The role of phonology on recognising interlingual homographs in L1 sentence context

Merit Huisman
s1081186
MA Language Variation and Multilingualism, Radboud University
Master's Thesis
Sybrine Bultena and Gerrit Jan Kootstra
30-06-2025

The work submitted here is the sole responsibility of the undersigned, who has neither committed plagiarism nor colluded in its production.

Signed:
Merit Huisman

Word count: 13033

Table of contents

Abstract

In research on bilingualism, it is assumed that the mental lexicon contains all words from both languages and that its lexical representations can be activated simultaneously, meaning that bilingual word recognition is non-selective with respect to language. This is supported by evidence from interlingual homograph effects, in both isolation and sentence context, in both L1 and L2. Although attention had been paid to the role of orthographic and semantic overlap in bilingual word recognition, the role of phonological overlap was less clear. The present study therefore investigated how full phonological overlap in interlingual homographs influences bilingual word recognition in L1 sentence context. This was examined using a self-paced reading paradigm, in which Dutch-English bilinguals were presented with Dutch sentences containing an interlingual homograph or a Dutch control word. The interlingual homographs were phonologically identical or phonologically non-identical between Dutch and English. Reading times for target words and post-target words were measured. Interlingual homographs and control words did not show significantly different reading times, neither did the difference in reading times between these word types across the phonologically identical and non-identical conditions. These results indicate that cross-linguistic activation may be subject to certain limitations.

Keywords: Bilingual word recognition, interlingual homographs, sentence context, phonological overlap, self-paced reading

1 Background

1.1 Introduction

Words that appear to be identical between languages can be a source of clever wordplay. A striking example is a cap with the Dutch phrase '*HUISDIER in het Engels*' ('*pet in English*') printed on it, see Figure 1. At first glance, this may seem like an odd or even nonsensical statement. However, for bilingual speakers of Dutch and English, the phrase plays on the fact that the English translation of '*huisdier'* is '*pet'*, while in Dutch, '*pet'* refers to a cap.

Figure 1. *A cap stating 'HUISDIER in het Engels' which is 'pet', a pun on the fact the Dutch word for a cap is 'pet'.*

This example highlights the existence of interlingual homographs. Interlingual homographs, like '*pet'*, are words that are orthographically identical but semantically different between languages. Furthermore, some words are not only orthographically, but also semantically identical between languages, such as the word '*lip*' in Dutch and English, which are known as cognates. Because of their cross-linguistic overlap, interlingual homographs and cognates are widely used to study how bilinguals recognise words, providing insights into how the bilingual lexicon is organised and accessed.

Two hypotheses about bilingual word recognition are prominent in the field of bilingual research. According to the language selective access hypothesis, only word representations from the target language are activated during word recognition. This hypothesis is often accompanied by the idea that bilinguals have two separate lexicons, one lexicon for the first language (L1) and one lexicon for the second language (L2). Consequently, the presentation of words with cross-linguistic overlap, such as interlingual homographs and cognates, should not lead to co-activation of these words (Gerard &

Scarborough, 1989). A contrasting hypothesis is the language non-selective access hypothesis, which assumes that word representations from both languages are simultaneously activated during word recognition. This hypothesis is commonly accompanied by the idea that bilinguals have one integrated lexicon, containing all words from both languages. As a result, cross-linguistically overlapping words will be co-activated during word recognition (Beauvillain & Grainger, 1987).

Studies have shown that interlingual homographs exhibit various effects in bilingual word recognition (Dijkstra et al. 1998; Dijkstra et al. 2000). They have been found to show interference, reflected in slower reaction times. This effect can be modulated or even shift to facilitation, reflected in faster reaction times, depending on the specific task and relative frequency of the interlingual homograph in each of the languages. In addition, cognates are typically processed faster and more accurately, thus showing facilitation in bilingual word recognition (Dijkstra et al., 1999; Van Hell & Dijkstra, 2002). Since these cross-linguistically overlapping words were compared to words that exist exclusively in one language, i.e. control words, these findings suggest that words with cross-linguistic overlap are processed differently. This is seen as evidence that bilingual word recognition is non-selective with respect to language (see Bultena & Dijkstra, 2023; Palma & Titone, 2020, for overviews).

