p = .045$]. Also at an uncorrected $\alpha$ 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 [compared to the cognates: $\chi^2(1) = 4.971$, $p = .026$; compared to the pseudohomophones: $\chi^2(1) = 4.360$, $p = .037$]. None of the other interactions were significant (all *p*s > .2).

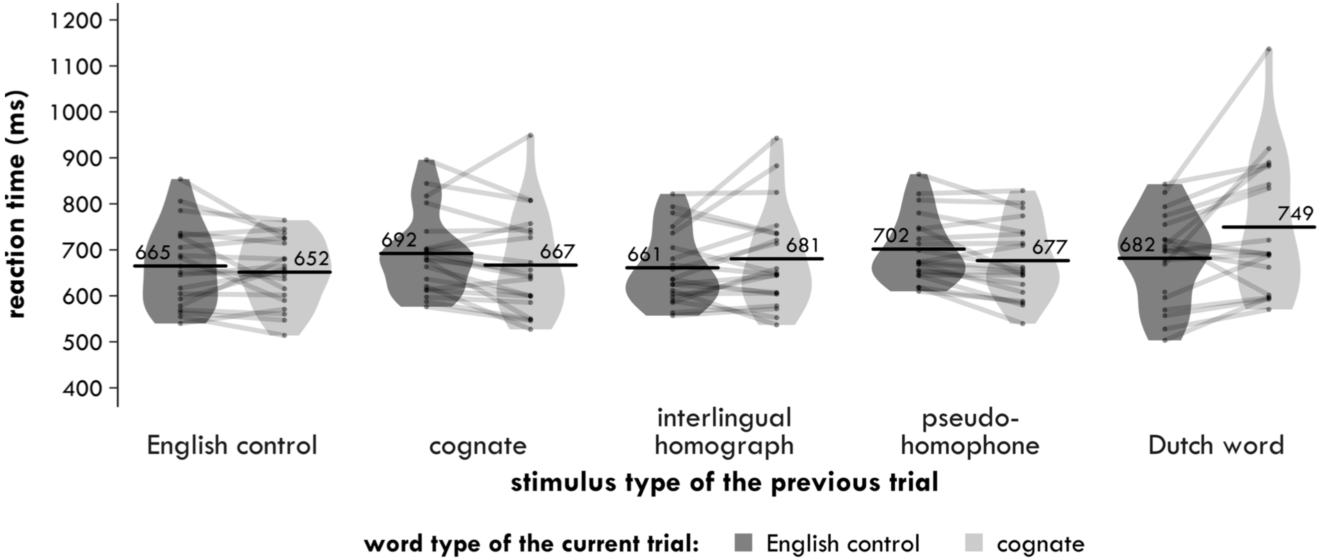


**Figure 3-7:** Experiment 2. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by stimulus type of the preceding trial (cognate, English control, interlingual homograph, pseudohomophone, Dutch word; x-axis) and word type of the current trial (English control, dark grey; cognate, light grey). Each point represents a condition mean for a participant with lines connecting word type means of the current trial from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees. *Note*. Three participants did not have any data for at least one condition. Their data is not plotted, but was included in the analysis.

# 4 General discussion

The aim of Experiment 1 and 2 was to determine whether the cognate facilitation effect in bilingual lexical decision is affected by the other types of stimuli included in the experiment. In Experiment 1, cognates in the standard version of the 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., 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). 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. 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 cognate effect in the mixed version was significantly smaller than the effect in the standard version. 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. These findings also suggest that it is indeed the case that the size and direction of the cognate effect can be influenced by stimulus list composition.

Specifically, as Vanlangendonck (2012) also found, 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 mixed and +DW versions), 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, the Dutch words also affected the cognates more directly on a trial-by-trial basis: when the preceding trial had been a Dutch word, cognates were recognised more slowly than the English controls, by 50 ms and 49 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, there was 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, Experiment 1) and Dijkstra et al. (1999, Experiment 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, these findings are fully consistent with the idea that the participants constructed a response strategy specifically to account for and respond accurately to the stimuli they encountered during the experiment. It appears that in single-language lexical decision tasks that do not include non-target language words (such as the 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 Chapter 1). Importantly, this is only possible because in such tasks, the participants merely need to decide whether the stimuli they see are word-like or familiar.

In contrast, when single-language lexical decision tasks *do* include non-target language words to which the participants should respond "no" (such as in the 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. 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. It appears 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.

These 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 (Experiment 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 (Experiment 2). Indeed, one of the exploratory analyses replicates this finding. In the +IH version of Experiment 2, which did not include any Dutch words, there was no significant difference between the interlingual homographs and the English controls (although there was a 8 ms trend towards inhibition). In contrast, there was a significant interlingual homograph inhibition effect of 43 ms and 24 ms in the mixed versions of both Experiment 1 and 2, respectively, which did include Dutch words. The interaction between word type and version between the +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 was the case for the cognates in this experiment, 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 these experiments.

In contrast, the 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 there is 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), it seems 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 contrast, these findings do not fit within the distributed framework of the DFM as it is currently proposed. However, these findings could be incorporated in this model by including a task system like the BIA+ model.

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. However, 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.

# CHAPTER 4: Studies of cross-lingual long-term priming

An earlier version of this chapter is available as a pre-print on the PsyArXiv server:

Poort, E. D., & Rodd, J. M. (2017, May 30). Studies of cross-lingual long-term priming. *PsyArXiv*. doi: 10.17605/OSF.IO/ERT8K. Retrieved from psyarxiv.com/ert8k.

# 1 Introduction

As discussed in Chapter 1, the size of the cognate facilitation effect is greater for identical cognates like “wolf”–“wolf” than for non-identical cognates like “kat”–“cat” (Comesaña et al., 2015; Dijkstra et al., 2010; Duyck et al., 2007; Font, 2001; Van Assche et al., 2009). However, there is little agreement in the literature as to whether there is a linear or non-linear relationship between the degree of orthographic similarity and the size of the facilitation effect. Consequently, researchers differ in their interpretations of how non-identical cognates are stored in the bilingual mental lexicon. While both the localist and distributed connectionist account agree that identical cognates must share their orthographic representation, proponents of the localist distributed BIA+ model (Dijkstra & Van Heuven, 2002) argue that non-identical cognates must have two separate, language-specific orthographic representations (e.g. Comesaña et al., 2015; Dijkstra et al., 2010). In other words, this account claims that there must be a *qualitative* difference between how identical and non-identical cognates are stored in the bilingual mental lexicon.

Dijkstra et al. (2010) base their conclusion on the finding that in Experiment 1 of a set of three experiments, there was an additional benefit for identical cognates compared to non-identical cognates, over and above a linear effect of orthographic similarity. In this experiment, 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 very similar (“metaal”–“metal”) to identical (“menu”–“menu”). Overall, the results showed that as orthographic similarity increased, the lexical decision reaction times decreased. On top of that, Dijkstra et al. (2010) observed a steep decline in the reaction times going from non-identical cognates like “metaal”–“metal” to identical cognates like “menu”–“menu”. They conclude that this must mean that non-identical cognates consist of two orthographic nodes: the lateral inhibition between these two orthographic representations would have cancelled out a large part of the facilitation effect that is due to the non-identical cognates’ shared semantic representation.

Comesaña et al. (2015) draw the same conclusion. They only found evidence for a cognate facilitation effect for identical cognates and not for non-identical cognates. In two experiments, they 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; Experiment 1a) or only to non-identical cognates and Spanish control words (Experiment 2). In Experiment 1a, they observed an overall significant facilitation effect for the identical and non-identical cognates together compared to the control words, but the follow-up analyses revealed only a significant facilitation effect for the identical cognates and not the non-identical cognates. Furthermore, in Experiment 2, they found a significant inhibition effect of approximately 20 ms for the non-identical cognates compared to the control words. Comesaña et al. (2015) conclude on the basis of these findings 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 second experiment.

In contrast, data collected by Duyck et al. (2007) and Van Assche et al. (2011) suggest a more graded nature for the bilingual lexicon. For example, Van Assche et al. (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 (Experiment 2).

The results of these two experiments, in contrast to Dijkstra et al.'s (2010) and Comesaña et al.'s (2015) results, indicate that there is a *quantitative* difference in the representations of identical and non-identical cognates. Indeed, their results are more in line with the view of the Distributed Feature Model (De Groot, 1992, 1993, 1995; De Groot et al., 1994; Van Hell, 1998; Van Hell & De Groot, 1998), which suggests that the only difference between identical and non-identical cognates is that their orthographic patterns are less similar. In other words, while the orthographic pattern of an identical cognates would be completely shared between the two languages, the two language-specific patterns of orthographic features for a non-identical cognate would be highly similar and overlapping (but not separate).

The cross-lingual long-term priming paradigm employed 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. As discussed previously, the Dutch–English participants in Poort et al.'s (2016) experiment made English lexical decisions to identical cognates and interlingual homographs that they had seen only 16 minutes before in Dutch sentences (e.g. "Alleen vrouwelijke bijen en wespen hebben een **angel**", i.e. "Only female bees and wasps have a **sting**"). 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 the effect of this cross-lingual long-term priming was beneficial for the cognates but disruptive for the interlingual homographs.

Because this kind of priming is thought to modify and strengthen the connections between word forms and meanings (Rodd et al., 2016; Rodd et al., 2013), this method can be used to shed a light on the issue of the representation of non-identical cognates. For the identical cognates in Poort et al.'s (2016) experiment, priming was facilitative because identical cognates share both their form and meaning. In contrast, the identical interlingual homographs only shared their form and not their meaning. When these words were primed in Dutch, the connection between the interlingual homograph's form and its Dutch meaning was strengthened. Because the participants had to access the English meaning in the English lexical decision task, priming was then disruptive.

If non-identical cognates, like identical cognates, share an orthographic representation, cross-lingual long-term priming should be facilitative. If instead, non-identical cognates do not share a form representation, cross-lingual long-term priming could be either ineffective or disruptive. In Rodd et al.'s (2013) Experiment 3, long-term priming was not effective when a synonym (e.g. "supporter") was used to prime the ambiguous word (e.g. "fan"), according to Rodd et al. (2013) because synonyms only share their meaning and not their form representation. Depending on exactly how cross-lingual long-term priming works, however, this method could also be disruptive for the non-identical cognates, as it was for the interlingual homographs in Poort et al.'s (2016) experiment. Experiment 3 was designed to determine whether and how cross-lingual long-term priming affects non-identical cognates. After finding only weak evidence for a priming effect based on the results of Experiment 3, the aim of Experiment 4 was to determine whether the size of the priming effect is influenced by the presence of interlingual homographs in the experiment.

# 2 Experiment 3: Cross-lingual long-term priming of identical and non-identical cognates in lexical decision

## 2.1 Introduction

The procedure for Experiment 3 was identical to the procedure employed by Poort et al. (2016), except that this experiment did not include any interlingual homographs so that more identical and non-identical cognates could be included. In the priming phase, Dutch–English bilinguals read Dutch sentences that contained either an identical cognate or a non-identical cognate. A set of translation equivalents was also primed in Dutch to create a semantic priming control condition. After a filler task of approximately 20 minutes, the identical cognates, non-identical cognates and translation equivalents were presented again in isolation in an English lexical decision task.

In line with Poort et al. (2016), a facilitative effect of priming was predicted for the identical cognates. If 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 priming is ineffective or disruptive for the non-identical cognates, this would suggest that they do not share a form representation. Based on research in the monolingual domain with synonyms (Rodd et al., 2013), the priming manipulation was not predicted to be successful 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 overwhelming amount of evidence in favour of a universal cognate facilitation effect, a large cognate facilitation effect was predicted for the identical cognates compared to the translation equivalents and a smaller effect for the non-identical cognates, or even no effect at all.[11]

## 2.2 Methods

### 2.2.1 Participants

The aim was to recruit at least 30 participants, in line with Poort et al. (2016). Participants were recruited for a lab-based experiment that took place in London and were eligible to participate in the experiment if they were a native speaker of Dutch and fluent speaker of English, with no diagnosis of a language disorder. They also had to be between the ages of 18 and 50 and of Dutch or Belgian nationality. In the end, a total of 33 participants was recruited through social media and word-of-mouth. The participants gave informed consent and were paid £6 for their participation in the experiment.

[11] Note that Experiment 3 included Dutch words in the English lexical decision task, so it may seem strange that a facilitation effect for the cognates was still predicted. However, this experiment was chronologically conducted before Experiment 1 and 2, so these predictions were formulated prior to the discovery that including non-target language (Dutch) words in the English lexical decision task reduces the cognate facilitation effect.

Despite stating the eligibility criteria upfront, one participant indicated afterwards in the demographics questionnaire that they were not a native speaker of Dutch. This participant did score very highly on the Dutch version of the LexTALE (96.3%), so this participant was not excluded from the analysis. Similarly, another participant indicated they were not a fluent speaker of English, but their score on the English version of the LexTALE was high (88.8%) so they were also not excluded. The data from two other 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 [$t(28) = 4.209$, $p < .001$]. These self-ratings were confirmed by their high LexTALE scores in both languages, which were also higher in Dutch, though a two-sided paired $t$-test showed this difference was not significant [Dutch: $M$ = 92.8%, $SD$ = 5.5%; English: $M$ = 90.3%, $SD$ = 7.9%; $t(28) = 1.405$, $p = .171$].

## 2.2.2 Materials

Table 4-1 lists the number of items of each stimulus type included in the two main tasks of the experiment. The tables in Appendix A indicate which identical cognates, non-identical cognates and translation equivalents were included in Experiment 3.

#### 2.2.2.1 Words & sentences

The software package Match (Van Casteren & Davis, 2007) was used to select 58 identical cognates, 58 non-identical cognates and 58 translation equivalents from the pre-tested materials from Chapter 2. Matching was based on Dutch and English log-transformed word frequency (weights: 1.5), the number of letters of the Dutch and English words (weights: 1.0) and orthographic complexity of the words in Dutch and English using the words' mean orthographic Levenshtein distance to its 20 closest neighbours (OLD20; Yarkoni et al., 2008; weights: 0.5). Table 4-2 lists means and standard deviations per word type for each of these measures (and raw word frequency), as well as the meaning, spelling and pronunciation similarity ratings obtained from the pre-tests and prime sentence length.

| | N | | identical cognates | non-identical cognates | interlingual homographs | translation equivalents | pseudo-homophones | Dutch words | task durations | average delay |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | **NUMBER OF ITEMS PER STIMULUS TYPE** | | | | | | | |
| **Experiment 3** | 28 | semantic relatedness task (priming task) | 58 | 58 | 0 | | – | – | 12:13 | 19:27 |
| | | lexical decision task (testing task) | 58 | 58 | 0 | 58 | 144 | 30 | 13:36 | |
| **Experiment 4** | | | | | | | | | | |
| –interlingual homographs | 33 | semantic relatedness task (priming task) | 50 | 50 | 0 | 0 | – | – | 07:13 | 17:13 |
| | | lexical decision task (testing task) | 50 | 50 | 0 | 50 | 125 | 25 | 14:02 | |
| +interlingual homographs | 34 | semantic relatedness task (priming task) | 50 | 50 | 50 | 0 | – | – | 10:20 | 17:54 |
| | | lexical decision task (testing task) | 50 | 50 | 50 | 50 | 166 | 34 | 17:08 | |

**Table 4-1:** Experiment 3 & 4. Overview of the types and numbers of stimuli included in (each version of) Experiment 3 and 4, 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 (for that version).

| | Dutch characteristics | | | | English characteristics | | | | similarity ratings | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | meaning | spelling | pronunciation | sentence length |
| identical cognates | 37.0 (56.3) | 2.90 (0.49) | 4.57 (1.11) | 1.61 (0.42) | 41.5 (54.0) | 3.07 (0.47) | 4.57 (1.11) | 1.63 (0.35) | 6.83 (0.22) | 7.00 (0.01) | 5.91 (0.67) | 9.59 (1.63) |
| non-identical cognates | 31.4 (34.8) | 2.89 (0.48) | 4.76 (0.98) | 1.50 (0.34) | 44.0 (50.7) | 3.14 (0.43) | 4.72 (0.95) | 1.58 (0.34) | 6.87 (0.19) | 5.35 (0.55) | 5.05 (0.76) | 9.50 (1.88) |
| translation equivalents | 29.7 (31.4) | 2.90 (0.45) | 4.74 (0.91) | 1.46 (0.30) | 34.2 (31.1) | 3.07 (0.41) | 4.67 (0.98) | 1.62 (0.33) | 6.88 (0.15) | 1.14 (0.25) | 1.12 (0.21) | 9.50 (1.49) |

**Table 4-2:** Experiment 3. Means (and standard deviations) for the Dutch and English characteristics, similarity ratings and prime sentence length for the identical and non-identical cognates and the translation equivalents. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; sentence length refers to the length of the Dutch prime sentences shown during the semantic relatedness task.

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 OLD20 [$t(104.0) = 2.200$, $p = .030$]. No other differences between the word types on these matching criteria were significant (all *p*s > .1), nor were the differences in prime sentence length significant (all *p*s > .7). All three word types were significantly different from each other in terms of spelling and pronunciation similarity ratings (all *p*s < .001), but not with respect to the meaning similarity ratings (all *p*s >.1). This confirmed the word types' intended status.

#### 2.2.2.2 Non-words

The English lexical decision task included the same number of non-words as words. The 174 non-words comprised 144 English-sounding pseudohomophones (e.g. "mistaik") selected from Rodd (2000) and 30 Dutch words (e.g. "vijand") of a similar frequency as the target items, selected pseudo-randomly from the SUBTLEX-NL database (Keuleers et al., 2010). The Dutch words were included because Poort et al. (2016) included some Dutch words as well (which they did to ensure participants only responded "yes" to English words). Similarly, pseudohomophones were used instead of regular non-words because Poort et al. (2016) had as well (they did this to to encourage relatively deep processing; see Rodd et al., 2002). The non-words were of a similar length as the target items.

### 2.2.3 Design and procedure

This experiment employed a mixed design. Priming was a within-participants/within-items factor: 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 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. Word type was a within-participants/between-items factor: participants saw words from all three word types, but each word of course belonged to only one word type.

The experiment comprised five separate tasks: (1) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012), (2) the Dutch semantic relatedness task, (3) the English digit span task, (4) the English lexical decision task and (5) the English version of the LexTALE (Lemhöfer & Broersma, 2012). At the end of the experiment, participants completed a self-report language background survey in Dutch. The experiment was created using MATLAB (version R2012a; The Mathworks Inc., 2012) and conducted at the Department of Experimental Psychology at University College London. 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.

#### 2.2.3.1 Dutch semantic relatedness task

This task served to prime the identical cognates, non-identical cognates and translation equivalents. The 58 sentences for each word type were pseudo-randomly divided into two sets of 87 sentences in total, matched for all key variables, for use in the two versions of the experiment. To ensure they processed the prime sentences, participants were asked to indicate by means of a button press, as quickly and accurately as possible, whether a subsequent probe was semantically related to the preceding sentence. Half of the sentences in each version were paired with related probes and half with unrelated probes. The probes were either very strongly related or completely unrelated to the sentence. The same native speaker of Dutch that had proofread the sentences during stimulus development also confirmed that the probes were (un)related to the sentence as intended.

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

#### 2.2.3.2 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 switching cost on the lexical decision task. Each string of digits comprised four to eight digits. Participants saw five practice strings followed by 36 experimental strings divided into two blocks. A 10-second break was enforced after the first block. Each digit was presented for 500 ms. Depending on the string length, participants had between 4000 ms and 6000 ms to recall the sequence.  The inter-trial interval was 1000 ms.

#### 2.2.3.3 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 a button press, as quickly and accurately as possible, whether it was a *real* English word or not (emphasis was also present in the instructions). A practice block of 24 strings was followed by eight blocks of 44 or 43 experimental stimuli.

The 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 item. 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, with a five-second break after each block. Each item remained on screen until the participant responded or until 2000 ms passed. The inter-trial interval was 500 ms.

## 2.3 Results

Although Experiment 3 was not pre-registered, the data were analysed following same procedures as for the confirmatory analyses conducted for Experiment 4, which was pre-registered. All analyses were carried out in R (version 3.3.2; R Core Team, 2016) using the lme4 package (version 1-1.12; Bates et al., 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. Reaction times were analysed using the lmer() function with the default optimiser; accuracy data were analysed using the glmer() function with the bobyqa optimiser. Again, it should be noted that the graphs in the figures display 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 summary.

In addition to frequentist statistics, as per the pre-registration, Bayes factors are also reported. These were computed using the following formula suggested by Wagenmakers (2007): $BF_{10} = e^{\frac{BIC_{null}-BIC_{alternative}}{2}}$. Jeffrey's (1961) guidelines were followed when interpreting the Bayes factors: values between 1 and 3 were considered as anecdotal evidence for the alternative hypothesis, values between 3 and 10 as moderate evidence, between 10 and 30 as strong evidence, between 30 and 100 as very strong evidence and greater than 100 as decisive. In contrast, values between 1/3 and 1 were considered as anecdotal evidence for the null hypothesis, values between 1/10 and 1/3 as moderate evidence, between 1/30 and 1/10 as strong evidence, between 1/100 and 1/30 as very strong evidence and less than 1/100 as decisive. However, since these Bayes factors have an uninformative prior (i.e. they assume an effect of zero), all conclusions are based on the frequentist statistics.

Three items (the identical cognate "fort"–"fort", the non-identical cognate "daad"–"deed" and the translation equivalent "mier"–"ant") were excluded from the analyses, 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 items' word type mean (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.

### 2.3.1 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.

### 2.3.2 Semantic relatedness task: 'Confirmatory' analyses

High accuracy scores (*M* = 93.6%, *SD* = 3.8%, *range* = 85.1%–98.9%) confirmed the participants had processed the sentence meanings. To determine whether any of the observed effects of priming in the lexical decision task could have been due to differences between the word types at the time of priming, an analysis with the fixed factor word type (3 within-participants/between-items levels: identical cognate, non-identical cognate, translation equivalent) was conducted on the accuracy data. The maximal model converged and included a correlated random intercept and random slope for word type by participants and a random intercept by items. This analysis revealed that the main effect of word type was not significant [$\chi^2(2) = 1.225$, $p = .542$, $BF_{10} < 0.001$; $M_{\text{identical cognates}} = 98.1\%$, $M_{\text{non-identical cognates}} = 97.0\%$, $M_{\text{English controls}} = 97.7\%$]. The Bayes factor also provided decisive evidence for the null hypothesis.

### 2.3.3 Lexical decision task: 'Confirmatory' analyses

#### 2.3.3.1 Analysis procedure

The same analysis procedure was employed for the reaction times and accuracy data. In all cases, positive effects of priming indicate a facilitative effect of priming (i.e. faster reaction times and higher accuracy for primed items), while negative effects indicate a disruptive effect of priming (i.e. slower reaction times and lower accuracy for primed items). Positive (negative) effects of word type indicate an advantage (disadvantage) for the first-named word type over the second-named word type.

Two fixed factors were included in the main 2×3 analysis: priming (2 within-participant/within-items levels: unprimed, primed) and word type (3 within-participant/between-items levels: identical cognate, non-identical cognate, translation equivalents). The maximal random effects structure of this model included a correlated random intercept and random slopes for word type, priming and their interaction by participants and a correlated random intercept and random slope for priming by items. This maximal model converged for the accuracy analysis but not the reaction time analysis, for which a model without correlations between the random effects converged.

To examine the effect of priming for each of the three word types separately, three simple effects analyses were conducted. The maximal model random effects structure of this model included a correlated random intercept and random slope for priming by both participants and items. This model converged for the reaction times analysis but not the accuracy analysis, for which a model without correlations between the random effects converged. The *p*-values for these three analyses were compared against a Bonferroni-corrected $\alpha$ of .017.

Finally, three pairwise comparisons were conducted on the unprimed data only, comparing the identical cognates, non-identical cognates and translation equivalents to each other, in an attempt to replicate the findings from Experiment 1 and 2. The maximal model converged for both the reaction times and accuracy analysis and included a correlated random intercept and random slope for word type by participants and a random intercept by items. The *p*-values for these three analyses were also compared against a Bonferroni-corrected $\alpha$ of .017.

#### 2.3.3.2 Reaction times

Reaction times are shown in Figure 4-1. Reaction times (RTs) faster than 300ms or slower than 1500ms were discarded (2.8% of the data), as were RTs for incorrect trials and trials that participants had not responded to (1.5% of the remaining data). The RTs were inverse-transformed (inverse-transformed RT = 1000/raw RT), as a histogram of the residuals and a predicted-vs-residuals plot for the main 2×3 analysis showed that the assumptions of normality and homoscedasticity were violated. After inverse-transforming the RTs, inverse-transformed RTs were removed that were more than three standard deviations above or below a participant's mean inverse-transformed RT (0.2% of the remaining data).

***2×3***

In the 2×3 analysis, the main effect of priming was marginally significant [$\chi^2(1) = 3.261$, $p = .071$, $BF_{10} = 0.074$], with a facilitative effect of 7 ms. The main effect of word type was not significant [$\chi^2(2) = 4.123$, $p = .127$, $BF_{10} = 0.002$], nor was the interaction between word type and priming [$\chi^2(2) = 0.226$, $p = .893$, $BF_{10} < 0.001$]. All Bayes factors provided strong or even decisive evidence for the null hypothesis.

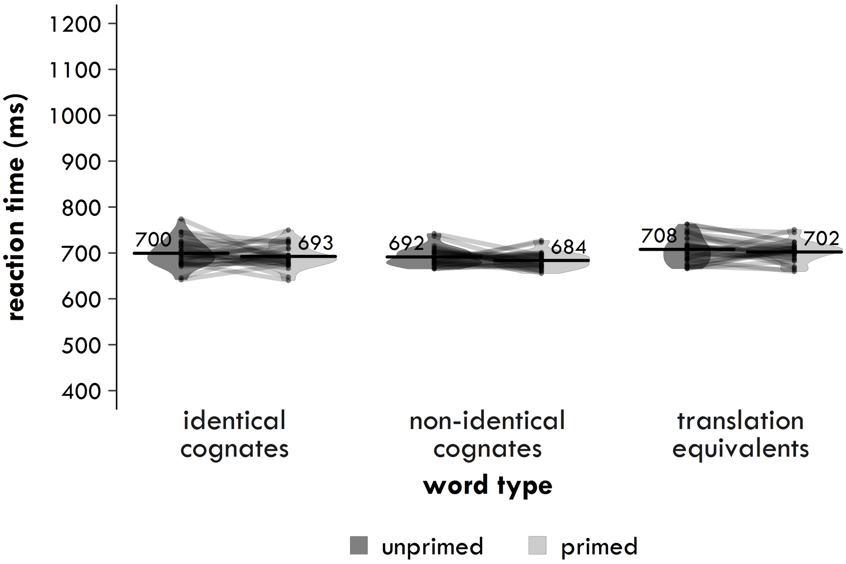


**Figure 4-1:** Experiment 3. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by word type (identical cognates, non-identical cognates, translation equivalents; x-axis) and priming (unprimed, dark grey; primed, light grey). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

***Simple effects***

In line with these findings, none of the simple effects of priming were significant [for the identical cognates: $\chi^2(1) = 0.852$, $p = .356$, $BF_{10} = 0.039$, $\Delta = 6$ ms; for the non-identical cognates: $\chi^2(1) = 2.053$, $p = .152$, $BF_{10} = 0.070$, $\Delta = 9$ ms; for 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.

***Pairwise comparisons***

None of the three pairwise comparisons on the unprimed trials 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[12]: $\chi^2(1) = 0.569$, $p = .451$, $BF_{10} = 0.034$, $\Delta = 8$ ms; non-identical cognates vs translation equivalents: $\chi^2(1) = 2.636$, $p = .105$, $BF_{10} = 0.094$, $\Delta = 20$ ms]. The Bayes factors again provided strong evidence for the null hypothesis.

---

[12] 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, the analysis was not re-run with the Dutch OLD20 measure included as a covariate.

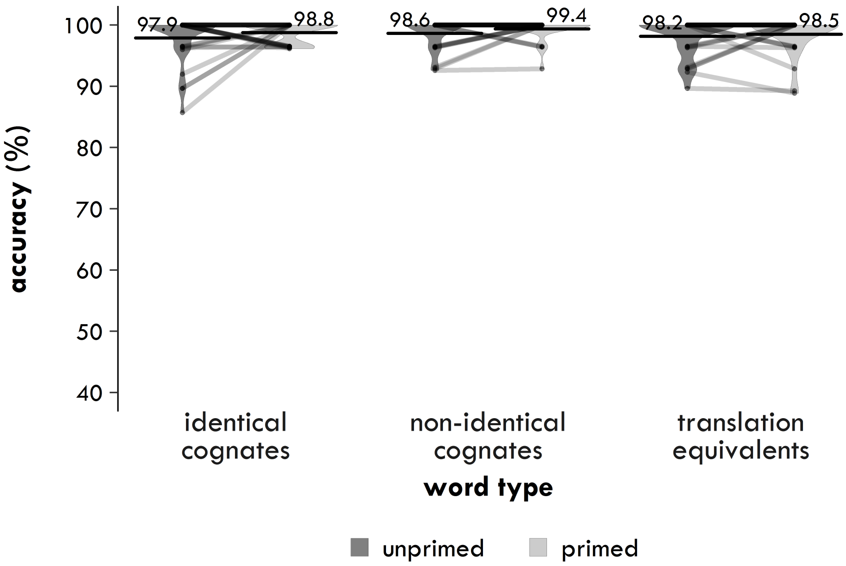


**Figure 4-2:** Experiment 3. Participant means of lexical decision accuracy (percentages correct) by word type (identical cognates, non-identical cognates, translation equivalents; x-axis) and priming (unprimed, dark grey; primed, light grey). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

#### 2.3.3.3 Accuracy

Accuracy is shown in Figure 4-2. In line with the trimming procedure for the reaction times, any trials with RTs faster than 300ms or slower than 1500ms were removed. Broadly speaking, the analyses on the accuracy data revealed a similar pattern of results as the analyses on the reaction time data.

***2×3***

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

***Simple effects***

Unsurprisingly, none of the simple effects of priming were significant [for the identical cognates: $\chi^2(1) = 0.339$, $p = .560$, $BF_{10} = 0.029$, $\Delta = 0.1\%$; for the non-identical cognates: $\chi^2(1) = 1.669$, $p = .196$, $BF_{10} = 0.057$, $\Delta = 0.2\%$; for the 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.

***Pairwise comparisons***

Again, none of the three pairwise comparisons on the unprimed trials 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 equivalents: $\chi^2(1) = 0.047$, $p = .828$, $BF_{10} = 0.026$, $\Delta = 0.1\%$]. The Bayes factors also provided very strong evidence for the null hypothesis.

## 2.4 Discussion

In contrast to the predictions, there was 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, there was also no evidence for a cognate facilitation effect in the unprimed trials, either for the identical cognates (a facilitative effect of 8 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, although this pattern of results is not in line with the predictions, it is fully consistent with the experiments presented in Chapter 3. However, the fact that there was no 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 this experiment were relatively easy words, so the 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, as they may have done in Experiment 2 (Chapter 3). Experiment 4 was designed to determine whether priming is indeed affected by the presence of interlingual homographs in the experiment.

# 3 Experiment 4: Cross-lingual long-term priming of identical and non-identical cognates and identical interlingual homographs in lexical decision

Experiment 4 was again pre-registered as part of the Center for Open Science's Preregistration Challenge (www.cos.io/prereg). The stimuli, data and processing and analysis scripts can be found on the Open Science Framework (www.osf.io/dytp7), as well as an Excel document with detailed results from all of the analyses. The preregistration can be retrieved from www.osf.io/33r86 (Poort & Rodd, 2017, January 24). Where applicable, deviations from the preregistration will be noted.

## 3.1 Introduction

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. The translation equivalents were not primed in this experiment, as this made it possible to include more items of the critical word types without increasing the experiment's duration. 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, it is called 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 3, the following predictions were made regarding the priming effect: (1) there will be evidence for a cross-lingual long-term priming effect in the +IH version, (2) there will be no evidence for a cross-lingual long-term priming effect in the –IH version and (3) priming in the +IH version will be facilitative for the identical cognates and disruptive for the interlingual homographs. 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, with a largely shared form representation. In contrast, if priming is ineffective or disruptive for the non-identical cognates, this would indicate that they do not share their form representation.

For the unprimed trials, the following predictions are made: (1) there will not be a cognate facilitation effect for the identical or non-identical cognates in either version of the experiment but (2) there will be 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 2 (Chapter 3) and Experiment 3 of this chapter. The prediction that there will be evidence for an interlingual homograph inhibition effect also follows from Experiment 1 and 2 (Chapter 3), as well as the many other experiments discussed in that chapter (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.

# 3.2 Methods

## 3.2.1 Participants

Given the uncertainty surrounding the size of the priming effect, the aim was again to recruit at least 32 participants per version, consistent with Experiment 3. To be able to recruit participants more easily, participants were again recruited that were resident in the Netherlands or Belgium at the time of the experiment for a web-based experiment. In the end, a total of 74 participants were recruited through Prolific Academic (Damer & Bradley, 2014), social media and personal contacts resident in the Netherlands. 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.

As in the other web-based experiments (Experiment 1 and 2), excluding participants happened in two stages. First, while testing was still on-going, participants who scored less than 80% correct on the semantic relatedness task or on the lexical decision task and/or scored less than 50% on either of the two language proficiency measures were excluded and replaced. Participants were also excluded and replaced if their average delay between priming and testing was more than 30 minutes[13]. Four participants (three in the –IH version and one in the +IH version) 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 (all in the –IH version) could not be used, so these participants were also replaced. Finally, it should be noted that although the intention was to exclude participants whose internet connection was not fast or stable enough, the software frequently filtered out participants whose internet connection appeared to be fine when tested with an online internet speed test, so the internet

[13] Gorilla saved timestamps of when each item was presented during the priming task and during the testing task, which were used to calculate the priming delay for each item. The average delay for each participant was simply the average of their delay across all primed items.

connection filter was disabled 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, 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 was compared to the mean of all participants who completed that version to determine whether any more participants needed to be excluded. All participants had 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 as 9.5 out of 10 in Dutch ($SD = 0.6$) and 8.7 in English ($SD = 0.9$). A two-sided paired *t*-test showed this difference to be significant [$t(66) = 8.246, p < .001$]. 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 [$M = 85.2\%$, $SD = 8.1\%$; $t(60.3) = -2.368$, $p = .021$]. No other differences between the two versions with respect to the demographic variables reported here were significant (as shown by additional independent-samples Welch's *t*-tests and chi-square tests where appropriate; all *p*s > .08).

### 3.2.2 Materials

Table 4-1 lists the number of items of each stimulus type included in the two main tasks of the two versions of the experiment. The tables in Appendix A indicate which identical cognates, non-identical cognates, interlingual homographs and English controls were included in Experiment 4.

| | Dutch characteristics | | | | English characteristics | | | | Similarity ratings | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | frequency | log10(fre-quency) | word length | OLD20 | frequency | log10(fre-quency) | word length | OLD20 | meaning | spelling | pronun-ciation | sentence length |
| identical cognates | 39.0 (59.4) | 2.92 (0.49) | 4.48 (0.97) | 1.58 (0.40) | 44.8 (57.0) | 3.10 (0.48) | 4.48 (0.97) | 1.59 (0.31) | 6.85 (0.22) | 7.00 (0.00) | 5.87 (0.68) | 9.46 (1.64) |
| non-identical cognates | 32.2 (35.4) | 2.90 (0.49) | 4.66 (0.98) | 1.45 (0.34) | 45.4 (51.6) | 3.15 (0.44) | 4.62 (0.92) | 1.54 (0.30) | 6.89 (0.18) | 5.34 (0.57) | 5.04 (0.73) | 9.28 (1.92) |
| interlingual homographs | 39.2 (82.9) | 2.70 (0.68) | 4.30 (0.93) | 1.34 (0.35) | 56.2 (108) | 2.98 (0.66) | 4.30 (0.93) | 1.46 (0.32) | 1.15 (0.28) | 7.00 (0.01) | 5.53 (0.79) | 9.20 (1.81) |
| translation equivalents | 27.8 (26.3) | 2.88 (0.46) | 4.68 (0.87) | 1.43 (0.31) | 30.2 (26.0) | 3.02 (0.41) | 4.46 (0.93) | 1.57 (0.31) | 6.86 (0.19) | 1.12 (0.24) | 1.11 (0.21) | – |

**Table 4-3:** Experiment 4. Means (and standard deviations) for the Dutch and English characteristics, similarity ratings and prime sentence length for the identical and non-identical cognates, interlingual homographs and the translation equivalents. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; sentence length refers to the length of the Dutch prime sentences shown during the semantic relatedness task.

#### 3.2.2.1 Words & sentences

Fifty interlingual homographs were selected from the set of 56 interlingual homographs included in Experiment 1 and 2 to match a subset of 50 identical cognates and 50 non-identical cognates selected from the stimuli included in Experiment 3. A subset of 50 items was used for each word type as it was possible to achieve better matching with the interlingual homographs with a smaller set of items. A set of 50 English control words was also selected from the sets included in Experiment 1 and 2 and in Experiment 3 to serve as matched fillers in the lexical decision task[14]. 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 4-3 lists means and standard deviations per word type for each of these measures (and raw word frequency) for both English and Dutch, as well as the meaning, spelling and pronunciation similarity ratings obtained from the pre-tests and prime sentence length. Minor changes were made 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. All modified sentences were proofread by the same native speaker of Dutch who proofread the sentences for the rating experiments. Only the identical and non-identical cognates and the English controls were included in the –IH version, for a total of 150 words included in that version; the +IH version also included the 50 interlingual homographs for a total of 200 words.

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 > .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 ($p = .299$). 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 ($p = .109$). 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

[14] All but one of the English controls selected for Experiment 4 had also been used in Experiment 1 and 2. The one item that had not been used in Experiment 1 and 2 had been used in Experiment 3, as had most other items.

significantly from each other in terms of pronunciation similarity ratings (all *p*s $<$ .03). This confirmed the word types' intended status.

#### 3.2.2.2 Non-words

Each version of the experiment again included the same number of non-words as words. For the largest part, the same materials were used as for Experiment 3. Where necessary, additional pseudohomophones were selected from the ARC Non-Word and Pseudohomophone database (Rastle et al., 2002) or newly created; some additional Dutch words were also selected from the SUBTLEX-NL database. This was in order to ensure that, in all versions, each word was matched in terms of length to a non-word.

### 3.2.3 Design and procedure

This experiment employed a mixed design again. Version was a between-participants/within-items factor: participants were randomly assigned to one of the two versions of the experiment, but each item was included in both versions (except the interlingual homographs which were, of course, only included in the +IH version). As in Experiment 3, word type was a within-participants/between-items factor and priming was a within-participants/within-items factor. Only the identical cognates, non-identical cognates and interlingual homographs were primed. Two subversions of each experimental version (–IH, +IH) were created such that participants saw each experimental item only once but across participants within each experimental version, items occurred in both the primed and unprimed condition.

As in Experiment 3, the experiment comprised five separate tasks: (1) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012), (2) the Dutch semantic relatedness task, (3) the Towers of Hanoi task (with instructions presented in English), (4) the English lexical decision task and (5) the English version of the LexTALE (Lemhöfer & Broersma, 2012). At the start of the experiment, the participants completed a self-report language background survey in Dutch to verify their eligibility to take part in the experiment. The experiment was created and conducted using Gorilla online experimental software (Evershed & Hodges, 2016), with the design of each task (except for the Towers of Hanoi task) being the virtually same as in Experiment 3. On average across participants, 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 [$t(58.2) = 1.232$, $p = .223$]. The five tasks were presented separately with no indication that they were linked.

#### 3.2.3.1 Dutch semantic relatedness task

The 50 sentences for the identical cognates and the 50 sentences for the non-identical cognates were pseudo-randomly divided into two sets, matched for all key variables, for a total of 50 sentences in each of the two subversions of the –IH version. Similarly, the 50 sentences for the interlingual homographs were also pseudo-randomly divided into two sets, matched for all key variables and prime sentence length, and combined with the two sets of identical and non-identical cognates for a total of 75 sentences in each of the two subversions of the +IH version. Again, half the sentences in each subversion of the experimental versions were paired with related items and half with unrelated items.

A practice block of four (–IH version) or six (+IH version) sentences was followed by two blocks of 25 (–IH version) or 37 or 38 (+IH version) experimental sentences. The order of the items within blocks was randomised for each participant, but the order of the blocks was fixed and counterbalanced across participants (making eight subversions of the experiment in total). 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. Each sentence remained on screen for 4000 ms; each probe until the participant responded or until 2500 ms passed. The inter-trial interval was 1000 ms.

#### 3.2.3.2 The Towers of Hanoi task

Like the digit span task in Experiment 3, this task served to introduce a delay between priming and testing, while minimising exposure to additional linguistic material. The instructions were presented in English to minimise any general language switching cost on the lexical decision task.

The Towers of Hanoi task is a puzzle task 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, a time limit was imposed on the Towers 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 that 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.

#### 3.2.3.3 English lexical decision task

Participants saw all 300 (–IH version) or 400 (+IH version) experimental stimuli. 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. As in Experiment 3, the 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 order of the items within blocks was randomised for each participant, as was the order of the blocks within each part. Five (–IH version) or six (+IH version) fillers were presented at the beginning of each block, with a 30-second break at the end of each block.

## 3.3 Results

This section reports the results of the planned confirmatory analyses. No exploratory analyses are reported. All analyses were carried out in R (version 3.3.2; R Core Team, 2016) using the lme4 package (version 1-1.12; Bates et al., 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. Bayes Factors are also reported again and were calculated in the same manner as for Experiment 3. Reaction times were analysed using the lmer() function with the default optimiser; accuracy data were analysed using the glmer() function with the bobyqa optimiser. Detailed results of all analyses for Experiment 4 can be found in the Excel document analysisResults.xlsx in the Analysis scripts component of the OSF project (www.osf.io/6qnfc). Again, the graphs in the figures display 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 summary.

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 for those items on the lexical decision task (88.1%, 81.5% and 85.1%, respectively) were more than three standard deviations below the items' word type mean (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.

### 3.3.1 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). This confirms the participants' engagement with the task.

### 3.3.2 Semantic relatedness task: Confirmatory analyses

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 the participants had processed the sentence meanings. To determine whether any of the observed effects of priming 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, a 2×2 analysis on the accuracy data was conducted with the fixed 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 correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for version by items. This analysis revealed that neither the main effect of word type [$\chi^2(1) = 0.852$, $p = .356$, $BF_{10} = 0.027$; $M_{\text{identical cognates}} = 98.8\%$, $M_{\text{non-identical cognates}} = 98.1\%$] nor the main effect of version [$\chi^2(1) = 1.149$, $p = .284$, $BF_{10} = 0.031$; $M_{\text{–IH version}} = 98.8\%$, $M_{\text{+IH version}} = 98.2\%$] 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, three pairwise comparisons were conducted on the data for the +IH version, to compare the identical cognates, non-identical cognates and interlingual homographs to each other. The maximal model also converged for these three analyses and included a correlated random intercept and random slope for word type by participants and a random intercept by items. The $p$-values for pairwise comparisons were compared against a Bonferroni-corrected $\alpha$ of .017. Again, these analyses revealed no significant differences between the word types [identical cognates ($M$ = 98.7%) vs non-identical cognates ($M$ = 98.4%): $\chi^2(1) = 0.114$, $p = .736$, $BF_{10} = 0.026$; identical cognates ($M$ = 98.5%) vs interlingual homographs ($M$ = 98.6%): $\chi^2(1) = 0.004$, $p = .950$, $BF_{10} = 0.024$; non-identical cognates ($M$ = 97.6%) vs interlingual homographs ($M$ = 98.0%): $\chi^2(1) = 0.150$, $p = .699$, $BF_{10} = 0.026$]. The Bayes factors also provided very strong evidence for the null hypothesis.

### 3.3.3 Lexical decision task: Confirmatory analyses

#### 3.3.3.1 Analysis procedure

The same analysis procedure was employed for the reaction times and accuracy data. In all cases, positive effects of priming indicate a facilitative effect of priming (i.e. faster reaction times and higher accuracy for primed items), while negative effects indicate a disruptive effect of priming (i.e. slower reaction times and lower accuracy for primed items). Positive (negative) effects of word type indicate an advantage (disadvantage) for the first-named word type over the second-named word type. Positive (negative) effects of version indicate an advantage (disadvantage) for the +IH version over the –IH version.

Three fixed factors were included in the main 2×2×2 analysis: priming (2 within-participant/within-items levels: unprimed, primed), word type (2 within-participant/between-items levels: identical cognate, non-identical cognate), and version (2 between-participants/within-items levels: –IH, +IH). The maximal random effects structure of this model included a correlated random intercept and random slopes for word type, priming and their interaction by participants and a correlated random intercept and random slope for priming by items. The maximal model did not converge for either the reaction times or accuracy analysis, so the correlations between the random effects were removed.

In addition, a 2×2 analysis was conducted for the –IH version, to compare the effect of priming for the identical and the non-identical cognates. Another three 2×2 analysis were conducted for the +IH version, to compare 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 maximal random effects structure for these models also included a correlated random intercept and random slope for word type, priming and their interaction by participants and a correlated random intercept and random slope for priming by items. 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. For these four 2×2 analyses, only the interactions between word type and priming were of interest, so significance was only determined for these effects and the resulting *p*-values were compared against a Bonferroni-corrected $\alpha$ of .013.

To examine the effect of priming for each of the three word types in the two version separately, five simple effects analyses were conducted. The maximal model converged for the reaction times analysis and included a correlated random intercept and random slope for priming by both participants and items, but not for the accuracy analysis, for which a model without the correlations between the random effects converged. The *p*-values for these five analyses were compared against a Bonferroni-corrected $\alpha$ of .01.

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

#### 3.3.3.2 Reaction times

Reaction times are shown in Figure 4-3. As in Experiment 3, reaction times (RTs) faster than 300ms or slower than 1500ms were discarded (1.8% of the data), as were RTs for incorrect trials and trials that participants had not responded to (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 and the two 2×2s that included the (non-)identical cognates and the interlingual homographs in the +IH version showed that the assumptions of normality and homoscedasticity were violated. After inverse-transforming the RTs, any inverse-transformed RTs were removed that were 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).

##### *2×2×2*

In the 2×2×2 analysis, the main effect of priming was not significant [$\chi^2(1) < 0.001$, $p = .999$, $BF_{10} = 0.013$, $\Delta = -0.01$ ms], nor was the main effect of word type [$\chi^2(1) = 0.188$, $p = .665$, $BF_{10} = 0.014$, $\Delta = -3$ ms]. The main effect of version was also not significant[15] [$\chi^2(1) = 2.254$, $p = .133$, $BF_{10} = 0.039$, $\Delta = 26$ ms]. The interaction 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 were significant (all $p$s > .8, all BF10s = 0.013). All Bayes factors provided strong or even very strong evidence for the null hypothesis.

##### *2×2 interactions*

In line with these results, none of the interactions between word type and priming in the four 2×2 analyses were significant and all Bayes factors provided very strong evidence for the null hypothesis (all $p$s > .3, all BF10s < 0.03).

[15] There was a small but significant difference between the participants in 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, the analysis was not conducted again with the Dutch LexTALE scores included as a covariate.

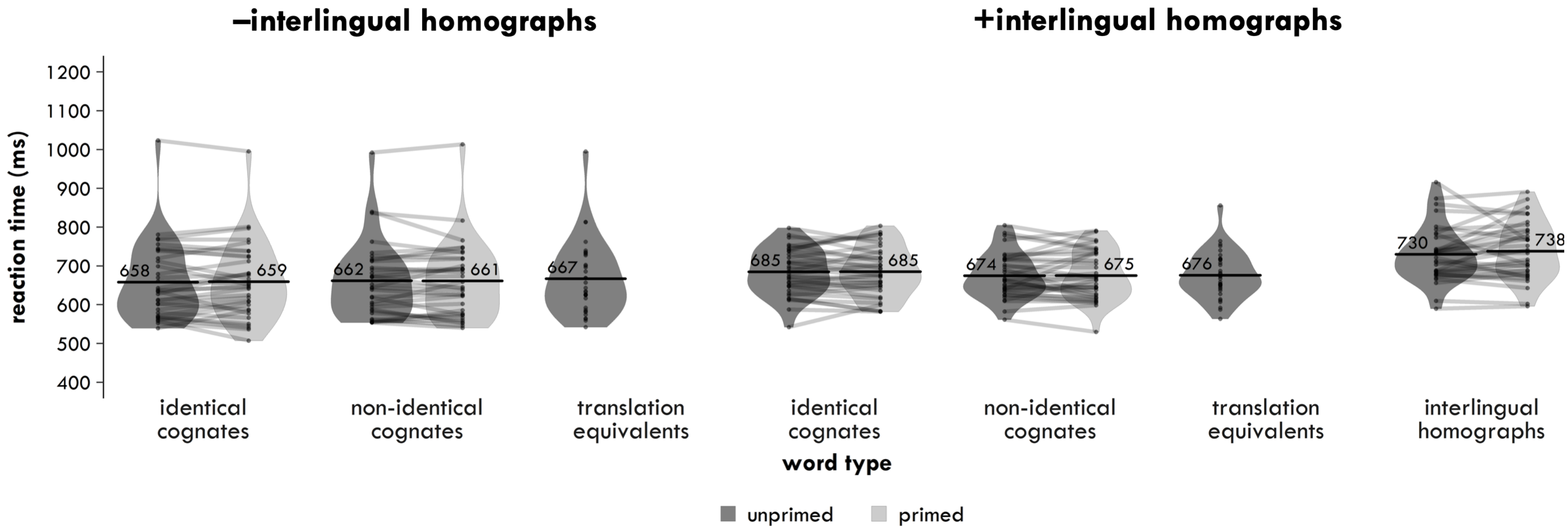


**Figure 4-3:** Experiment 4. Harmonic participant means of the inverse-transformed lexical decision reaction times (in milliseconds) by word type (identical cognates, non-identical cognates, translation equivalents, interlingual homographs; x-axis), priming (unprimed, dark grey; primed, light grey) and version (–interlingual homographs, +interlingual homographs). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

***Simple effects***

Unsurprisingly, none of the simple effects of priming were significant in either the –IH version [for the identical cognates: $\chi^2(1) = 0.023$, $p = .880$, $BF_{10} = 0.026$, $\Delta = -1$ ms; for the non-identical cognates: $\chi^2(1) < 0.001$, $p = .988$, $BF_{10} = 0.025$, $\Delta = 0.1$ ms]. None of the simple effects of priming in the +IH version were significant either [for the identical cognates: $\chi^2(1) < 0.001$, $p = .982$, $BF_{10} = 0.025$, $\Delta = -0.1$ ms; for the non-identical cognates: $\chi^2(1) = 0.034$, $p = .855$, $BF_{10} = 0.025$, $\Delta = -1$ ms; for 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.

***Pairwise comparisons***

As in Experiment 3, none of the pairwise comparisons were significant in the –IH version [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]. 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.

#### 3.3.3.3 Accuracy

Accuracy is shown in Figure 4-4. In line with the trimming procedure for the reaction times, any trials with RTs faster than 300ms or slower than 1500ms were removed.

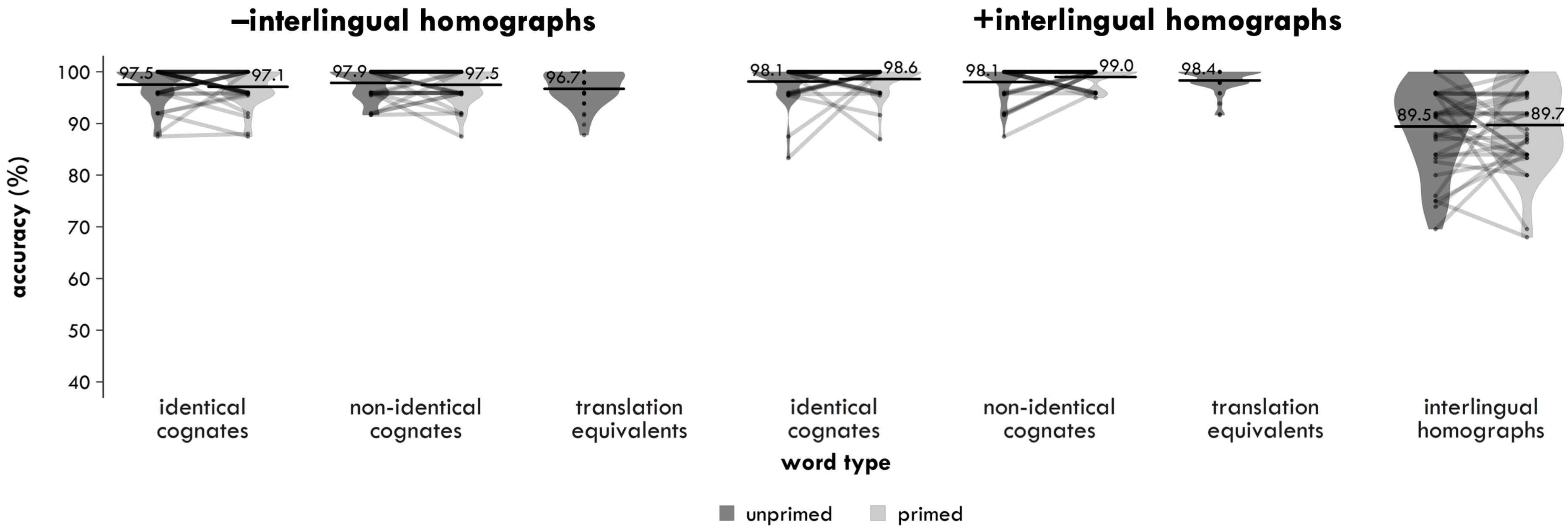


**Figure 4-4:** Experiment 4. Participant means of lexical decision accuracy (percentages correct) by word type (identical cognates, non-identical cognates, translation equivalents, interlingual homographs; x-axis), priming (unprimed, dark grey; primed, light grey) and version (–interlingual homographs, +interlingual homographs). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

Broadly speaking, the results for the accuracy data are essentially the same as those for the reaction time data.

### *2×2×2*

In the 2×2×2 analysis, the main effect of priming was not significant [$\chi^2(1) = 0.835$, $p = .361$, $BF_{10} = 0.019$, $\Delta = 0.2\%$], nor was the main effect of word type [$\chi^2(1) = 0.301$, $p = .583$, $BF_{10} = 0.014$, $\Delta = 0.1\%$]. 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.[16] In contrast, all Bayes factors provided moderately to very strong evidence for the null hypothesis. None of the two-way interactions nor 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).

### *2×2 interactions*

In line with these results, none of the interactions between word type and priming in the 2×2 analyses were significant and all Bayes Factors provided strong or even very strong evidence for the null hypothesis (all *p*s > .1, all BF10s < .05).

### *Simple effects*

Again unsurprisingly, none of the simple effects of priming were significant in the –IH version [for the identical cognates: $\chi^2(1) = 0.092$, $p = .762$, $BF_{10} = 0.026$, $\Delta = -0.1\%$; for the non-identical cognates: $\chi^2(1) = 0.030$, $p = .862$, $BF_{10} = 0.025$, $\Delta = -0.1\%$], nor were they significant in the +IH version [for the identical cognates: $\chi^2(1) = 0.254$, $p = .615$, $BF_{10} = 0.028$, $\Delta = 0.1\%$; for the non-identical cognates: $\chi^2(1) = 1.004$, $p = .316$, $BF_{10} = 0.041$, $\Delta = 0.3\%$; for the 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.

### *Pairwise comparisons*

Again, as in Experiment 3, 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: $\chi^2(1) = 0.474$, $p = .491$,

---

[16] Because the effect of version was significant and there was a small but significant difference in the Dutch LexTALE scores between the two versions, an exploratory analysis was conducted including this variable as a covariate in the 2×2×2. The effect of the covariate was not significant [$\chi 2(1) = 0.325$, $p = .569$, BF10 = 0.015], but with it included the main effect of version was only marginally significant, though still of the same size [$\chi 2(1) = 3.665$, $p < .056$, BF10 = 0.078, $\Delta = 0.5\%$]. The priming by version interaction became marginally significant [$\chi 2(1) = 3.264$, $p = .071$, BF10 = 0.064]. Both Bayes factors still provided strong or very strong evidence for the null hypothesis. None of the other significance levels changed. The random effects structure had to be reduced to an intercepts-only model.

$BF_{10} = 0.022$, $\Delta = 0.4\%$; non-identical cognates vs translation equivalents: $\chi^2(1) = 1.286$, $p = .257$, $BF_{10} = 0.034$, $\Delta = 0.6\%$]. 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: $\chi^2(1) = 27.95, p < .001$, $BF_{10} = 2.1 \cdot 10^4$, $\Delta = 4.4\%$; non-identical cognates vs interlingual homographs: $\chi^2(1) = 25.46$, $p < .001$, $BF_{10} = 5.9 \cdot 10^3$, $\Delta = 4.4\%$; translation equivalents vs interlingual homographs: $\chi^2(1) = 25.45$, $p < .001$, $BF_{10} = 5.9 \cdot 10^3$, $\Delta = 4.2\%$]. 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 Experiment 3 and 4 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, there was no evidence for the expected priming effect in either of the two experiments, so it is impossible to provide a definitive answer to the research question. In Experiment 3, 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 the predictions, Experiment 4 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 the data provided strong evidence for the null hypothesis that there was no effect of cross-lingual long-term word-meaning priming.

One likely explanation for the fact that there was no 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 (possibly at the cost of any other meanings associated with that word form). The lexical decision tasks used in this experiment, however, did not necessarily require the participants to access the meaning of the stimuli, as lexical decisions can be made on the basis of a general sense of familiarity with a stimulus. If the 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 there was no priming effect.

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 this is why they observed an effect of priming.

Although there was no 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 "bark" (i.e. using the "cover of a tree" meaning as opposed to the "sound a dog makes" meaning) is sufficient to bias participants' future interpretation of the word "bark" 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; Experiment 1 and 2). They also 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 (Experiment 3 & 4). Finally, Gilbert, Davis, Gaskell, and Rodd (2018) further showed that the priming effect transfers across modalities, from the visual domain to the auditory domain and vice versa.

With regards to the predictions for the unprimed trials, the data from Experiment 3 did not confirm the hypotheses, but were in line with the data from Experiment 1 and 2. The data from Experiment 4 did confirm the hypotheses. In both experiments, there was no 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 the 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), the experiments presented in Chapter 3 indicate that this effect can disappear when the experiment includes non-target language words (like the Dutch words in these experiments) that the participants are required to respond "no" to. The data of these experiments are fully consistent with these findings and further show that the same appears to be true for non-identical cognates.

Second, as predicted, there was evidence for an interlingual homograph inhibition effect in both the reaction time data and the accuracy data of the +IH version of Experiment 4, with participants responding 50-60 ms more slowly and approximately 4% less accurately to the interlingual homographs than to any of the three other word types. This pattern of results was confirmed by both the traditional frequentist *p*-values and the Bayes factors. These findings are in line with research discussed in Chapter 3 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 et al., 2000; Dijkstra et al., 2000; Dijkstra et al., 1998; Von Studnitz & Green, 2002).

In summary, there was no evidence for a cross-lingual long-term word-meaning priming effect for any of the word types that were examined, so it is impossible to answer the primary research question. The fact that there was 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 the data. Therefore, it appears that 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, it seems 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, like the semantic relatedness task that Gilbert et al. (2018) used.

# CHAPTER 5: Towards a distributed connectionist account of cognates and interlingual homographs: Evidence from semantic relatedness tasks

# 1 Introduction

The experiments included in Chapter 3 and Chapter 4 both indicated that the use of a lexical decision task to study (bilingual) language processing is problematic. Experiment 1 and 2 (Chapter 3) showed that including words from the non-target language in a lexical decision task reduced the cognate facilitation effect but increased the interlingual homograph inhibition effect. Most likely these effects were caused by response competition between the two readings of the cognates and interlingual homographs. This makes it difficult, if not impossible, to determine whether the facilitation and inhibition effects traditionally observed in lexical decision tasks for cognates and interlingual homographs originate in the lexicon or whether they arise at the decision stage. Furthermore, the results from Experiment 3 and 4 (Chapter 4) suggest that a lexical decision task is unlikely to show reliable effects of cross-lingual long-term priming, despite earlier success (Poort et al., 2016). Since the cross-lingual long-term word-meaning priming effect is largely semantic in nature, this could explain why no effect of priming was found in either Experiment 3 or 4.

In particular, Experiment 1 and 2 (Chapter 3) showed that the specificity of the representation of a word that is retrieved during lexical decision depends on the other types of stimuli included in the experiment. When the lexical decision task in those experiments did not include non-target language words, participants were able to respond based on a general sense of familiarity or 'word-likeness' of the stimuli. When the lexical decision task did include such words, however, they had to decide whether a given stimulus was a word in a particular language or not. In these tasks, participants could only respond accurately if they accessed a specific representation of the word in their mental lexicon, which includes information about the word's (written and spoken) form, its meaning and its language membership.

Indeed, while research in the monolingual domain suggests that lexical decision tasks may involve some level of semantic processing, as shown by studies demonstrating effects of semantic priming in lexical decision tasks (e.g. Becker, 1979, 1980; Bentin, McCarthy, & Wood, 1985; Dannenbring & Briand, 1982; Holcomb, 1988; Holcomb & Neville, 1990; Meyer & Schvaneveldt, 1971; Perea & Rosa, 2002), accessing the meaning of a letter string is not always necessary in order to decide whether it is a real word or not (Piercey & Joordens, 2000). James (1975), for example, was one of the first to demonstrate that concrete words are processed more quickly than abstract words in lexical decision tasks, but in his experiment this was only the case when the non-words were pronounceable (compared to when they were not). Azuma and Van Orden (1997) similarly showed that the relatedness among a word's meanings only influenced lexical decision reaction times when the non-words were difficult. Such examples suggest that the meaning of a word plays a role in lexical decision tasks in some cases, but not always.

In other words, in a lexical decision task, bilinguals can use different sources of information or access different levels of representation to decide whether stimuli are words or not, depending on the level of specificity required by the task. The main issue with using lexical decision tasks, then, is that for any given experiment it is difficult if not impossible to know which level of representation a particular lexical decision task is tapping in to. Furthermore, some of these sources of information (like information about a word's form or meaning) will exert facilitatory or inhibitory influences within the lexicon, while others (like information about a word's language membership) will play a role in decision processes. A second issue with using lexical decision tasks is that it is difficult (or again even impossible) to separate these decision-level processes from lexicon-based processes. This indicates that lexical decision tasks should be used with caution and in combination with other tasks to obtain converging evidence regarding the representation of cognates and interlingual homographs in the bilingual mental lexicon. Using many different types of tasks that tap into different levels of representation to study the processing of cognates and interlingual homographs provides an opportunity to unpick lexicon-based processes and task-based processes.

Therefore, the experiments presented in this chapter use a semantic relatedness judgements task to study the processing of cognates and interlingual homographs. In its most basic form, during such a task, participants see pairs of words and are asked to decide as quickly and accurately as possible whether the words in each pair are related to each other in meaning or not. For example, participants may see the word "goat" followed by the word "sheep" (for a related response) or the word "wardrobe" (for an unrelated response). There are several advantages of using a semantic relatedness task to examine the processing and representation of cognates and interlingual homographs.

The obvious advantage of using a semantic relatedness task is that it allows researchers to examine word representations at the semantic level. For example, Cai, Gilbert, Davis, Gaskell, Farrar, Adler, and Rodd (2017) used a semantic relatedness judgements task to show that accent cues can bias the interpretation of an ambiguous word towards one of its meanings. To illustrate, the word "bonnet" in British English more often refers to the engine cover of a car (the "hood" in American English), while in American English it is more commonly used to refer to a type of hat. In a semantic relatedness judgements task (Experiment 4), the British participants responded more quickly and accurately to words like "bonnet" when the probe was related to the British meaning of the word (e.g. when the probe was "car") and especially when they were spoken in a British accent compared to when they were spoken in an American accent. In contrast, when the probe was related to the American meaning (e.g. when the probe was "hat"), the British participants responded more quickly and accurately when "bonnet" had been pronounced with an American accent.

Another advantage of using a semantic relatedness task is that, in such a task, participants are more likely to use the same sources of information regardless of the demands of the task or the other stimuli included in it. To decide whether two words are related to each other in meaning, a participant must always access a representation of both the forms of these words and, through these forms, the meanings of those words. The characteristics of the other stimuli included in the task may make the decision of whether those two words are related or not more difficult (e.g. if some pairs are less related than others), but those characteristics do not influence whether only the form representation or also the meaning representation is accessed and used to inform the participants' decisions, as can be the case in a lexical decision task. (Although this may still affect whether all parts of the meaning representation are accessed.) In this way, a semantic relatedness task is also more similar to natural language processing than a lexical decision task. In particular, although the task may still be artificial in nature, the end point of lexical access is the same: information about a word's form is used in order to access the relevant information about its meaning. Therefore, a semantic relatedness task is more likely to be a true reflection of the processes that occur in the mental lexicon during natural language processing than a lexical decision task.

Furthermore, decisions in a semantic relatedness tasks are less likely to be based on language membership information, the main driving factor for the emergence of response competition in lexical decision tasks. In single-language lexical decision tasks that include non-target language words, language membership information is inherent to the decision that participants are required to make. However, even in single-language lexical decision tasks that do not include non-target language words, participants may rely extensively on language membership information, because they are instructed to decide if the stimulus is a word in a particular language.

In a semantic relatedness task, it is less clear what role language membership plays. In particular, when participants see an interlingual homograph (e.g. "angel") and an English probe (e.g. "heaven"), they may use information about language membership to allow them to ignore information about the interlingual homograph's Dutch meaning (e.g. "insect's sting"). However, they may also base their decisions purely on the relationship between the meaning of the target and the probe and this is arguably more likely. In this case, response competition in a semantic relatedness task would reflect lexico-semantic competition (i.e. the fact that the concept of "insect's sting" is not related to the concept of "heaven", which would warrant a "no"-response, but the concept of "spiritual being" is, which requires a "yes"-response).

It seems, however, that semantic relatedness tasks have not been commonly used in bilingual research. As discussed previously, most experiments that have investigated cognate and interlingual homograph processing have used lexical decision tasks, although some other tasks have been used as well. Only one study appears to have used a semantic relatedness task to examine cognate processing. In one of their tasks, Yudes, Macizo, and Bajo (2010) recorded EEG signals while Spanish–English bilingual participants decided whether pairs of Spanish words were semantically related or not. The first word in each pair was either a Spanish–English cognate or a Spanish-only control word. Their data showed that participants did not respond significantly more quickly or slowly to pairs that included a cognate than to pairs that did not. Furthermore, there were no differences in the N400 component between the cognates and the control words when analysing the ERP data either. However, Yudes et al. (2010) included mainly non-identical cognates in this experiment: only eight of the 100 cognates were identically spelled in Spanish and English. Since research has shown that the cognate facilitation effect is greater for identical cognates than for non-identical cognates (Comesaña et al., 2015; Dijkstra et al., 2010; Duyck et al., 2007; Van Assche et al., 2011), Yudes et al. (2010) may not have found evidence for a cognate facilitation effect because they used almost exclusively non-identical cognates.

Similarly, the interlingual homograph inhibition effect has been demonstrated mostly in lexical decision tasks. As with cognates, it appears that only one study has used a semantic relatedness task to examine interlingual homograph processing. Macizo, Bajo, and Cruz Martín (2010) asked Spanish–English bilinguals to make semantic relatedness judgements in *English* to identical interlingual homographs paired with probes that were related to the *Spanish* meanings of the interlingual homographs but unrelated to the English meanings (e.g. "pie"–"toe", where "pie" means "foot" in Spanish). The participants were slower to respond on interlingual homograph trials than on control trials (e.g. "log"–"toe"), which suggests that the participants accessed the non-target language (Spanish) meaning of the interlingual homographs and then inhibited this. This study provides converging evidence that the disadvantage for interlingual homographs in lexical decision is not solely an artefact of using a lexical decision task. However, all of the interlingual homographs were paired with probes that were related to the non-target language meaning (i.e. "pie"–"toe"), which will have made the participants rely more on language membership information. This suggests that response competition based on language membership information played a role in this effect.

In the two experiments presented here, participants were asked to judge whether a target word — either a cognate, English control or interlingual homograph — was semantically related to a subsequent probe. All targets were paired with related probes because a "yes"-response in that case clearly signals that the participant accessed the relevant, target-language meaning of both the target and the probe. In addition, none of the interlingual homographs were paired with probes that were related only to the non-target language (Dutch) meaning, to avoid participants linking the non-target language meanings of the interlingual homographs to the "no"-response and, thus, eliciting language membership-based response competition. As all targets were paired with related probes, additional filler items were included that were paired with unrelated probes. The targets were presented for a short amount of time (200 ms) and the probe appeared almost immediately after the target disappeared (50 ms after target offset). The target was presented before the probe to prevent biasing the interpretation of the targets towards the meaning of the probe, which could have negated any effects. The target was presented only briefly and almost immediately followed by the probe so that processing of the target would not be finished before the probe was presented. Consequently, the task would still reflect processing of the targets even though participants were responding to the probes. To further minimise the role that language membership information would play in the semantic relatedness task, Experiment 5 was advertised and conducted entirely in the language of the task (English), so that the Dutch–English bilinguals would not assume that their knowledge of the stimuli in Dutch would be relevant to the task.

Assuming that the cognate facilitation effect in lexical decision (and other tasks) points to a special status of cognates in the bilingual lexicon, when the participants see a cognate in this task, they will easily and quickly access its meaning. When they next see the probe, it will be relatively easy and quick to decide that the probe is related to the cognate. In other words, a cognate facilitation effect would be expected in this task because the probe would be related to the cognate in both Dutch and English, which would make the decision easier compared to deciding whether an English control is related to its probe. In contrast, when participants encounter an interlingual homograph, initially they will access both its Dutch and English meaning. This will slow them down and make them more prone to mistakes when the probe is then presented, as this is only related to the English meaning. So an interlingual homograph effect would be expected because the participants would not have settled on an interpretation of the interlingual homograph yet by the time the probe appeared, so that both meanings would still be active and would compete for selection.

Using a semantic relatedness task also offers advantages when it comes to investigating cross-lingual long-term priming of cognates and interlingual homographs. In the monolingual domain, semantic relatedness tasks have been used successfully to replicate effects of long-term word-meaning priming. Initial experiments using this paradigm used a word association task to show that a single encounter with an ambiguous word's less-used (subordinate) meaning can bias future interpretation of that word towards that meaning (Betts, Gilbert, Cai, Okedara, & Rodd, 2017; Rodd et al., 2016; Rodd et al., 2013). Gilbert et al. (2018) used a semantic relatedness task to replicate these findings and to show that long-term word-meaning priming transfers across modalities (from visual to auditory and vice versa). The successful use of a semantic relatedness task in the monolingual domain to show effects of long-term priming offers another compelling reason to use such a task to study cross-lingual long-term priming.

As discussed in Chapter 4, in the bilingual domain, Poort et al. (2016) demonstrated that this long-term word-meaning priming effect transfers across languages in Dutch–English bilinguals. A single encounter with a cognate or interlingual homograph in Dutch affected how quickly these words were processed in a subsequent English lexical decision task. The experiments presented in Chapter 4, however, did not replicate these findings. The long-term word-meaning priming effect is thought to be due to a strengthening of the connections between the orthographic representation of a word and its semantic representation (Rodd et al., 2013). Furthermore, based on the assumption that word meanings are represented by patterns of distributed features, Rodd et al. (2016) note that it is also possible that priming results in changes to the connections between the semantic units of the primed meaning, making it a more stable representation relative to that of the unprimed meaning. Since semantic relatedness tasks require access to these meanings, an effect of priming is more likely to be observed in such tasks than in lexical decision.

In particular, when the bilingual participants encounter a cognate in Dutch, this will strengthen the connection between the cognate's form representation and its meaning representation. When the participant then sees that same cognate during the English semantic relatedness task, it will be even easier to access its meaning and decide it is related to the probe than if the participant had not seen the cognate before. For the interlingual homographs, however, the encounter with its Dutch meaning will have strengthened the connection between its (Dutch) form representation and its Dutch meaning. This will make it easier to access the Dutch meaning again during the English lexical decision task, which is likely to slow participants down (and make them more prone to errors) when deciding whether that meaning is related to the English probe or not.

In sum, there are reasons to think that lexical decision tasks are problematic to study the representation of cognates and interlingual homographs in the bilingual mental lexicon, as well as to examine effects of recent experience with these words in one language on subsequent processing in another language. Semantic relatedness tasks have the advantage of requiring access to meaning representations. Furthermore, the specificity of the representation that is accessed during the task is less likely to depend on task demands and stimulus list composition. Finally, the decision is also less likely to involve language membership information and the task has been used successfully in the monolingual domain to replicate effects of long-term priming. As such, the experiments presented in this chapter are similar to the experiments from the previous chapters, except they use a semantic relatedness judgements task instead of a lexical decision task. Specifically, Experiment 5 was similar to Experiment 1 and 2 and focussed on the cognate facilitation and interlingual homograph inhibition effect. Experiment 6 was similar to Experiment 3 and 4 and attempted to replicate the cross-lingual long-term priming effect initially observed by Poort et al. (2016).

# 2 Experiment 5: The cognate facilitation and interlingual homograph inhibition effect in semantic relatedness

Experiment 5 was again pre-registered as part of the Center for Open Science's Preregistration Challenge (www.cos.io/prereg). The stimuli, data and processing and analysis scripts can be found on the Open Science Framework (www.osf.io/ndb7p), as well as an Excel document with detailed results from all of the analyses. The preregistration can be retrieved from www.osf.io/u2fyk (Poort & Rodd, 2017, December 7). Where applicable, deviations from the pre-registration will be noted.

## 2.1 Introduction

The aim of Experiment 5 was to determine whether the cognate facilitation effect and interlingual homograph inhibition effect observed in lexical decision are a result of how these words are stored in the bilingual lexicon or whether they are merely task artefacts. Experiment 5 was similar in design to Experiment 1 and 2, but it used a semantic relatedness task instead of a lexical decision task. In this task, Dutch–English bilingual participants were asked to judge whether a target word — either a cognate, English control or interlingual homograph — was semantically related to a subsequent probe. A group of native monolingual British English speakers performed the same experiment to rule out the possibility that the effects seen in the bilingual group were due to pre-existing differences between the three word types in the relatedness of the targets and their probes.

The predictions were based on the assumption that lexical decision tasks involve some degree of semantic processing (e.g. Azuma & Van Orden, 1997; James, 1975); in other words, that the facilitation and inhibition effects observed in lexical decision tasks for cognates and interlingual homographs are a consequence of how these words are represented at both the orthographic and semantic level in the bilingual mental lexicon and not of decision-level processes. As such, a similar pattern of results was predicted for the semantic relatedness task used in this experiment: the task was predicted to show both an effect of cognate facilitation and interlingual homograph inhibition for the bilinguals but not for the monolinguals.

## 2.2 Methods

### 2.2.1 Participants

The aim was to recruit at least 30 participants for each group, about half more than for each of the versions of Experiment 1 and 2. The bilingual participants had to be between the ages of 18 and 35, of Dutch or Belgian nationality and resident in the Netherlands or Belgium at the time of the experiment. Their native language had to be Dutch and they had to be fluent speakers of English, with no diagnosis of any language disorders. Recruitment proceeded only through Prolific Academic (Damer & Bradley, 2014) and was entirely in English, to ensure that the bilingual participants did not know that their knowledge of Dutch would be relevant to the experiment. For this reason, the criteria regarding native language, nationality and country of residence were not stated upfront, but instead the experiment was accessible only to participants who had previously indicated on Prolific that these criteria applied to them. A total of 31 bilingual participants who met these criteria was recruited. (Two other bilingual participants did not meet these criteria.)

The monolingual participants were recruited next and efforts were made to match the two groups in terms of age and educational profile. The monolingual participants had to be native speakers of British English who spoke no other languages fluently and were not diagnosed with any language disorders. They had to be between the ages of 18 and 31 (the age range of the bilingual participants), of British nationality and be resident in the United Kingdom at the time of the experiment. To match on education, the bilingual participants were classified as having obtained a 'high' educational degree if they had completed or were currently enrolled in a HBO[17] or higher degree. If they had only completed secondary school

[17] HBO stands for 'Hoger Beroepsonderwijs' in Dutch or 'Higher Vocational Education'/'University of Professional Education' in English. A HBO bachelor's degree is considered one tier below a university bacherlor's degree in the Netherlands and Belgium, but courses taught at HBO-level in the Netherlands and Belgium (e.g. journalism) are often taught at university in the UK. For this reason, HBO degrees were included in the category of 'high' education.

or an MBO[18] degree, they were classified as having obtained a 'low' educational degree. Monolingual participants were then recruited in roughly equal proportions, with participants who had indicated on the Prolific platform that they had completed a university undergraduate or higher degree classified as having obtained a 'high' educational degree and those who indicated they had only completed their GCSEs or A-levels as having obtained a 'low' educational degree. A total of 31 monolingual participants that met these criteria was recruited. (Three other participants did not meet these criteria.) All participants gave informed consent and were paid £3 for their participation in the experiment.

Excluding participants again happened in two stages. First, while testing was still ongoing, participants who scored less than 80% correct on the semantic relatedness task and/or less than 50% on either of the two language proficiency measures were excluded and replaced. Two participants (one in each group) that met these criteria were excluded and two new participants tested in their stead. Second, after testing had finished and a total of 60 useable datasets had been gathered, each participant's performance on the targets (cognates, interlingual homographs and English controls) included in the semantic relatedness task was compared to the mean of their group to determine whether any more participants needed to be excluded. One bilingual participant had performed worse than three standard deviations below their group's mean (68.0%, $M$ = 89.2%, $SD$ = 6.9%) and was excluded.

The remaining 29 bilingual participants (18 males; $M_{age}$ = 22.4 years, $SD_{age}$ = 3.8 years) had started learning English from an average age of 7.4 ($SD$ = 2.3 years) and so had an average of 15.0 years of experience with English ($SD$ = 4.2 years). They rated their proficiency as 9.3 out of 10 in Dutch ($SD$ = 0.9) and as 8.5 in English ($SD$ = 0.7). A two-sided paired $t$-test showed this difference to be significant [$t(28) = 5.010$, $p < .001$]. These self-ratings were confirmed by their high LexTALE scores in both languages, which a two-sided paired $t$-test showed were also higher in Dutch [Dutch: $M$ = 87.2%, $SD$ = 6.3%; English: $M$ = 82.5%, $SD$ = 9.5%; $t(28) = 2.406$ $p = .023$].

The remaining 30 monolingual participants (10 males; $M_{age}$ = 26.2 years, $SD_{age}$ = 3.7 years) had scored 8.4% higher on the English LexTALE than the bilingual participants. A two-sided independent-samples Welch's $t$-test showed this difference to be significant [bilinguals: $M$ = 82.5%, $SD$ = 9.5%; monolinguals: $M$ = 91.0%, $SD$ = 6.1%; $t(47.7) = -4.029$ $p < .001$]. With respect to the efforts made to match the two groups of participants, there were participants in both groups of roughly equal proportions of 'high' and 'low' educational degrees. Twenty-three (out of 29) bilingual and twenty-five (out of 30) monolingual

[18] MBO stands for 'Middelbaar Beroepsonderwijs' in Dutch. Courses taught at MBO level in the Netherlands are aimed at teaching a profession and are often taught at Further Education colleges in the UK. For this reason, MBO degrees were included in the category of 'low' education.

| | **ENGLISH CHARACTERISTICS OF THE PROBES** | | | |
|---|---|---|---|---|
| | frequency | log10(fre-quency) | word length | OLD20 |
| cognates | 66.3 (113) | 3.04 (0.69) | 5.34 (1.39) | 1.92 (0.62) |
| interlingual homographs | 60.7 (117) | 2.98 (0.67) | 5.14 (1.37) | 1.80 (0.46) |
| English controls | 55.1 (109) | 2.98 (0.64) | 5.34 (1.53) | 1.90 (0.61) |
| fillers | 50.1 (113) | 2.89 (0.63) | 5.11 (1.23) | 1.85 (0.49) |

**Table 5-1:** Experiment 5 & 6. Means (and standard deviations) of the English characteristics of the probes for the semantic relatedness task for the cognates, interlingual homographs, English controls and the fillers. *Note*. OLD20 information was not available for the probe "logo" for the interlingual homograph "brand"–"brand".

participants had obtained high education degrees according to the classification. A chi-square test showed this difference to be non-significant [$\chi^2(1) = 0.152$, $p = .697$; $\alpha = .01$]. In terms of age, there was a small but significant difference of 3.8 years between the two groups, with the monolingual participants being older than the bilingual participants [$t(56.7) = -3.891$, $p < .001$; $\alpha = .01$). As per the preregistration, all analyses that involved group as a factor were conducted both with age included as a covariate and without.