Although much attention had been paid to the influence of orthography and semantics in bilingual word recognition, the role of phonology was less clear (Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004). Studies on cognates found an inhibitory role of phonological overlap, (Frances et al., 2021; Tiffin-Richards, 2024), while a study on interlingual homographs showed that identical phonological overlap was associated with reduced interference (Kerkemeijer, 2024). However, these findings were obtained in studies that investigated these cross-linguistically overlapping words out of context, i.e. in isolation. How identical phonological overlap affects bilingual word recognition when interlingual homographs are embedded in sentences has not yet been studied. Therefore, the aim of the current study is to investigate the influence of full phonological overlap on the processing of interlingual homographs in L1 sentence context.

In the remainder of this chapter, the relevant literature will be reviewed, covering models of bilingual word recognition, the recognition of interlingual homographs, the role of phonology, word recognition in sentence context, and methods for sentence reading. Chapter 2 will present the methodology of the current study, while Chapter 3 will focus on the results. Finally, in Chapter 4, the results will be discussed in relation to the existing literature.

1.2 Previous research

1.2.1 Models of bilingual word recognition

Interlingual homographs have shown to exhibit various effects, such as interference, and cognates have shown to exhibit facilitation in bilingual word recognition, supporting the language non-selective access hypothesis (see Bultena & Dijkstra, 2023; Palma & Titone, 2020, for overviews). Based on this assumption, Dijkstra and Van Heuven (1998, as cited in Dijkstra & Van Heuven, 2002) developed the Bilingual Interactive Activation (BIA) model for bilingual word recognition, see Figure 2.

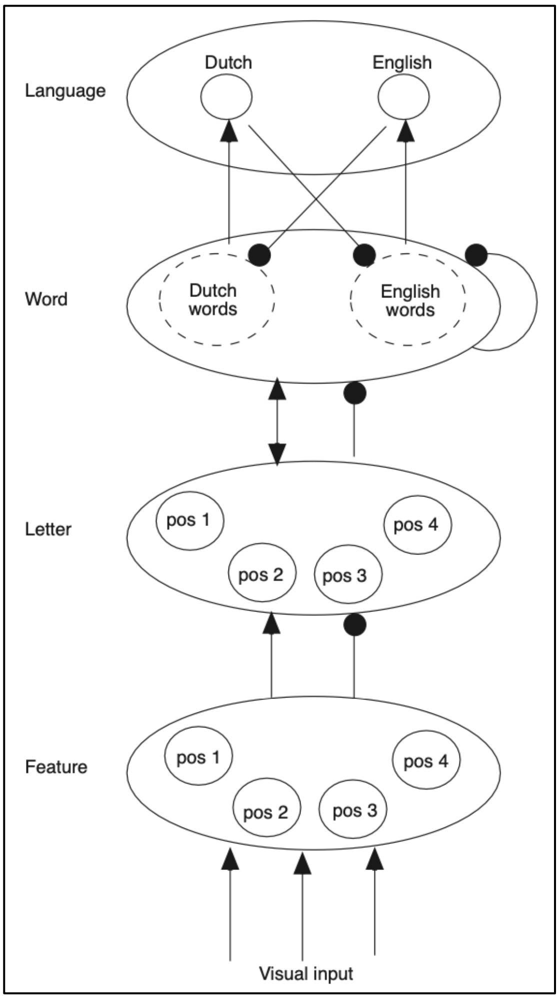

Figure 2. *The BIA model for bilingual word recognition (Dijkstra & Van Heuven, 1998, as cited in Dijkstra & Van Heuven, 2002).*

Note. Arrowheads indicate excitatory connections, black filled circled indicate inhibitory connections.

The BIA model (Dijkstra & Van Heuven, 1998, as cited in Dijkstra & Van Heuven, 2002) predicts the recognition of words based on their orthographic representations. When presenting a word to the BIA model, features at each letter position are activated first, which in turn activate letters that contain these features and inhibit letters that do not contain these features. The activated letters then activate words in both languages that contain the activated letter at the position in question and inhibit all other words. In addition, at the word level, all words inhibit each other, irrespective of the language they belong to. The co-activated words that belong to the same language activate their corresponding language node. This language node, in turn, inhibits activated words that belong to the other language, providing top-down facilitation of the selection of words from the target language. The placement of the language nodes after the activation of words makes the model fundamentally non-selective with respect to language.

The BIA model, however, appeared to have limitations, as Dijkstra and Van Heuven (2002) noticed. For instance, they mentioned that semantic and phonological representations of words were lacking and the representations of interlingual homographs and cognates were underspecified. Therefore, Dijkstra and Van Heuven (2002) updated their initial BIA model and called this the Bilingual Interactive Activation Plus (BIA+) model for bilingual word recognition, see Figure 3.