### 2.2.2 Materials

The tables in Appendix A indicate which cognates, interlingual homographs and English controls were included in Experiment 5.

#### 2.2.2.1 Targets & probes

The same set of 50 identical cognates, 50 identical interlingual homographs and 50 English controls included in Experiment 4 was used for this experiment. Each of these items was assigned a related probe for the semantic relatedness task, making sure none of the probes were also targets. The probes were of roughly equal frequency, length and orthographic complexity as the targets themselves. They ranged in frequency from 0.98 and 866.04 (in occurrences per million according to the SUBTLEX-US database; Brysbaert & New, 2009), were between 3 and 9 letters long and had OLD20 values between 1 and 3.75. Means and standard deviations per word type for each of these measures (in both Dutch and English) can be found in Table 4-3. Table 5-1 lists the English characteristics of the probes. The sets of probes for the three word types did not significantly differ from each other in terms of log-transformed frequency, word length or OLD20 (all *p*s > .2).

#### 2.2.2.2 Fillers

The experiment included the same number of fillers as targets. All fillers were paired with unrelated probes. The 150 fillers comprised an additional 15 identical cognates, 15 identical interlingual homographs and 15 English controls from the remaining pre-tested materials (see Chapter 2) and 105 regular fillers. The additional cognates, interlingual homographs and English controls were included and assigned unrelated probes to ensure that the participants would not assume that any cognate or interlingual homograph always required a "yes"-response.

#### 2.2.2.3 Pilot experiment

To verify that each item was indeed related or unrelated to its chosen probe as intended, a pilot of the semantic relatedness task was conducted with a group of 16 monolinguals who did not take part in the main experiment. Two participants were excluded, as they had scored less than 80% correct on the task. The data from the remaining 14 participants (2 male; $M_{age}$ = 33.1 years, $SD_{age}$ = 8.8 years) indicated that overall accuracy for the related trials was 93.4% and for the unrelated trials was 96.7%. The probes for three items (two related probes and one unrelated probe) with a percentage correct of less than 70% were changed.

### 2.2.3 Design and procedure

The experiment employed a mixed design. Word type was a within-participants/between-items factor: participants saw all words of each word type, but each word of course belonged to only one word type. Group was a between-participants/within-items factor: participants belonged either to the bilingual or the monolingual group, but each item was seen by both groups.

The experiment was created and conducted using Gorilla online experimental software (Evershed & Hodges, 2016). It comprised two (for the monolinguals) or three (for the bilinguals) separate tasks: (1) the English semantic relatedness task, (2) the English version of the LexTALE (Lemhöfer & Broersma, 2012) and, for the bilingual participants only, (3) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012). At the end of the experiment, the participants completed a self-report language background survey in Dutch (for the bilinguals) or English (for the monolinguals).

The semantic relatedness task was identical for the bilingual and monolingual participants. During the semantic relatedness task, the participants saw all 150 related target-probe pairs ("yes"-responses) and all 150 unrelated filler-probe pairs ("no"-responses) and were asked to indicate, by means of a button press, as quickly and accurately as possible, whether the word they saw first was related in meaning to the word they saw second. A practice block of 6 pairs was followed by 6 blocks of 50 experimental pairs. The order of the pairs within blocks was randomised for each participant, as was the order of the blocks. Three filler pairs (additional to the 105 experimental filler pairs) were presented at the beginning of each block, with a 15-second break after each block. The target or filler item would appear first on screen and remain for 200 ms. After a blank screen lasting 50 ms, the probe appeared. The probe remained on screen until the participant responded or until 2500 ms passed. A warning was presented to the participant that they were responding too slowly if they had not responded 2000 ms after the probe first appeared. The warning remained on screen for 500 ms, during which time the participant could still respond. The inter-trial interval was 1000 ms.

As mentioned previously, the bilingual participants were not told that they were being recruited because of their status as native Dutch speakers or that Dutch would be in any way relevant to the experiment until after they had completed the semantic relatedness task. To achieve this, the consent form at the start of the experiment was in English, as was the study description in Prolific. Nevertheless, four participants indicated at the end that they did think Dutch would be relevant and a further five said they were not sure but suspected so because they noticed the name of the researcher (included in the consent form) was Dutch.

## 2.3 Results

All analyses were carried out in R (version 3.4.3; R Core Team, 2017) using the lme4 package (version 1.1-13; Bates et al., 2015), following Barr et al.'s (2013) guidelines (with some personal amendments[19]) for confirmatory hypothesis testing and using Type III Sums of Squares likelihood ratio tests to determine significance. Reaction times were analysed using the lmer() function with the default optimiser; accuracy data were analysed using the glmer() function with the bobyqa optimiser. Detailed results of all analyses for Experiment 5 can be found in the Excel document experimentOverview.xlsx in the OSF project (www.osf.io/ndb7p). Finally, the graphs in the figures display the (harmonic) participant means, while the effects reported in the text were derived from the estimates of the fixed effects provided by the model summary.

---

[19] A model with a maximal random effects structure converged for all analyses. If a maximal model had not converged, however, the random effects structure would have been reduced in the following manner: 1) the correlations between the random effects were removed, 2) the correlations were reintroduced and the slope that had the smallest variance in the maximal model was removed, 3) the correlations were removed again, 4) the final slope was removed, which left only the random intercepts. This approach differs from Barr et al. (2013) in that they would recommend to try to remove the random intercepts at some point, retaining as many random slopes as possible. A model without random intercepts, however, would not take into account the within-participants and within-items nature of this experiment's design.

One item (the cognate "type"–"type") was excluded from the analysis, as its percentage correct for each of the two groups (bilinguals: 58.6%; monolinguals: 50.0%) was more than three standard deviations below its word type mean for both groups ($M_{bilinguals}$ = 92.6%, $SD_{bilinguals}$ = 8.7%; $M_{monolinguals}$ = 91.7%, $SD_{monolinguals}$ = 8.4%). Excluding this item did not affect the matching of the word types. An additional six items (the interlingual homographs "slot"–"slot", "stem"–"stem" and "strand"–"strand" and the English controls "emmer"–"bucket", "lijm"–"glue" and "moeras"–"swamp") were outliers for their word type for either of the two groups, but as per the pre-registration these items were not excluded.

## 2.3.1 Confirmatory analyses

### 2.3.1.1 Analysis procedure

The same analysis procedure was employed for the reaction times and accuracy data. In all cases, positive effects of word type indicate a facilitative effect for the first-named word type over the second (i.e. faster reaction times and higher accuracy), while negative effects indicate inhibitory effects (i.e. slower reaction times and lower accuracy). Positive (negative) effects of group indicate an advantage (disadvantage) for the bilinguals over the monolinguals.

Two fixed factors were included in the main 3×2 analysis: word type (3 within-participant/between-items levels: cognate, English control, interlingual homograph) and group (2 between-participant/within-items levels: bilingual, monolingual). The maximal random effects structure included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for group by items. This maximal model converged for both the reaction times and accuracy analyses.

To examine more closely whether there was evidence for a cognate facilitation effect, a 2×2 analysis similar to the 3×2 analysis was conducted, but which included data for only the cognates and English controls (for both groups). The maximal random effects structure again included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for group by items and converged for both the reaction times and accuracy analyses. In addition, two pairwise comparisons were conducted, comparing the cognates and English controls separately for both the bilinguals and the monolinguals. The maximal random effects structure for these analyses included a correlated random intercept and random slope for word type by participants and a random intercept by items. These models also converged for both the reaction time and accuracy analysis.

Similarly, to examine more closely whether there was evidence for an interlingual homograph inhibition effect, a 2×2 analysis was conducted which included the data for only the interlingual homographs and English controls (for both groups). The maximal random effects structure again included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for group by items and converged for both the reaction times and accuracy analyses. In addition, two pairwise comparisons were conducted, comparing the interlingual homographs and English controls separately for both the bilinguals and the monolinguals. The maximal random effects structure for these analyses included a correlated random intercept and random slope for word type by participants and a random intercept by items. These models converged for both the reaction time and accuracy analysis.

Finally, a 2×2 analysis was conducted which included the data for only the cognates and interlingual homographs (for both groups), as well as two pairwise comparisons that compared the cognates and interlingual homographs separately for both the bilinguals and the monolinguals. The maximal random effects structure for the 2×2 analysis again included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for group by items and converged for both the reaction times and accuracy analyses. The maximal random effects structure for the pairwise comparison included a correlated random intercept and random slope for word type by participants and a random intercept by items. These models converged for both the reaction time and accuracy analysis.

For the 2×2 analyses, the *p*-values were compared against a Bonferroni-corrected $\alpha$ of .0167. The *p*-values for the two sets of three pairwise comparisons were compared against a Bonferroni-corrected $\alpha$ of .0167. As per the pre-registration, the 3×2 analysis and the three 2×2 analyses were each conducted once without age included as a covariate and once with. In most cases, including age as a covariate did not change the significance level of the analysis; in cases when the significance level did change, this will be noted. Finally, age was centred to have a mean of zero prior to including it in the model. Random slopes were not included for age.

#### 2.3.1.2 Reaction times

Reaction times are shown in Figure 5-1. Trials with reaction times (RTs) of less than 300 ms or more than 2000 ms were discarded (0.5% of the data), as were incorrect trials and trials that participants had not responded to (8.5% 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 3×2 analysis showed that the assumptions of normality and homoscedasticity were violated. After transforming the RTs,

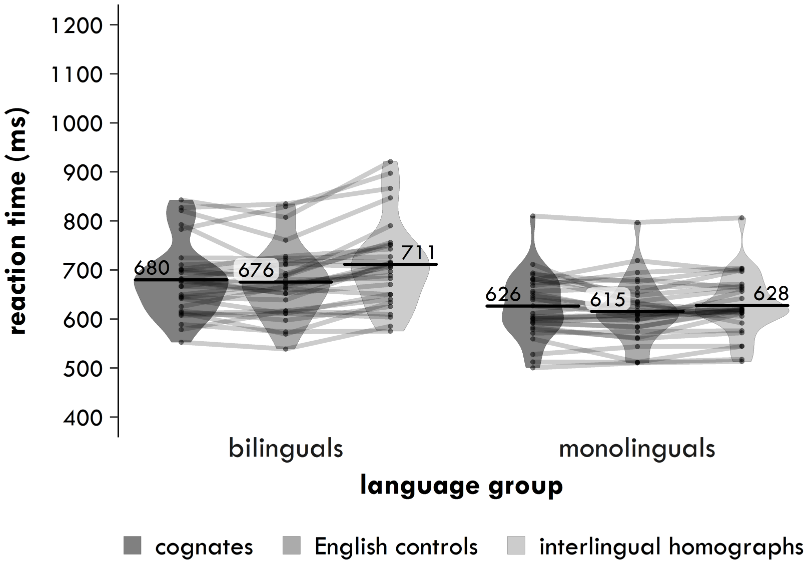


**Figure 5-1:** Experiment 5. Harmonic participant means of the inverse-transformed semantic relatedness reaction times (in milliseconds) by group (bilinguals, monolinguals; x-axis) and word type (cognates, darkest grey; English controls, medium grey; interlingual homographs, lightest grey). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

any inverse-transformed RTs were removed that were more than three standard deviations above or below a participant's mean inverse-transformed RT (0.2% of the remaining data).

### *3×2*

In the 3×2, the main effect of word type was significant [$\chi^2(2) = 6.068$, $p = .048$], indicating that there was a difference in reaction times between the three word types. The main effect of group was also significant [$\chi^2(1) = 12.74$, $p < .001$], with the bilingual participants responding on average 65 ms more slowly than the monolingual participants. Crucially, the interaction between word type and group was significant [$\chi^2(2) = 7.173$, $p = .028$].

### *Cognates versus English controls*

In the 2×2 that included the cognates and English controls, the main effect of word type was not significant [$\chi^2(1) = 0.678$, $p = .411$, $\Delta = -8$ ms], nor was the interaction between word type and group [$\chi^2(1) = 0.758$, $p = .384$]. The main effect of group was significant again [$\chi^2(1) = 10.00$, $p = .002$], with the bilingual participants responding on average 56 ms more slowly than the monolingual participants. The two pairwise comparisons told a similar story. There was no significant advantage for cognates compared to English controls for either the bilinguals [$\chi^2(1) = 0.120$, $p = .730$, $\Delta = -4$ ms] or the monolinguals [$\chi^2(1) = 1.334$, $p = .248$, $\Delta = -11$ ms].

#### *Interlingual homographs versus English controls*

In the 2×2 that included the interlingual homographs and English controls, the main effect of word type was only significant at an uncorrected $\alpha$ of .05 [$\chi^2(1) = 5.711$, $p = .017$, $\Delta = -25$ ms]. The interaction between word type and group was not significant [$\chi^2(1) = 3.588$, $p = .058$], but the main effect of group was again significant [$\chi^2(1) = 14.58$, $p < .001$], with the bilingual participants responding on average 72 ms more slowly than the monolingual participants. Although the interaction was not significant, the two pairwise comparisons suggest a difference in how the bilinguals and monolinguals responded to the interlingual homographs. The bilinguals responded 37 ms significantly more slowly to the interlingual homographs than the English controls [$\chi^2(1) = 7.915$, $p = .005$], while for the monolinguals the difference between these two word types was only 13 ms and not significant [$\chi^2(1) = 1.936$, $p = .164$].

#### *Cognates versus interlingual homographs*

In the 2×2 that included the cognates and the interlingual homographs, the main effect of word type was not significant [$\chi^2(1) = 2.452$, $p = .117$, $\Delta = 17$ ms]. Crucially, the interaction between word type and group was significant [$\chi^2(1) = 6.072$, $p = .014$]. The main effect of group was also significant [$\chi^2(1) = 12.97$, $p < .001$], with the bilingual participants responding on average 68 ms more slowly than the monolingual participants. Again, the two pairwise comparisons do suggest a difference in how the bilinguals and monolinguals responded to the cognates and the interlingual homographs. The bilinguals responded 33 ms significantly more quickly to the cognates compared to the interlingual homographs [$\chi^2(1) = 6.033$, $p = .014$], but the difference between these word types of 3 ms for the monolinguals was not significant [$\chi^2(1) = 0.084$, $p = .772$].

### 2.3.1.3 Accuracy

Accuracy is shown in Figure 5-2. In line with the trimming procedure for the reaction times, any trials with RTs faster than 300ms or slower than 2000ms were removed.

#### *3×2*

In the 3×2, the main effect of word type was again significant [$\chi^2(2) = 14.00$, $p = .001$]. The main effect of group, in contrast, was not significant [$\chi^2(1) = 0.029$, $p = .864$, $\Delta = 0.2\%$]. Crucially, the interaction between word type and group was again significant [$\chi^2(2) = 6.205$, $p = .045$]. When age was included as a covariate, however, the interaction became marginally

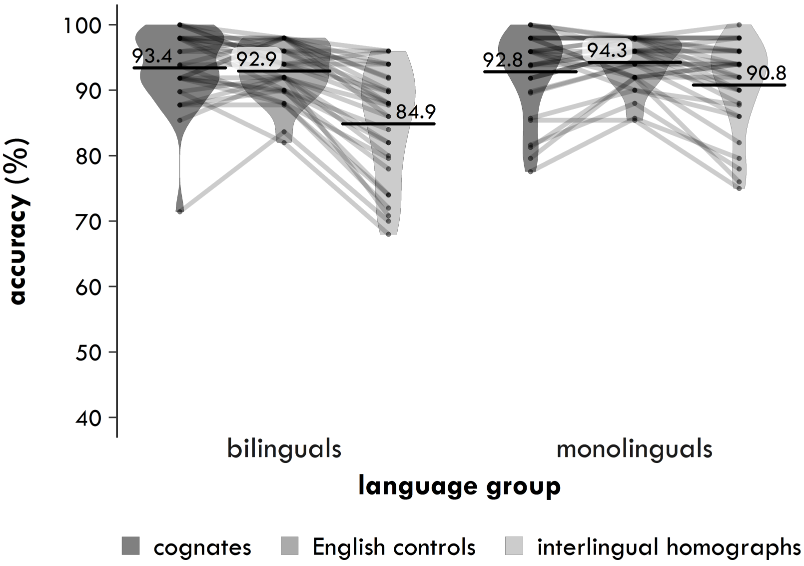


**Figure 5-2:** Experiment 5. Participant means of semantic relatedness accuracy (percentages correct) by group (bilinguals, monolinguals; x-axis) and word type (cognates, darkest grey; English controls, medium grey; interlingual homographs, lightest grey). Each point represents a condition mean for a participant with lines connecting means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

significant [$\chi^2(1) = 5.478$, $p = .065$]. The effect of age itself was not significant [$\chi^2(1) = 0.491$, $p = .484$].[20]

### *Cognates versus English controls*

In the 2×2 that included the cognates and English controls, the main effect of word type again was not significant [$\chi^2(1) = 0.110$, $p = .740$, $\Delta = 0.2\%$], nor was the main effect of group [$\chi^2(1) = 0.443$, $p = .506$, $\Delta = -0.6\%$] or the interaction between word type and group [$\chi^2(1) = 1.258$, $p = .262$]. The two pairwise comparisons told a similar story. There was no significant advantage for cognates compared to English controls for either the bilinguals [$\chi^2(1) = 0.620$, $p = .431$, $\Delta = 0.9\%$] or the monolinguals [$\chi^2(1) = 0.250$, $p = .617$, $\Delta = 0.5\%$].

### *Interlingual homographs versus English controls*

In the 2×2 that included the interlingual homographs and English controls, the main effect of word type was significant [$\chi^2(1) = 9.552$, $p = .002$], with participants making on average 3.5% more mistakes with pairs that included an interlingual homograph. The interaction between word type and group was not significant [$\chi^2(1) = 1.608$, $p = .205$], nor was the main

[20] When age was included, the maximal model would not converge, but a model without correlations between the random effects did.

effect of group [$\chi^2(1) = 0.294$, $p = .588$, $\Delta = 0.7\%$]. As for the reaction times, although the interaction was not significant, the two pairwise comparisons suggest a difference in how the bilinguals and monolinguals responded to the interlingual homographs. The bilinguals made 4.8% significantly more mistakes with the interlingual homographs than the English controls [$\chi^2(1) = 7.889$, $p = .005$], while the much smaller difference of 2.3% for the monolinguals was only significant at an uncorrected $\alpha$ of .05 [$\chi^2(1) = 4.085$, $p = .043$].

#### *Cognates versus interlingual homographs*

In the 2×2 that included the cognates and the interlingual homographs, the main effect of word type was significant [$\chi^2(1) = 10.79$, $p = .001$], with participants making on average 3.7% fewer mistakes with pairs that included a cognate. The interaction between word type and group was also significant [$\chi^2(1) = 5.750$, $p = .016$], but the main effect of group was not significant [$\chi^2(1) = 0.108$, $p = .743$, $\Delta = -0.4\%$]. Again, the two pairwise comparisons do suggest a difference in how the bilinguals and monolinguals responded to the cognates and the interlingual homographs. There was a significant difference between these two word types of 5.5% for the bilinguals [$\chi^2(1) = 10.91$, $p = .001$], but the difference of 1.8% for the monolinguals was not significant [$\chi^2(1) = 2.560$, $p = .110$].

### 2.3.2 Exploratory analyses

An exploratory analysis was conducted to compare the data from the English semantic relatedness task and the data from the lexical decision tasks from Experiment 2. As discussed in Chapter 3, in lexical decision tasks without an element of response competition (e.g. due to the inclusion of non-target language words), cognates elicit facilitation, but interlingual homographs are generally not recognised any more slowly (or only a little more slowly) than English controls. As a reminder, this is indeed what Experiment 2 showed: there was a significant cognate facilitation effect of 46 ms in the standard version but only a small, non-significant interlingual homograph inhibition effect of 8 ms in the +IH version. Two 2×2 analyses were conducted to compare these effects to the effects obtained in Experiment 5, to determine whether the size of cognate facilitation effect and the interlingual homograph inhibition effect was modulated by the type of task.

The first 2×2 analysis focussed on the cognate facilitation effect. This analysis included the data from the bilingual group of Experiment 5 ($N = 29$) and the data from the standard version of Experiment 2 ($N = 21$). Only the cognates ($n = 49$) and English controls ($n = 49$) that had been included in both experiments were included in this analysis.[21] Two fixed factors

---

[21] The cognates that were removed were "amber"–"amber", "ego"–"ego", "instinct"–"instinct", "jury"–"jury", "lens"–"lens", "tennis"–"tennis" and "type"–"type". The first six of these items had not been included in

were included in this analysis: word type (2 within-participant/between-items levels: cognate, English control) and task (2 between-participant/within-items levels: semantic relatedness, lexical decision). The maximal model converged and included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for task by items. The main effect of task was not significant [$\chi^2(1) = 0.129$, $p = .719$, $\Delta = -9$ ms]. (Positive effects of task indicate a benefit for lexical decision over semantic relatedness.) There was a significant main effect of word type [$\chi^2(1) = 9.403$, $p = .002$], with cognates being responded to 22 ms more quickly than English controls on average across both tasks. Crucially, there was also a significant interaction between word type and task [$\chi^2(1) = 10.93$, $p < .001$], indicating that the cognate facilitation effect was significantly smaller in the semantic relatedness task than in the lexical decision task.

The second 2×2 analysis was conducted in a similar manner as the first but focussed on the interlingual homograph inhibition effect. This analysis included the data from the bilingual group of Experiment 5 ($N = 29$) and the data from the +IH version of Experiment 2 ($N = 20$). Again, only the interlingual homographs ($n = 50$) and English controls ($n = 49$) that had been included in both experiments were included in this analysis.[22] As before, two fixed factors were included in this analysis: word type (2 within-participant/between-items levels: interlingual homograph, English control) and task (2 between-participant/within-items levels: semantic relatedness, lexical decision). The maximal model converged again and included a correlated random intercept and random slope for word type by participants and a correlated random intercept and random slope for task by items. The main effect of task was again not significant [$\chi^2(1) = 0.888$, $p = .346$, $\Delta = 24$ ms]. There was a significant main effect of word type [$\chi^2(1) = 6.967$, $p = .008$], with interlingual homographs being responded to 23 ms more slowly than English controls on average across both tasks. The interaction between word type and task was marginally significant [$\chi^2(1) = 3.303$, $p = .069$], suggesting that the interlingual homograph inhibition effect was larger in the semantic relatedness task than in the lexical decision task.

Experiment 5 at all; “type”–“type” was included in both experiments but had been excluded from the confirmatory analyses of Experiment 5. The English controls that were removed were “gedicht”–“poem”, “geweer”–“rifle”, “griep”–“flu”, “kunst”–“art”, “lucht”–“sky”, “plicht”–“duty”, “verdrag”–“treaty” and “wreed”–“cruel”. The item “lucht”–“sky” had only been included in Experiment 5, whereas the other items had only been included in Experiment 2.

[22] The interlingual homographs that were removed were “fee”–“fee”, “hoop”–“hoop”, “lever”–“lever”, “mate”–“mate”, “pal”–“pal” and “toe”–“toe”. These items had all only been included in Experiment 2. The same English controls were removed for this analysis as for the other 2×2 analysis.

## 2.4 Discussion

The aim of Experiment 5 was to examine semantic processing of cognates and interlingual homographs in bilinguals (and monolinguals). A subsidiary aim was to determine whether the cognate facilitation and interlingual homograph inhibition effects observed in lexical decision tasks are a reflection of how these words are stored in the bilingual mental lexicon or whether these effects are artefacts of such tasks. In line with the predictions, there was a significant interaction between word type and language group in the main analysis, which indicates that the bilinguals processed the three word types differently than the monolinguals did. Specifically, this interaction between word type and language group was significant in the analysis that included the cognates and interlingual homographs. The bilingual participants responded 33 ms more slowly and 5.5% less accurately to the interlingual homographs than to the cognates, while the monolingual participants did not (with effects of 3 ms and 1.8%).

The other findings were a little less clear-cut, however, which makes it difficult to definitively pinpoint whether this difference between the cognates and interlingual homographs for the bilinguals was due to a facilitation effect for the cognates, an inhibition effect for the interlingual homographs or both. In line with the predictions, the analysis showed that the bilingual participants responded 37 ms more slowly to the interlingual homographs than to the English controls. In contrast, the difference between these two word types for the monolingual participants was only 13 ms. Despite a difference of 24 ms between these two effects, the interaction between word type and language group was not significant. Similarly, the interaction between word type and language group was not significant in the accuracy data, despite the fact that the bilingual participants made 4.8% more mistakes with the interlingual homographs, while the monolingual participants made only 2.3% more mistakes. Regarding the cognates, both groups of participants processed the cognates slightly more slowly than the English controls, in contrast to the predictions. For the monolingual participants, this difference was 11 ms, while for the bilinguals it was 4 ms. Neither of these effects was significant, however, nor was the interaction between them. The accuracy data told a similar story, although these effects were in the expected direction: the monolingual participants made 0.5% fewer mistakes with the cognates, while the bilingual participants made 0.9% fewer mistakes.

In short, it seems that the interaction between word type and language group in the main analysis was driven by an interlingual homograph inhibition effect and not a cognate facilitation effect (or both), but the statistics do not provide enough evidence to back up this claim. When comparing the data to the lexical decision data from Experiment 2, it appeared that for bilingual participants the cognate facilitation effect was significantly larger in lexical decision, while the interlingual homograph inhibition effect was marginally significantly bigger in semantic relatedness. This suggests that these two effects are to some extent task-dependent. Before discussing the results of Experiment 5 and the differences with lexical decision in more detail, however, Experiment 6 attempts to replicate these findings, in addition to examining whether recent experience with a cognate or interlingual homograph in one's native language affects subsequent processing of those words in one's second language.

# 3 Experiment 6: Cross-lingual long-term priming of identical cognates and interlingual homographs in semantic relatedness

Experiment 6 was again pre-registered as part of the Center for Open Science's Preregistration Challenge (www.cos.io/prereg). The stimuli, data and processing and analysis scripts can be found on the Open Science Framework (www.osf.io/2at49), as well as an Excel document with detailed results from all of the analyses. The preregistration can be retrieved from www.osf.io/y6phs (Poort & Rodd, 2018, February 28). Where applicable, deviations from the pre-registration will be noted.

## 3.1 Introduction

The aim of Experiment 6 was investigate, again, whether bilinguals processing cognates and interlingual homographs in their second language are affected by recent experience with these words in their native language. As discussed in Chapter 4 and the Introduction of this chapter, despite earlier success in demonstrating a cross-lingual long-term priming effect (Poort et al., 2016), subsequent efforts were unsuccessful. The most likely explanation for the absence of a priming effect in Experiment 3 and 4 is the use of a lexical decision task to probe an effect that is driven by semantics. Lexical decision tasks require only minimal access to semantics for participants to be able to respond accurately. Semantic relatedness tasks, in contrast, require access to a particular meaning of a word. Since the long-term priming effect is largely semantic in nature (Rodd et al., 2016; Rodd et al., 2013), an effect of priming is more likely to be observed in such tasks than in lexical decision. Indeed, as mentioned previously, this task has been used successfully in the monolingual domain to demonstrate effects of long-term priming (Betts, 2018; Gilbert et al., 2018).

To determine whether the semantic relatedness task can be used successfully to demonstrate cross-lingual long-term priming, Experiment 6 used the same design again as Poort et al. (2016) except instead of using a lexical decision task as the testing task, it used the same semantic relatedness task as Experiment 5. As in Experiment 3 and 4, participants first read Dutch sentences that contained either an identical cognate, an interlingual homograph or the Dutch forms of the translation equivalents. After a brief and non-linguistic filler task that created a delay of approximately 15 minutes, participants then completed the semantic relatedness task. Because cognates share their meaning across languages, cross-lingual long-term priming was predicted to have a facilitative effect on the cognates. In contrast, because the two meanings of an interlingual homograph are unrelated, priming was predicted to be disruptive for the interlingual homographs. Since the translation equivalents had the same meaning in Dutch as in English but did not share their form, a priming effect was not predicted for the translation equivalents, as before. Finally, by analysing the unprimed trials, it was possible to determine whether the pattern of results observed in Experiment 5 replicated (although this time only for a group of bilingual participants).

# 3.2 Methods

## 3.2.1 Participants

### 3.2.1.1 Power analysis

To maximise the chance of finding an effect of priming, a smallest-effect-size-of-interest power analysis was conducted in R (R Core Team, 2017) using the simr package (Green & MacLeod, 2016). The simr package runs power analyses using simulations based on an existing dataset. The dataset that was used to run these power simulations was the data from Experiment 5, since this dataset was expected to be most similar to the data for Experiment 6. Experiment 5, of course, did not include a priming manipulation, so participants and items were post-hoc ‘assigned’ to one of the two priming versions as planned for Experiment 6. Priming effects of approximately 20 ms for the cognates and the interlingual homographs were then simulated by subtracting or adding 20 ms to the reaction times for the ‘primed’ items.

A power analysis was conducted only for the analyses that had been specified in the preregistration as being directly related to the hypotheses: the two simple effects analyses of priming for the cognates and the interlingual homographs as well as the two 2×2 analyses that included the cognates and translation equivalents and the interlingual homographs and the translation equivalents. The significance level for these analyses was set at the Bonferroni-corrected level of .0167, as per the preregistration; the reaction times were also inverse-transformed (as the assumptions of normality and homogeneity were expected to be violated) and a maximal random effects structure was used for the linear mixed effects models (assuming that this would converge for these analyses). Power curves for these simulations can be found in Appendix C. Each power curve was based on 50 simulations each for 10 hypothetical sample sizes between 0 and 100. These power curves indicated that at least 60 to 80 participants were required for the two simple effects analyses and upwards of 100 for the two 2×2 analyses. As it seemed not feasible to recruit more than 100 participants, the target was set at 100 participants.

#### 3.2.1.2 Participant characteristics

In the end, a total of 114 participants was recruited through personal contacts resident in the Netherlands, the University of Ghent SONA system and social media. The eligibility criteria were the same as for Experiment 2 and 4 and the bilingual participants of Experiment 5. The participants gave informed consent and received a gift card worth €10 (roughly £8) for their participation in the experiment.

Excluding participants was again carried out in two stages. First, while testing was still on-going, participants who scored less than 80% correct on either of the semantic relatedness tasks and/or less than 50% on either of the two language proficiency measures were excluded and replaced. As in Experiment 4, participants were also excluded and replaced if their average delay between the two presentations of the primed items was more than 30 minutes. Nine participants that met these criteria were excluded and nine new participants tested in their stead. Due to a technical error, the data from four other participants could not be used, so these participants were also replaced. Second, after testing had finished and a total of 101 useable datasets had been gathered, each participant's performance on the targets (cognates, interlingual homographs and translation equivalents) included in the English semantic relatedness task was compared to the grand mean of all participants to determine whether any more participants needed to be excluded. All participants had performed within three standard deviations from the mean ($M$ = 88.1%, $SD$ = 5.6%), so none were excluded at this stage.

The 101 included participants (30 males, 70 females, 1 non-binary; $M_{age}$ = 24.7 years, $SD_{age}$ = 6.6 years) had started learning English from an average age of 8.1 ($SD$ = 3.0 years) and so had an average of 16.6 years of experience with English ($SD$ = 6.5 years). The participants rated their proficiency as 9.7 out of 10 in Dutch ($SD$ = 0.5) and as 8.7 in English

($SD = 0.9$). A two-sided paired $t$-test showed this difference to be significant [$t(100) = 11.12$, $p < .001$]. These self-ratings were confirmed by their high LexTALE scores in both languages, which a two-sided paired $t$-test showed were also higher in Dutch [Dutch: $M = 88.5\%$, $SD = 6.7\%$; English: $M = 85.8\%$, $SD = 9.3\%$; $t(100) = 2.767$ $p = .007$].

### 3.2.2 Materials

The same items were used as for Experiment 5. Minor changes were made to approximately one-fifth of the prime sentences for these items, for example when an item's probe for the English semantic relatedness task was also present in its prime sentence or any of the other prime sentences. Some sentences were also changed to make sure, as much as possible, that no translations of any of the probes from the English semantic relatedness task occurred in any of the Dutch prime sentences. Furthermore, none of those translations were used as probes in the Dutch semantic relatedness task either.

Many of the related probes that had originally been paired with the prime sentences to use in the Dutch semantic relatedness task were now being used in the English semantic relatedness task, as they were considered the best related probes for those items in English. These probes were replaced with new probes. Some prime sentences had to be modified or rewritten entirely if the choice of these new probes was limited. All modified and rewritten sentences were proofread by the same native speaker of Dutch who proofread the sentences for the rating experiments in Chapter 2. Independent-samples two-tailed Welch's $t$-tests showed that the differences between the word types in terms of prime sentence length were not significant (all $p$s > .6).

The 15 cognate, 15 interlingual homograph and 15 English control fillers now served a double purpose. First, they were included to ensure that the participants would not assume that any cognate or interlingual homograph always required a "yes"-response. Second, half of them were primed so that, as primed items that were assigned unrelated probes in the English task, they would also ensure that the participants would not assume that any words they had seen during the Dutch semantic relatedness task would always require a "yes"-response.

### 3.2.3 Design and procedure

This experiment employed a similar mixed design as in Experiment 3. Priming was a within-participants/within-items factor: for each participant, half the targets and eight of the fillers of each word type were primed (i.e. appeared during the priming phase) while half were unprimed (i.e. only occurred in the 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 same fillers were primed for all participants. Word type was a within-participants/between-items factor: participants saw words from all three word types, but each word of course belonged to only one word type.

The experiment was again created and conducted using Gorilla online experimental software (Evershed & Hodges, 2016). The experiment comprised five separate tasks: (1) the Dutch version of the LexTALE (Lemhöfer & Broersma, 2012), (2) the Dutch semantic relatedness task (mean duration in mm:ss: 08:38), (3) the Towers of Hanoi task (with instructions presented in English; maximum duration set to four minutes), (4) the English semantic relatedness task (mean duration: 10:53) and (5) the English version of the LexTALE (Lemhöfer & Broersma, 2012). Across participants, English semantic relatedness judgements to primed items were made on average 14 minutes and 27 seconds after they were primed in the Dutch semantic relatedness task. 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.

#### 3.2.3.1 Dutch semantic relatedness task

As in Experiment 3 and 4, this task served to prime the cognates, interlingual homographs and translation equivalents. Due to the large number of items that were primed, the design of the task was changed in some minor ways compared to Experiment 3 and 4. The 50 target sentences for each of the three word types were again pseudorandomly divided into two sets of 75, matched for all key variables and prime sentence length, for use in the two versions of the experiment. Including the 24 primed filler items, participants read a total of 101 experimental sentences. Half of the sentences in each version were paired with related probes and half with unrelated probes. A practice block of six sentences was followed by four blocks of 24 or 25 sentences (mixed targets and fillers). The order of the items within a block was randomised for each participant, as was the order of the blocks. Of all the items of each word type that were assigned to a block (6 or 7 items), 1 (or 2) item(s) was/were assigned to each of the six blocks of the English semantic relatedness task. This was done to ensure that the (variation in) duration of the delay would be similar for each of the three word types. A 15-second break was enforced after the first block and three fillers (additional to the 24 cognate, interlingual homograph and translation equivalent fillers) were presented at the start of each block. To shorten the duration of the task, participants in this experiment were allowed to read the sentences at their own pace, with a minimum presentation time of 1000 ms and a maximum of 4000 ms. Participants pressed the spacebar after they had read the sentence, after which the probe appeared on the screen. Each probe remained on the screen until the participant responded or until 2000ms passed. The inter-trial interval was 500ms.

#### 3.2.3.2 The Towers of Hanoi task

This task again served to introduce a delay between priming and testing, while minimising exposure to additional linguistic material. As in Experiment 4, the instructions were presented in English to minimise any general language switching cost on the English semantic relatedness task. Participants were given four minutes to complete as many puzzles as they could within this time limit, again 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.

#### 3.2.3.3 English semantic relatedness task

The English semantic relatedness task was identical to that used for Experiment 5, except that participants were given only 2000 ms to respond, to reduce the overall duration of this task. A warning was presented to the participant that they were responding too slowly if they had not responded 1500 ms after the probe first appeared. The warning remained on screen for 500 ms, during which time the participant could still respond. The response window was reduced to only 1500 ms as the experiment would otherwise likely have taken too long to complete, again due to the larger than usual number of sentences included in the priming task.

## 3.3 Results

All analyses were again carried out in R (version 3.4.4; R Core Team, 2018) using the lme4 package (version 1.1-17; Bates et al., 2015), following Barr et al.'s (2013) guidelines (with the same personal amendments as in Experiment 5) for determining the random effects structure of the models and using Type III Sums of Squares likelihood ratio tests to determine significance. Reaction times were analysed using the lmer() function with the default optimiser; accuracy data were analysed using the glmer() function with the bobyqa optimiser. Detailed results of all analyses for Experiment 6 can be found in the Excel document experimentOverview.xlsx in the OSF project (www.osf.io/2at49). Again, the graphs in the figures display 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 summary.

Four items (the cognate "nest"–"nest", the interlingual homograph "slot"–"slot" and the translation equivalents "emmer"–"bucket" and "moeras"–"swamp") were excluded from the analyses, as the percentages correct on the English semantic relatedness task for those items (61.4%, 23.8%, 35.6% and 55.4%, respectively) were more than three standard deviations below the items' word type mean (for the cognates: $M$ = 92.1%, $SD$ = 7.8%; for the interlingual homographs: $M$ = 81.3%, $SD$ = 17.9%; for the translation equivalents:

*M* = 90.8%, *SD* = 11.6%). Excluding these items did not affect the matching of the word types.

### 3.3.1 The Towers of Hanoi Task

Participants completed on average 2.3 puzzles (*mode* = 2, *range* = 1-3), confirming task engagement.

### 3.3.2 Dutch semantic relatedness task: Confirmatory analyses

High accuracy (*M* = 93.6%, *SD* = 3.9%, *range* = 80.8%–99.0%) confirmed participants had processed the sentence meanings. To determine whether any of the observed effects of priming in the English semantic relatedness task could have been due to differences between the word types or priming versions at the time of priming, a 3×2 analysis was conducted on the accuracy data with the fixed factors word type (3 within-participants/between-items levels: cognate, interlingual homograph, translation equivalent) and priming version (2 between-participants/within-items levels: version 1, version 2). 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 the main effect of word type was not significant [$\chi^2(2) = 1.226$, $p = .542$; $M_{\text{cognates}} = 97.6\%$, $M_{\text{interlingual homographs}} = 98.2\%$, $M_{\text{translation equivalents}} = 97.6\%$,], but the main effect of version was [$\chi^2(1) = 5.824$, $p = .016$], with participants in version 1 (*M* = 98.4%) on average outperforming those in version 2 (*M* = 96.8%) by 1.6%. This difference between the two versions could not have affected any of the key findings, which concern the differences between priming for the different word types. The interaction was also not significant [$\chi^2(1) = 2.999$, $p = .223$].

In addition, three pairwise comparisons were conducted, to compare the cognates, interlingual homographs and translation equivalents to 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 the pairwise comparisons were compared against a Bonferroni-corrected $\alpha$ of .0167. Again, these analyses revealed no significant differences between the word types [cognates (*M* = 97.5%) vs translation equivalents (*M* = 97.9%): $\chi^2(1) = 0.240$, $p = .625$; interlingual homographs (*M* = 98.1%) vs translation equivalents (*M* = 97.5%): $\chi^2(1) = 0.704$, $p = .402$; cognates (*M* = 97.1%) vs interlingual homographs (*M* = 98.2%): $\chi^2(1) = 1.854$, $p = .173$].

### 3.3.3 English semantic relatedness task: Confirmatory analyses

#### 3.3.3.1 Analysis procedure

The same analysis procedure was employed for the reaction times and accuracy data. In all cases, positive effects of priming indicate a facilitative effect of priming (i.e. faster reaction times and higher accuracy for primed items), while negative effects indicate a disruptive effect of priming (i.e. slower reaction times and lower accuracy for primed items). Positive (negative) effects of word type indicate an advantage (disadvantage) for the first-named word type over the second-named word type.

Two fixed factors were included in the main 2×3 analysis: priming (2 within-participant/within-items levels: unprimed, primed) and word type (3 within-participant/between-items levels: cognate, interlingual homograph, translation equivalent). The maximal random effects structure of this model included a correlated random intercept and random slope for word type, priming and their interaction by participants and a correlated random intercept and random slope for priming by items. This maximal model converged for the accuracy analysis but not the reaction time analysis, for which a model without correlations between the random effects converged.

In addition, three 2×2 analyses were conducted comparing the effect of priming for the cognates and translation equivalents, the interlingual homographs and translation equivalents and the cognates and interlingual homographs. The maximal random effects structure for these models also included a correlated random intercept and random slope for word type, priming and their interaction by participants and a correlated random intercept and random slope for priming by items. Again, the maximal model only converged for the accuracy analysis; a model without correlations between the random effects converged for the reaction times analysis. The $p$-values for these three analyses were compared against a Bonferroni-corrected $\alpha$ of .0167.

To examine the effect of priming for each of the three word types separately, three simple effects analyses were conducted. The maximal model converged for both the reaction times and accuracy analysis and included a correlated random intercept and random slope for priming by both participants and items. The $p$-values for these three analyses were compared against a Bonferroni-corrected $\alpha$ of .0167.

Finally, three pairwise comparisons were conducted on the unprimed data only, comparing the cognates, interlingual homographs and translation equivalents to each other as in Experiment 5. The maximal model converged for these three models for both the reaction times and accuracy analysis and included a correlated random intercept and random

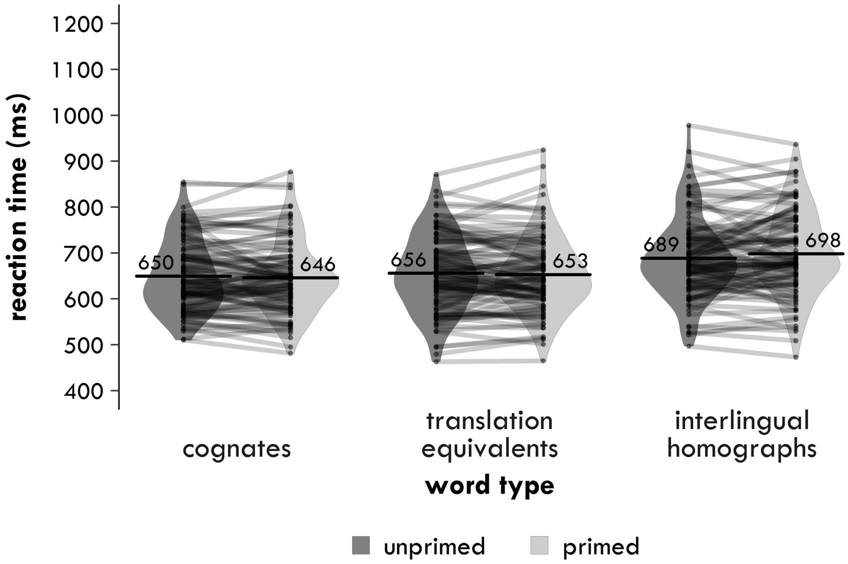


**Figure 5-3.** Experiment 6. Harmonic participant means of the inverse-transformed English semantic relatedness task reaction times (in milliseconds) by word type (cognates, translation equivalents, interlingual homographs; x-axis) and priming (unprimed, dark grey; primed, light grey). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

slope for word type by participants and a random intercept by items. The *p*-values for these three analyses were compared against a Bonferroni-corrected $\alpha$ of .0167.

#### 3.3.3.2 Reaction times

Reaction times are shown in Figure 5-3. Reaction times (RTs) faster than 300ms or slower than 1500ms were discarded (0.9% of the data), as were RTs for incorrect trials and trials that participants had not responded to (10.4% 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 2×3 analysis showed that the assumptions of normality and homoscedasticity were violated. After inverse-transforming the RTs, any inverse-transformed RTs were removed that were more than three standard deviations above or below a participant's mean inverse-transformed RT (0.1% of the remaining data).

***2×3***

In the 2×3 analysis, the main effect of priming was not significant [$\chi^2(1) = 0.115$, $p = .734$, $\Delta = -1$ ms]. The main effect of word type was significant, however [$\chi^2(2) = 24.66$, $p < .001$], indicating that there was a difference in reaction times between the three word types.

Critically, the interaction between word type and priming was also significant [$\chi^2(2) = 6.703$, $p = .035$].

***2×2s***

In the 2×2 analysis that included the cognates and translation equivalents, the main effect of priming was not significant [$\chi^2(1) = 1.725$, $p = .189$, $\Delta = 3$ ms], nor was the main effect of word type [$\chi^2(1) = 0.280$, $p = .597$, $\Delta = 5$ ms] or the interaction between word type and priming [$\chi^2(1) = 0.278$, $p = .605$]. In the 2×2 analysis that included the interlingual homographs and translation equivalents, the main effect of priming was not significant again [$\chi^2(1) = 1.707$, $p = .191$, $\Delta = -4$ ms] nor was the interaction between word type and priming [$\chi^2(1) = 3.234$, $p = .072$]. The main effect of word type was significant [$\chi^2(1) = 17.26$, $p < .001$], with participants responding on average 44 ms more slowly to the interlingual homographs than the translation equivalents. Finally, in the 2×2 analysis that included the cognates and interlingual homographs, the main effect of priming was not significant either [$\chi^2(1) = 0.480$, $p = .488$, $\Delta = -2$ ms]. The main effect of word type was significant [$\chi^2(1) = 19.37$, $p < .001$], with participants responding on average 49 ms more quickly to the cognates than the interlingual homographs. Crucially, the interaction between word type and priming was significant [$\chi^2(1) = 6.412$, $p = .011$].

***Simple effects***

In line with these findings, none of the simple effects of priming were significant at the Bonferroni-corrected $\alpha$ of .0167, though the effect of priming was significant at an uncorrected $\alpha$ of .05 for the interlingual homographs [for the cognates: $\chi^2(1) = 1.739$, $p = .187$, $\Delta = 5$ ms; for the interlingual homographs: $\chi^2(1) = 4.334$, $p = .037$, $\Delta = -10$ ms; for the translation equivalents: $\chi^2(1) = 0.308$, $p = .579$, $\Delta = 2$ ms].

***Pairwise comparisons***

The pairwise comparisons on the unprimed trials revealed a significant disadvantage of 37 ms for the interlingual homographs compared to the English controls [$\chi^2(1) = 11.72$, $p = .001$,], but no significant advantage for the cognates [$\chi^2(1) = 0.149$, $p = .699$, $\Delta = 4$ ms]. There was also a significant difference between the cognates and interlingual homographs of 41 ms [$\chi^2(1) = 13.34$, $p < .001$]. These results are fully consistent with the bilingual data from Experiment 5.

#### 3.3.3.3 Accuracy

Accuracy is shown in Figure 5-4. In line with the trimming procedure for the reaction times, any trials with RTs faster than 300ms or slower than 1500ms were removed. Overall, the

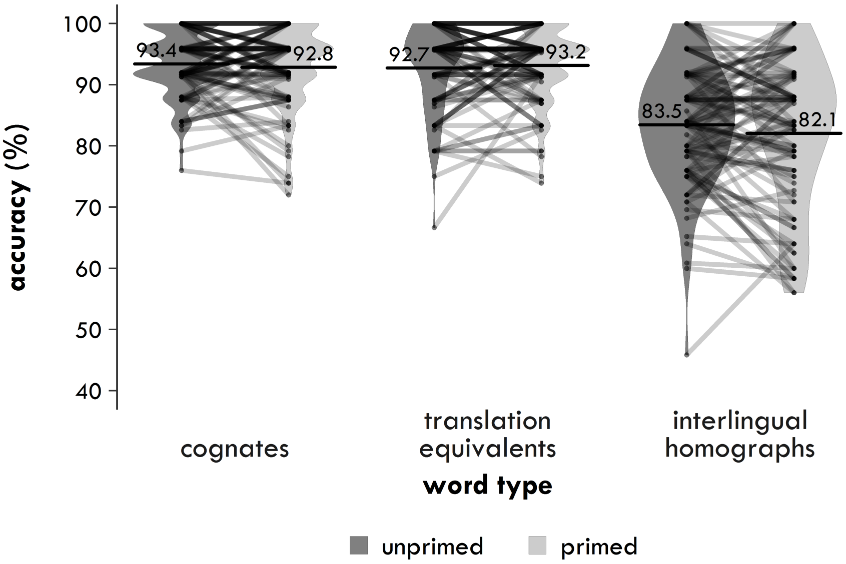


**Figure 5-4.** Experiment 6. Participant means of English semantic relatedness accuracy (percentages correct) by word type (cognates, translation equivalents, interlingual homographs; x-axis) and priming (unprimed, dark grey; primed, light grey). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

results for the accuracy data are essentially the same as those for the reaction time data, except that any hints at priming effects present in the reaction time data were not present in the accuracy data.

### *2×3*

In the 2×3 analysis, the main effect of priming was not significant [$\chi^2(1) = 0.194$, $p = .660$, $\Delta = 0.2\%$], but the main effect of word type was [$\chi^2(2) = 20.99$, $p < .001$]. The interaction between word type and priming was also not significant [$\chi^2(2) = 0.254$, $p = .881$].

### *2×2s*

In the 2×2 analysis that included the cognates and translation equivalents, the main effect of priming was also not significant [$\chi^2(1) = 0.340$, $p = .560$, $\Delta = 0.3\%$], nor was the main effect of word type [$\chi^2(1) = 0.0002$, $p = .990$, $\Delta = -0.01\%$], or the interaction between word type and priming [$\chi^2(1) = 0.005$, $p = .946$]. In the 2×2 analysis that included the interlingual homographs and translation equivalents, the main effect of priming was again not significant [$\chi^2(1) = 0.062$, $p = .803$, $\Delta = 0.2\%$], nor was the interaction between word type and priming [$\chi^2(1) = 0.088$, $p = .767$]. The main effect of word type was significant [$\chi^2(1) = 13.97$, $p < .001$], with participants responding on average 6.2% less accurately to the interlingual homographs than the English controls. Finally, in the 2×2 analysis that included the cognates and interlingual homographs, the main effect of priming was not significant [$\chi^2(1) = 0.028$, $p = .868$, $\Delta = 0.1\%$], nor was the interaction between word type and priming [$\chi^2(1) = 0.131$, $p = .717$]. The main effect of word type was significant [$\chi^2(1) = 13.45$, $p < .001$], with participants responding on average 6.2% more accurately to the cognates than the interlingual homographs.

#### *Simple effects*

In line with these findings, none of the simple effects of priming were significant [for the cognates: $\chi^2(1) = 0.098$, $p = .755$, $\Delta = 0.2\%$; for the interlingual homographs: $\chi^2(1) = 0.022$, $p = .883$, $\Delta = -0.2\%$; for the translation equivalents: $\chi^2(1) = 0.121$, $p = .728$, $\Delta = 0.3\%$].

#### *Pairwise comparisons*

Again, in line with the results from Experiment 5, the pairwise comparisons on the unprimed trials again revealed a significant disadvantage of 6.5% for the interlingual homographs compared to the English controls [$\chi^2(1) = 13.80$, $p < .001$], but no significant advantage for the cognates [$\chi^2(1) = 0.059$, $p = .808$, $\Delta = -0.3\%$]. There was also a significant difference between the cognates and interlingual homographs of 6.3% [$\chi^2(1) = 12.76$, $p < .001$].

## 3.3.4 English semantic relatedness task: Exploratory analyses

An exploratory analysis was conducted on the data from the English semantic relatedness task, to determine whether the effect of priming was influenced by speed of processing. As discussed in Chapter 4, priming may have been especially strong in Poort et al.'s (2016) experiment because their participants had unlimited time to respond during the lexical decision task. This exploratory analysis was conducted after a pattern was observed in the data that indeed suggested that priming was more or even only effective for participants who were generally slow to respond during this task, as can be seen in Figure 5-5. It may have been the case that priming was only effective for the slow responders because these participants were more likely to access the Dutch readings of the cognates and interlingual homograph, which may then have facilitated or disrupted their processing, while the fast responders may have responded based solely on the English reading of these words. To investigate this further, a median split analysis was conducted, comparing the size of the priming effect in fast and slow responders.

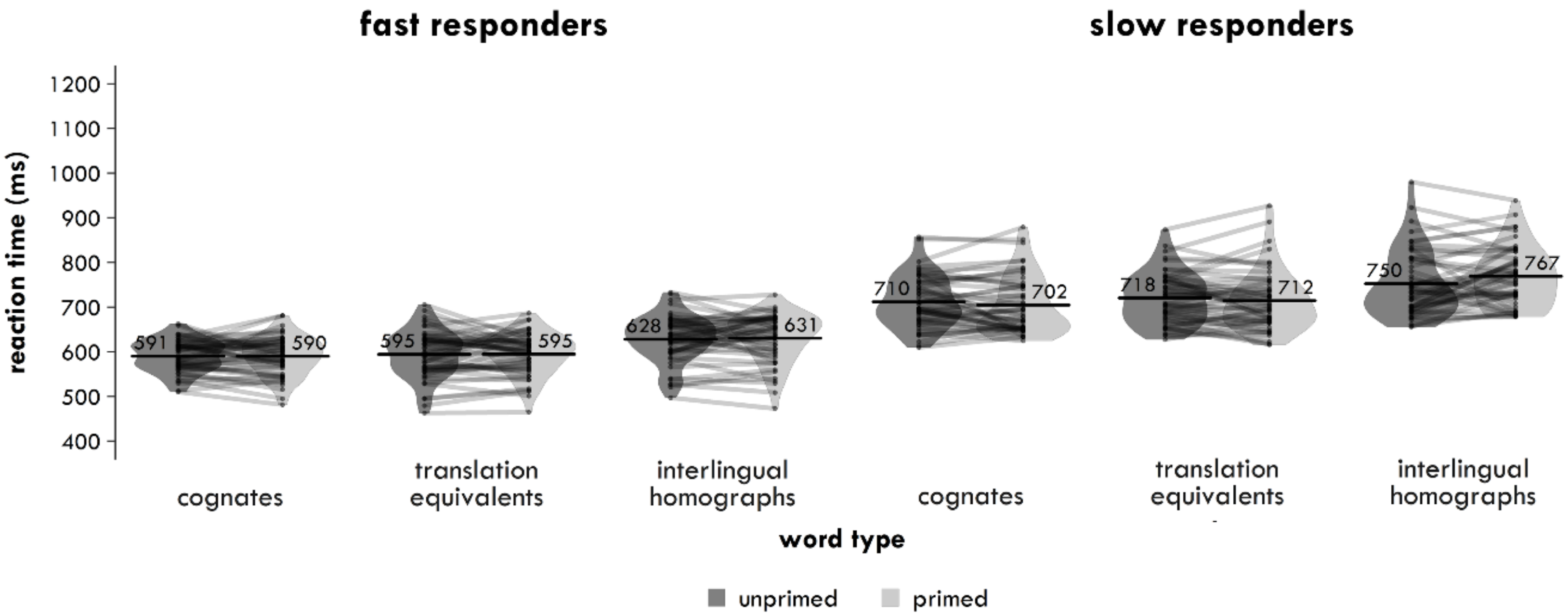


**Figure 5-5.** Experiment 6. Harmonic participant means of the inverse-transformed English semantic relatedness task reaction times (in milliseconds) by word type (cognates, translation equivalents, interlingual homographs; x-axis), priming (unprimed, dark grey; primed, light grey) and speed of responding (fast responders, slow responders). Each point represents a condition mean for a participant with lines connecting unprimed and primed means from the same participant. Each bar provides the mean across all participants in that condition. The violin is a symmetrical density plot rotated by 90 degrees.

Using the data from the 2×3, a mean inverse-transformed reaction time was calculated for each participant. Note that when working with inverse-transformed RTs, larger values indicate fast responses and smaller values indicate slow responses. Accordingly, participants were classified as 'fast' responders ($N = 51$) if their mean inverse-transformed RT was greater than or equal to the median inverse-transformed RT of all participants and as 'slow' responders ($N = 50$) if their mean was less than the median. Speed of responding was then entered into the 2×3 and the three simple effects analyses as an additional between-participants/within-items factor on its own as well as in an interaction with priming. In addition to the random effects that were already included in these models for the confirmatory analyses, these models included a random slope for speed of responding and for the interaction between priming and speed of responding by items. The correlations between the random effects were removed as the models would otherwise not converge.

In the 2×3 analysis, unsurprisingly, the main effect of speed of responding was highly significant [$\chi^2(1) = 92.23, p < .001$], with fast responders responding on average 122 ms more quickly than slow responders. The interaction between priming and speed of responding was not significant [$\chi^2(1) = 0.119, p = .731$], however. Including speed of responding as a factor in this model also did not change whether the effects of word type, priming or the interaction between word type and priming were significant. The main effect of speed was also highly significant in the three simple effects analyses [for the cognates: $\chi^2(1) = 95.28$, $p < .001$, $\Delta = 115$ ms; for the interlingual homographs: $\chi^2(1) = 78.08$, $p < .001$, $\Delta = 131$ ms; for the translation equivalents: $\chi^2(1) = 81.34$, $p < .001$, $\Delta = 121$ ms]. (Positive effects indicate an advantage for the fast responders.) As in the 2×3 analysis, the interaction between speed of responding and priming was not significant for any of the three simple effects, despite the priming effect being numerically larger for the slow responders in each of these three analyses [for the cognates: $\chi^2(1) = 0.140, p = .709, \Delta = 3$ ms vs 7 ms; for the interlingual homographs: $\chi^2(1) = 1.089$, $p = .297$, $\Delta = -4$ ms vs -17 ms; for the translation equivalents: $\chi^2(1) = 1.415$, $p = .234$, $\Delta = -2$ ms vs 8 ms]. In sum, although it appears that the effect of priming was numerically stronger for participants who responded more slowly, the analysis indicates that this was not statistically significant.

## 3.4 Discussion

The aim of Experiment 6 was two-fold: (1) to attempt to replicate the pattern of results in the bilingual group from Experiment 5, who showed an effect of interlingual homograph inhibition but no effect of cognate facilitation and (2) to determine whether cross-lingual long-term priming affects bilingual language processing. In terms of the first aim, the analysis of the unprimed trials of Experiment 6 revealed the same pattern of results as Experiment 5: significant disadvantages of 37 ms and 6.5% for the interlingual homographs compared to the English controls (i.e. the unprimed translation equivalents) but non-significant differences of only 4 ms and 0.3% between the cognates and the English controls. (Note that, although the reaction time effect for the cognates was still not significant, it was now in the expected direction. The effect in the accuracy data was in the opposite direction to what was predicted and what was found in Experiment 5.) Also consistent with the results from Experiment 5, the interlingual homographs were again processed more slowly and less accurately than the cognates, by 41 ms and 6.3%. In short, these results indicate that the findings from Experiment 5 are reliable.

In terms of the second aim, as in Experiment 3 and 4, there was weak evidence for a strong effect of cross-lingual long-term priming in Experiment 6. There were no effects of priming in the accuracy data at all. In the reaction time data, however, there was a significant interaction between word type and priming in the main analysis, as predicted. Crucially, the effect of priming was significantly different for the cognates compared to the interlingual homographs. As in Experiment 5, however, the other findings are less straightforward, making it difficult to determine whether this difference was due to a facilitatory effect of priming for the cognates or a disruptive effect for the interlingual homographs. Although the participants responded 5 ms more quickly to primed cognates than unprimed cognates, this effect was not significant and much smaller than the *a priori* effect size of interest (20 ms). Neither was the 10 ms disruptive effect of priming significant for the interlingual homographs (though note that it was significant at an uncorrected $\alpha$ of .05). In line with the predictions, the 2 ms 'facilitative' effect for the translation equivalents was also not significant. Unsurprisingly given the sizes of the effects, but again in contrast to the predictions, the priming effects for the cognates and for the interlingual homographs were not significantly different from the effect for the translation equivalents.

In sum, despite using a task that requires participants to access the meanings of the words they see and so was predicted to show a priming effect, a strong effect of priming was not observed. This is striking, as research in the monolingual domain has shown an effect of long-term priming when using a similar task (Betts, 2018; Gilbert et al., 2018). It is unclear whether these findings speak more to the mechanism that underlies *cross-lingual* long-term priming or to the representation of cognates and interlingual homographs in the bilingual mental lexicon. The findings of Experiment 6 will be discussed in more detail along with the findings of Experiment 5 in the General discussion.

# 4 General discussion

## 4.1 Processing cognates and interlingual homographs: Task effects

One of the aims of the experiments included in this chapter was to examine how bilinguals process cognates and interlingual homographs in a task that requires semantic access. In line with previous research using mainly lexical decision tasks, both Experiment 5 and 6 showed that bilinguals processed cognates differently than interlingual homographs. In particular, pairs that included an interlingual homograph were responded to 30-40 ms more slowly and 5-6% less accurately than pairs that included a cognate. Critically, in Experiment 5, this was not the case for the monolinguals, which indicates that the bilinguals processed these words differently because of their knowledge of them in Dutch. Unexpectedly, however, the experiments did not indicate that the bilinguals found the cognates easier to process than the English controls, which may call into question how researchers traditionally view these words. They did demonstrate that the bilingual participants found the interlingual homographs more difficult to process than the English controls. They were more than 35 ms slower on trials where they encountered an interlingual homograph and made approximately 5-6% more mistakes on these trials. The interaction with the monolingual participants was not quite significant, however, so some caution is advised when interpreting this finding.

Nevertheless, the presence of an interlingual homograph inhibition effect fits in with previous research showing such an effect for interlingual homographs in lexical decision tasks (Dijkstra et al., 1999; Dijkstra et al., 1998; Lemhöfer & Dijkstra, 2004; Van Heuven et al., 2008), sentence processing (Libben & Titone, 2009; Titone et al., 2011), auditory word recognition (Lagrou et al., 2011; Schulpen et al., 2003) and word production (Jared & Szucs, 2002; Smits et al., 2006). Furthermore, this finding is in line with research conducted by Macizo et al. (2010) who also used a semantic relatedness task to investigate the processing of interlingual homographs. As mentioned in the Introduction, in their experiment, Spanish–English bilinguals were slower to respond in English to identical interlingual homographs that were paired with probes that were related only to the Spanish meanings of the interlingual homographs (e.g. "pie"–"toe"; "pie" means "foot" in Spanish) than to control words paired with an unrelated probe (e.g. "log"–"toe").

The fact that an interlingual homograph inhibition effect was observed in a semantic relatedness task also appears to fit in the framework of the Bilingual Interactive Activation plus (BIA+) model of the bilingual mental lexicon (Dijkstra & Van Heuven, 2002). As mentioned in the Introduction, this model claims that interlingual homographs consist of two orthographic nodes that are each connected to their own semantic node (Kerkhofs et al., 2006). The lateral inhibition between the two orthographic nodes could also have been the cause of the disadvantage for interlingual homographs observed in the semantic relatedness task used in this experiment, although whether this would last long enough to survive the processing of the probe, which was presented after the interlingual homograph, is unclear. Alternatively, the model could be modified to allow competition between the two meanings of interlingual homographs.

Finally, it should be noted that the semantic relatedness task, like any task, involves decision processes and so response competition could have played a role in this experiment, especially for the interlingual homographs. However, whether this response competition was caused by language membership information or by information about the degree of relatedness between the target and probe is unclear. Specifically, when responding to pairs that included an interlingual homograph (e.g. "angel"–"heaven"), the fact that the unrelated meaning of the interlingual homograph (i.e. "insect's string") was the *Dutch* meaning of the interlingual homograph may have tempted the participants to respond "no". They may also have been tempted to respond "no" in this case solely because "insect's sting" is not semantically related to "heaven", however. Importantly, Experiment 5 was advertised and conducted entirely in English so that the participants would not assume that their knowledge of the stimuli in Dutch would be relevant to the task. Indeed, only four of the 29 bilingual participants had indicated that they had assumed that their knowledge of Dutch would be relevant (and another five said they had suspected so). Furthermore, unlike in the experiment conducted by Macizo et al. (2010), none of the probes in this experiment were exclusively related to the Dutch meanings of the interlingual homographs.

In other words, it is possible that the interlingual homograph inhibition effect observed here was (partially) the result of response competition. In that case, however, it seems more likely that this response competition was lexico-semantic in nature and arose because the "insect's sting" meaning was unrelated to the concept of "heaven" and competed with the related "spiritual being" meaning. It seems less likely that response competition emerged at the decision stage because the "insect's sting" meaning was known to be the Dutch meaning and participants had strategically linked the Dutch meanings of the interlingual homographs to the "no"-response. Finally, it should also be noted that the interlingual homograph inhibition effect was not much bigger in Experiment 6 than in Experiment 5, despite the fact that participants in Experiment 6 were aware that their knowledge of Dutch was relevant to the experiment. This also suggests that participants did not rely on language membership information during the semantic relatedness task and that any response competition that may have arisen was really due to competition between the two meanings of an interlingual homograph.

The lack of a cognate benefit (compared to the English controls) seems not to fit in with the large body of research that suggests that the cognate facilitation effect is very robust and transfers across tasks. As discussed in Chapter 3 and the Introduction to this chapter, the cognate facilitation effect has been observed mainly in visual lexical decision tasks (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), but it has also been found in word production tasks like picture naming (e.g. Costa et al., 2000) and single-word reading out loud (e.g. Schwartz et al., 2007). It has even been found when cognates were embedded in sentences during single sentence reading (Duyck et al., 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche et al., 2009; Van Hell & De Groot, 2008) and during reading of an entire novel (Cop, Dirix, Van Assche, Drieghe, & Duyck, 2016), although the facilitation effect is (much) smaller in these studies.

These findings are line with the single bilingual study conducted by Yudes et al. (2010) that used a semantic relatedness task to examine cognate processing. Yudes et al. (2010) found that Spanish–English bilingual participants completing a semantic relatedness task in Spanish did not respond more quickly or slowly to word pairs that included a cognate than word pairs that did not. In addition, there were no differences in the N400 component for pairs that included a cognate than those that did not. Although this may have been due to the fact that they used almost exclusively non-identical cognates, it remains a fact that Yudes et al. (2010) also found no evidence for a cognate facilitation effect.

Perhaps more problematically, these findings also do not seem to fit in the framework of the BIA+ model (Dijkstra & Van Heuven, 2002). Cognates, contrary to interlingual homographs, consist of a single orthographic node that is connected to a single semantic node (Peeters et al., 2013). Feedback looping back and forth from the semantic node to the orthographic node is thought to be the cause of the cognate facilitation effect. It seems the only way the current findings could be incorporated into the localist connectionist framework provided by the BIA+ model is to allow for separate (but identical) semantic representations for a cognate, one for each language. Furthermore, there should be no competition between these two semantic representations, otherwise a disadvantage would presumably have been observed in the current task (and likely in lexical decision tasks as well). However, if there were facilitation between these two semantic representations, this experiment would have shown evidence for a cognate facilitation effect, so it appears that these separate semantic representations would have to be completely independent, which is against the principles of an interactive activation model.

Although response competition may have played a role in the interlingual homograph inhibition effect, it seems unlikely that it could account for the absence of a cognate facilitation effect. For the cognates, which share their meaning across languages, the correct response would always have been "yes", as the probe was related to the cognate's meaning in both Dutch and English. Although there may be variation in how the cognates in this experiment are used in Dutch and English in daily life, the probes were not chosen to be related only to a specific sense of the cognates in English that they did not share with Dutch. It is, therefore, implausible that response competition (partially) cancelled out a facilitation effect as in the lexical decision tasks in Experiment 1 and 2.

To sum up, so far, Experiment 5 and 6 both suggest a disadvantage for the interlingual homographs compared to the English control words in a semantic relatedness task, but no difference in the processing of the cognates and English controls. Based on the assumption that tasks like lexical decision involve semantic processing to some degree, both an interlingual homograph inhibition effect and a cognate facilitation effect were predicted to be found in the semantic relatedness task. Research from the monolingual literature may be able to explain these contradicting findings. Using lexical decision tasks, many studies in the monolingual domain (e.g. Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007; Rodd et al., 2002) have shown that polysemous words — words with many related senses like "twist" which means "to make into a coil", "to operate by turning", "to misconstrue the meaning of", etc. — are often processed more quickly than unambiguous words like "dance". In contrast, in such task there is often no difference in processing times or only a small disadvantage for homonyms — words with multiple unrelated meanings like "bark" which means either "the cover of a tree" or "the sound a dog makes" — compared to unambiguous words. In tasks that involve more semantic processing, however, homonyms are often associated with a processing disadvantage (Hino, Pexman, & Lupker, 2006), while polysemous words are not processed any more quickly or slowly than unambiguous control words.

Rodd et al. (2004) developed a distributed connectionist model to account for the effects of different types of ambiguity (homonymy and polysemy) on processing times in lexical decision tasks. In this distributed connectionist model, homonyms consist of a single orthographic pattern that is connected to two separate semantic patterns. These semantic patterns form deep and narrow attractor basins in different regions of the semantic space. Settling into one of these attractor basins is difficult for the model and this is thought to cause the disadvantage for homonyms. Polysemous words, in contrast, consist of a single orthographic pattern that is connected to multiple overlapping semantic patterns. These patterns are close together in semantic space and, because of this, they form a broad and shallow attractor basin with multiple stable states. It is relatively easy for the network to settle into such an attractor basin and this results in a processing advantage for polysemous words compared to unambiguous words.

The settling dynamics of such a distributed connectionist model can also account for the different patterns of facilitation and inhibition observed for homonyms and polysemes in different kinds of tasks (but see Hino et al., 2006, for an account of these findings in terms of decision-making processes). Indeed, Rodd et al.'s (2004) simulations already revealed that in the later stages of settling, there was an advantage for the unambiguous words over the polysemous words. This suggests that in tasks in which access to a specific sense of the word is required, such as semantic categorisation or a semantic relatedness task, there may not be an advantage for polysemous words. Armstrong and Plaut (2008) also showed that the settling dynamics of such a model predict differences in the presence or absence of a polysemy advantage and a homonymy disadvantage in lexical decision versus semantic relatedness tasks. Their model showed that tasks that require little semantic activation would show a polysemy advantage but no homonymy disadvantage, while tasks that require high levels of semantic activation and specifically the retrieval of a precise semantic pattern would show no effect of polysemy but only a homonymy disadvantage. Using a lexical decision task with different levels of non-word difficulty, which affects the level of semantic specificity that is required, they supported the predictions from the computation model: when the non-words were easy and participants did not need to access a specific semantic representation of the words, a polysemy advantage was found; when the non-words were difficult and participants did have to retrieve a specific semantic representation, a homonymy disadvantage was observed.

There is a remarkable similarity between the pattern of results described for homonyms and polysemes in these monolingual studies and the contradiction between the pattern of results observed for the cognates and interlingual homographs in this experiment and the pattern of results found in the lexical decision tasks in Experiment 2. In the monolingual domain, the disadvantage for homonyms is often only small in lexical decision tasks, as it is for interlingual homographs in the bilingual domain when those tasks do not include non-target language words. In tasks that involve more semantic processing, homonyms often show a strong disadvantage, as the interlingual homographs did in this experiment. Indeed, the exploratory analysis for Experiment 5 revealed that the interlingual homograph inhibition effect was larger in the semantic relatedness task used in Experiment 5 than in the lexical decision task used in the +IH version of Experiment 2, although this interaction was only marginally significant. This striking similarity between the monolingual and bilingual domain suggests that interlingual homographs are not just etymologically the bilingual equivalent of homonyms, but also behaviourally.

Similarly, there is a compelling resemblance between the processing of polysemous words by monolinguals and the processing of cognates by bilinguals. In lexical decision tasks, monolinguals often process polysemes more quickly than unambiguous words, as bilinguals process cognates more quickly than control words (at least when the task does not include non-target language words). In tasks that require access to a specific meaning or sense of a word, polysemes are not processed differently than unambiguous words in the monolingual domain, the same way cognates were not processed differently than the English controls in this bilingual experiment. Again, the exploratory analysis for Experiment 5 revealed that the cognate facilitation effect was larger in the lexical decision task used in the standard version of Experiment 2 than in the semantic relatedness task used Experiment 5. This suggests that cognates may be processed by bilinguals and represented in the bilingual mental lexicon in a similar manner as polysemous words are processed by monolinguals and represented in the monolingual lexicon.

In the monolingual domain, the differences in how participants process homonyms and polysemes in different tasks appear best accounted for by the settling dynamics of a distributed connectionist network (Armstrong & Plaut, 2008, 2016). It seems, then, that such a model may also be more appropriate to explain the findings from Experiment 5 than a localist connectionist model like the BIA+ model. In a bilingual distributed connectionist model like the DFM, interlingual homographs would be represented in a similar manner as homonyms in a monolingual model: they would consist of a single orthographic pattern that is associated with multiple distinct semantic patterns. Cognates would likely have a similar representation as polysemous words: one single orthographic pattern that is connected to multiple overlapping semantic patterns, each of which represents a single sense. As mentioned in Chapter 1, the extent to which these patterns overlap would depend on the extent to which the cognate is used in mostly similar or slightly different contexts in the two languages it belongs to. The cognate "alarm", for example, would consist of highly overlapping patterns that represent the senses of "warning" and "warning system", which are shared in Dutch and English. It would also be associated with a pattern that represents its use in English to mean "alarm clock" and this pattern would overlap less with the other patterns, since in Dutch the word "wekker" would be used to refer to an "alarm clock". This presents another advantage of a distributed connectionist model over a localist model like the BIA+ model, as the BIA+ model assumes that cognates have exactly the same meaning across different languages.

Before continuing to discuss the findings from Experiment 6 that relate to priming, it seems necessary to revisit the alternative explanation that was offered for the effects observed in Chapter 3, namely that the lack of a cognate facilitation effect in the mixed version and the version with Dutch words (the +DW version) was because those versions tapped into a later stage of processing. If such a settling dynamics account, like the one proposed by Armstrong and Plaut (2008, 2016), is indeed the most appropriate account to explain differences in performance between lexical decision tasks and semantic relatedness tasks, then it stands to reason that this account can (and perhaps must) also be applied to differences between lexical decision tasks of varying degrees of difficulty. Such an account assumes that, by including stimuli that made the task more difficult (like the Dutch words), participants need more time to make a decision. Furthermore, it predicts that in the later stages of processing, there would be no advantage for cognates over control words. Indeed, Armstrong and Plaut (2008) support the predictions of their connectionist model with data from a lexical decision task with different levels of difficulty. In other words, the settling dynamics account proposed by Armstrong and Plaut (2008, 2016), when applied to cognates, predicts that there would be no advantage for cognates over control words in lexical decision tasks that require access to a (language-)specific representation of a cognate. This could have been the case in the mixed and +DW versions of Experiment 2, in which participants had to decide specifically whether the letter string they saw was a word in English or not. In the other versions of Experiment 2, participants could respond instead on the basis of a general sense of 'word-likeness'.

This explanation of the results of Experiment 2 seems unlikely to fully account for those results, however. First, it should be noted again that a large cognate facilitation effect was observed in the standard version of both Experiment 1 and 2. In Experiment 2, participants in this version were generally the slowest responders. Furthermore, a large disadvantage was observed for cognates when these words were immediately preceded by a Dutch word. This suggests that it was not the higher difficulty of the task in the mixed and +DW versions that caused a reduction of the cognate effect, but specifically the response competition caused by the presence of the Dutch words. This indicates that a settling dynamics account alone cannot explain the results of Experiment 1 and 2. In other words, it is most likely that both the settling dynamics of the mental lexicon and the decision criteria required by the task determine performance on a task.

## 4.2 Effects of cross-language long-term priming on the processing of cognates and interlingual homographs

The second aim of the experiments included in this chapter was to determine whether recent experience with a cognate or interlingual homograph in one's native language affects subsequent processing of these words in one's second language. Experiment 6 indeed demonstrated that there was some evidence for a cross-lingual long-term priming effect. Importantly, the effect of priming was in the predicted direction for both the cognates and the interlingual homographs, although the effects themselves were smaller than expected: for the cognates, priming was facilitative but the difference between primed and unprimed cognates was only 5 ms, while for the interlingual homographs it was disruptive and the difference was only 10 ms. The analyses also revealed that these two effects were significantly different from each other, as predicted, although neither effect was significant on its own nor significantly different compared to the 2 ms 'effect' for the translation equivalents. In sum, although cross-lingual long-term priming did not have as strong an effect as expected, Experiment 6 did broadly replicate Poort et al.'s (2016) findings that a single encounter with a cognate or interlingual homograph in one's native language can influence subsequent processing in one's second language and indicates that, crucially, it is the relatedness between such a words' meanings that determines how the bilinguals processed them.

On the assumption that long-term priming is caused by a strengthening of the connection between a word's form and its meaning (see Rodd et al., 2016; Rodd et al., 2013), these findings also suggest that interlingual homographs may share their form representation in the bilingual lexicon. The BIA+ model states that interlingual homographs have separate, language-specific form representations that are connected to the two language-specific meaning representations. In that case, it seems unlikely that priming would have been observed for the interlingual homographs as there is no reason to assume that strengthening the connection between the Dutch form representation and the Dutch meaning representation would interfere with accessing the English representations. It is difficult to draw any firm conclusions based on the results from Experiment 6, however, since the effect of priming for the interlingual homographs was not significant or significantly different to the effect for the translation equivalents.

Indeed, it remains a fact that the cross-lingual long-term priming manipulation had a weaker effect than predicted. It is unclear exactly why this was the case, but several interpretations are possible. First, it is possible that there is no such thing as a *cross-lingual* long-term priming effect or, more likely, that the *cross-lingual* long-term priming effect is not as strong as the *within-language* long-term priming effect in the monolingual domain. As mentioned previously, this long-term priming effect is thought to be the result of strengthened connections between a word's form and its primed meaning, which makes the primed meaning easier to access when the word form is encountered again. In the monolingual domain, many experiments (Betts, 2018; Betts et al., 2017; Gilbert et al., 2018; Rodd et al., 2016) have successfully replicated the original findings described by Rodd et al. (2013). In the bilingual domain, so far only Poort et al. (2016) have successfully demonstrated an effect of *cross-lingual* long-term priming.

One reason the cross-lingual long-term priming effect may be less strong is that the underlying mechanism is more akin to semantic priming than is the case for the long-term word-meaning priming effect in the monolingual domain. In their original study, Rodd et al. (2013) compared long-term word-meaning priming (e.g. using the prime sentence "The man accepted the **post** in the accountancy firm" to prime "post" in a subsequent word association task) to semantic priming (by replacing "post" with a synonym like "job" in the prime sentence). They found that the priming effect was less strong and more transient in this latter semantic priming condition than the word-meaning priming condition, which led them to the conclusion that long-term word-meaning priming is not a case of purely semantic priming. Similarly, Poort et al. (2016) primed a set of English control words with Dutch translations of those words (e.g. using the prime sentence "De schrijver zat achter zijn **bureau** te schrijven" to prime "desk", the translation of "bureau") in an effort to rule out semantic priming as an alternative explanation for cross-lingual long-term priming. However, they found that the priming effect was of roughly the same size for those 'semantic controls' as for their cognates. Indeed, although the priming effects were much smaller in Experiment 6 and Experiment 3, it appears that priming was to some extent effective for the translation equivalents as well. It may, therefore, be the case that cross-lingual long-term priming relies more heavily on a similar, perhaps more transient, mechanism as semantic priming.