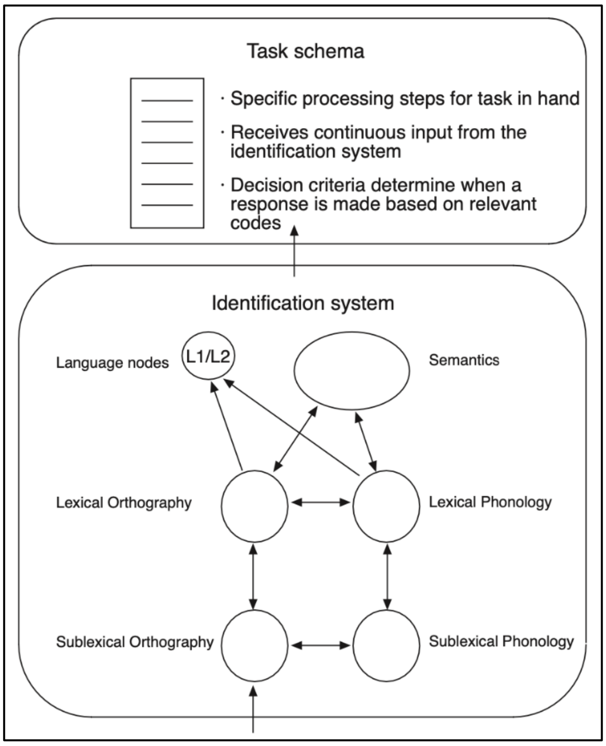

Figure 3. *The BIA+ model for bilingual word recognition (Dijkstra & Van Heuven, 2002).*

The BIA+ model (Dijkstra & Van Heuven, 2002) predicts the recognition of words using an identification system, in which phonological and semantic representations have been added, as well as a task schema. When presenting a word to the BIA+ model, the first stages of word recognition follow the same process as in the BIA model. This means that word recognition starts with the activation of sublexical orthographic representations, consisting of features and letters. These, in turn, activate lexical orthographic representations, which are words. Phonological and semantic representations are only activated once the sublexical and lexical orthographic representations have been activated. Consequently, phonological and semantic representations are activated later than orthographic representations during word recognition. Activation can also be influenced by other factors, such as relative frequency in the two languages. Since L2 representations generally have a lower relative frequency than L1 representations, L2 representations are activated more slowly. Unlike in the BIA model, language nodes in the BIA+ model serve as representations of language membership rather than functioning as a language filter that affects word activation levels. However, like in the BIA model, the placement of the language node after the activation of the lexical representations makes the model fundamentally non-selective with respect to language. The identification system is linked to a task schema to account for non-linguistic factors during bilingual word recognition, such as task demands or instructions.

Relatively recently, Dijkstra et al. (2019) developed the Multilink model, which builds on the BIA+ model (Dijkstra & Van Heuven, 2002). The Multilink model suggests that words with semantic representations related to the presented word also receive activation. This thus concerns an extension at the level of semantic representations in bilingual word recognition. Since the focus of the current study is on phonology and not on semantics, the Multilink model will not be discussed further.

1.2.2 Recognition of interlingual homographs

Interlingual homographs are characterised by slower processing than control words (see Bultena & Dijkstra, 2023, for a review), but other effects may arise due to other factors, such as task demands, stimulus list composition, and relative word frequency.

A study that has examined how the recognition of interlingual homographs can be affected by the testing conditions in which they appear, is a study by Dijkstra et al. (1998). Dutch-English bilinguals participated in three lexical decision tasks in which the relative frequency of interlingual homographs in both languages was manipulated, and the effect of different task contexts was examined. There were four frequency categories: high-frequent in Dutch and high-frequent in English (HFD-HFE); high-frequent in Dutch and low-frequent in English (HFD-LFE); low-frequent in Dutch and high-frequent in English (LFD-HFE); and low-frequent in Dutch and low-frequent in English (LFD-LFE). Experiment 1 involved an English lexical decision task, in which participants had to indicate whether the string of letters presented on a screen was an existing English word. The stimuli consisted of interlingual homographs, English control words, and pseudowords. No significant differences in response times were found between interlingual homographs and English control words in all frequency categories. Experiment 2 also involved an English lexical decision task. The same stimuli as in Experiment 1 were used, but Dutch control words were added, which had to be treated as nonwords. Depending on the relative frequency of the interlingual homographs in both languages, responses to interlingual homographs were significantly slower than to English control words, and the size of the interference effect depended on the frequency in Dutch (the non-target language). For instance, interlingual homographs in the HFD-LFE frequency category, meaning they are high-frequent in the non-target language but low-frequent in the target language, showed more interference than the other frequency conditions. Experiment 3 involved a generalised lexical decision task, in which participants had to indicate whether the string of letters presented on a screen was an existing word in either language. This experiment used the same set of stimuli as in Experiment 2, but now Dutch words had to be treated as words. Responses to interlingual homographs were significantly faster than English control words, but not faster than Dutch control words, which themselves were faster than the English control words. This study thus showed that the recognition of interlingual homographs is influenced by task demands and stimulus list composition.

Another study that investigated the effects of task demands and relative frequency is a study by Dijkstra et al. (2000). Dutch-English bilinguals took part in three experiments, all of which used the same set of stimuli, containing interlingual homographs in three frequency categories: LFD-HFE; HFD-LFE; LFD-LFE. Additionally, the set of stimuli contained English control words and Dutch control words. In Experiment 1, participants performed a language decision task, pressing one button when they identified an English word and another button when they identified a Dutch word. Language choice was influenced by the interlingual homograph's relative frequency in Dutch and English. Participants chose the language in which the interlingual homograph is more frequent: LFD-HFE elicited more 'English' responses and HFD-LFE elicited more 'Dutch' responses. In addition, response times for interlingual homographs were slower than for control words and response times were slower for interlingual homographs that are less frequent in the selected response language than in the other language. Experiment 2 involved an English go/no-go paradigm, in which participants only pressed a button when they identified an English word and not when they identified a word from the non-target language Dutch. Experiment 3 involved a Dutch go/no-go paradigm, in which participants only pressed a button when they identified a Dutch word and not when they identified a word from the non-target language English. Experiment 2 and Experiment 3 showed similar results, namely that interlingual homographs were responded to more slowly than to control words, indicating an interference effect. This even applied to the Dutch go/no-go task, which the participants performed in their L1. Analogous to Dijkstra et al. (1998), the interference effect was influenced by the relative frequency of the interlingual homograph in both languages. When the interlingual homograph was high-frequent in the non-target language and low-frequent in the target language, there was a stronger interference effect. Thus, this study also shows that task demands and relative frequency affect the recognition of interlingual homographs.

These findings reveal how interlingual homographs and cognates are represented in the bilingual lexicon. The BIA+ model (Dijkstra & Van Heuven, 2002) assumes that there are two representations for interlingual homographs, one for each language, at both the orthographic, phonological and semantic representational levels. For cognates, it is proposed that there are two representations, one for each language, at the orthographic and phonological representational levels, but only one semantic representation. The findings that relative frequency in the two languages affects the recognition of interlingual homographs (Dijkstra et al., 1998; Dijkstra et al., 2000) supports the assumption of the BIA+ model (Dijkstra & Van Heuven, 2002) that interlingual homographs have two separate representations in the bilingual lexicon, with one representation for each language. Indeed, if representations were shared between languages, recognition should be characterised by the cumulative frequency of the interlingual homograph rather than the relative frequency in the two languages.

1.2.3 Role of phonology in bilingual word recognition

According to the BIA+ model (Dijkstra & Van Heuven, 2002), representations at both orthographic, semantic, and phonological levels contribute to bilingual word recognition. Studies on language production found that phonological representations of both languages were co-activated, highlighting the activation of phonology in the bilingual lexicon (Jared & Kroll, 2001; Jared & Szucs, 2002). However, in language comprehension, studies have mainly looked at the effect of representational overlap at the orthographic and semantic level in bilingual word recognition, while the role of representational overlap at the phonological level has often been overlooked. It is important to consider the role of phonology, as such overlap between languages is likely to affect the degree of cross-linguistic activation and, consequently, bilingual word recognition (Dijkstra et al., 1999).