In addition, in her discussion of the mechanism behind the long-term priming effect, Betts (2018) posits that there is probably also an episodic memory component to the effect, aside from the more direct effect priming appears to have on retuning of the lexical-semantic representations of (ambiguous) words. That is, the episodic memory of having encountered the word previously may be part of the reason why long-term priming is effective. Betts (2018) further suggests that the episodic traces must be consolidated into the lexicon before they can affect lexical-semantic processing. It could be the case that in cross-lingual long-term priming the episodic trace is relied upon less (e.g. because the language of presentation at the two time points is different) or that it takes longer for this trace to become consolidated into the lexicon and facilitate or interfere with processing. Alternatively, lexical-semantic representations may not be retuned in the same manner or as quickly for a bilingual, perhaps because words are encountered in many more different (language) contexts and updating the representation of those words each time they are encountered would be generally less beneficial. Finally, it is also possible that the different language-specific senses or meanings of a cognate or interlingual homograph are less tightly intertwined, so that priming one sense or meaning may not facilitate or interfere with processing of another sense or meaning as strongly as for a polyseme or homonym.

Second, it is possible that cross-lingual long-term priming operates on the same principles as within-language long-term priming, but that a strong effect of priming was not observed for some other reason to do with the design of the experiment or the differences between cognates and interlingual homographs on the one hand and polysemes and homonyms on the other. Priming was predicted for the cognates and the interlingual homographs on the assumption that they share their form across languages. However, cognates and interlingual homographs are often not pronounced identically in both languages (indeed, the mean pronunciation similarity rating for the cognates was 5.87 and for the interlingual homographs was 5.53), so if participants accessed the phonological form of these words as well as their orthographical form, this slight mismatch in pronunciation in Dutch and English may have interfered with priming.

Furthermore, as mentioned previously, cognates are unlikely to have exactly the same meaning in all of the languages they belong to, as they would naturally be used in slightly different contexts in those languages. The extent to which these meanings or senses are similar and overlap in the mental lexicon may also interact with the priming manipulation. When the participants encountered the cognates in Dutch, the connections between the cognates' form and the Dutch-specific senses were strengthened. This may not have benefitted the participants when they then encountered the words in English and had to retrieve the English-specific patterns. In fact, it is possible that for some cognates, the different senses were too different, which may have caused a disruptive effect of priming for these cognates.

In addition, it could also be the case that priming had a relatively small effect because the 'dominant' native-language meaning was primed and not the 'subordinate' second-language meaning. Since the participants were more fluent in Dutch (as demonstrated both by higher subjective proficiency ratings and higher objective LexTALE scores in Dutch) and were resident in the Netherlands or Belgium at the time of the experiment, the Dutch interpretations of the cognates and interlingual homographs were most likely dominant for these participants, while the English interpretations were subordinate. Research in the monolingual domain suggests that priming with a single instance of the dominant meaning does not affect subsequent processing presumably because this meaning is already strongly represented such that one additional encounter has minimal impact (Betts, 2018). Therefore, priming the cognates and interlingual homographs in Dutch may not have influenced subsequent access to the English meanings of these items. Other research carried out by Betts (2018) suggests that multiple instances with a word (albeit with its subordinate meaning) increase the long-term priming effect. This suggests that perhaps with multiple exposures to the items in Dutch, priming would have been observed.

This account of the findings could also provide an alternative interpretation of the presence of a strong priming effect for the interlingual homographs in Poort et al.'s (2016) experiment, despite their use of a lexical decision task. Their Dutch–English participants were living in the United Kingdom at the time of testing, which suggests that English was their dominant language and therefore the English meanings of the interlingual homographs were the dominant interpretations. Betts (2018) has shown that priming with the subordinate meaning of a homonym can interfere with the availability of the dominant meaning (albeit only after three exposures to this subordinate meaning). In other words, the priming manipulation in Poort et al.'s (2016) experiment may have been more effective because in their experiment the Dutch meanings that were used to prime the dominant English meanings were the subordinate meanings. It is less clear how this theory could explain why Poort et al. (2016) found priming for the cognates as well, however. Furthermore, in Experiment 3 the participants were also Dutch–English bilinguals living in the United Kingdom, yet no effect of priming was observed in this experiment. Further research would have to examine how the direction of priming may affect the size of the cross-lingual long-term priming effect (as in cross-lingual semantic priming and translation priming, see Schoonbaert et al., 2009; Wen & Van Heuven, 2017).

Thirdly and finally, it is also possible that speed of processing interacts with priming. This option was already explored in Chapter 4, where it was suggested that perhaps Poort et al. (2016) found priming because they did not set a time-out during their lexical decision task. Indeed, the exploratory analysis of Experiment 6 provided hints that participants who generally responded more slowly to the stimuli showed a greater effect of priming than participants who generally responded more quickly, although this effect was not significant. At the moment of priming, the encounter with the cognates and interlingual homographs in Dutch would have strengthened the connections between these words' orthographic patterns and their (Dutch-specific) semantic patterns. In the English task, however, these features may not have become active until a later point in processing, to the extent that they only facilitated or interfered with the processing of the primed cognates and interlingual homographs in English when participants responded more slowly. Further research would be required to determine whether this could indeed be the case.

# 5 Conclusion

Together, the findings from Experiment 5 and 6 suggest that a distributed connectionist approach, like that taken by the DFM (De Groot, 1992, 1993, 1995; De Groot et al., 1994; Van Hell, 1998; Van Hell & De Groot, 1998), Rodd et al. (2004) or Armstrong and Plaut (2008), may be the most appropriate approach to modelling the bilingual mental lexicon. In such a model, cognates would consist of a single orthographic representation, two highly overlapping phonological representations and two highly similar semantic representations within a single broad and stable attractor basin. Interlingual homographs, in contrast, would have two separate and distinct semantic representations that each form deep and narrow attractor basins (but also a single orthographic and two highly overlapping phonological representations).

The settling dynamics of such a model would by design be able to account for the different patterns of null effects, facilitation and inhibition for cognates and interlingual homographs in tasks like lexical decision and semantic relatedness. In addition, the learning rate of the model could be adjusted downwards if further research shows that cross-lingual long-term priming indeed is not as strong as monolingual long-term priming. By allowing non-target language (semantic) features to become active less quickly than target-language features, the settling dynamics of the model may also play a role in modulating the strength of the cross-lingual long-term priming effect for both cognates and interlingual homographs. Furthermore, the highly-overlapping but non-identical phonological representations for cognates and interlingual homographs may turn out to be a critical factor in determining the size of the cross-lingual long-term priming effect. The specific absence of a strong cross-lingual long-term priming effect could also be accommodated in this model by assuming that strengthening a native-language form-to-meaning connection has little effect on subsequent processing of the second-language interpretation.

# CHAPTER 6: Concluding remarks

# 1 Theoretical contributions

The aim of the experiments presented in this thesis was to shed light on the representation of cognates and interlingual homographs in the bilingual mental lexicon, by investigating both how cognates and interlingual homographs are processed in different tasks and when different types of stimuli are included in the same task and how they are processed following recent experience with them in another language. Table 6-1 on page 179 presents an overview of the experiments included in this thesis, their design and their main findings.

This thesis has made three important theoretical contributions to the field of bilingualism. First, the aim of the experiments presented in Chapter 3 (Experiment 1 and 2) was to determine whether cognates, like interlingual homographs, are subject to response competition (specifically, response competition elicited by the task that is influenced by the composition of the stimulus list). These experiments confirmed that this is indeed the case for interlingual homographs and further demonstrated that response competition plays a role in the processing of cognates as well. Therefore, it is crucial for models of the bilingual mental lexicon to include a mechanism that models the decision processes involved in the task that is used to study the processing of cognates and interlingual homographs.

Second, this thesis attempted to determine whether a bilingual is influenced by recent experience with a cognate or interlingual homograph in his or her first language when he or she subsequently comes across these words in their second language. Several experiments attempted to answer this question, to varying degrees of success. In Experiment 3 and 4, no evidence was found in favour of this cross-lingual long-term priming effect. In fact, the Bayes factors provided strong evidence against an effect of cross-lingual long-term priming. In broad terms, however, Experiment 6 replicated the initial findings reported by Poort et al. (2016): cross-lingual long-term priming had the expected facilitative effect on cognates and the expected disruptive effect on interlingual homographs, although the effects were smaller than expected.

Thirdly and finally, the experiments presented in Chapter 5 attempted to determine whether the cognate facilitation effect and the interlingual homograph inhibition effect are (partly) due to task artefacts or due to how these words are stored in the bilingual mental lexicon. Experiment 5 and 6 demonstrated that these effects are to some extent task-based, as there was no evidence for a cognate facilitation effect in a semantic relatedness task. However, it seems unlikely that the cognate facilitation effect and the interlingual homograph inhibition effect are merely artefacts of lexical decision. Indeed, Chapter 5 proposed that the settling dynamics of a distributed connectionist model would be able to account for the different and apparently contradictory patterns of facilitation and inhibition for cognates and interlingual homographs in lexical decision tasks (including those presented in Chapter 2) and the semantic relatedness tasks used in Experiment 5 and 6 (and by Macizo et al., 2010; Yudes et al., 2010). In other words, although the type of task does affect the size of the cognate facilitation effect and the interlingual homograph inhibition effect, Chapter 5 concluded that this is because different tasks tap into different levels of representation.

## 1.1 Contributions regarding the representation of cognates and interlingual homographs

In Chapter 1, several viewpoints were discussed that offer accounts of how cognates and interlingual homographs could be represented in the bilingual mental lexicon. Throughout this thesis, most attention was paid to the localist connectionist account offered by the BIA+ model (Dijkstra & Van Heuven, 2002) and to a speculative distributed connectionist account similar to the Distributed Feature Model (DFM; De Groot, 1992, 1993, 1995; De Groot et al., 1994; Van Hell, 1998; Van Hell & De Groot, 1998) or a bilingual extension of Rodd et al.'s (2004) or Armstrong and Plaut's (2008) models of semantic ambiguity resolution in the monolingual domain. Both the localist and distributed connectionist account offer a view on cognates as well as interlingual homographs, but Dijkstra et al. (2010) also identified two additional views on cognates: the single-morpheme view proposed by Kirsner and colleagues (Cristoffanini et al., 1986; Kirsner, Lalor, & Hird, 1993; Lalor & Kirsner, 2000) and Sanchéz-Casas and colleagues (Sánchez-Casas & García-Albea, 2005; Sánchez-Casas et al., 1992) and the associative-links account of the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994).

The data presented in this thesis offer new insight into the processing and representation of cognates and interlingual homographs. This in turn has consequences for the different models and viewpoints. For example, the fact that the composition of the stimulus list can affect the size of the cognate facilitation effect and the interlingual homograph inhibition effect can currently only be explained by the BIA+ model (Dijkstra & Van Heuven, 2002). Unlike the three other viewpoints, the BIA+ model includes a task system that allows it to take into account the decision-level processes involved in different tasks and different instantiations of the same task. As such, it is the only model that incorporates the response competition mechanism that decreased the cognate facilitation effect and increased the interlingual homograph inhibition effect in Experiment 1 and 2. It should be noted, however, that the task system proposed by the BIA+ model has not yet been implemented computationally, so future efforts should focus on this aspect of the model and/or incorporating a similar mechanism into other accounts.

Despite having a task system, the BIA+ appears unable to explain the lack of a cognate facilitation effect in the semantic relatedness judgement tasks used in Experiment 5 and 6, as discussed in Chapter 6. This finding also appears difficult to reconcile with the single-morpheme view, which posits that (non-identical) cognates share a single morphological representation. The difficulty lies largely in the fact that this account does not provide a well-specified explanation for the cognate facilitation effect in the first place. If the cognate facilitation effect depends on accessing a shared morphological representation, however, then it would seem only reasonable to assume that this effect should also be found in a semantic relatedness judgement task, assuming that morphological representations are accessed during such a task. The same reasoning applies to the account posited by the RHM. In this model, the cognate facilitation effect is the result of faster retrieval of the cognate's shared meaning representation through a strong associative link between the two language-specific lexical representations of a cognate, which is stronger than the associative links between the lexical representations of a translation equivalent. There is no reason to assume that this strong link would not result in faster retrieval of the cognate's meaning in a semantic relatedness task.

Indeed, as discussed previously, it appears that only a distributed connectionist account would be able to account for this finding. Furthermore, such an account would be the only account that would be able to explain any effects of cross-lingual long-term priming, as was discussed in the General discussion of Chapter 5. As Poort et al. (2016) originally noted, the BIA+ does not currently incorporate a learning mechanism that could support long-term priming, so it would need to be modified to allow it to account for effects of cross-lingual long-term priming. The same is true for the single-morpheme account and the RHM, neither of which is currently capable of learning from experience. Furthermore, in the case of the single-morpheme view, it is unclear whether priming would be explained by a strengthening of the connection between the cognate's form and its morphological representation, the connection between its morphological representation and its meaning representation or both. In the case of the RHM, the existence of an associative link between the language-specific lexical representations of cognates as well as translation equivalents would predict that cross-lingual long-term priming would be equally effective for cognates as for translation equivalents. Based on Rodd et al.'s (2013) findings with synonyms, this would not be expected, but both Poort et al.'s (2016) original data and the data from Experiment 3 and 6 suggest that cross-lingual priming may be effective for translation equivalents as well. In sum, it seems that most of the findings presented in this thesis would fit best in a distributed connectionist model, but even such a model would benefit from incorporating aspects of the other models.

| | DESIGN | | FACILITATION & INHIBITION EFFECTS | | | PRIMING EFFECTS | | | |
|---|---|---|---|---|---|---|---|---|---|
| | main task | key manipulation(s) | identical cognates | non-identical cognates | interlingual homographs | identical cognates | non-identical cognates | interlingual homographs | translation equivalents |
| Experiment 1 (Chapter 3) | lexical decision | stimulus list composition with two versions:<br>(a) standard<br>(b) mixed | (a) yes (+31 ms)<br>(b) no (-8 ms) | (a) –<br>(b) – | (a) –<br>(b) yes (-43 ms) | (a) –<br>(b) – | (a) –<br>(b) – | (a) –<br>(b) – | (a) –<br>(b) – |
| Experiment 2 (Chapter 3) | lexical decision | stimulus list composition with five versions:<br>(a) standard<br>(b) mixed<br>(c) +Dutch words<br>(d) +interlingual homographs<br>(e) pseudohomophones | (a) yes (+46 ms)<br>(b) no (+13 ms)<br>(c) no (+6 ms)<br>(d) yes (+22 ms)<br>(e) yes (+30 ms) | (a) –<br>(b) –<br>(c) –<br>(d) –<br>(e) – | (a) –<br>(b) yes (-24 ms)<br>(c) –<br>(d) no (-8 ms)<br>(e) – | (a) –<br>(b) –<br>(c) –<br>(d) –<br>(e) – | (a) –<br>(b) –<br>(c) –<br>(d) –<br>(e) – | (a) –<br>(b) –<br>(c) –<br>(d) –<br>(e) – | (a) –<br>(b) –<br>(c) –<br>(d) –<br>(e) – |
| Experiment 3 (Chapter 4) | lexical decision | cross-lingual long-term priming | no (+8 ms) | no (+20 ms) | – | no (+6 ms) | no (+9 ms) | – | no (+5 ms) |
| Experiment 4 (Chapter 4) | lexical decision | cross-lingual long-term priming & stimulus list composition with two versions:<br>(a) –interlingual homographs<br>(b) +interlingual homographs | (a) no (+10 ms)<br>(b) no (-8 ms) | (a) no (+7 ms)<br>(b) no (+1 ms) | (a) –<br>(b) yes (-61 ms) | (a) no (-1 ms)<br>(b) no (-0.1 ms) | (a) no (+0.1 ms)<br>(b) no (-1 ms) | (a) –<br>(b) no (-9 ms) | (a) –<br>(b) – |
| Experiment 5 (Chapter 5) | semantic relatedness | word type comparisons in two language groups:<br>(a) bilinguals<br>(b) monolinguals | (a) no (-4 ms)<br>(b) no (-11 ms) | (a) –<br>(b) – | (a) yes (-37 ms)<br>(b) no (-13 ms) | (a) –<br>(b) – | (a) –<br>(b) – | (a) –<br>(b) – | (a) –<br>(b) – |
| Experiment 6 (Chapter 5) | semantic relatedness | cross-lingual long-term priming | no (+4 ms) | – | yes (-37 ms) | no (+5 ms) | – | no (-10 ms) | no (+2 ms) |

**Table 6-1:** Summary of main experiments (reaction time data only). Effects that were significant are indicated as 'yes', the rest as 'no'. The facilitation (+) and inhibition (-) effects are for that word type compared to the English controls. Facilitative priming is indicated with +, disruptive priming with -.

# 2 Methodological contributions

Aside from the theoretical contributions this thesis has made, there are also several methodological remarks worth noting. A first remark concerns the need to use different tasks in order to form a complete picture of bilingual language processing. In the bilingual literature, most studies that have examined cognate and interlingual homograph processing have used lexical decision tasks. As discussed in Chapter 3 and Chapter 5, response competition plays a major role in these tasks. This makes it difficult to determine whether effects of facilitation and inhibition in these tasks are due to how cognates and interlingual homographs are represented in the bilingual lexicon or whether these effects arise during decision making. While no single task is without some kind of decision component, only by integrating findings from different tasks and being explicit about the expectations and strategies that a particular task may elicit in participants can we make progress with respect to the representation of words in the mental lexicon.

Furthermore, many of the tasks that have been used to study bilingual language processing in the past do not require participants to access and process the meanings of the words they see (but for experiments examining cognate processing using a word association task, see Van Hell & De Groot, 1998; Van Hell & Dijkstra, 2002). In recent years, research has started investigating the processing of cognates and interlingual homographs in sentence contexts (Duyck et al., 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Assche et al., 2009; Van Hell & De Groot, 2008), yet there is still a lack of research that focusses on processing of cognates and interlingual homographs at a semantic level. The studies that have used eye-tracking have demonstrated that even in the context of a sentence, bilinguals still process cognates more quickly than single-language control words and interlingual homographs more slowly. These studies have also shown, however, that the cognate facilitation effect and interlingual homograph inhibition effect are often smaller when these words are embedded in sentences (both in an absolute sense and relative to the average processing times). More research is needed to determine whether this is due to the context provided by the sentence, which may bias the interpretation of these words, or because sentence reading naturally involves semantic processing, which Experiment 5 and 6 have shown can affect the size of the facilitation and inhibition effects observed traditionally.

A second remark concerns the benefits of conducting web-based research. Of the experiments presented here, only Experiment 3 was conducted in the lab. Experiment 1, 2, 4, 5 and 6, as well as both rating experiments, were conducted online either using Qualtrics

(Qualtrics, 2015; Rating experiment 1 and 2) and the Qualtrics Reaction Time Engine (Barnhoorn et al., 2014; Experiment 1 and 2) or Gorilla (Evershed & Hodges, 2016; Experiment 4, 5 and 6). For this thesis, the most obvious advantage of conducting these experiments online was that a much larger and more representative sample of Dutch–English bilinguals could be recruited than would have been possible had all experiments been conducted in the London-based lab. Furthermore, conducting these experiments online meant that participants could be recruited who were resident in the Netherlands (or Belgium). This is especially important when considering the need to compare the results obtained in these experiments to the bilingual literature, which is mostly comprised of research examining processing of the second, lesser-used language.

In general, it appears that both reaction-time based and questionnaire-based experiments conducted online provide data of a similar quality as lab-based experiments (e.g. Barnhoorn et al., 2014; Casler, Bickel, & Hackett, 2013; Gosling, Vazire, Srivastava, & John, 2004; Hilbig, 2016; McGraw, Tew, & Williams, 2000). Nevertheless, some people have argued that online research is inferior to lab-based research, particularly with respect to the accuracy of the reaction time measurements (Plant, 2016). Indeed, in these experiments there was considerably less between-participant variation in average reaction times in Experiment 3 compared to the other experiments, but of course this may have been due to the more homogenous composition of the participants that were tested in Experiment 3. Furthermore, reassuringly, the size of the cognate facilitation effects and the interlingual homograph inhibition effects found in the lexical decision tasks used in Experiment 1, 2 and 4 were of a similar size as the effects usually reported in the literature. This suggests that the quality of the data gathered in Experiment 1, 2, 4, 5 and 6 was high and certainly not inferior to the data gathered in the lab for Experiment 3.

A common argument against web-based research is that it is difficult, if not impossible, to verify participant eligibility. Indeed, Experiment 1, 2, 4, 5 and 6 involved a little more effort and up-front thinking to ensure that the participants met the eligibility requirements, but relatively few participants were excluded because they had not performed the task well enough. In addition, combining web-based experiments with recruitment platforms such as Prolific (Damer & Bradley, 2014) offers the great advantage of being able to more easily recruit participants without making them aware of the study's aims, as was the case for the bilingual participants that were recruited for Experiment 5. Some of the other advantages of conducting web-based experiments include the fact that recruitment is often faster online than in the lab and can be cheaper than conducting lab-based experiments, as participants are less likely to come into the lab for a £3 reward than they are to spend half an hour of their time at home doing an online experiment. All in all, it seems that web-based research offers a promising alternative to lab-based research.

A third point to note concerns the benefits of the Open Science movement for individual researchers. The well-known study conducted by the Open Science Collaboration (2015) showed that the reproducibility of psychological research is poor: of the 100 studies they attempted to replicate, 97% of the original experiments found a significant effect, while only 36% of the replications did. Average effect sizes were also half as small in the replication studies as in the originals. Some causes of this poor replication rate have been identified and these include publication bias, hypothesising after the results are known (i.e. 'HARKing'), failure to control for (experimenter) bias, low statistical power, and *p*-hacking (Munafò, Nosek, Bishop, Button, Chambers, Percie du Sert, Simonsohn, Wagenmakers, Ware, & Ioannidis, 2017). One solution that has been proposed to counter many of these threats to reproducible science is preregistration of study protocols.

In its most basic form, a preregistration involves specifying up-front what research question the study will attempt to answer, what the hypotheses are, how the experiment has been designed (including what variables will be measured and/or manipulated), how the data will be collected (including rationales for the chosen sample size) and how the data will be analysed. While some have argued that preregistration limits creativity and encourages selective publishing based on reputation (Scott, 2013), others have argued the direct opposite, that Open Science is liberating and fosters creativity and collaboration (Frankenhuis & Nettle, 2018, Feb 18). Given the many benefits of preregistration for both the field of psychological research (Mellor, 2017, Sept 19; Nosek, Ebersole, DeHaven, & Mellor, 2018) and the individual researcher (McKiernan, Bourne, Brown, Buck, Kenall, Lin, McDougall, Nosek, Ram, Soderberg, Spies, Thaney, Updegrove, Woo, & Yarkoni, 2016), four of the experiments in this thesis (Experiment 2, 4, 5 and 6) were pre-registered as part of the Center for Open Science's Preregistration Challenge.

Writing the preregistration document for the first of these (Experiment 2) did take some time, but the time spent narrowing down the analysis and deciding on the exclusion criteria up-front offered the great advantage of analysing the data in a theory-driven manner — with the aim of answering the particular question this experiment asked in the best manner possible — instead of a data-driven manner — the aim of which is usually to produce the most impactful and clean narrative. The process also did not feel stifling but rather reassuring, in that it increased my confidence in the quality of the design and the 'truth value' of the confirmatory outcomes. Indeed, one major benefit of this process was that any flaws in the experiment's design could be identified and rectified prior to starting data collection. Additionally, analysing the data in a data-driven manner could have been particularly troublesome for this experiment, given the complicated design of this experiment and the many different analyses that could have been conducted. Importantly, these 'start-up' costs were much smaller for the subsequent preregistrations, to the point that conducting Experiment 6 — including writing the preregistration — took hardly any more time than if this experiment had not been pre-registered.

Furthermore, there were considerable advantages to uploading the stimuli, (raw) data and processing and analysis scripts for all pre-registered experiments to the Open Science Framework as well. Again, for the first of these experiments it took quite some time to ensure that these materials were in a format that could be used relatively easily by other researchers. The great advantage of doing this, however, is that the materials are now forever well-documented and can be re-used and re-analysed by anyone at any point in the future. This process was improved greatly by relying more and more on R for the processing and analysis of the data, to the point that I have almost entirely automatised these aspects of conducting research. The skills that this has taught me are applicable to any research project and may be the most valuable skills I have learnt during my PhD.

# 3 Future directions

This thesis has made valuable contributions to the field of bilingualism. Not all of the questions that the experiments in this thesis attempted to address have been answered, however. One of these questions concerns the conditions under which recent experience with cognates and interlingual homographs in one languages affects processing of these words in another. As discussed in Chapter 5, there are various reasons that could account for why the cross-lingual long-term priming effect in Experiment 6 was smaller than the within-language long-term priming effect observed in the monolingual domain. For example, there may be differences in the mechanisms that underlie cross-lingual long-term priming and within-language long-term priming. Further research that directly compares cross-lingual and within-language long-term priming (ideally in the same group of participants) would be especially informative to resolve this issue. Furthermore, it is also important to compare the direction of priming (from first to second language and vice versa) and how this influences the size of the effect. It may have been the case that priming was less effective in Experiment 6 than in Poort et al.'s (2016) original experiment because the language that was used to prime the cognates and interlingual homographs was the participants' dominant language. The other question that remains open concerns the representation of non-identical cognates in the bilingual mental lexicon. In Experiment 3, cross-lingual long-term priming appeared to affect the non-identical cognates in a similar manner as it affected the identical cognates. If future work convincingly manages to replicate Poort et al.'s (2016) original findings, this method may still prove to useful to study the representation of non-identical cognates.

Altogether, the research presented in this thesis indicates that it is necessary to explore the viability of a distributed connectionist account of cognates and interlingual homographs. The most important next step is to run computational simulations to determine if such a model can indeed account for the findings presented in this thesis. In addition, when developing this distributed connectionist account, it is important to keep in mind that, to be a comprehensive model of the bilingual mental lexicon, this model must eventually also be able to account for many of the other findings in the field of bilingualism. Special attention must also be given to finding computational means of incorporating the effects of response competition found in Experiment 1 and 2 in such a model, for example by way of including a task system as in the Bilingual Interactive Activation plus (BIA+) model (Dijkstra & Van Heuven, 2002).

Finally, while running these computational simulations, it is also crucial not to forget that cognates and interlingual homographs are in essence just semantically ambiguous words. Nevertheless, it seems that few researchers in the bilingual domain treat them as such. As researchers in the bilingual domain, therefore, we may be able to learn a lot from our colleagues in the monolingual domain who study polysemes and homonyms. Indeed, as discussed in the General discussion of Chapter 5, bilinguals may process cognates and interlingual homographs in a similar manner as monolinguals process polysemes and homonyms. At the same time, there may be distinct differences between how monolinguals process semantic ambiguity within a language and how bilinguals process ambiguity that crosses language boundaries. There may even be differences in how monolinguals and bilinguals process polysemes and homonyms in a single language. Perhaps most importantly, this thesis underscores the argument made by Degani and Tokowicz (2009) that bilingual and monolingual researchers should work together in their quest to determine how people resolve semantic ambiguity.

# REFERENCES


Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 802–814. doi: 10.1037/0278-7393.15.5.802.

Armstrong, B. C., & Plaut, D. C. (2008). *Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence.* Paper presented at the 30th Annual Meeting of the Cognitive Science Society.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience, 31*(7), 940–966. doi: 10.1080/23273798.2016.1171366.

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*(4), 484–504. doi: 10.1006/jmla.1997.2502.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. doi: 10.3758/BF03193014.

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. doi: 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. doi: 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. doi: 10.18637/jss.v067.i01.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5*(2), 252–259. doi: 10.1037/0096-1523.5.2.252.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8*(6), 493–512. doi: 10.3758/bf03213769.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343–355. doi: 10.1016/0013-4694(85)90008-2.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research, 24*(1), 57–65. doi: 10.1016/j.cogbrainres.2004.12.006.

Betts, H. N. (2018). *Retuning lexical-semantic representations on the basis of recent experience.* (Ph.D. dissertation, University College London, London, United Kingdom).

Betts, H. N., Gilbert, R. A., Cai, Z. G., Okedara, Z. B., & Rodd, J. M. (2017). Retuning of lexical-semantic representations: Repetition and spacing effects in word-meaning priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi: 10.1037/xlm0000507.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition, 25*(4), 447–457. doi: 10.3758/BF03201121.

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. doi: 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. doi: 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. doi: 10.1037/xhp0000159.

Bultena, S., Dijkstra, T., & Van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology, 67*(6), 1214–1241. doi: 10.1080/17470218.2013.853090.

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology, 98*, 73–101. doi: 10.1016/j.cogpsych.2017.08.003.

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society, 13*(4), 212–214. doi: 10.3758/BF03335062.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*(6), 2156–2160. doi: 10.1016/j.chb.2013.05.009.

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. doi: 10.1037/xlm0000065.

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2016). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition, 20*(4), 1–23. doi: 10.1017/S1366728916000213.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296. doi: 10.1037/0278-7393.26.5.1283.

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. doi: 10.1080/14640748608401604.

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

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 36*(3), 435–444. doi: 10.1037/h0080650.

De Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects on interlingual homograph recognition: Evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition, 4*(2), 155–168. doi: 10.1017/S1366728901000256.

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1001–1018. doi: 10.1037/0278-7393.18.5.1001.

De Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27–51). Amsterdam, The Netherlands: John Benjamins.

De Groot, A. M. B. (1995). Determinants of bilingual lexicosemantic organisation. *Computer Assisted Language Learning, 8*(2–3), 151–180. doi: 10.1080/0958822940080204.

De Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language, 33*(5), 600–629. doi: 10.1006/jmla.1994.1029.

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. doi: 10.1080/713755891.

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*(1), 90–123. doi: 10.1016/0749-596X(91)90012-9.

Degani, T., & Tokowicz, N. (2009). Semantic ambiguity within and across languages: An integrative review. *The Quarterly Journal of Experimental Psychology, 63*(7), 1266–1303. doi: 10.1080/17470210903377372.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). New York, NY: Oxford University Press.

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. doi: 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. doi: 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. doi: 10.1016/j.jml.2009.12.003.

Dijkstra, T., Timmermans, M., & Schriefers, H. J. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*(4), 445–464. doi: 10.1006/jmla.1999.2697.

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

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. doi: 10.1017/S1366728902003012.

Dijkstra, T., & Van Heuven, W. J. B. (2012). Word recognition in the bilingual brain. In M. Faust (Ed.), *The handbook of the neuropsychology of language* (pp. 451–471). Chichester, UK: Blackwell Publishing.

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. doi: 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. doi: 10.1037/0278-7393.33.4.663.

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.* (Ph.D. dissertation, Université Paul Valery, Montpellier, France).

Frankenhuis, W., & Nettle, D. (2018, Feb 18). Open science is liberating and can foster creativity. *OSF Preprints*. doi: 10.17605/osf.io/edhym.

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2018). Listeners and readers generalize their experience with word meanings across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication, doi: 10.1037/xlm0000532.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*(2), 93–104. doi: 10.1037/0003-066X.59.2.93.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29*(2), 228–244. doi: 10.1016/0749-596X(90)90074-A.

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. *Advances in Psychology, 83*, 207–220. doi: 10.1016/S0166-4115(08)61496-X.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81. doi: 10.1017/S1366728998000133.

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. doi: 10.1111/2041-210X.12504.

Hilbig, B. E. (2016). Reaction time effects in lab- versus web-based research: Experimental evidence. *Behavior Research Methods, 48*(4), 1718–1724. doi: 10.3758/s13428-015-0678-9.

Hino, Y., Pexman, P. M., & Lupker, S. J. (2006). Ambiguity and relatedness effects in semantic tasks: Are they due to semantic coding? *Journal of Memory and Language, 55*(2), 247–273. doi: 10.1016/j.jml.2006.04.001.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*(1), 66–85. doi: 10.1016/0093-934X(88)90101-0.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*(4), 281–312. doi: 10.1080/01690969008407065.

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology, 41*(6), 401–410. doi: 10.1037/h0056020.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1*(2), 130–136. doi: 10.1037/0096-1523.1.2.130.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language, 44*(1), 2–31. doi: 10.1006/jmla.2000.2747.

Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition, 5*(3), 225–239. doi: 10.1017/S1366728902003024.

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*(1), 170–183. doi: 10.1016/j.brainres.2005.10.087.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627–633. doi: 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. doi: 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. doi: 10.1080/17470218.2015.1022560.

Kirsner, K., Lalor, E., & Hird, K. (1993). The bilingual lexicon: Exercise, meaning and morphology. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 215–248). Amsterdam, The Netherlands: John Benjamins Publishing.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20*(1), 1–24. doi: 10.1016/j.jneuroling.2006.02.001.

Kroll, J. F., Dijkstra, A., Janssen, N., & Schriefers, H. J. (1999). *Cross-language lexical activity during production: Evidence from cued picture naming.* Paper presented at the 11th Congress of the European Society for Cognitive Psychology.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174. doi: 10.1006/jmla.1994.1008.

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), 952–965. doi: 10.1037/a0023217.

Lalor, E., & Kirsner, K. (2000). Cross-lingual transfer effects between English and Italian cognates and noncognates. *International Journal of Bilingualism, 4*(3), 385–398. doi: 10.1177/13670069000040030501.

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. doi: 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. doi: 10.3758/BF03195845.

Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes, 19*(5), 585–611. doi: 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. doi: 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. doi: 10.1037/a0014875.

Macizo, P., Bajo, T., & Cruz Martín, M. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish–English interlexical homographs. *Journal of Memory and Language, 63*(2), 232–244. doi: 10.1016/j.jml.2010.04.002.

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173–193. doi: 10.1017/S0142716403000092.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review, 88*(5), 375–407. doi: 10.1037/0033-295X.88.5.375.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11*(6), 502–506. doi: 10.1111/1467-9280.00296.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., . . . Yarkoni, T. (2016). How open science helps researchers succeed. *eLife, 5*, e16800. doi: 10.7554/eLife.16800.

Mellor, D. (2017, Sept 19). Preregistration and increased transparency will benefit science. *OSF Preprints*. doi: 10.20316/ESE.2017.43.018.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234. doi: 10.1037/h0031564.

Midgley, K. J., Holcomb, P. J., Van Heuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research, 1246*, 123–135. doi: 10.1016/j.brainres.2008.09.078.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 1–9. doi: 10.1038/s41562-016-0021.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. doi: 10.1073/pnas.1708274114.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi: 10.1126/science.aac4716.

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. doi: 10.1016/j.jml.2012.12.003.

Perea, M., & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research, 66*(3), 180–194. doi: 10.1007/s00426-002-0086-5.

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition, 28*(4), 657–666. doi: 10.3758/bf03201255.

Plant, R. R. (2016). A reminder on millisecond timing accuracy and potential replication failure in computer-based psychology experiments: An open letter. *Behavior Research Methods, 48*(1), 408–411. doi: 10.3758/s13428-015-0577-0.

Poort, E. D., & Rodd, J. M. (2016, February 8). Does the cognate facilitation effect depend on task demands? Preregistration retrieved from www.osf.io/9b4a7.

Poort, E. D., & Rodd, J. M. (2017, December 7). Investigating the cognate facilitation effect and the interlingual homograph inhibition effect with a semantic relatedness task. Preregistration retrieved from www.osf.io/u2fyk.

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. Preregistration retrieved from www.osf.io/33r86.

Poort, E. D., & Rodd, J. M. (2018, February 28). Does recent experience with identical cognates and interlingual homographs in one's native language affect processing of these words in one's second language? Preregistration retrieved from www.osf.io/y6phs.

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. doi: 10.1017/S1366728915000395.

Qualtrics. (2015). Qualtrics Survey Software. Provo, Utah, USA: Qualtrics. Available at www.qualtrics.com.

R Core Team. (2015). R: A language and environment for statistical computing (Version 3.2.1). Vienna, Austria: R Foundation for Statistical Computing. Available at www.R-project.org.

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 www.R-project.org.

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.3). Vienna, Austria: R Foundation for Statistical Computing. Available at www.R-project.org.

R Core Team. (2018). R: A language and environment for statistical computing (Version 3.4.4). Vienna, Austria: R Foundation for Statistical Computing. Available at 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*(4), 1339–1362. doi: 10.1080/02724980244000099.

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

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. doi: 10.1016/j.jml.2015.10.006.

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*(2), 180–198. doi: 10.1016/j.jml.2012.08.002.

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*(2), 245–266. doi: 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. doi: 10.1016/j.cogsci.2003.08.002.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception, Part II: The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*(1), 60–94. doi: 10.1037/0033-295X.89.1.60.

Sánchez-Casas, R. M., & García-Albea, J. E. (2005). The representation of cognate and noncognate words in bilingual memory: Can cognate status be characterized as a special kind of morphological relation? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 226–250). New York, USA: Oxford University Press.

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. doi: 10.1080/09541449208406189.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition, 37*(5), 569–586. doi: 10.3758/mc.37.5.569.

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. doi: 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. doi: 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 and Cognitive Processes, 22*(1), 106–129. doi: 10.1080/01690960500463920.

Scott, S. (2013, July 25). Pre-registration would put science in chains. Times Higher Education. Retrieved from http://www.timeshighereducation.com/comment/opinion/pre-registration-would-put-science-in-chains/2005954.article

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*(3), 281–297. doi: 10.1017/S136672890600263X.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281–284. doi: 10.1111/1467-9280.00151.

The Mathworks Inc. (2012). Matlab (Version R2012a). Natick, Massachusetts, United States.

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(6), 1412–1431. doi: 10.1037/a0024492.

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. doi: 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. doi: 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. doi: 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. doi: 10.3758/BF03192992.

Van Hell, J. G. (1998). *Cross-language processing and bilingual memory organization.* (Ph.D. dissertation, Universiteit van Amsterdam, Amsterdam, the Netherlands).

Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193–211. doi: 10.1017/S1366728998000352.

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. doi: 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. doi: 10.3758/BF03196335.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458–483. doi: 10.1006/jmla.1998.2584.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. doi: 10.1080/17470218.2013.850521.

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. doi: 10.1093/cercor/bhn030.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*(3), 181–198. doi: 10.3758/bf03197716.

Vanlangendonck, F. (2012). *Conflict in the bilingual brain: The case of cognates and false friends.* (Master's thesis, Radboud Universiteit Nijmegen, Nijmegen, The Netherlands). Retrieved from http://www.ru.nl/master/cns/journal/archive/volume-7-issue-1/print-edition/.

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. doi: 10.1017/S1366728902000111.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review, 14*(5), 779–804. doi: 10.3758/bf03194105.

Wen, Y., & Van Heuven, W. J. B. (2017). Non-cognate translation priming in masked priming lexical decision experiments: A meta-analysis. *Psychonomic Bulletin & Review, 24*(3), 879–886. doi: 10.3758/s13423-016-1151-1.

Yarkoni, T., Balota, D., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*(5), 971–979. doi: 10.3758/PBR.15.5.971.

Yudes, C., Macizo, P., & Bajo, T. (2010). Cognate effects in bilingual language comprehension tasks. *NeuroReport, 21*(7), 507–512. doi: 10.1097/WNR.0b013e328338b9e1.