Dijkstra et al. (1999) therefore investigated the role of representational overlap in word recognition by Dutch-English bilinguals. Stimuli were created by systematically manipulating the degree of orthographic, semantic and phonological overlap between words in Dutch and English. This resulted in six categories: SOP, SO, SP, OP, O and P, where S refers to similar semantics, O to similar orthography and P to similar phonology. Categories SOP (‘*hotel*’), SO (‘*fruit*’) and SP (‘*news*’ – ‘*nieuws*’) were classified as cognates, and categories OP (‘*step*’), O (‘*stage*’) and P (‘*leaf*’ – ‘*lief*’) as interlingual homographs and homophones. The stimuli were presented in two experiments. In Experiment 1, participants performed a progressive demasking task, in which the presentation of a target word was alternated with a mask and the presentation time of the target word increased while that of the mask decreased progressively. It was the participants’ task to press a button once they identified the target word. In Experiment 2, participants performed an English lexical decision task. Results in both tasks showed that words with orthographic and semantic overlap were recognised faster than control words, whereas words with phonological overlap showed interference compared to control words. The results of this study thus suggest that phonological overlap affects bilingual word recognition in addition to orthographic and semantic overlap by exhibiting an inhibitory effect.

Lemhöfer and Dijkstra (2004) further explored the role of representational overlap in word recognition by Dutch-English bilinguals. In contrast to Dijkstra et al. (1999), they examined the three different categories of cognates (SOP, SO, SP) and interlingual homographs and homophones (OP, O, P) separately, in both English lexical decision tasks and generalised lexical decision tasks. In Experiment 1 participants performed an English lexical decision task involving interlingual homographs and in Experiment 2 they performed an English lexical decision involving cognates. Results showed that responses to words in the O condition were significantly faster than words in the OP condition. There was no significant difference in response times between SOP and SO words. In Experiment 3 participants performed a generalised lexical decision task involving interlingual homographs, and in Experiment 4 they performed a generalised lexical decision task involving cognates. There were no significant differences between the OP and O conditions and the SOP and SO conditions, demonstrating no effect of phonological overlap on bilingual word recognition in generalised lexical decision. This contrasts with the findings of Dijkstra et al.'s (1999) study, which did find an inhibitory effect of phonology. However, when looking at the stimuli in the OP condition, one will notice that many of these words did not fully overlap phonologically (e.g. '*star*', /stɑ:/, /stɑr/; '*rover*', /rəʊvə/, /ro:vər/; '*stout*', /staʊt/, /stɑut/), while Lemhöfer and Dijkstra mentioned that for these words, there was "almost complete phonological overlap" (p. 538). This could potentially account for the divergent results in the two studies, highlighting that the degree of phonological overlap may play a crucial role in bilingual word recognition.

More recent studies have investigated how varying degrees of phonological overlap influence bilingual word recognition, primarily by focusing on cognates. For instance, Arana et al. (2022) examined whether the presence of orthographically identical cognates contributed to the observation of a cognate facilitation effect. Portuguese-English bilinguals participated in four English lexical decision tasks, each containing a different proportion of identical cognates relative to non-identical cognates, with ratios of: 50-50; 25-75; 12-88; and 0-100, respectively. The first two experiments with identical cognate ratios of 50-50 and 25-75 showed a cognate facilitation effect that was caused by identical cognates, although the effect was smaller when the ratio was 25-75. There was no cognate facilitation effect in the last two experiments with identical cognate ratios of 12-88 and 0-100. This study thus revealed that the cognate facilitation effect diminished progressively and eventually vanished as the proportion of identical cognates decreased, suggesting that identical cognates play a central role in driving the cognate facilitation effect. Although Arana et al. aimed to examine how identical cognates influenced the cognate facilitation effect, the so-called identical cognates were not phonologically identical and did not differ significantly from the non-identical cognates. For example, identical cognates such as '*regime*' differ phonologically between English and Portuguese (/reɪˈʒiːm/ vs /ʁɨˈʒi.mɨ/), as do non-identical cognates like '*regular*' (/ˈrɛg.jʊ.lər/ vs /ʁeguˈlaɾ/). As phonological overlap did not seem to be a distinguishing factor between the two cognate types, no conclusions can be drawn regarding the role of phonological overlap in bilingual word recognition based on these findings.

In contrast, Frances et al. (2021) were able to distinguish between varying degrees of phonological overlap in cognates in order to investigate its role in bilingual word recognition. In their study, Spanish-English bilinguals completed both visual and auditory lexical decision tasks in their L2 (English) using cognates that varied in orthographic and phonological overlap. Participants were presented with cognates that varied systematically in orthographic and phonological overlap, with both factors fully crossed (high and low), as well as perfect cognates, which were defined as having identical orthography and high phonological overlap. While the study addressed effects across modalities, of particular relevance here are the findings on the degree of phonological overlap in the visual modality. In the visual lexical decision task, high orthographic overlap led to faster responses while high phonological overlap led to slower responses. Perfect cognates showed larger facilitation effects in the visual modality due to the identical orthography of these cognates. These results support the inhibitory role of phonological overlap found by Dijkstra et al. (1999) and potentially clarifies why Lemhöfer and Dijkstra (2004) might not have found such effects, as a high degree of phonological overlap appears to be required.