# APPENDIX A: Database of possible stimuli

This Appendix includes the full database of possible stimuli that resulted from the two rating experiments discussed in Chapter 2.

| | | | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
| accent | accent | 1, 2, 3, 4, 5, 6 | 11.64 | 2.71 | 6 | 2.05 | 12.18 | 2.79 | 6 | 1.85 | 1.00 |
| alarm | alarm | 1, 2, 3, 4, 5, 6 | 34.78 | 3.18 | 5 | 1.95 | 29.84 | 3.18 | 5 | 1.95 | 1.00 |
| amber | amber | 1, 2, 3 | 8.14 | 2.55 | 5 | 1.85 | 9.27 | 2.68 | 5 | 1.85 | 1.00 |
| blind | blind | 1, 2, 3, 4, 5, 6 | 35.61 | 3.19 | 5 | 1.75 | 45.82 | 3.37 | 5 | 1.70 | 1.00 |
| bus | bus | 1, 2, 3, 4, 5, 6 | 64.83 | 3.45 | 3 | 1.00 | 74.18 | 3.58 | 3 | 1.15 | 1.00 |
| campus | campus | 1, 2, 3, 4, 5, 6 | 8.12 | 2.55 | 6 | 2.20 | 10.71 | 2.74 | 6 | 1.95 | 1.00 |
| chaos | chaos | 1, 2, 3, 4, 5, 6 | 15.80 | 2.84 | 5 | 2.00 | 9.39 | 2.68 | 5 | 1.20 | 1.00 |
| circus | circus | 1, 2, 3, 4, 5, 6 | 13.77 | 2.78 | 6 | 2.25 | 17.06 | 2.94 | 6 | 2.25 | 1.00 |
| code | code | 1, 2, 3, 4, 5, 6 | 46.70 | 3.31 | 4 | 1.40 | 53.12 | 3.43 | 4 | 1.05 | 1.00 |
| coma | coma | 1, 2, 3, 4, 5, 6 | 14.09 | 2.79 | 4 | 1.70 | 12.27 | 2.80 | 4 | 1.60 | 1.00 |
| concept | concept | 1, 2, 3, 4, 5, 6 | 5.88 | 2.41 | 7 | 2.25 | 10.84 | 2.74 | 7 | 1.95 | 1.00 |
| contact | contact | 3 | 92.52 | 3.61 | 7 | 2.40 | 64.80 | 3.52 | 7 | 2.05 | 1.00 |
| crisis | crisis | 1, 2, 3, 4, 5, 6 | 11.89 | 2.72 | 6 | 2.25 | 16.65 | 2.93 | 6 | 2.00 | 1.00 |
| detail | detail | 1, 2, 3, 4, 5, 6 | 9.26 | 2.61 | 6 | 2.35 | 19.39 | 3.00 | 6 | 1.85 | 1.00 |
| duel | duel | 1, 2, 3, 4, 5, 6 | 4.85 | 2.33 | 4 | 1.70 | 2.35 | 2.08 | 4 | 1.70 | 1.00 |
| echo | echo | 1, 2, 3, 4, 5, 6 | 8.92 | 2.59 | 4 | 1.95 | 6.86 | 2.54 | 4 | 1.85 | 1.00 |
| ego | ego | 1, 2, 3 | 7.98 | 2.54 | 3 | 1.80 | 7.49 | 2.58 | 3 | 1.85 | 1.00 |
| film | film | 1, 2, 3, 4, 5, 6 | 174.28 | 3.88 | 4 | 1.85 | 65.25 | 3.52 | 4 | 1.75 | 1.00 |
| fort | fort | 1, 2, 3, 4, 5, 6 | 9.38 | 2.61 | 4 | 1.05 | 15.43 | 2.90 | 4 | 1.20 | 1.00 |
| fruit | fruit | 1, 2, 3, 4, 5, 6 | 12.94 | 2.75 | 5 | 1.70 | 21.73 | 3.04 | 5 | 1.90 | 1.00 |
| gas | gas | 1, 2, 3, 4, 5, 6 | 26.41 | 3.06 | 3 | 1.00 | 67.78 | 3.54 | 3 | 1.30 | 1.00 |
| golf | golf | 1, 2, 3, 4, 5, 6 | 17.54 | 2.89 | 4 | 1.70 | 25.53 | 3.12 | 4 | 1.80 | 1.00 |
| hand | hand | 1, 2, 3, 4, 5, 6 | 199.91 | 3.94 | 4 | 1.25 | 279.65 | 4.15 | 4 | 1.35 | 1.00 |
| hotel | hotel | 1, 2, 3, 4, 5, 6 | 88.73 | 3.59 | 5 | 1.70 | 103.22 | 3.72 | 5 | 1.85 | 1.00 |
| instinct | instinct | 1, 2, 3 | 9.81 | 2.63 | 8 | 2.50 | 7.65 | 2.59 | 8 | 2.60 | 1.00 |
| jeep | jeep | 1, 2, 3, 4, 5, 6 | 7.71 | 2.53 | 4 | 1.40 | 10.27 | 2.72 | 4 | 1.65 | 1.00 |
| jury | jury | 1, 2, 3 | 31.17 | 3.13 | 4 | 1.90 | 42.76 | 3.34 | 4 | 1.95 | 1.00 |

| | | | Dutch characteristics | | | | English characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(fre-quency) | word length | OLD20 | frequency | log10(fre-quency) | word length | OLD20 | orthographic similarity |
| lamp | lamp | 1, 2, 3, 4, 5, 6 | 13.88 | 2.78 | 4 | 1.45 | 12.88 | 2.82 | 4 | 1.30 | 1.00 |
| lens | lens | 1, 2, 3 | 4.14 | 2.26 | 4 | 1.00 | 4.67 | 2.38 | 4 | 1.40 | 1.00 |
| lip | lip | 1, 2, 3, 4, 5, 6 | 5.92 | 2.42 | 3 | 1.00 | 10.75 | 2.74 | 3 | 1.00 | 1.00 |
| menu | menu | 1, 2, 3, 4, 5, 6 | 6.63 | 2.46 | 4 | 1.75 | 9.96 | 2.71 | 4 | 1.85 | 1.00 |
| mild | mild | 1, 2, 3, 4, 5, 6 | 2.47 | 2.04 | 4 | 1.70 | 4.80 | 2.39 | 4 | 1.55 | 1.00 |
| model | model | 1, 2, 3, 4, 5, 6 | 20.24 | 2.95 | 5 | 1.60 | 32.06 | 3.21 | 5 | 1.65 | 1.00 |
| moment | moment | 1, 2, 3, 4, 5, 6 | 253.97 | 4.05 | 6 | 2.00 | 187.04 | 3.98 | 6 | 1.95 | 1.00 |
| motto | motto | 1, 2, 3, 4, 5, 6 | 5.17 | 2.36 | 5 | 1.80 | 5.10 | 2.42 | 5 | 1.30 | 1.00 |
| nest | nest | 1, 2, 3, 4, 5, 6 | 11.27 | 2.69 | 4 | 1.00 | 11.10 | 2.75 | 4 | 1.30 | 1.00 |
| oven | oven | 1, 2, 3, 4, 5, 6 | 9.90 | 2.64 | 4 | 1.05 | 8.88 | 2.66 | 4 | 1.60 | 1.00 |
| park | park | 1, 2, 3, 4, 5, 6 | 30.87 | 3.13 | 4 | 1.25 | 72.12 | 3.57 | 4 | 1.20 | 1.00 |
| pen | pen | 1, 2, 3, 4, 5, 6 | 21.66 | 2.98 | 3 | 1.00 | 24.73 | 3.10 | 3 | 1.00 | 1.00 |
| plan | plan | 1, 2, 3, 4, 5, 6 | 143.34 | 3.80 | 4 | 1.25 | 145.73 | 3.87 | 4 | 1.50 | 1.00 |
| plant | plant | 1, 2, 3, 4, 5, 6 | 11.69 | 2.71 | 5 | 1.35 | 27.61 | 3.15 | 5 | 1.55 | 1.00 |
| rat | rat | 1, 2, 3, 4, 5, 6 | 22.73 | 3.00 | 3 | 1.00 | 32.61 | 3.22 | 3 | 1.00 | 1.00 |
| rib | rib | 1, 2, 3, 4, 5, 6 | 2.17 | 1.98 | 3 | 1.55 | 5.90 | 2.48 | 3 | 1.35 | 1.00 |
| ring | ring | 1, 2, 3, 4, 5, 6 | 52.34 | 3.36 | 4 | 1.20 | 92.75 | 3.68 | 4 | 1.30 | 1.00 |
| sofa | sofa | 1, 2, 3, 4, 5, 6 | 3.13 | 2.14 | 4 | 1.80 | 5.86 | 2.48 | 4 | 1.80 | 1.00 |
| storm | storm | 1, 2, 3, 4, 5, 6 | 29.61 | 3.11 | 5 | 1.70 | 30.86 | 3.20 | 5 | 1.75 | 1.00 |
| taxi | taxi | 1, 2, 3, 4, 5, 6 | 50.84 | 3.35 | 4 | 1.75 | 25.84 | 3.12 | 4 | 1.85 | 1.00 |
| tennis | tennis | 1, 2, 3 | 4.64 | 2.31 | 6 | 1.75 | 13.63 | 2.84 | 6 | 2.30 | 1.00 |
| tent | tent | 3 | 40.93 | 3.25 | 4 | 1.00 | 17.49 | 2.95 | 4 | 1.00 | 1.00 |
| test | test | 1, 2, 3, 4, 5, 6 | 43.15 | 3.28 | 4 | 1.00 | 84.08 | 3.63 | 4 | 1.25 | 1.00 |
| ticket | ticket | 1, 2, 3, 4, 5, 6 | 11.73 | 2.71 | 6 | 1.90 | 45.57 | 3.37 | 6 | 1.70 | 1.00 |
| type | type | 1, 2, 3, 4, 5, 6 | 38.08 | 3.22 | 4 | 1.65 | 60.65 | 3.49 | 4 | 1.95 | 1.00 |
| villa | villa | 1, 2, 3, 4, 5, 6 | 7.59 | 2.52 | 5 | 1.95 | 4.39 | 2.35 | 5 | 1.80 | 1.00 |
| water | water | 1, 2, 3, 4, 5, 6 | 244.50 | 4.03 | 5 | 1.50 | 225.06 | 4.06 | 5 | 1.50 | 1.00 |

| | | | Dutch characteristics | | | | English characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(fre-quency) | word length | OLD20 | frequency | log10(fre-quency) | word length | OLD20 | orthographic similarity |
| west | west | 1, 2, 3, 4, 5, 6 | 13.40 | 2.77 | 4 | 1.00 | 60.55 | 3.49 | 4 | 1.30 | 1.00 |
| wild | wild | 1, 2, 3, 4, 5, 6 | 25.41 | 3.05 | 4 | 1.50 | 57.31 | 3.47 | 4 | 1.55 | 1.00 |
| winter | winter | 1, 2, 3, 4, 5, 6 | 22.36 | 2.99 | 6 | 1.65 | 26.22 | 3.13 | 6 | 1.65 | 1.00 |
| wolf | wolf | 1, 2, 3, 4, 5, 6 | 20.26 | 2.95 | 4 | 1.40 | 20.27 | 3.01 | 4 | 1.90 | 1.00 |

**Table A-2:** The full set of 58 identical cognates. For each item, the third column indicates in which (if any) experiment the item was included. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; orthographic similarity refers to the measure of objective orthographic similarity (on a scale from 0 to 1), calculated as the Levenshtein distance between the Dutch and English forms of the words divided by the length of the longest of the two forms.

| | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| Dutch word | English word | Dutch (priming) sentence | sentence length | meaning | spelling | pronunciation |
| accent | accent | Volgens de leraar Duits heeft hij een beetje een Nederlands **accent.** | 11 | 7.00 (0.00) | 7.00 (0.00) | 5.46 (0.78) |
| alarm | alarm | Elke eerste maandag van de maand gaat in Nederland het **alarm** af. | 12 | 6.62 (1.12) | 7.00 (0.00) | 6.69 (0.48) |
| amber | amber | Alles in the Barnsteenkamer van de Tsaar was van **amber** gemaakt. | 11 | 6.31 (1.70) | 7.00 (0.00) | 5.85 (0.55) |
| blind | blind | Het is een welbekend misverstand, maar vleermuizen zijn niet **blind**. | 10 | 6.79 (0.80) | 7.00 (0.00) | 4.93 (1.33) |
| bus | bus | Hij nam die dag de **bus** naar zijn werk. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.62 (0.77) |
| campus | campus | In hun eerste jaar wonen alle geneeskunde studenten op **campus**. | 10 | 7.00 (0.00) | 7.00 (0.00) | 5.77 (0.60) |
| chaos | chaos | Na de revolutie in 1789 verkeerde Frankrijk in **chaos**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.08 (0.90) |
| circus | circus | De acrobaat was zeer te spreken over de directeur van het **circus**. | 12 | 7.00 (0.00) | 7.00 (0.00) | 6.17 (0.72) |
| code | code | De hacker had slechts een uur nodig om de **code** te kraken. | 12 | 6.85 (0.55) | 7.00 (0.00) | 5.77 (0.73) |
| coma | coma | Na bijna zeven jaar ontwaakte de patiënt alsnog uit zijn **coma**. | 11 | 7.00 (0.00) | 7.00 (0.00) | 6.77 (0.44) |
| concept | concept | Voor rijke mensen is “geen geld” een vreemd **concept**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.92 (0.51) |
| contact | contact | Het HIV-virus wordt onder andere verspreid door seksueel **contact**. | 9 | 6.47 (1.36) | 7.00 (0.00) | 6.00 (0.65) |
| crisis* | crisis | De werkloosheid nam toe door de economische **crisis**. | 8 | 6.92 (0.29) | 6.92 (0.29) | 5.58 (0.51) |
| detail | detail | De managementassistente heeft echt oog voor **detail**. | 7 | 6.79 (0.58) | 7.00 (0.00) | 4.57 (1.34) |
| duel | duel | Vroeger losten mannen hun ruzies vaak op met een **duel**. | 10 | 7.00 (0.00) | 7.00 (0.00) | 5.50 (0.67) |
| echo | echo | In een anechoische kamer hoor je geen **echo**. | 8 | 6.50 (1.61) | 7.00 (0.00) | 5.14 (1.51) |
| ego | ego | Dat niemand zijn schilderijen mooi vond, kwetste zijn **ego**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.38 (0.96) |
| film | film | Heb je de nieuwste James Bond **film** al gezien? | 9 | 6.38 (1.39) | 7.00 (0.00) | 6.77 (0.44) |
| fort | fort | Het museum is gevestigd in een oud **fort**. | 8 | 6.92 (0.28) | 7.00 (0.00) | 6.46 (0.66) |
| fruit | fruit | In de zeventiende eeuw schilderde men veel stillevens van **fruit**. | 10 | 7.00 (0.00) | 7.00 (0.00) | 4.77 (0.73) |
| gas* | gas | In de scheikunde spreekt men van drie fasen: vast, vloeibaar en **gas**. | 12 | 7.00 (0.00) | 7.00 (0.00) | 5.25 (0.87) |
| golf | golf | In haar vrije tijd speelt ze graag een spelletje **golf**. | 10 | 7.00 (0.00) | 7.00 (0.00) | 6.08 (0.51) |
| hand | hand | Het is onverstandig met een schaar in de **hand** te rennen. | 11 | 7.00 (0.00) | 7.00 (0.00) | 5.83 (0.39) |
| hotel | hotel | Tijdens hun vakantie verbleven ze in een duur **hotel**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 6.75 (0.45) |
| instinct | instinct | Hij moet leren meer te vertrouwen op zijn **instinct**. | 9 | 6.86 (0.36) | 7.00 (0.00) | 6.21 (0.58) |
| jeep | jeep | Op oneven terrein kun je het beste met een **jeep** rijden. | 11 | 7.00 (0.00) | 7.00 (0.00) | 7.00 (0.00) |
| jury | jury | De kandidate die won, bedankte de **jury**. | 7 | 6.60 (1.06) | 7.00 (0.00) | 5.60 (0.83) |

| Dutch word | English word | Dutch (priming) sentence | sentence length | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| | | | | meaning | spelling | pronunciation |
| lamp | lamp | In de kelder hing slechts één **lamp**. | 7 | 7.00 (0.00) | 7.00 (0.00) | 6.00 (0.41) |
| lens | lens | Het belangrijkste onderdeel van het oog is de **lens**. | 9 | 6.67 (0.49) | 6.92 (0.29) | 6.42 (0.51) |
| lip | lip | Alle leden van de rockband hadden een piercing door hun **lip**. | 11 | 6.73 (0.70) | 7.00 (0.00) | 6.73 (0.46) |
| menu | menu | Zodra ze aan tafel zaten, kregen ze van de ober het **menu**. | 12 | 6.93 (0.26) | 7.00 (0.00) | 5.07 (0.80) |
| mild | mild | Ondanks de grote hoeveelheid chilipoeder was het gerecht **mild**. | 9 | 6.27 (1.33) | 7.00 (0.00) | 5.13 (0.64) |
| model | model | Het meisje wilde haar hele leven lang al **model** worden. | 10 | 6.53 (1.06) | 7.00 (0.00) | 5.27 (1.16) |
| moment | moment | Zijn bruiloft was het gelukkigste **moment** van zijn leven. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.38 (0.77) |
| motto | motto | “Geef nooit op!” zou ieders **motto** moeten zijn. | 8 | 6.80 (0.41) | 7.00 (0.00) | 6.13 (0.74) |
| nest* | nest | Eekhoorns bouwen elk jaar een nieuw **nest**. | 7 | 6.86 (0.36) | 7.00 (0.00) | 6.79 (0.58) |
| oven | oven | Het gerecht moest bijna een uur in de **oven**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.85 (0.69) |
| park | park | Omdat het zulk mooi weer was, gingen ze picknicken in het **park**. | 12 | 6.62 (0.96) | 7.00 (0.00) | 6.54 (0.66) |
| pen | pen | Examens moeten met een blauwe **pen** ingevuld worden. | 8 | 6.93 (0.26) | 7.00 (0.00) | 6.80 (0.56) |
| plan | plan | In menig stripverhaal heeft de slechterik een duivels **plan**. | 9 | 6.20 (1.66) | 7.00 (0.00) | 5.67 (0.62) |
| plant | plant | De enige decoratie in haar kantoor was een halfdode **plant**. | 10 | 6.92 (0.28) | 7.00 (0.00) | 5.62 (0.65) |
| rat | rat | Het stekelvarken is een knaagdier, net als de **rat**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.33 (0.62) |
| rib | rib | Het paard was zo mager dat je elke **rib** kon zien. | 11 | 6.92 (0.28) | 7.00 (0.00) | 6.08 (0.64) |
| ring | ring | Bij een traditionele verloving geeft de man de vrouw een **ring**. | 11 | 6.46 (1.66) | 7.00 (0.00) | 5.85 (0.99) |
| sofa | sofa | Het eerste dat ze kochten voor hun nieuwe huis was een **sofa**. | 12 | 6.71 (0.73) | 7.00 (0.00) | 6.57 (0.51) |
| storm* | storm | De boot kwam vast te zitten in een **storm**. | 9 | 6.92 (0.28) | 7.00 (0.00) | 6.54 (0.66) |
| taxi | taxi | Na het feestje deelden ze een **taxi**. | 7 | 7.00 (0.00) | 7.00 (0.00) | 5.85 (0.55) |
| tennis | tennis | Wimbledon is één van de belangrijkste toernooien voor wie van **tennis** houdt. | 12 | 7.00 (0.00) | 7.00 (0.00) | 6.85 (0.38) |
| tent | tent | Echte kampeerders slapen niet in een camper maar in een **tent**. | 11 | 6.86 (0.36) | 7.00 (0.00) | 6.71 (0.47) |
| test | test | Daans broer hielp hem zich voor te bereiden voor de wiskunde **test**. | 12 | 6.85 (0.38) | 7.00 (0.00) | 6.92 (0.28) |
| ticket* | ticket | Twee dagen voor het concert was haar **ticket** nog steeds niet aangekomen. | 12 | 6.64 (1.08) | 7.00 (0.00) | 6.86 (0.36) |
| type | type | Ze was gewoon zijn **type** niet. | 6 | 6.50 (0.85) | 7.00 (0.00) | 4.21 (1.31) |
| villa | villa | Haar grootouders wonen in een grote **villa**. | 7 | 7.00 (0.00) | 7.00 (0.00) | 6.13 (0.83) |
| water* | water | In een gemiddeld bad gaat zo'n tachtig liter **water.** | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.38 (1.04) |

| Dutch word | English word | Dutch (priming) sentence | sentence length | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| | | | | meaning | spelling | pronunciation |
| west | west | Mijn moeder zegt altijd: "Oost **west**, thuis best!" | 8 | 7.00 (0.00) | 7.00 (0.00) | 6.31 (0.48) |
| wild* | wild | Er leven nog maar weinig neushoorns in het **wild**. | 9 | 6.69 (0.85) | 7.00 (0.00) | 5.15 (1.46) |
| winter | winter | Van kinds af aan was zijn favoriete seizoen al **winter**. | 10 | 7.00 (0.00) | 7.00 (0.00) | 6.23 (0.73) |
| wolf | wolf | De hond is een gedomesticeerde ondersoort van de **wolf**. | 9 | 7.00 (0.00) | 7.00 (0.00) | 5.36 (1.08) |

**Table A-4:** The Dutch prime sentences, sentence lengths and average similarity ratings (and standard deviations) for the full set of 58 identical cognates. The similarity ratings were provided on a scale from 1 (not at all similar) to 7 ((almost) identical). For the 7 items (marked with *) that were included in the second rating experiment as fillers, only the average rating (and standard deviation) from the first experiment is given.

| | | | Dutch characteristics | | | | English characteristics | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
| actie | action | – | 39.81 | 3.24 | 5 | 1.75 | 61.08 | 3.49 | 6 | 1.85 | 0.67 |
| advies | advice | – | 33.23 | 3.16 | 6 | 2.10 | 47.98 | 3.39 | 6 | 2.30 | 0.67 |
| appel | apple | 3, 4 | 10.20 | 2.65 | 5 | 1.70 | 23.67 | 3.08 | 5 | 1.80 | 0.60 |
| auteur | author | – | 4.12 | 2.26 | 6 | 1.90 | 7.94 | 2.61 | 6 | 2.40 | 0.67 |
| baard | beard | 3, 4 | 11.64 | 2.71 | 5 | 1.35 | 12.61 | 2.81 | 5 | 1.65 | 0.80 |
| bakker | baker | 3, 4 | 4.00 | 2.25 | 6 | 1.10 | 13.69 | 2.84 | 5 | 1.40 | 0.83 |
| bal | ball | 3, 4 | 80.63 | 3.55 | 3 | 1.00 | 104.96 | 3.73 | 4 | 1.00 | 0.75 |
| banaan | banana | 3 | 5.33 | 2.37 | 6 | 1.90 | 10.73 | 2.74 | 6 | 2.40 | 0.67 |
| bod | bid | 3, 4 | 10.61 | 2.67 | 3 | 1.00 | 12.59 | 2.81 | 3 | 1.00 | 0.67 |
| boek | book | 3, 4 | 150.93 | 3.82 | 4 | 1.00 | 176.98 | 3.96 | 4 | 1.20 | 0.75 |
| bruid | bride | 3, 4 | 21.13 | 2.97 | 5 | 1.80 | 24.22 | 3.09 | 5 | 1.55 | 0.60 |
| cirkel | circle | 3 | 12.44 | 2.74 | 6 | 1.90 | 21.51 | 3.04 | 6 | 2.25 | 0.50 |
| daad | deed | 3 | 16.65 | 2.86 | 4 | 1.45 | 9.31 | 2.68 | 4 | 1.30 | 0.50 |
| dans | dance | 3, 4 | 37.82 | 3.22 | 4 | 1.15 | 148.04 | 3.88 | 5 | 1.65 | 0.60 |
| debat | debate | 3, 4 | 5.74 | 2.40 | 5 | 1.90 | 9.29 | 2.68 | 6 | 1.80 | 0.83 |
| dief | thief | 3, 4 | 29.96 | 3.12 | 4 | 1.35 | 24.27 | 3.09 | 5 | 1.95 | 0.60 |
| dokter | doctor | – | 244.07 | 4.03 | 6 | 1.75 | 263.94 | 4.13 | 6 | 2.10 | 0.67 |
| domein | domain | – | 3.11 | 2.14 | 6 | 1.85 | 2.59 | 2.12 | 6 | 2.55 | 0.83 |
| eind | end | 3, 4 | 83.17 | 3.56 | 4 | 1.55 | 265.86 | 4.13 | 3 | 1.50 | 0.75 |
| fataal | fatal | 3, 4 | 4.64 | 2.31 | 6 | 1.85 | 7.10 | 2.56 | 5 | 1.95 | 0.83 |
| glas | glass | 3, 4 | 57.15 | 3.40 | 4 | 1.40 | 60.71 | 3.49 | 5 | 1.75 | 0.80 |
| goud | gold | 3, 4 | 61.99 | 3.43 | 4 | 1.50 | 78.94 | 3.61 | 4 | 1.35 | 0.75 |
| grond | ground | 3, 4 | 110.25 | 3.68 | 5 | 1.70 | 72.47 | 3.57 | 6 | 1.80 | 0.83 |
| hart | heart | – | 196.37 | 3.93 | 4 | 1.00 | 244.18 | 4.10 | 5 | 1.65 | 0.80 |
| honger | hunger | – | 92.45 | 3.61 | 6 | 1.80 | 5.88 | 2.48 | 6 | 1.85 | 0.83 |
| inkt | ink | 3, 4 | 5.17 | 2.36 | 4 | 1.60 | 7.49 | 2.58 | 3 | 1.25 | 0.75 |
| kat | cat | 3, 4 | 52.85 | 3.36 | 3 | 1.00 | 66.33 | 3.53 | 3 | 1.00 | 0.67 |

| | | | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
| klimaat | climate | 3, 4 | 4.12 | 2.26 | 7 | 2.45 | 3.53 | 2.26 | 7 | 2.45 | 0.57 |
| klok | clock | 3, 4 | 23.90 | 3.02 | 4 | 1.15 | 58.63 | 3.48 | 5 | 1.45 | 0.60 |
| koffie | coffee | – | 133.30 | 3.77 | 6 | 1.95 | 144.53 | 3.87 | 6 | 2.55 | 0.67 |
| koord | cord | 3, 4 | 2.26 | 2.00 | 5 | 1.60 | 7.02 | 2.56 | 4 | 1.30 | 0.60 |
| kroon | crown | 3, 4 | 14.32 | 2.80 | 5 | 1.45 | 13.69 | 2.84 | 5 | 1.50 | 0.60 |
| leider | leader | – | 45.76 | 3.30 | 6 | 1.60 | 31.16 | 3.20 | 6 | 1.60 | 0.83 |
| licht | light | 3, 4 | 103.64 | 3.66 | 5 | 1.45 | 165.20 | 3.93 | 5 | 1.35 | 0.80 |
| maan | moon | 3, 4 | 42.10 | 3.27 | 4 | 1.00 | 49.96 | 3.41 | 4 | 1.25 | 0.50 |
| masker | mask | 3, 4 | 19.23 | 2.93 | 6 | 1.55 | 19.80 | 3.00 | 4 | 1.45 | 0.67 |
| massa | mass | 3 | 9.17 | 2.60 | 5 | 1.85 | 17.25 | 2.95 | 4 | 1.05 | 0.80 |
| melk | milk | 3, 4 | 39.70 | 3.24 | 4 | 1.55 | 42.53 | 3.34 | 4 | 1.50 | 0.75 |
| metaal | metal | 3, 4 | 9.54 | 2.62 | 6 | 1.85 | 19.45 | 3.00 | 5 | 1.75 | 0.83 |
| mijl | mile | 3, 4 | 15.32 | 2.83 | 4 | 1.35 | 21.00 | 3.03 | 4 | 1.00 | 0.50 |
| motief | motive | 3 | 15.50 | 2.83 | 6 | 1.85 | 13.24 | 2.83 | 6 | 2.05 | 0.67 |
| nek | neck | 3, 4 | 57.72 | 3.40 | 3 | 1.00 | 59.51 | 3.48 | 4 | 1.70 | 0.75 |
| nobel | noble | 3 | 5.90 | 2.41 | 5 | 1.85 | 14.59 | 2.87 | 5 | 1.90 | 0.60 |
| paniek | panic | – | 39.86 | 3.24 | 6 | 2.00 | 21.84 | 3.05 | 5 | 1.90 | 0.67 |
| parel | pearl | 3, 4 | 3.02 | 2.12 | 5 | 1.60 | 15.67 | 2.90 | 5 | 1.70 | 0.60 |
| peper | pepper | 3, 4 | 3.80 | 2.22 | 5 | 1.45 | 8.80 | 2.65 | 6 | 1.85 | 0.83 |
| pijp | pipe | 3, 4 | 13.81 | 2.78 | 4 | 1.45 | 19.39 | 3.00 | 4 | 1.45 | 0.50 |
| piraat | pirate | 3, 4 | 6.40 | 2.45 | 6 | 1.90 | 7.35 | 2.58 | 6 | 1.90 | 0.67 |
| prijs | price | 3, 4 | 86.60 | 3.58 | 5 | 1.55 | 53.37 | 3.44 | 5 | 1.55 | 0.60 |
| prins | prince | 3 | 45.26 | 3.30 | 5 | 1.75 | 45.08 | 3.36 | 6 | 1.85 | 0.67 |
| publiek | public | – | 41.87 | 3.26 | 7 | 2.45 | 71.08 | 3.56 | 6 | 2.55 | 0.71 |
| rivier | river | – | 52.87 | 3.36 | 6 | 1.90 | 55.47 | 3.45 | 5 | 1.55 | 0.83 |
| roos | rose | 3, 4 | 11.71 | 2.71 | 4 | 1.00 | 53.02 | 3.43 | 4 | 1.00 | 0.50 |
| saus | sauce | 3, 4 | 9.56 | 2.62 | 4 | 1.65 | 15.59 | 2.90 | 5 | 1.75 | 0.60 |

| | | | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
| schip | ship | 3, 4 | 115.28 | 3.70 | 5 | 1.50 | 98.88 | 3.70 | 4 | 1.45 | 0.80 |
| schoen | shoe | 3, 4 | 13.45 | 2.77 | 6 | 1.55 | 30.39 | 3.19 | 4 | 1.40 | 0.67 |
| schouder | shoulder | – | 18.57 | 2.91 | 8 | 1.80 | 26.20 | 3.13 | 8 | 2.45 | 0.75 |
| slaaf | slave | 3, 4 | 17.75 | 2.89 | 5 | 1.35 | 18.43 | 2.97 | 5 | 1.50 | 0.60 |
| sneeuw | snow | 3, 4 | 24.31 | 3.03 | 6 | 1.90 | 31.35 | 3.20 | 4 | 1.50 | 0.50 |
| soep | soup | 3, 4 | 17.84 | 2.89 | 4 | 1.20 | 25.20 | 3.11 | 4 | 1.70 | 0.75 |
| sok | sock | 3, 4 | 3.11 | 2.14 | 3 | 1.00 | 8.98 | 2.66 | 4 | 1.15 | 0.75 |
| straat | street | 3 | 100.34 | 3.64 | 6 | 1.80 | 148.18 | 3.88 | 6 | 1.95 | 0.67 |
| thee | tea | 3, 4 | 58.79 | 3.41 | 4 | 1.65 | 58.63 | 3.48 | 3 | 1.40 | 0.50 |
| tijger | tiger | – | 11.69 | 2.71 | 6 | 1.75 | 18.53 | 2.98 | 5 | 1.70 | 0.83 |
| tong | tongue | – | 31.90 | 3.14 | 4 | 1.35 | 31.16 | 3.20 | 6 | 2.20 | 0.67 |
| totaal | total | – | 48.66 | 3.33 | 6 | 1.90 | 37.65 | 3.28 | 5 | 1.90 | 0.83 |
| trein | train | 3, 4 | 73.15 | 3.51 | 5 | 1.65 | 95.06 | 3.69 | 5 | 1.55 | 0.80 |
| troon | throne | – | 12.10 | 2.72 | 5 | 1.65 | 8.65 | 2.65 | 6 | 1.90 | 0.50 |
| vaas | vase | 3, 4 | 4.57 | 2.30 | 4 | 1.00 | 3.84 | 2.29 | 4 | 1.60 | 0.50 |
| vallei | valley | 3, 4 | 10.98 | 2.68 | 6 | 1.95 | 25.00 | 3.11 | 6 | 1.80 | 0.83 |
| vinger | finger | 3, 4 | 28.91 | 3.10 | 6 | 1.70 | 36.67 | 3.27 | 6 | 1.60 | 0.83 |
| wiel | wheel | 3, 4 | 7.04 | 2.49 | 4 | 1.00 | 27.06 | 3.14 | 5 | 1.90 | 0.60 |
| zeil | sail | 3, 4 | 6.91 | 2.48 | 4 | 1.25 | 13.75 | 2.85 | 4 | 1.15 | 0.50 |
| zilver | silver | 3, 4 | 9.95 | 2.64 | 6 | 2.00 | 31.75 | 3.21 | 6 | 1.90 | 0.83 |
| zomer | summer | – | 42.90 | 3.27 | 5 | 1.55 | 78.67 | 3.60 | 6 | 1.80 | 0.50 |
| zweet | sweat | 3, 4 | 12.87 | 2.75 | 5 | 1.05 | 21.86 | 3.05 | 5 | 1.70 | 0.60 |

**Table A-6:** The full set of 76 non-identical cognates. For each item, the third column indicates in which (if any) experiment the item was included. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; orthographic similarity refers to the measure of objective orthographic similarity (on a scale from 0 to 1), calculated as the Levenshtein distance between the Dutch and English forms of the words divided by the length of the longest of the two forms.

| Dutch word | English word | Dutch (priming) sentence | sentence length | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| | | | | meaning | spelling | pronunciation |
| actie | action | De NAVO verzocht zijn leden dringend om **actie** te nemen. | 10 | 6.92 (0.29) | 5.33 (0.65) | 4.67 (0.78) |
| advies | advice | De verdachte negeerde het **advies** van zijn advocaat. | 8 | 6.80 (0.56) | 4.87 (1.06) | 4.80 (0.94) |
| appel | apple | Van God mochten Adam en Eva alles eten, behalve de **appel**. | 11 | 6.92 (0.28) | 5.92 (0.28) | 5.54 (0.66) |
| auteur | author | Voor de signeersessie zal de **auteur** een lezing geven. | 9 | 6.92 (0.28) | 4.92 (0.49) | 4.23 (1.09) |
| baard | beard | De Kerstman heeft een lange, witte **baard**. | 7 | 7.00 (0.00) | 5.54 (0.66) | 4.62 (1.12) |
| bakker | baker | Het is een vermoeiend beroep, dat van **bakker**. | 8 | 6.57 (0.94) | 5.64 (0.50) | 4.64 (1.22) |
| bal* | ball | Het jongetje gooide de **bal** helemaal ernaast. | 7 | 6.83 (0.39) | 6.00 (0.00) | 6.17 (0.39) |
| banaan | banana | Je kunt heerlijke koekjes maken van slechts wat havermout en een **banaan**. | 12 | 6.93 (0.26) | 5.53 (0.64) | 4.60 (0.83) |
| bod | bid | Bij een veiling wint de persoon met het hoogste **bod**. | 10 | 6.85 (0.55) | 4.85 (1.07) | 4.15 (1.52) |
| boek | book | Ernest Hemingway zei, “Er is geen vriend zo loyaal als een **boek**.” | 12 | 7.00 (0.00) | 5.80 (0.41) | 6.13 (0.83) |
| bruid | bride | Op haar bruiloft was ze één en al blozende **bruid**. | 10 | 7.00 (0.00) | 4.62 (0.65) | 4.38 1.33) |
| cirkel | circle | Met een passer kun je heel precies een **cirkel** tekenen. | 10 | 6.86 (0.36) | 5.21 (0.70) | 5.50 (1.02) |
| daad | deed | De multimiljonair gaf de helft van zijn geld weg als goede **daad**. | 12 | 6.33 (1.54) | 4.87 (0.74) | 4.07 (0.96) |
| dans | dance | De choreograaf had maanden gewerkt aan die **dans**. | 8 | 6.77 (0.83) | 5.38 (0.51) | 5.69 (0.75) |
| debat | debate | De Tweede Kamer houdt vanavond een belangrijk **debat**. | 8 | 6.85 (0.38) | 6.00 (0.00) | 5.23 (0.83) |
| dief | thief | Robin Hood is waarschijnlijk de bekendste **dief** allertijden. | 8 | 6.93 (0.27) | 5.21 (0.58) | 5.29 (0.83) |
| dokter | doctor | Op maandag had ze een afspraak met haar **dokter**. | 9 | 6.83 (0.39) | 5.58 (0.67) | 6.00 (0.43) |
| domein | domain | Hier weet Jan veel van; dit is echt zijn **domein**. | 10 | 6.58 (0.79) | 5.75 (0.45) | 5.75 (0.45) |
| eind | end | Ze was doodop aan het **eind** van de wandeling. | 9 | 6.92 (0.29) | 5.67 (0.49) | 5.25 (0.45) |
| fataal | fatal | De tweede hersenbloeding werd haar helaas **fataal**. | 7 | 6.92 (0.28) | 5.92 (0.28) | 4.92 (0.76) |
| glas* | glass | De schoonmaker schrok van de hond en liet het **glas** vallen. | 11 | 6.67 (0.82) | 5.87 (0.00) | 5.13 (0.92) |
| goud | gold | Ze kreeg van haar man op hun trouwdag een ketting van **goud**. | 12 | 7.00 (0.00) | 4.80 (0.36) | 3.93 (1.28) |
| grond | ground | Hij sprong van het dak af en viel op de **grond**. | 11 | 6.93 (0.27) | 5.71 (0.38) | 4.50 (1.16) |
| hart | heart | Met een stethoscoop kun je je eigen **hart** horen kloppen. | 10 | 7.00 (0.00) | 5.92 (0.69) | 6.15 (0.80) |
| honger | hunger | Het gerecht was zo klein dat het mijn **honger** niet stilde. | 11 | 6.31 (1.44) | 6.00 (1.21) | 5.00 (0.82) |
| inkt | ink | Mijn vulpen is leeg; ik moet dringend nieuwe **inkt** kopen. | 10 | 6.86 (0.36) | 5.86 (1.03 | 5.50 (0.94) |
| kat* | cat | Haar ouders hebben een dikke, grijze **kat**. | 7 | 7.00 (0.00) | 5.85 (0.90) | 5.46 (0.88) |

| Dutch word | English word | Dutch (priming) sentence | sentence length | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| | | | | meaning | spelling | pronunciation |
| klimaat | climate | In het gebied tussen de keerkringen heerst een tropisch **klimaat**. | 10 | 7.00 (0.00) | 5.15 (0.69) | 4.38 (1.04) |
| klok | clock | Vorig jaar erfde ze van haar grootvader een antieke **klok**. | 10 | 6.86 (0.53) | 4.93 (1.21) | 5.71 (1.20) |
| koffie | coffee | De assistente van de minister drinkt haar **koffie** zwart. | 9 | 7.00 (0.00) | 4.83 (1.03) | 5.83 (0.39) |
| koord | cord | Om de parachute te openen moet je aan het **koord** trekken. | 11 | 6.58 (0.67) | 5.08 (0.90) | 6.08 (0.51) |
| kroon | crown | Een tiara is een soort **kroon**. | 6 | 7.00 (0.00) | 4.15 (1.52) | 4.31 (1.25) |
| leider | leader | Tijdens de groepsopdracht bleek wel dat ze een goede **leider** was. | 11 | 7.00 (0.00) | 5.33 (0.72) | 4.80 (0.56) |
| licht | light | Een brandende kaars geeft maar weinig **licht**. | 7 | 6.85 (0.38) | 5.54 (0.52) | 3.62 (1.19) |
| maan | moon | In 1969 zette NASA de eerste man op de **maan**. | 10 | 7.00 (0.00) | 5.42 (0.67) | 5.00 (0.85) |
| masker | mask | Tijdens het Venetiaans carnaval heeft iedereen een **masker** op. | 9 | 7.00 (0.00) | 4.87 (0.92) | 4.33 (1.29) |
| massa | mass | De SI-eenheid van **massa** is de kilogram. | 7 | 6.69 (0.63) | 5.77 (0.60) | 4.15 (1.07) |
| melk | milk | Bij de lunch drinkt hij altijd een halve liter **melk**. | 10 | 7.00 (0.00) | 5.71 (0.61) | 5.14 (1.23) |
| metaal | metal | Het meubel is leverbaar in plastic en **metaal**. | 8 | 6.85 (0.55) | 5.92 (0.28) | 5.23 (0.60) |
| mijl* | mile | In Groot-Brittanië is de maat voor afstanden niet de kilometer de **mijl**. | 12 | 7.00 (0.00) | 5.33 (0.89) | 5.67 (0.65) |
| motief | motive | Men weet nog steeds niet wat het **motief** van de moordenaar was. | 12 | 6.64 (0.63) | 4.93 (0.83) | 4.86 (1.03) |
| nek | neck | Van alle zoogdieren heeft de giraffe de langste **nek**. | 9 | 6.93 (0.26) | 5.87 (0.35) | 6.80 (0.56) |
| nobel | noble | Hij wilde meer vrijwilligerswerk doen, een zeer **nobel** streven. | 9 | 6.67 (0.62) | 5.93 (0.26) | 6.67 (0.62) |
| paniek | panic | Bij het zien van de bom raakten alle omstanders in **paniek**. | 11 | 6.64 (0.63) | 5.50 (0.52) | 4.93 (1.14) |
| parel | pearl | Geschat wordt dat in één op de vijftienduizend oesters een **parel** zit. | 12 | 6.93 (0.27) | 4.57 (1.34) | 3.64 (1.60) |
| peper | pepper | De chef-kok hield helemaal niet van **peper**. | 7 | 6.87 (0.52) | 5.87 (0.35) | 5.00 (0.76) |
| pijp* | pipe | De loodgieter moest onder de wastafel een **pijp** vervangen. | 9 | 6.38 (1.66) | 5.23 (0.44) | 5.46 (0.78) |
| piraat | pirate | Het meisje wilde zich het liefste verkleden als **piraat**. | 9 | 7.00 (0.00) | 5.54 (0.52) | 4.69 (0.85) |
| prijs* | price | Voor groepen van tien of meer is de **prijs** per persoon lager. | 12 | 7.00 (0.00) | 5.00 (0.85) | 5.33 (1.07) |
| prins | prince | In de meeste sprookjes wordt de prinses door de **prins** gered. | 11 | 7.00 (0.00) | 5.75 (0.45) | 6.33 (0.49) |
| publiek | public | Het kasteel was de hele dag geopend voor het **publiek**. | 10 | 6.14 (1.23) | 5.21(0.58) | 4.86 (1.10) |
| rivier | river | Met bijna zevenduizend kilometer is de Nijl 's werelds langste **rivier**. | 11 | 7.00 (0.00) | 5.79 (0.43) | 4.93 (0.83) |
| roos | rose | Tijdens de scheikundeles doopte de lerares een **roos** in vloeibaar stikstof. | 11 | 6.79 (0.58) | 5.36 (0.63) | 5.50 (1.09) |
| saus | sauce | Bij de vis werd een heerlijke **saus** geserveerd. | 8 | 7.00 (0.00) | 5.42 (0.67) | 5.25 (0.75) |

| Dutch word | English word | Dutch (priming) sentence | sentence length | SIMILARITY RATINGS meaning | spelling | pronunciation |
|---|---|---|---|---|---|---|
| schip | ship | Omstreeks het jaar 1000 reisde Leif Eriksson met een **schip** naar Noord- | 12 | 6.87 (0.35) | 5.73 (0.46) | 5.07 (1.10) |
| schoen | shoe | Als Sinterklaas aangekomen is, mogen alle kinderen hun **schoen** zetten. | 10 | 7.00 (0.00) | 4.53 (0.99) | 3.67 (1.23) |
| schouder | shoulder | Toen de skileraar viel, schoot z'n **schouder** uit de kom. | 10 | 7.00 (0.00) | 4.92 (0.28) | 4.23 (0.83) |
| slaaf | slave | Nu nog steeds worden ontelbare mensen als **slaaf** gebruikt. | 9 | 7.00 (0.00) | 5.00 (0.91) | 4.69 (1.11) |
| sneeuw* | snow | De weerman voorspelde dat er vijftien centimeter **sneeuw** zou vallen. | 10 | 7.00 (0.00) | 4.13 (0.99) | 4.07 (1.22) |
| soep | soup | Mijn grootmoeder maakt van aardappelen de lekkerste **soep** ooit. | 9 | 7.00 (0.00) | 5.85 (0.38) | 6.46 (0.52) |
| sok | sock | De wasmachine had een **sok** opgeslokt. | 6 | 6.92 (0.28) | 5.77 (0.44) | 6.08 (0.86) |
| straat | street | In de meeste musicals is het volstrekt normaal te zingen op **straat**. | 12 | 7.00 (.00) | 5.31 (0.63) | 4.54 (1.33) |
| thee | tea | Bij binnenkomst vroeg ze aan iedereen of ze een kopje **thee** wilden. | 12 | 7.00 (.00) | 4.00 (1.07) | 4.20 (0.94) |
| tijger | tiger | Er zijn veel prachtige documentaires over de **tijger**. | 8 | 7.00 (0.00) | 5.92 (0.29) | 5.08 (0.67) |
| tong | tongue | Hij beet per ongeluk op z'n **tong** en schreeuwde. | 9 | 7.00 (0.00) | 5.38 (0.51) | 5.77 (0.60) |
| totaal | total | De rekening kwam uit op tweehonderd euro in **totaal**. | 9 | 6.93 (0.26) | 5.93 (0.26) | 5.13 (0.74) |
| trein | train | Net toen hij een belangrijke afspraak had, was zijn **trein** vertraagd. | 11 | 7.00 (0.00) | 5.85 (0.38) | 5.15 (0.90) |
| troon | throne | Het Nederlandse Koninkrijk heeft geen officiële **troon**. | 7 | 6.62 (0.87) | 4.85 (0.80) | 5.15 (1.21) |
| vaas | vase | Op de bijzettafel stond een grote **vaas** bloemen. | 8 | 7.00 (0.00) | 5.62 (0.51) | 5.15 (1.0) |
| vallei | valley | Het dorpje lag in een afgelegen **vallei**. | 7 | 6.92 (0.28) | 5.62 (0.65) | 4.54 (1.05) |
| vinger | finger | Het Latijnse woord dactylus betekent **vinger**. | 6 | 6.85 (0.55) | 6.08 (0.28) | 5.08 (0.95) |
| wiel | wheel | Rond 3500 voor Christus is het **wiel** uitgevonden. | 8 | 6.83 (0.39) | 4.42 (1.08) | 5.75 (0.62) |
| zeil | sail | Soms moet de bemanning reven: het oppervlakte van het **zeil** verkleinen. | 11 | 6.00 (1.57) | 4.43 (1.09) | 4.29 (1.49) |
| zilver | silver | Haar favoriete paar oorbellen was van brons en **zilver**. | 9 | 6.92 (0.28) | 5.92 (0.49) | 5.62 (0.96) |
| zomer | summer | We gaan altijd op vakantie in de **zomer**. | 8 | 7.00 (0.00) | 4.42 (0.90) | 4.50 (1.17) |
| zweet | sweat | Na de marathon was het shirt van de atleet doorweekt met **zweet**. | 12 | 7.00 (0.00) | 4.50 (1.00) | 4.75 (1.29) |

**Table A-8:** The Dutch prime sentences, sentence lengths and average similarity ratings (and standard deviations) for the full set of 76 non-identical cognates. The similarity ratings were provided on a scale from 1 (not at all similar) to 7 ((almost) identical). For the 7 items (marked with *) that were included in the second rating experiment as fillers, only the average rating (and standard deviation) from the first experiment is given.

| Dutch word | English word | included in Experiment | Dutch characteristics: frequency | log10(frequency) | word length | OLD20 | English characteristics: frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
|---|---|---|---|---|---|---|---|---|---|---|---|
| angel | angel | 1, 2, 4, 5, 6 | 10.63 | 2.67 | 5 | 1.75 | 78.27 | 3.60 | 5 | 1.85 | 1.00 |
| bad | bad | – | 42.58 | 3.27 | 3 | 1.00 | 545.18 | 4.44 | 3 | 1.00 | 1.00 |
| bang | bang | 1, 2, 4, 5, 6 | 477.21 | 4.32 | 4 | 1.00 | 19.98 | 3.01 | 4 | 1.20 | 1.00 |
| beer | beer | 1, 2, 4, 5, 6 | 25.45 | 3.05 | 4 | 1.00 | 75.49 | 3.59 | 4 | 1.30 | 1.00 |
| boom | boom | 1, 2, 4, 5, 6 | 52.25 | 3.36 | 4 | 1.10 | 21.80 | 3.05 | 4 | 1.30 | 1.00 |
| boot | boot | 1, 2, 4, 5, 6 | 95.93 | 3.62 | 4 | 1.00 | 11.14 | 2.76 | 4 | 1.00 | 1.00 |
| brand | brand | 1, 2, 4, 5, 6 | 44.39 | 3.29 | 5 | 1.75 | 13.96 | 2.85 | 5 | 1.55 | 1.00 |
| breed | breed | 1, 2, 4, 5, 6 | 6.22 | 2.44 | 5 | 1.65 | 6.33 | 2.51 | 5 | 1.55 | 1.00 |
| brief | brief | 1, 2, 4, 5, 6 | 73.84 | 3.51 | 5 | 1.75 | 14.35 | 2.87 | 5 | 1.85 | 1.00 |
| den | den | – | 4.76 | 2.32 | 3 | 1.00 | 6.12 | 2.50 | 3 | 1.05 | 1.00 |
| drop | drop | 1, 2, 4, 5, 6 | 1.83 | 1.91 | 4 | 1.35 | 130.61 | 3.82 | 4 | 1.75 | 1.00 |
| fee | fee | 1, 2 | 5.67 | 2.40 | 3 | 1.15 | 9.69 | 2.70 | 3 | 1.20 | 1.00 |
| file | file | 1, 2, 4, 5, 6 | 4.94 | 2.34 | 4 | 1.75 | 44.04 | 3.35 | 4 | 1.15 | 1.00 |
| gang | gang | 1, 2, 4, 5, 6 | 110.80 | 3.69 | 4 | 1.20 | 30.14 | 3.19 | 4 | 1.50 | 1.00 |
| glad | glad | 1, 2, 4, 5, 6 | 8.00 | 2.55 | 4 | 1.80 | 171.37 | 3.94 | 4 | 1.70 | 1.00 |
| gulp | gulp | 1, 2, 4, 5, 6 | 0.85 | 1.58 | 5 | 1.55 | 0.98 | 1.71 | 5 | 1.80 | 1.00 |
| honk | honk | 1, 2, 4, 5, 6 | 4.39 | 2.29 | 4 | 1.75 | 2.39 | 2.09 | 4 | 1.65 | 1.00 |
| hoop | hoop | 1, 2 | 367.83 | 4.21 | 4 | 1.05 | 2.69 | 2.14 | 4 | 1.50 | 1.00 |
| kin | kin | – | 7.80 | 2.53 | 3 | 1.00 | 4.27 | 2.34 | 3 | 1.05 | 1.00 |
| kind | kind | 1, 2, 4, 5, 6 | 333.30 | 4.16 | 4 | 1.50 | 590.69 | 4.48 | 4 | 1.45 | 1.00 |
| lap | lap | – | 3.43 | 2.18 | 3 | 1.00 | 13.47 | 2.84 | 3 | 1.00 | 1.00 |
| last | last | – | 53.49 | 3.37 | 4 | 1.00 | 723.10 | 4.57 | 4 | 1.15 | 1.00 |
| leek | leek | – | 95.47 | 3.62 | 4 | 1.00 | 0.29 | 1.20 | 4 | 1.60 | 1.00 |
| lever | lever | 1, 2 | 16.35 | 2.85 | 5 | 1.00 | 3.20 | 2.22 | 5 | 1.35 | 1.00 |
| lid | lid | – | 33.18 | 3.16 | 3 | 1.20 | 4.92 | 2.40 | 3 | 1.15 | 1.00 |
| list | list | 1, 2, 4, 5, 6 | 3.77 | 2.22 | 4 | 1.00 | 80.59 | 3.61 | 4 | 1.30 | 1.00 |
| log | log | – | 1.03 | 1.66 | 3 | 1.00 | 11.96 | 2.79 | 3 | 1.05 | 1.00 |

| Dutch word | English word | included in Experiment | DUTCH CHARACTERISTICS frequency | log10(frequency) | word length | OLD20 | ENGLISH CHARACTERISTICS frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
|---|---|---|---|---|---|---|---|---|---|---|---|
| map | map | 1, 2, 4, 5, 6 | 4.53 | 2.30 | 3 | 1.00 | 31.82 | 3.21 | 3 | 1.00 | 1.00 |
| mate | mate | 1, 2 | 3.84 | 2.23 | 4 | 1.20 | 29.24 | 3.17 | 4 | 1.00 | 1.00 |
| mop | mop | – | 7.04 | 2.49 | 3 | 1.00 | 4.14 | 2.33 | 3 | 1.30 | 1.00 |
| mug | mug | – | 2.26 | 2.00 | 3 | 1.55 | 6.84 | 2.54 | 3 | 1.30 | 1.00 |
| nut | nut | 1, 2, 4, 5, 6 | 13.86 | 2.78 | 3 | 1.00 | 15.63 | 2.90 | 3 | 1.40 | 1.00 |
| pad | pad | – | 41.99 | 3.26 | 3 | 1.25 | 8.14 | 2.62 | 3 | 1.00 | 1.00 |
| pal | pal | 1, 2 | 4.16 | 2.26 | 3 | 1.00 | 57.59 | 3.47 | 3 | 1.00 | 1.00 |
| peer | peer | – | 1.76 | 1.89 | 4 | 1.00 | 1.53 | 1.90 | 4 | 1.20 | 1.00 |
| perk | perk | – | 1.99 | 1.94 | 4 | 1.40 | 1.41 | 1.86 | 4 | 1.45 | 1.00 |
| pet | pet | 1, 2, 4, 5, 6 | 13.19 | 2.76 | 3 | 1.00 | 20.18 | 3.01 | 3 | 1.00 | 1.00 |
| pink | pink | 1, 2, 4, 5, 6 | 2.88 | 2.10 | 4 | 1.15 | 28.47 | 3.16 | 4 | 1.15 | 1.00 |
| pool | pool | 1, 2, 4, 5, 6 | 3.54 | 2.19 | 4 | 1.15 | 46.98 | 3.38 | 4 | 1.45 | 1.00 |
| prop | prop | 1, 2, 4, 5, 6 | 2.38 | 2.02 | 4 | 1.45 | 3.69 | 2.28 | 4 | 1.35 | 1.00 |
| put | put | – | 11.69 | 2.71 | 3 | 1.00 | 828.45 | 4.63 | 3 | 1.05 | 1.00 |
| ramp | ramp | 1, 2, 4, 5, 6 | 25.89 | 3.05 | 4 | 1.35 | 2.88 | 2.17 | 4 | 1.30 | 1.00 |
| roof | roof | 1, 2, 4, 5, 6 | 2.26 | 2.00 | 4 | 1.30 | 35.65 | 3.26 | 4 | 1.55 | 1.00 |
| room | room | 1, 2, 4, 5, 6 | 7.59 | 2.52 | 4 | 1.00 | 439.51 | 4.35 | 4 | 1.40 | 1.00 |
| rooster | rooster | 1, 2, 4, 5, 6 | 8.35 | 2.56 | 7 | 1.65 | 3.86 | 2.30 | 7 | 1.85 | 1.00 |
| rug | rug | – | 80.79 | 3.55 | 3 | 1.25 | 10.41 | 2.73 | 3 | 1.15 | 1.00 |
| rust | rust | 1, 2, 4, 5, 6 | 75.40 | 3.52 | 4 | 1.00 | 2.49 | 2.11 | 4 | 1.15 | 1.00 |
| sip | sip | – | 1.88 | 1.92 | 3 | 1.00 | 5.10 | 2.42 | 3 | 1.00 | 1.00 |
| slang | slang | 1, 2, 4, 5, 6 | 21.59 | 2.98 | 5 | 1.70 | 1.39 | 1.86 | 5 | 1.65 | 1.00 |
| slap | slap | 1, 2, 4, 5, 6 | 7.11 | 2.49 | 4 | 1.15 | 12.47 | 2.80 | 4 | 1.15 | 1.00 |
| slim | slim | 1, 2, 4, 5, 6 | 111.55 | 3.69 | 4 | 1.45 | 11.86 | 2.78 | 4 | 1.50 | 1.00 |
| slip | slip | 1, 2, 4, 5, 6 | 1.26 | 1.75 | 4 | 1.00 | 25.88 | 3.12 | 4 | 1.30 | 1.00 |
| slot | slot | 1, 2, 4, 5, 6 | 52.46 | 3.36 | 4 | 1.25 | 5.49 | 2.45 | 4 | 1.10 | 1.00 |
| smart | smart | 1, 2, 4, 5, 6 | 1.99 | 1.94 | 5 | 1.65 | 96.25 | 3.69 | 5 | 1.90 | 1.00 |

| Dutch word | English word | included in Experiment | Dutch characteristics: frequency | Dutch characteristics: log10(frequency) | Dutch characteristics: word length | Dutch characteristics: OLD20 | English characteristics: frequency | English characteristics: log10(frequency) | English characteristics: word length | English characteristics: OLD20 | orthographic similarity |
|---|---|---|---|---|---|---|---|---|---|---|---|
| spin | spin | 1, 2, 4, 5, 6 | 7.80 | 2.53 | 4 | 1.45 | 14.63 | 2.87 | 4 | 1.50 | 1.00 |
| spit | spit | 1, 2, 4, 5, 6 | 1.33 | 1.77 | 4 | 1.10 | 19.35 | 3.00 | 4 | 1.50 | 1.00 |
| spot | spot | 1, 2, 4, 5, 6 | 7.57 | 2.52 | 4 | 1.20 | 61.57 | 3.50 | 4 | 1.45 | 1.00 |
| stadium | stadium | 1, 2, 4, 5, 6 | 3.70 | 2.21 | 7 | 2.70 | 6.12 | 2.50 | 7 | 2.80 | 1.00 |
| stage | stage | 1, 2, 4, 5, 6 | 2.93 | 2.11 | 5 | 1.60 | 45.57 | 3.37 | 5 | 1.45 | 1.00 |
| stand | stand | 1, 2, 4, 5, 6 | 15.71 | 2.84 | 5 | 1.55 | 226.20 | 4.06 | 5 | 1.70 | 1.00 |
| star | star | 1, 2, 4, 5, 6 | 8.03 | 2.55 | 4 | 1.25 | 81.35 | 3.62 | 4 | 1.20 | 1.00 |
| steel | steel | 1, 2, 4, 5, 6 | 8.62 | 2.58 | 5 | 1.10 | 18.45 | 2.97 | 5 | 1.70 | 1.00 |
| stem | stem | 1, 2, 4, 5, 6 | 86.53 | 3.58 | 4 | 1.30 | 2.24 | 2.06 | 4 | 1.70 | 1.00 |
| strand | strand | 1, 2, 4, 5, 6 | 40.16 | 3.24 | 6 | 1.85 | 1.84 | 1.98 | 6 | 1.90 | 1.00 |
| teen | teen | 1, 2, 4, 5, 6 | 7.39 | 2.51 | 4 | 1.00 | 4.10 | 2.32 | 4 | 1.55 | 1.00 |
| toe | toe | 1, 2 | 580.41 | 4.40 | 3 | 1.00 | 12.69 | 2.81 | 3 | 1.00 | 1.00 |
| trap | trap | 1, 2, 4, 5, 6 | 52.28 | 3.36 | 4 | 1.45 | 23.84 | 3.09 | 4 | 1.50 | 1.00 |
| vast | vast | 1, 2, 4, 5, 6 | 0.57 | 1.42 | 7 | 1.00 | 6.10 | 2.49 | 7 | 1.45 | 1.00 |
| vet | vet | 1, 2, 4, 5, 6 | 18.52 | 2.91 | 3 | 1.00 | 5.80 | 2.47 | 3 | 1.20 | 1.00 |
| wand | wand | 1, 2, 4, 5, 6 | 3.29 | 2.16 | 4 | 1.15 | 3.08 | 2.20 | 4 | 1.40 | 1.00 |
| war | war | 1, 2, 4, 5, 6 | 3.43 | 2.18 | 3 | 1.00 | 174.75 | 3.95 | 3 | 1.00 | 1.00 |
| wet | wet | 1, 2, 4, 5, 6 | 80.45 | 3.55 | 3 | 1.00 | 39.22 | 3.30 | 3 | 1.00 | 1.00 |

**Table A-10:** The full set of 72 identical interlingual homographs. For each item, the third column indicates in which (if any) experiment the item was included. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; orthographic similarity refers to the measure of objective orthographic similarity (on a scale from 0 to 1), calculated as the Levenshtein distance between the Dutch and English forms of the words divided by the length of the longest of the two forms.

| | | | | **SIMILARITY RATINGS** | | |
|---|---|---|---|---|---|---|
| Dutch word | English word | Dutch (priming) sentence | sentence length | meaning | spelling | pronunciation |
| angel | angel | Alleen vrouwelijke bijen en wespen hebben een **angel**. | 8 | 1.03 (0.11) | 6.97 (0.11) | 3.84 (1.17) |
| bad | bad | Voor het slapen gaan moest Maria nog even in **bad.** | 10 | 1.55 (1.81) | 7.00 (0.00) | 5.18 (0.87) |
| bang | bang | Heel haar leven lang is Heleen al **bang** voor honden. | 10 | 1.00 (0.00) | 7.00 (0.00) | 5.50 (0.52) |
| beer | beer | Zijn favoriete knuffel was een schattige zwarte **beer**. | 8 | 1.00 (0.00) | 7.00 (0.00) | 5.83 (0.72) |
| boom | boom | Het jongetje had zich achter een **boom** verstopt. | 8 | 1.03 (0.08) | 7.00 (0.00) | 4.81 (1.17) |
| boot | boot | Het eiland is alleen te bereiken per **boot**. | 8 | 1.00 (0.00) | 7.00 (0.00) | 4.82 (1.08) |
| brand | brand | Er woedde een hevige **brand** in het bos. | 8 | 1.00 (0.00) | 7.00 (0.00) | 5.09 (0.98) |
| breed | breed | De tafel paste niet door de deur; hij was te **breed**. | 11 | 1.00 (0.00) | 7.00 (0.00) | 4.75 (0.87) |
| brief | brief | Haar oma stuurt haar nog elke week een **brief.** | 9 | 1.09 (0.33) | 7.00 (0.00) | 5.81 (0.75) |
| den | den | Omdat hij het uitzicht blokkeerde, werd de **den** gekapt. | 9 | 1.46 (1.00) | 7.00 (0.00) | 6.69 (0.66) |
| drop | drop | Johan is helemaal verzot op zoute **drop**. | 7 | 1.00 (0.00) | 7.00 (0.00) | 5.83 (0.58) |
| fee | fee | In een heleboel kinderverhalen komt een **fee** voor. | 7 | 1.01 (0.06) | 6.96 (0.17) | 4.69 (1.25) |
| file | file | Olivier staat elke dag op weg naar zijn werk in de **file**. | 12 | 1.00 (0.00) | 7.00 (0.00) | 3.83 (1.40) |
| gang | gang | De leerling verstoorde de les en werd de **gang** opgestuurd. | 10 | 1.00 (0.00) | 7.00 (0.00) | 4.40 (1.29) |
| glad | glad | Door de ijzel waren de straten erg **glad**. | 8 | 1.03 (0.11) | 7.00 (0.00) | 4.43 (1.19) |
| gulp | gulp | Hij kwam er pas later achter dat zijn **gulp** open stond. | 11 | 1.00 (0.00) | 7.00 (0.00) | 5.33 (0.65) |
| honk | honk | De slagman haalde de derde **honk**. | 6 | 2.20 (2.53) | 7.00 (0.00) | 6.18 (0.75) |
| hoop | hoop | In moeilijke tijden is het belangrijk **hoop** te hebben. | 9 | 1.00 (0.00) | 7.00 (0.00) | 4.91 (1.04) |
| kin | kin | Ze kreeg altijd veel complimentjes over het kuiltje in haar **kin**. | 11 | 1.00 (0.00) | 7.00 (0.00) | 6.85 (0.42) |
| kind | kind | Grietje was een bijzonder lief **kind**. | 6 | 1.00 (0.00) | 7.00 (0.00) | 5.09 (0.70) |
| lap | lap | Voor het poetsen van zijn schoenen gebruikt Lucas een oude **lap**. | 11 | 1.18 (0.60) | 7.00 (0.00) | 5.91 (0.30) |
| last | last | Het voorzitterschap is een zware **last** voor Frederiek. | 8 | 1.08 (0.29) | 7.00 (0.00) | 5.33 (1.15) |
| leek | leek | Op het gebied van quantum fysica was Irene een volslagen **leek**. | 11 | 1.64 (1.80) | 7.00 (0.00) | 5.09 (0.70) |
| lever | lever | Eén van de meest veelzijdige organen is de **lever**. | 9 | 2.09 (2.43) | 7.00 (0.00) | 5.08 (1.00) |
| lid | lid | De voetballer was al twintig jaar **lid** van de club. | 10 | 1.07 (0.18) | 7.00 (0.00) | 5.99 (0.89) |
| list | list | De gevangene wist te ontsnappen door middel van een slimme **list**. | 11 | 1.24 (0.76) | 7.00 (0.00) | 6.70 (0.55) |
| log | log | De olifant is een nogal **log** beest. | 7 | 1.18 (0.60) | 7.00 (0.00) | 5.00 (1.00) |

| | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| Dutch word | English word | Dutch (priming) sentence | sentence length | meaning | spelling | pronunciation |
| map | map | Het document zit in die **map** daar. | 7 | 1.45 (1.10) | 7.00 ().00) | 5.45 (0.79) |
| mate | mate | De meeste mensen kennen de slogan " Geniet, maar drink met **mate**." | 11 | 1.00 (0.00) | 7.00 (0.00) | 4.09 (0.70) |
| mop | mop | De beroemde caberetier vertelde een saaie **mop**. | 7 | 1.00 (0.00) | 7.00 (0.00) | 6.36 (0.81) |
| mug | mug | Malaria is een ziekte die verspreid wordt door een **mug**. | 10 | 1.09 (0.00) | 7.00 (0.00) | 4.81 (1.14) |
| nut | nut | Ze kon er echt het **nut** niet van inzien. | 9 | 1.00 (0.00) | 7.00 (0.00) | 5.64 (0.87) |
| pad | pad | Langs de weg liep een smal **pad** voor voetgangers. | 9 | 1.00 (0.00) | 7.00 (0.00) | 5.18 (0.98) |
| pal | pal | De zwerver ging **pal** naast hem zitten. | 7 | 1.00 (0.00) | 7.00 (0.00) | 5.64 (1.03) |
| peer | peer | Pieter at elke morgen bij zijn ontbijt een **peer**. | 9 | 1.00 (0.00) | 6.91 (0.30) | 5.55 (1.37) |
| perk | perk | Er groeiden prachtige bloemen in dat **perk**. | 7 | 1.17 (0.39) | 7.00 (0.00) | 5.42 (0.67) |
| pet | pet | Toevallig stond de naam van haar favoriete band op zijn **pet**. | 11 | 1.01 (0.06) | 7.00 (0.00) | 6.61 (0.78) |
| pink | pink | Tijdens de training brak Tom zijn **pink**. | 7 | 1.00 (0.00) | 7.00 (0.00) | 6.82 (0.40) |
| pool | pool | Het wordt steeds kouder hoe dichterbij de **pool** je komt. | 10 | 1.00 (0.00) | 7.00 (0.00) | 4.73 (1.01) |
| prop | prop | Sanne maakte van de krant een grote **prop**. | 8 | 1.33 (1.15) | 7.00 (0.00) | 5.83 (0.58) |
| put | put | De gegijzelde werd in een diepe **put** gegooid. | 8 | 1.09 (0.30) | 7.00 (0.00) | 5.18 (0.87) |
| ramp | ramp | Hoewel ze veel gerepeteerd hadden, was het optreden een regelrechte **ramp**. | 11 | 1.01 (0.06) | 7.00 (0.00) | 5.18 (1.02) |
| roof | roof | De beveiliging wist nog steeds niet hoe de **roof** had kunnen gebeuren. | 12 | 1.00 (0.00) | 7.00 (0.00) | 4.57 (1.23) |
| room | room | Marc drinkt zijn koffie met een scheutje **room**. | 8 | 1.00 (0.00) | 7.00 (0.00) | 5.00 (1.04) |
| rooster | rooster | Koekjes moet je altijd even laten afkoelen op een **rooster.** | 10 | 1.09 (0.34) | 7.00 (0.00) | 4.93 (1.03) |
| rug | rug | Na de val had de turnster veel pijn in haar **rug**. | 11 | 1.07 (0.27) | 7.00 (0.00) | 4.79 (1.20) |
| rust | rust | Na al die drukte heb ik echt even wat **rust** nodig**.** | 12 | 1.01 (0.06) | 7.00 (0.00) | 5.49 (1.10) |
| sip | sip | Toen hij niet had gewonnen, keek Jim behoorlijk **sip**. | 9 | 1.00 (0.00) | 7.00 (0.00) | 7.00 (0.00) |
| slang | slang | Een oud woord voor **slang** is serpent. | 7 | 1.00 (0.00) | 7.00 (0.00) | 5.27 (1.01) |
| slap | slap | Hij vond de koffie uit de kantine veel te **slap**. | 10 | 1.09 (0.30) | 7.00 (0.00) | 5.27 (1.01) |
| slim | slim | Het spreekwoord luidt: "Wie niet sterk is, moet **slim** zijn." | 10 | 1.03 (0.11) | 7.00 (0.00) | 6.72 (0.54) |
| slip | slip | Ze moest zich uitkleden tot op haar BH en **slip**. | 10 | 1.17 (0.58) | 7.00 (0.00) | 6.83 (0.39) |
| slot | slot | In zijn haast was hij vergeten de deur op **slot** te doen. | 12 | 1.36 (0.69) | 7.00 (0.00) | 6.06 (0.88) |
| smart | smart | Tijden van grote vreugde en diepe **smart** wisselden elkaar af. | 10 | 1.17 (0.58) | 7.00 (0.00) | 6.08 (0.67) |

| | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| Dutch word | English word | Dutch (priming) sentence | sentence length | meaning | spelling | pronunciation |
| spin | spin | Op vakantie in Brazilië was hij gebeten door een giftige **spin**. | 11 | 1.00 (0.00) | 7.00 (0.00) | 6.91 (0.30) |
| spit | spit | Clara's lievelingsgerecht was kip van het **spit**. | 7 | 1.50 (1.73) | 7.00 (0.00) | 6.58 (0.67) |
| spot | spot | De leerlingen dreven de **spot** met hun leraar. | 8 | 1.09 (0.30) | 7.00 (0.00) | 6.36 (0.50) |
| stadium | stadium | De ontwikkeling van het nieuwe medicijn bleef steken in een vroeg **stadium**. | 12 | 2.17 (2.29) | 7.00 (0.00) | 5.67 (0.65) |
| stage | stage | In de vakantie liep ze een aantal weken **stage**. | 9 | 1.06 (0.17) | 7.00 (0.00) | 3.96 (1.23) |
| stand | stand | Vroeger was het een schande onder je **stand** te trouwen. | 10 | 1.73 (1.19) | 7.00 (0.00) | 5.45 (0.52) |
| star | star | Zijn eigenzinnigheid maakte hem veel te **star.** | 7 | 1.00 (0.00) | 7.00 (0.00) | 6.19 (0.75) |
| steel | steel | De oude bezem had een nogal gehavende **steel.** | 8 | 1.09 (0.32) | 6.99 (0.05) | 4.97 (0.94) |
| stem | stem | De artiest verloor drie dagen voor zijn optreden zijn **stem.** | 10 | 1.00 (0.00) | 7.00 (0.00) | 6.67 (0.49) |
| strand | strand | Als het mooi weer is, gaan veel mensen naar het **strand**. | 11 | 1.50 (1.00) | 7.00 (0.00) | 5.42 (0.51) |
| teen | teen | Elise stootte per ongeluk haar **teen.** | 6 | 1.55 (1.81) | 7.00 (0.00) | 5.18 (0.75) |
| toe | toe | Lars was echt aan een whisky **toe**. | 7 | 1.00 (0.00) | 7.00 (0.00) | 4.36 (0.81) |
| trap | trap | Met 11.674 treden is het de langste **trap** ter wereld. | 10 | 1.07 (0.23) | 7.00 (0.00) | 5.24 (0.86) |
| vast | vast | Die dop zat wel heel erg **vast**. | 7 | 1.50 (1.08) | 7.00 (0.00) | 5.60 (0.52) |
| vet | vet | Men zegt altijd dat plantaardige olie beter is dan dierlijk **vet**. | 11 | 1.03 (0.11) | 7.00 (0.00) | 6.36 (0.95) |
| wand | wand | Er hing een prachtig schilderij aan de **wand**. | 8 | 1.00 (0.00) | 6.97 (0.11) | 5.48 (1.02) |
| war | war | De dementerende patiënt was helemaal in de **war**. | 8 | 1.00 (0.00) | 7.00 (0.00) | 5.42 (0.51) |
| wet | wet | In de meeste landen is openbare dronkenschap bij de **wet** verboden. | 11 | 1.00 (0.00) | 7.00 (0.00) | 6.12 (0.76) |

**Table A-12:** The Dutch prime sentences, sentence lengths and average similarity ratings (and standard deviations) for the full set of 72 identical interlingual homographs. The similarity ratings were provided on a scale from 1 (not at all similar) to 7 ((almost) identical).

| | | | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
| afval | trash | 1, 2, 3, 4, 5, 6 | 12.44 | 2.74 | 5 | 1.80 | 22.47 | 3.06 | 5 | 1.70 | 0.00 |
| bot | rude | 1, 2, 3, 4, 5, 6 | 14.52 | 2.80 | 3 | 1.00 | 22.06 | 3.05 | 4 | 1.40 | 0.00 |
| doel | target | – | 77.45 | 3.53 | 4 | 1.00 | 37.96 | 3.29 | 6 | 1.95 | 0.17 |
| doos | box | 3 | 38.28 | 3.22 | 4 | 1.00 | 89.75 | 3.66 | 3 | 1.40 | 0.25 |
| eerlijk | honest | 3 | 172.38 | 3.88 | 7 | 1.85 | 72.33 | 3.57 | 6 | 1.95 | 0.14 |
| emmer | bucket | 1, 2, 3, 4, 5, 6 | 6.72 | 2.47 | 5 | 1.65 | 10.02 | 2.71 | 6 | 1.85 | 0.17 |
| fakkel | torch | 1, 2, 3, 4, 5, 6 | 2.20 | 1.99 | 6 | 1.90 | 4.98 | 2.41 | 5 | 1.90 | 0.00 |
| fiets | bike | 1, 2, 3, 4, 5, 6 | 21.75 | 2.98 | 5 | 1.45 | 25.88 | 3.12 | 4 | 1.50 | 0.20 |
| gazon | lawn | 1, 2, 3, 4, 5, 6 | 3.09 | 2.13 | 5 | 1.85 | 12.35 | 2.80 | 4 | 1.60 | 0.40 |
| gedicht | poem | 1, 2 | 14.73 | 2.81 | 7 | 1.60 | 13.65 | 2.84 | 4 | 1.85 | 0.00 |
| gemak | ease | 1, 2, 3, 4, 5, 6 | 18.04 | 2.90 | 5 | 1.75 | 19.10 | 2.99 | 4 | 1.40 | 0.20 |
| gevaar | danger | – | 91.54 | 3.60 | 6 | 1.80 | 43.67 | 3.35 | 6 | 1.65 | 0.17 |
| geweer | rifle | 1, 2 | 55.39 | 3.38 | 6 | 1.80 | 14.57 | 2.87 | 5 | 1.70 | 0.17 |
| grap | joke | 1, 2, 3, 4, 5, 6 | 51.02 | 3.35 | 4 | 1.50 | 73.02 | 3.57 | 4 | 1.45 | 0.00 |
| griep | flu | 1, 2 | 7.13 | 2.50 | 5 | 1.50 | 8.71 | 2.65 | 3 | 1.85 | 0.00 |
| grot | cave | 1, 2, 3, 4, 5, 6 | 17.45 | 2.88 | 4 | 1.05 | 13.98 | 2.85 | 4 | 1.00 | 0.00 |
| haai | shark | 1, 2, 3, 4, 5, 6 | 9.44 | 2.62 | 4 | 1.00 | 14.98 | 2.88 | 5 | 1.50 | 0.40 |
| heilig | sacred | 3 | 12.21 | 2.73 | 6 | 1.80 | 14.02 | 2.85 | 6 | 2.00 | 0.00 |
| heks | witch | 1, 2, 3, 4, 5, 6 | 26.76 | 3.07 | 4 | 1.55 | 27.65 | 3.15 | 5 | 1.50 | 0.00 |
| hout | wood | 1, 2, 3, 4, 5, 6 | 23.58 | 3.01 | 4 | 1.00 | 27.00 | 3.14 | 4 | 1.45 | 0.25 |
| huid | skin | 1, 2, 3, 4, 5, 6 | 39.56 | 3.24 | 4 | 1.55 | 44.04 | 3.35 | 4 | 1.35 | 0.25 |
| huur | rent | 3 | 28.88 | 3.10 | 4 | 1.40 | 34.55 | 3.25 | 4 | 1.20 | 0.00 |
| jammer | pity | 1, 2, 4, 5, 6 | 92.20 | 3.61 | 6 | 1.75 | 23.51 | 3.08 | 4 | 1.70 | 0.00 |
| jas | coat | 1, 2, 3, 4, 5, 6 | 48.11 | 3.32 | 3 | 1.00 | 42.08 | 3.33 | 4 | 1.35 | 0.25 |
| jeuk | itch | 1, 2, 3, 4, 5, 6 | 2.58 | 2.06 | 4 | 1.50 | 4.18 | 2.33 | 4 | 1.55 | 0.00 |
| jurk | dress | 1, 2, 3, 4, 5, 6 | 55.75 | 3.39 | 4 | 1.50 | 87.20 | 3.65 | 5 | 1.75 | 0.00 |
| keuze | choice | 3 | 66.45 | 3.46 | 5 | 1.80 | 97.55 | 3.70 | 6 | 1.95 | 0.17 |

| | | | DUTCH CHARACTERISTICS | | | | ENGLISH CHARACTERISTICS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dutch word | English word | included in Experiment | frequency | log10(frequency) | word length | OLD20 | frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
| kikker | frog | 1, 2, 4, 5, 6 | 8.23 | 2.56 | 6 | 1.25 | 11.82 | 2.78 | 4 | 1.80 | 0.00 |
| kogel | bullet | 3 | 43.72 | 3.28 | 5 | 1.65 | 38.24 | 3.29 | 6 | 1.80 | 0.17 |
| konijn | rabbit | 3 | 18.87 | 2.92 | 6 | 1.85 | 20.94 | 3.03 | 6 | 2.50 | 0.00 |
| kooi | cage | 1, 2, 3, 4, 5, 6 | 13.81 | 2.78 | 4 | 1.00 | 20.27 | 3.02 | 4 | 1.15 | 0.00 |
| koorts | fever | 3 | 14.77 | 2.81 | 6 | 1.90 | 19.94 | 3.01 | 5 | 1.75 | 0.00 |
| kruid | herb | 1, 2, 3, 4, 5, 6 | 2.15 | 1.98 | 5 | 1.45 | 4.98 | 2.41 | 4 | 1.65 | 0.00 |
| kunst | art | 1, 2 | 37.09 | 3.21 | 5 | 1.60 | 70.80 | 3.56 | 3 | 1.10 | 0.20 |
| lawaai | noise | 3 | 17.08 | 2.87 | 6 | 1.95 | 34.88 | 3.25 | 5 | 1.70 | 0.00 |
| leger | army | 1, 2, 3, 4, 5, 6 | 107.98 | 3.67 | 5 | 1.05 | 85.69 | 3.64 | 4 | 1.80 | 0.00 |
| lepel | spoon | 1, 2, 3, 4, 5, 6 | 5.01 | 2.34 | 5 | 1.55 | 7.61 | 2.59 | 5 | 1.75 | 0.00 |
| lied | song | 1, 2, 3, 4, 5, 6 | 21.27 | 2.97 | 4 | 1.25 | 93.69 | 3.68 | 4 | 1.45 | 0.00 |
| lijm | glue | 1, 2, 3, 4, 5, 6 | 3.98 | 2.24 | 4 | 1.40 | 5.88 | 2.48 | 4 | 1.65 | 0.00 |
| lucht | sky | 3, 4, 5, 6 | 89.00 | 3.59 | 5 | 1.50 | 44.80 | 3.36 | 3 | 1.65 | 0.00 |
| macht | power | 3 | 83.99 | 3.57 | 5 | 1.35 | 149.02 | 3.88 | 5 | 1.45 | 0.00 |
| mier | ant | 1, 2, 3, 4, 5, 6 | 2.54 | 2.05 | 4 | 1.00 | 5.35 | 2.44 | 3 | 1.20 | 0.00 |
| moeras | swamp | 1, 2, 3, 4, 5, 6 | 9.74 | 2.63 | 6 | 2.00 | 8.98 | 2.66 | 5 | 1.65 | 0.00 |
| muur | wall | 1, 2, 3, 4, 5, 6 | 66.89 | 3.47 | 4 | 1.40 | 70.69 | 3.56 | 4 | 1.20 | 0.00 |
| oorlog | war | 1, 2, 3, 4, 5, 6 | 178.96 | 3.89 | 6 | 2.05 | 174.75 | 3.95 | 3 | 1.00 | 0.17 |
| plicht | duty | 1, 2, 3, 4, 5, 6 | 28.10 | 3.09 | 6 | 1.75 | 50.96 | 3.42 | 4 | 1.90 | 0.00 |
| reus | giant | – | 7.45 | 2.51 | 4 | 1.20 | 27.06 | 3.14 | 5 | 1.90 | 0.00 |
| ridder | knight | 3 | 13.58 | 2.77 | 6 | 1.65 | 26.76 | 3.14 | 6 | 1.90 | 0.00 |
| roem | fame | 3 | 7.98 | 2.54 | 4 | 1.20 | 8.65 | 2.65 | 4 | 1.15 | 0.00 |
| ruil | swap | 1, 2, 3, 4, 5, 6 | 15.39 | 2.83 | 4 | 1.00 | 3.63 | 2.27 | 4 | 1.45 | 0.00 |
| saai | dull | 1, 2, 3, 4, 5, 6 | 31.47 | 3.14 | 4 | 1.05 | 12.08 | 2.79 | 4 | 1.15 | 0.17 |
| slager | butcher | 1, 2, 3, 4, 5, 6 | 6.63 | 2.46 | 6 | 1.65 | 8.51 | 2.64 | 7 | 2.35 | 0.29 |
| smaak | taste | 1, 2, 3, 4, 5, 6 | 29.18 | 3.11 | 5 | 1.55 | 51.31 | 3.42 | 5 | 1.60 | 0.00 |
| snoep | candy | 1, 2 | 12.55 | 2.74 | 5 | 1.40 | 35.78 | 3.26 | 5 | 1.65 | 0.00 |

| Dutch word | English word | included in Experiment | DUTCH CHARACTERISTICS frequency | log10(frequency) | word length | OLD20 | ENGLISH CHARACTERISTICS frequency | log10(frequency) | word length | OLD20 | orthographic similarity |
|---|---|---|---|---|---|---|---|---|---|---|---|
| spiegel | mirror | 1, 2, 3, 4, 5, 6 | 27.44 | 3.08 | 7 | 1.90 | 24.18 | 3.09 | 6 | 2.50 | 0.14 |
| steeg | alley | – | 8.74 | 2.58 | 5 | 1.60 | 16.29 | 2.92 | 5 | 1.65 | 0.20 |
| taart | pie | 1, 2 | 31.35 | 3.14 | 5 | 1.25 | 28.75 | 3.17 | 3 | 1.10 | 0.00 |
| tante | aunt | 1, 2, 3, 4, 5, 6 | 62.34 | 3.44 | 5 | 1.60 | 55.20 | 3.45 | 4 | 1.45 | 0.40 |
| tas | bag | 1, 2 | 58.36 | 3.41 | 3 | 1.00 | 94.04 | 3.68 | 3 | 1.00 | 0.33 |
| touw | rope | 1, 2, 3, 4, 5, 6 | 26.25 | 3.06 | 4 | 1.30 | 22.71 | 3.06 | 4 | 1.00 | 0.25 |
| tuin | garden | 1, 2, 3, 4, 5, 6 | 36.66 | 3.21 | 4 | 1.40 | 26.55 | 3.13 | 6 | 1.85 | 0.17 |
| twijfel | doubt | 3 | 34.03 | 3.17 | 7 | 2.25 | 62.84 | 3.51 | 5 | 1.95 | 0.00 |
| verdrag | treaty | 1, 2, 3, 4, 5, 6 | 7.00 | 2.49 | 7 | 1.90 | 4.69 | 2.38 | 6 | 1.90 | 0.14 |
| vloek | curse | 1, 2, 3, 4, 5, 6 | 18.77 | 2.91 | 5 | 1.65 | 18.22 | 2.97 | 5 | 1.60 | 0.00 |
| vogel | bird | 1, 2, 3, 4, 5, 6 | 32.27 | 3.15 | 5 | 1.80 | 45.45 | 3.37 | 4 | 1.75 | 0.00 |
| vrede | peace | 3 | 65.01 | 3.45 | 5 | 1.60 | 69.61 | 3.55 | 5 | 1.80 | 0.20 |
| vuil | dirt | 1, 2, 4, 5, 6 | 24.06 | 3.02 | 4 | 1.25 | 25.69 | 3.12 | 4 | 1.80 | 0.00 |
| wang | cheek | 1, 2, 3, 4, 5, 6 | 7.89 | 2.54 | 4 | 1.10 | 7.16 | 2.56 | 5 | 1.70 | 0.00 |
| woede | anger | 1, 2, 3, 4, 5, 6 | 24.95 | 3.04 | 5 | 1.40 | 19.43 | 3.00 | 5 | 1.65 | 0.00 |
| wortel | carrot | 1, 2, 3, 4, 5, 6 | 6.11 | 2.43 | 6 | 1.65 | 3.82 | 2.29 | 6 | 1.90 | 0.17 |
| wreed | cruel | 3 | 15.18 | 2.82 | 5 | 1.70 | 18.35 | 2.97 | 5 | 1.95 | 0.40 |
| zacht | soft | 1, 2, 4, 5, 6 | 22.85 | 3.00 | 5 | 1.30 | 32.02 | 3.21 | 4 | 1.65 | 0.20 |
| zwaar | heavy | 3 | 80.84 | 3.55 | 5 | 1.75 | 47.29 | 3.38 | 5 | 1.90 | 0.20 |
| zwakte | weakness | 3 | 5.90 | 2.41 | 6 | 1.65 | 8.90 | 2.66 | 8 | 2.50 | 0.38 |

**Table A-14:** The full set of 74 translation equivalents. For each item, the third column indicates in which (if any) experiment the item was included. Frequency refers to the SUBTLEX raw 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 of a word, expressed as its mean orthographic Levenshtein distance to its 20 closest neighbours; orthographic similarity refers to the measure of objective orthographic similarity (on a scale from 0 to 1), calculated as the Levenshtein distance between the Dutch and English forms of the words divided by the length of the longest of the two forms.

| | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| Dutch word | English word | Dutch (priming) sentence | sentence length | meaning | spelling | pronunciation |
| afval* | trash | Zet die kapotte stoel maar bij het **afval**. | 8 | 7.00 (0.00) | 1.17 (0.58) | 1.25 (0.62) |
| bot* | rude | De leraar was erg **bot** tegen zijn leerlingen. | 8 | 6.38 (0.87) | 1.08 (0.28) | 1.08 (0.28) |
| doel | target | De maffiabaas was het **doel** van een aanslag. | 8 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| doos | box | Op Valentijnsdag kreeg ze een **doos** bonbons. | 7 | 7.00 (0.00) | 1.27 (0.59) | 1.20 (0.56) |
| eerlijk | honest | Liegen is per definitie het tegenovergestelde van **eerlijk** zijn. | 9 | 7.00 (0.00) | 1.08 (0.28) | 1.08 (0.28) |
| emmer* | bucket | Voor dat deel van het zandkasteel hebben we een grote **emmer** nodig. | 12 | 6.79 (0.58) | 1.00 (0.00) | 1.00 (0.00) |
| fakkel | torch | De kinderen liepen naar het strand met ieder een **fakkel**. | 10 | 6.77 (0.60) | 1.00 (0.00) | 1.00 (0.00) |
| fiets | bike | Hij ging elke dag met de **fiets** naar school. | 9 | 6.77 (0.44) | 1.15 (0.55) | 1.15 (0.55) |
| gazon | lawn | In de tijd van de Tudors was het **gazon** een statussymbool. | 11 | 6.92 (0.28) | 1.08 (0.28) | 1.15 (0.38) |
| gedicht | poem | Bij de herdenking werd een prachtig **gedicht** voorgedragen. | 8 | 6.93 (0.27) | 1.00 (0.00) | 1.00 (0.00) |
| gemak | ease | De turnster beëindigde haar oefening met **gemak**. | 7 | 6.85 (0.38) | 1.00 (0.00) | 1.00 (0.00) |
| gevaar | danger | Het comité heeft aangetoond dat de President niet in **gevaar** was. | 11 | 6.93 (0.27) | 1.00 (0.00) | 1.00 (0.00) |
| geweer | rifle | De jager schoot het hert neer met zijn **geweer**. | 9 | 6.71 (0.61) | 1.00 (0.00) | 1.00 (0.00) |
| grap | joke | Helaas kon de directeur die **grap** niet waarderen. | 8 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| griep | flu | Op de kleuterschool van Carlijn heerste de **griep**. | 8 | 6.92 (0.29) | 1.00(0.00) | 1.00 (0.00) |
| grot | cave | De archeologen vonden een skelet diep in de **grot**. | 9 | 6.83 (0.39) | 1.17 (0.58) | 1.00 (0.00) |
| haai* | shark | Het best ontwikkelde zintuig van de **haai** is zijn reukvermogen. | 10 | 7.00 (0.00) | 1.08 (0.28) | 1.00 (0.00) |
| heilig | sacred | Delphi was voor de oude Grieken een **heilig** oord. | 9 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| heks | witch | Hans en Grietje werden gevangen gehouden door een gemene **heks**. | 10 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| hout | wood | De bouwvakker legde nog wat **hout** op de stapel. | 9 | 6.92 (0.29) | 1.75 (1.06) | 1.92 (1.16) |
| huid | skin | Naarmate ze ouder werd, werd haar **huid** steeds gevoeliger. | 9 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| huur | rent | Aan het begin van het jaar ging de **huur** weer omhoog. | 11 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| jammer | pity | In de herfst droeg ze al een dikke **jas**. | 9 | 6.33 (1.30) | 1.00 (0.00) | 1.00 (0.00) |
| jas* | coat | Dat het feest niet door ging, vonden we zeer **jammer**. | 10 | 6.79 (0.58) | 1.00 (0.00) | 1.07 (0.27) |
| jeuk | itch | Ik word helemaal gek van die vreselijke **jeuk**! | 8 | 6.92 (0.28) | 1.00 (0.00) | 1.00 (0.00) |
| jurk | dress | Voor het gala had ze een lange, groene **jurk** gekocht. | 10 | 6.85 (0.38) | 1.00 (0.00) | 1.00 (0.00) |
| keuze | choice | Het was voor haar een zeer moeilijke **keuze**. | 8 | 6.93 (0.26) | 1.93 (0.00) | 1.67 (1.45) |

| | | | | SIMILARITY RATINGS | | |
|---|---|---|---|---|---|---|
| Dutch word | English word | Dutch (priming) sentence | sentence length | meaning | spelling | pronunciation |
| kikker | frog | Elke avond hoorden ze een **kikker** kwaken. | 7 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| kogel* | bullet | Tijdens de strijd werd de soldaat geraakt door een **kogel**. | 10 | 6.93 (0.27) | 1.00 (0.00) | 1.00 (0.00) |
| konijn | rabbit | Op Eerste Kerstdag eten wij vaak fazant of **konijn.** | 9 | 7.00 (0.00) | 1.20 (0.77) | 1.20 (0.77) |
| kooi | cage | In Leningrad Zoo zitten alle dieren nog in een **kooi**. | 10 | 6.54 (1.66) | 1.46 (0.88) | 1.62 (0.87) |
| koorts* | fever | De arts schreef hem rust in bed voor tegen de **koorts**. | 11 | 7.00 (0.00) | 1.13 (0.52) | 1.13 (0.52) |
| kruid | herb | Kamille is een welriekend, geneeskrachtig **kruid**. | 6 | 6.87 (0.35) | 1.07 (0.26) | 1.07 (0.26) |
| kunst | art | Van kinds af aan was hij een grote liefhebber van **kunst**. | 11 | 6.92 (0.29) | 1.25 (0.87) | 1.17 (0.58) |
| lawaai | noise | Al die rotjes op Oudjaarsavond maken een geweldig **lawaai**. | 9 | 6.85 (0.55) | 1.00 (0.00) | 1.00 (0.00) |
| leger | army | De Zwitserse Garde is het **leger** van Vaticaanstad. | 8 | 7.00 (0.00) | 1.07 (0.26) | 1.07 (0.26) |
| lepel | spoon | Anna kocht op de antiekmarkt voor €15 een vergulde **lepel**. | 10 | 7.00 (0.00) | 1.08 (0.29) | 1.00 (0.00) |
| lied | song | De zangeres sloot het concert af met haar populairste **lied**. | 10 | 7.00 (0.00) | 1.15 (0.55) | 1.15 (0.55) |
| lijm | glue | Als je een fotoalbum maakt, moet je daar speciale **lijm** voor gebruiken. | 12 | 6.92 (0.29) | 1.17 (0.39) | 1.17 (0.58) |
| lucht | sky | Het was een prachtige dag met een staalblauwe **lucht**. | 9 | 6.87 (0.52) | 1.00 (0.00) | 1.00 (0.00) |
| macht | power | Van 1976 tot 2008 had Fidel Castro de **macht** in Cuba. | 11 | 6.80 (0.41) | 1.07 (0.26) | 1.07 (0.26) |
| mier* | ant | Weinig diersoorten kunnen overal ter wereld leven; de **mier** wel. | 10 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| moeras | swamp | Er is maar één soort krokodil die in het **moeras** leeft. | 11 | 6.92 (0.28) | 1.00 (0.00) | 1.00 (0.00) |
| muur | wall | In zijn kamer hing aan elke **muur** wel een poster. | 10 | 7.00 (0.00) | 1.08 (0.28) | 1.08 (0.28) |
| oorlog | war | Het Ministerie van Defensie heette vroeger het Ministerie van **Oorlog**. | 10 | 6.92 (0.28) | 1.00 (0.00) | 1.00 (0.00) |
| plicht | duty | De President beschermen was zijn **plicht**. | 6 | 6.86 (0.36) | 1.00 (0.00) | 1.00 (0.00) |
| reus* | giant | Roald Dahl schreef "De Grote Vriendelijke **Reus**". | 7 | 6.71 (0.47) | 1.07 (0.27) | 1.07 (0.27) |
| ridder* | knight | In de vroege Middeleeuwen was de rang van **ridder** een lage. | 11 | 7.00 (0.00) | 1.25 (0.87) | 1.08 (0.29) |
| roem | fame | Met zijn zeefdruk van Marilyn Monroe vergaarde Anthony Warhol eeuwige **roem**. | 11 | 6.92 (0.28) | 1.15 (0.38) | 1.15 (0.38) |
| ruil | swap | Beide jongens waren erg tevreden met de **ruil**. | 8 | 6.36 (1.39) | 1.00 (0.00) | 1.00 (0.00) |
| saai | dull | De leerlingen vonden de geschiedenislessen maar **saai**. | 7 | 6.62 (0.96) | 1.00 (0.00) | 1.00 (0.00) |
| slager | butcher | Eén van de oudste ambachten is dat van **slager**. | 9 | 7.00 (0.00) | 1.15 (0.38) | 1.15 (0.38) |
| smaak* | taste | In juni hebben pruimen een vollere **smaak** dan in februari. | 10 | 6.73 (0.46) | 1.07 (0.26) | 1.07 (0.26) |
| snoep | candy | Als je langs de deuren gaat met Halloween krijg je veel **snoep**. | 12 | 6.79 (0.43) | 1.00 (0.00) | 1.00 (0.00) |

| Dutch word | English word | Dutch (priming) sentence | sentence length | SIMILARITY RATINGS meaning | spelling | pronunciation |
|---|---|---|---|---|---|---|
| spiegel | mirror | Na de operatie durfde ze niet meer in de **spiegel** te kijken. | 12 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| steeg | alley | Die man wil je niet ’s avonds tegenkomen in een donkere **steeg**. | 12 | 6.92 (0.28) | 1.00 (0.00) | 1.00 (0.00) |
| taart | pie | Voor zijn afscheid hadden zijn collega’s een heerlijke **taart** gemaakt. | 10 | 6.23 (1.42) | 1.00 (0.00) | 1.00 (0.00) |
| tante | aunt | Ze gingen graag elk weekend bij hun **tante** logeren. | 9 | 6.86 (0.53) | 2.00 (1.30) | 1.64 (1.08) |
| tas | bag | Ze kon nooit iets vinden in haar rommelige **tas**. | 9 | 7.00 (0.00) | 1.92 (1.00) | 1.25 (0.87) |
| touw | rope | In de rechter keukenla lag een rol dik **touw**. | 9 | 6.92 (0.8) | 1.08 (0.28) | 1.08 (0.28) |
| tuin | garden | Als ze niet op haar werk was, was ze in de **tuin**. | 12 | 6.92 (0.28) | 1.00 (0.00) | 1.00 (0.00) |
| twijfel | doubt | Mijn oma zei altijd tegen me: “Bij **twijfel**, niet doen!” | 10 | 7.00 (0.00) | 1.17 (0.58) | 1.33 (1.15) |
| verdrag | treaty | In 1992 werd de Europese Unie opgericht met het **Verdrag** van Maastricht. | 12 | 6.92 (0.28) | 1.00 (0.00) | 1.00 (0.00) |
| vloek | curse | Jaren lang dacht men dat er op Toetanchamons tombe een **vloek** ruste. | 12 | 6.87 (0.52) | 1.07 (0.26) | 1.07 (0.26) |
| vogel | bird | De bijkolibrie is de kleinste **vogel** in de wereld. | 9 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| vrede | peace | Het was helaas weer mis, na vele jaren van **vrede**. | 10 | 7.00 (0.00) | 1.07 (0.26) | 1.07 (0.26) |
| vuil* | dirt | Mensen met smetvrees zijn bang voor **vuil**. | 7 | 6.75 (0.62) | 1.17 (0.58) | 1.25 (0.87) |
| wang | cheek | De hamster vervoert voedsel door het in zijn **wang** te stoppen. | 11 | 6.87 (0.52) | 1.00 (0.00) | 1.00 (0.00) |
| woede | anger | Sommige mensen uiten hun **woede** op ongezonde manieren. | 8 | 6.93 (0.27) | 1.00 (0.00) | 1.00 (0.00) |
| wortel | carrot | Een ezel kun je altijd blij maken met een **wortel**. | 10 | 6.93 (0.27) | 1.00 (0.00) | 1.07 (0.27) |
| wreed | cruel | Het is moeilijk te begrijpen hoe Hitler zo **wreed** kon zijn. | 11 | 7.00 (0.00) | 1.00 (0.00) | 1.00 (0.00) |
| zacht* | soft | Goede Italiaanse gnocchi zijn zo **zacht** als een kussentje. | 9 | 6.73 (0.80) | 1.80 (0.77) | 1.87 (0.99) |
| zwaar | heavy | Een ton veren en een ton bakstenen zijn even **zwaar**. | 10 | 7.00 (0.00) | 1.08 (0.28) | 1.08 (0.28) |
| zwakte | weakness | Oplichters weten goed gebruik te maken van je **zwakte**. | 9 | 7.00 (0.00) | 1.07 (0.27) | 1.07 (0.27) |

**Table A-16:** The Dutch prime sentences, sentence lengths and average similarity ratings (and standard deviations) for the full set of 74 translation equivalents. The similarity ratings were provided on a scale from 1 (not at all similar) to 7 ((almost) identical). For the 14 items (marked with *) that were included in the second rating experiment as fillers, only the average rating (and standard deviation) from the first experiment is given.

# APPENDIX B: Detailed results of the statistical analyses for Experiment 1 and 2

This Appendix includes detailed results of all of the analyses conducted on the data for Experiment 1 and 2 (Chapter 3).

# 1 Detailed results for Experiment 1

| | $\chi^2$ | $p$ | Δ (ms) | significant? |
|---|---|---|---|---|
| **2×2** | | | | |
| word type | 2.789 | .095 | 12 | marg. |
| version | 3.347 | .067 | 38 | marg. |
| word type × version | 15.01 | <.001 | - | sig. |
| **simple effects** | | | | |
| standard version | 13.52 | <.001 | 31 | sig. |
| mixed version | 0.744 | .388 | -8 | n.s. |
| **exploratory analysis: comparing interlingual homographs and English controls in the mixed version** | | | | |
| word type | 14.05 | <.001 | -43 | sig. |
| OLD20 | 0.071 | .791 | - | n.s. |

**Table B-1:** Reaction time data. All likelihood ratio tests had 1 degree of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in milliseconds. Positive effects indicate an advantage for the cognates or interlingual homographs compared to the English controls; negative effects indicate a disadvantage.

| | $\chi^2$ | $p$ | Δ (%) | significant? |
|---|---|---|---|---|
| **2×2** | | | | |
| word type | 0.157 | .692 | 0.2 | n.s. |
| version | 0.088 | .767 | -0.2 | n.s. |
| word type × version | 3.231 | .072 | - | marg. |
| **simple effects** | | | | |
| standard version | 1.415 | .234 | 1.0 | n.s. |
| mixed version | 0.651 | .420 | -0.4 | n.s. |

**Table B-2:** Accuracy data. All likelihood ratio tests had 1 degree of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in percentages. Positive effects indicate an advantage for the cognates or interlingual homographs compared to the English controls; negative effects indicate a disadvantage.

| | $\chi^2$ | $p$ | Δ (ms) | significant? |
|---|---|---|---|---|
| **simple effects** | | | | |
| preceding type: cognates | 0.174 | .677 | 7 | n.s. |
| preceding type: English controls | 0.307 | .580 | 9 | n.s. |
| preceding type: interlingual homographs | 0.529 | .467 | -12 | n.s. |
| preceding type: pseudohomophones | 0.144 | .705 | -4 | n.s. |
| preceding type: Dutch controls | 4.864 | .027 | -50 | n.s. |
| **2×2 interactions** | | | | |
| preceding types: Dutch controls & English controls | 5.516 | .019 | - | n.s. |
| preceding types: Dutch controls & cognates | 6.427 | .011 | - | n.s. |
| preceding types: Dutch controls & interlingual homographs | 2.850 | .091 | - | n.s. |
| preceding types: Dutch controls & pseudohomophones | 3.493 | .062 | - | n.s. |
| preceding types: English controls & cognates | 0.031 | .860 | - | n.s. |
| preceding types: English controls & interlingual homographs | 0.782 | .376 | - | n.s. |
| preceding types: English controls & pseudohomophones | 0.806 | .369 | - | n.s. |
| preceding types: cognates & interlingual homographs | 0.691 | .406 | - | n.s. |
| preceding types: cognates & pseudohomophones | 0.333 | .564 | - | n.s. |
| preceding types: interlingual homographs & pseudohomophones | 0.215 | .643 | - | n.s. |

**Table B-3:** Exploratory analysis of direct effects of the preceding trial. All 2×2 models included a maximal random effects structure with a random intercept and random slope for the word type of the current trial, stimulus type of the preceding trial and the interaction between these two factors by participants and a random intercept by items and. The by-participants random effects were not allowed to correlate. The simple effects models included only a random intercept and random slope for word type of the current trial by participants. All likelihood ratio tests had 1 degree of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in milliseconds. Positive effects indicate an advantage for the cognates compared to the English controls; negative effects indicate a disadvantage. The *p*-values were compared against a Bonferroni-corrected $\alpha$ of .01 and .005 for the simple effects and 2×2 interactions, respectively.

# 2 Detailed results for Experiment 2

| | $\chi^2$ | *p* | Δ (ms) | significant? |
|---|---|---|---|---|
| **2×5** | | | | |
| word type | 18.13 | <.001 | 23 | sig. |
| version* | 5.305 | .257 | - | n.s. |
| word type × version* | 46.65 | <.001 | - | sig. |
| **2×2 interactions** | | | | |
| standard vs mixed version | 16.23 | <.001 | - | sig. |
| standard vs +DC version | 23.83 | <.001 | - | sig. |
| standard vs +IH version | 6.657 | .010 | - | n.s. |
| standard vs +P version | 1.678 | .195 | - | n.s. |
| mixed vs +DC version | 0.878 | .349 | - | n.s. |
| mixed vs +IH version | 1.839 | .175 | - | n.s. |
| mixed vs +P version | 6.070 | .014 | - | n.s. |
| +DC vs +IH version | 4.463 | .035 | - | n.s. |
| +DC vs +P version | 10.31 | .001 | - | sig. |
| +IH vs +P version | 1.263 | .261 | - | n.s. |
| **simple effects** | | | | |
| standard version | 27.99 | <.001 | 46 | sig. |
| mixed version | 3.357 | .067 | 13 | n.s. |
| +DC version | 0.778 | .378 | 6 | n.s. |
| +IH version | 7.490 | .006 | 22 | sig. |
| +P version | 12.11 | <.001 | 30 | sig. |
| **exploratory analysis: interlingual homographs versus English controls** | | | | |
| mixed vs +IH version | 2.889 | .089 | - | marg. |
| *mixed version* | | | | |
| word type | 6.987 | .008 | -24 | sig. |
| OLD20 | 0.007 | .936 | - | n.s. |
| *+IH version* | | | | |
| word type | 0.693 | .405 | -8 | n.s. |
| OLD20 | 0.748 | .387 | - | n.s. |

**Table B-4:** Reaction time data, trimmed according to the reported trimming criteria. All likelihood ratio tests had 1 degree of freedom, except tests marked with *, which had 4 degrees of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in milliseconds. Positive effects indicate an advantage for the cognates or interlingual homographs compared to the English controls; negative effects indicate a disadvantage. The *p*-

values for the $2 \times 2$ and simple effects analyses were compared against a Bonferroni-corrected $\alpha$ of .005 and .01, respectively.

| | $\chi^2$ | $p$ | Δ (%) | significant? |
|---|---|---|---|---|
| 2×5 | | | | |
| word type | 1.234 | .165 | 0.3 | n.s. |
| version* | 9.575 | .048 | - | sig. |
| word type × version* | 6.885 | .142 | - | n.s. |
| **2×2 interactions** | | | | |
| standard vs mixed version | 0.053 | .818 | - | n.s. |
| standard vs +DC version | 0.258 | .612 | - | n.s. |
| standard vs +IH version | 2.411 | .121 | - | n.s. |
| standard vs +P version | 1.060 | .303 | - | n.s. |
| mixed vs +DC version | 1.029 | .311 | - | n.s. |
| mixed vs +IH version | 1.468 | .226 | - | n.s. |
| mixed vs +P version | 0.703 | .402 | - | n.s. |
| +DC vs +IH version | 5.928 | .015 | - | n.s. |
| +DC vs +P version | 4.101 | .043 | - | n.s. |
| +IH vs +P version | 0.165 | .684 | - | n.s. |
| **simple effects** | | | | |
| standard version | 0.034 | .854 | 0.1 | n.s. |
| mixed version | 0.357 | .550 | 0.4 | n.s. |
| +DC version | 2.289 | .093 | -0.7 | n.s. |
| +IH version | 2.964 | .085 | 1.0 | n.s. |
| +P version | 3.319 | .069 | 1.4 | n.s. |
| **exploratory analysis: pairwise comparisons for version** | | | | |
| standard vs mixed version | 0.053 | .818 | - | n.s. |
| standard vs +DC version | 0.258 | .612 | - | n.s. |
| standard vs +IH version | 2.411 | .121 | - | n.s. |
| standard vs +P version | 1.060 | .303 | - | n.s. |
| mixed vs +DC version | 1.029 | .311 | - | n.s. |
| mixed vs +IH version | 1.468 | .226 | - | n.s. |
| mixed vs +P version | 0.703 | .402 | - | n.s. |
| +DC vs +IH version | 5.928 | .015 | - | n.s. |
| +DC vs +P version | 4.101 | .043 | - | n.s. |
| +IH vs +P version | 0.165 | .684 | - | n.s. |

**Table B-5:** Accuracy data. All likelihood ratio tests had 1 degree of freedom, except tests marked with *, which had 4 degrees of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in percentages. Positive effects indicate an advantage for the cognates compared to the English controls; negative effects indicate a disadvantage. The p-values for the 2×2s and pairwise comparisons were compared against a Bonferroni-corrected $\alpha$ of .005; the *p*-values for the simple effects were compared against an $\alpha$ of .01. The models for the pairwise comparisons for version included a fixed effect for version, a random intercept by participants and an uncorrelated random intercept and slope for version by items.

| | $\chi^2$ | *p* | *Δ (ms)* | **significant?** |
|---|---|---|---|---|
| **simple effects** | | | | |
| preceding type: cognates | 3.237 | .072 | 25 | n.s. |
| preceding type: English controls | 0.635 | .426 | 11 | n.s. |
| preceding type: interlingual homographs | 0.541 | .462 | -10 | n.s. |
| preceding type: pseudohomophones | 6.011 | .014 | 25 | n.s. |
| preceding type: Dutch controls | 6.722 | .010 | -49 | sig. |
| **2×2 interactions** | | | | |
| preceding types: Dutch controls & English controls | 5.572 | .018 | - | n.s. |
| preceding types: Dutch controls & cognates | 10.70 | .001 | - | n.s. |
| preceding types: Dutch controls & interlingual homographs | 4.037 | .045 | - | n.s. |
| preceding types: Dutch controls & pseudohomophones | 10.65 | .001 | - | n.s. |
| preceding types: English controls & cognates | 0.463 | .496 | - | n.s. |
| preceding types: English controls & interlingual homographs | 1.107 | .293 | - | n.s. |
| preceding types: English controls & pseudohomophones | 0.891 | .345 | - | n.s. |
| preceding types: cognates & interlingual homographs | 4.971 | .026 | - | n.s. |
| preceding types: cognates & pseudohomophones | 0.008 | .927 | - | n.s. |
| preceding types: interlingual homographs & pseudohomophones | 4.360 | .037 | - | n.s. |

**Table B-6:** Exploratory analysis of direct effects of the preceding trial. All 2×2 models included a maximal random effects structure with a random intercept and random slope for the word type of the current trial, stimulus type of the preceding trial and the interaction between these two factors by participants and a random intercept by items and. The by-participants random effects were not allowed to correlate. The simple effects models included only a random intercept and random slope for word type of the current trial by participants. All likelihood ratio tests had 1 degree of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in milliseconds. Positive effects indicate an advantage for the cognates compared to the English controls; negative effects indicate a disadvantage. The *p*-values were compared against a Bonferroni-corrected $\alpha$ of .01 and .005 for the simple effects and 2×2 interactions, respectively.

As mentioned in the main text, the pre-registered trimming criteria for Experiment 2 were adjusted after the data had been collected, since when using the original trimming criteria, two data points remained that were less than 300ms. As these responses were considered to be accidental key presses, they were removed. This did not affect the significance level of any of the analyses, but there were some minor changes in the results for the confirmatory analyses. The following table lists the results of the confirmatory analyses using the original trimming criteria. The data for these analyses can also be found as a separate component of the project on the Open Science Framework.

| | $\chi^2$ | *p* | *Δ (ms)* | **significant?** |
|---|---|---|---|---|
| **2×5** | | | | |
| word type | 18.75 | <.001 | 24 | sig. |
| version* | 5.205 | .267 | - | n.s. |
| word type × version* | 43.08 | <.001 | - | sig. |
| **2×2 interactions** | | | | |
| standard vs mixed version | 14.35 | <.001 | - | sig. |
| standard vs +DC version | 24.12 | <.001 | - | sig. |
| standard vs +IH version | 7.113 | .007 | - | n.s. |
| standard vs +P version | 1.940 | .164 | - | n.s. |
| mixed vs +DC version | 1.439 | .230 | - | n.s. |
| mixed vs +IH version | 0.994 | .319 | - | n.s. |
| mixed vs +P version | 4.645 | .031 | - | n.s. |
| +DC vs +IH version | 4.463 | .035 | - | n.s. |
| +DC vs +P version | 10.31 | .001 | - | sig. |
| +IH vs +P version | 1.263 | .261 | - | n.s. |
| **simple effects** | | | | |
| standard version | 28.15 | <.001 | 47 | sig. |
| mixed version | 4.314 | .038 | 15 | n.s. |
| +DC version | 0.778 | .378 | 6 | n.s. |
| +IH version | 7.490 | .006 | 22 | sig. |
| +P version | 12.11 | <.001 | 30 | sig. |

**Table B-7:** Reaction time data, trimmed according to the original pre-registered trimming criteria. All likelihood ratio tests had 1 degree of freedom, except tests marked with *, which had 4 degrees of freedom. All effects (Δ) were derived from the estimates of the fixed effects provided by the model and are in milliseconds. Positive effects indicate an advantage for the cognates compared to the English controls; negative effects indicate a disadvantage. The *p*-values for the 2×2 and simple effects analyses were compared against a Bonferroni-corrected $\alpha$ of .005 and .01, respectively.

# APPENDIX C: Power curves for Experiment 6

This Appendix includes the power curves from the power simulations conducted to determine the required sample size for Experiment 6 (Chapter 5).

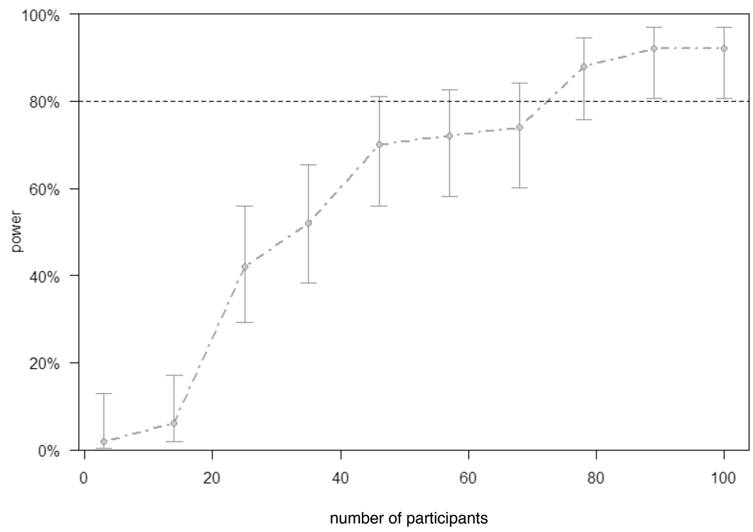


**Figure C-1:** Power curve for the simple effect of priming for the cognates (simulated to be approx. 20 ms). The level of power is displayed on the y-axis. The number of participants required to achieve a certain level of power is displayed on the x-axis. The dashed black line indicates 80% power. The simulation calculated the average estimated level of power (and 95% confidence interval) for 10 sample sizes between 0 and 100, based on 50 simulations each.

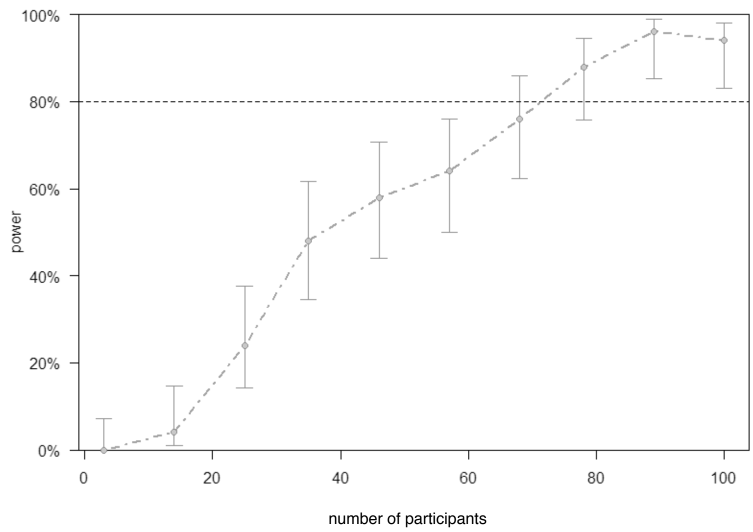


**Figure C-2:** Power curve for the simple effect of priming for the interlingual homographs (simulated to be approx. 20 ms). The level of power is displayed on the y-axis. The number of participants required to achieve a certain level of power is displayed on the x-axis. The dashed black line indicates 80% power. The simulation calculated the average estimated level of power (and 95% confidence interval) for 10 sample sizes between 0 and 100, based on 50 simulations each.

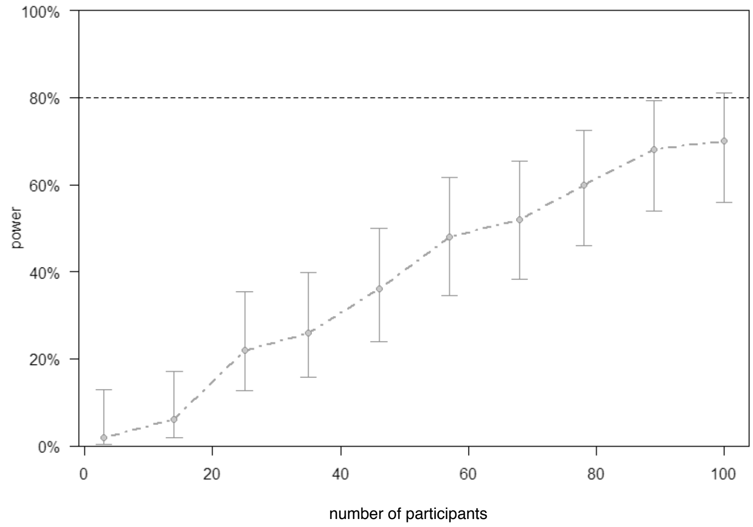


**Figure C-3:** Power curve for the 2×2 that compares the effect of priming for the cognates (simulated to be approx. 20 ms) and the translation equivalents (simulated to be approx. 0 ms). The level of power is displayed on the y-axis. The number of participants required to achieve a certain level of power is displayed on the x-axis. The dashed black line indicates 80% power. The simulation calculated the average estimated level of power (and 95% confidence interval) for 10 sample sizes between 0 and 100, based on 50 simulations each.

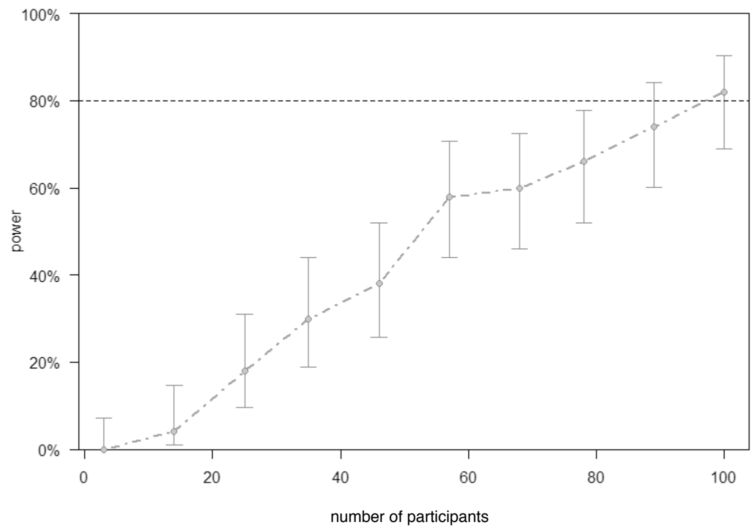


**Figure C-4:** Power curve for the 2×2 that compares the effect of priming for the interlingual homographs (simulated to be approx. 20 ms) and the translation equivalents (simulated to be approx. 0 ms). The level of power is displayed on the y-axis. The number of participants required to achieve a certain level of power is displayed on the x-axis. The dashed black line indicates 80% power. The simulation calculated the average estimated level of power (and 95% confidence interval) for 10 sample sizes between 0 and 100, based on 50 simulations each.