Tiffin-Richards (2024) followed up on this line of research by investigating the cognate facilitation effect directly in relation to the degree of orthographic and phonological overlap, not only in the L2 but also in the L1 of the participants. This study examined German-English bilinguals' recognition of three cognate types: identical cognates, close-cognates, and non-cognates. Identical cognates had significantly higher orthographic and phonological overlap than close cognates, which in turn had higher orthographic and phonological overlap than non-cognates. Importantly, not all identical cognates were phonologically identical, though they were all orthographically identical. Participants completed lexical decision tasks in their L1 (German) and their L2 (English). A cognate facilitation effect was found only in the L2: both identical and close cognates were recognised faster than non-cognates, but there was no significant difference between identical and close cognates. Examination of the effects of orthographic and phonological overlap revealed that, in both L1 and L2, higher orthographic overlap led to faster responses, while higher phonological overlap led to slower responses. This further emphasises the inhibitory role of phonological overlap (Dijkstra et al., 1999) and that not just the presence, but the degree of phonological overlap is crucial.

Although the studies by Frances et al. (2021) and Tiffin-Richards (2024) examined the effect of varying degrees of phonological overlap, they did not specifically focus on full phonological overlap. Moreover, their studies were limited to cognates, leaving the effect of varying degrees of phonological overlap in interlingual homographs unclear, even though the interest in phonological overlap originated from the interlingual homograph stimuli used in Lemhöfer and Dijkstra's (2004) study. Kerkemeijer (2024) therefore introduced a distinction between identical and non-identical phonology and applied this to interlingual homographs. Dutch-English participants took part in a lexical decision task in their L1 Dutch, in which phonologically identical interlingual homographs, phonologically non-identical interlingual homographs, Dutch control words, and pseudowords were presented. Both phonologically identical and non-identical interlingual homographs showed slower response times, thus interference, compared to Dutch control words. When comparing the two types of interlingual homographs, it was found that phonologically non-identical interlingual homographs showed slower responses compared to phonologically identical interlingual homographs. These findings suggest that phonologically identical interlingual homographs elicit reduced interference compared to phonologically non-identical interlingual homographs. This is consistent with the interference effects for interlingual homographs with phonological overlap observed by Dijkstra et al. (1999), and the inhibitory role of high phonological overlap in cognates observed by Frances et al. (2021) and Tiffin-Richards (2024). It extended their work by demonstrating that full phonological overlap, as in phonologically identical interlingual homographs, results in reduced interference, potentially indicating a facilitatory effect.

Taken together, the studies discussed suggest that phonological overlap can have an inhibitory effect on bilingual word recognition, particularly when overlap is substantial but non-identical (Dijkstra et al., 1999; Frances et al., 2021; Tiffin-Richards, 2024). This interference is likely due to increased cross-linguistic activation of two different phonological representations, which requires stronger inhibition of the non-target phonological representation, thereby slowing down recognition. In contrast, when phonological overlap is fully identical, this inhibitory effect attenuates, indicating a facilitatory role. According to the BIA+ model (Dijkstra & Van Heuven, 2002), this can be explained by parallel activation of phonological representations in both languages, which leads to stronger overall activation and faster access to semantic representations, thereby speeding up recognition.

1.2.4 Bilingual word recognition in sentence context

The abovementioned studies investigated the recognition of cognates and interlingual homographs and the effect of phonological overlap in experiments in which these words were presented to participants in isolation. However, people rarely read words in isolation, as words are usually embedded in meaningful sentences. To draw more ecologically valid conclusions about bilingual word recognition, research on bilingual word recognition in sentence context has been carried out (see Van Assche et al., 2012, for a review).

Research into the processing of cognates in sentence context has been done by Duyck et al. (2007). They conducted two experiments involving word recognition in L2 sentences by Dutch-English bilinguals. In their first experiment, participants performed an English lexical decision task with cognates or English control words as final words of low semantic constraint sentences. In these low constraint sentences, the cognates and control words are not predictable from the sentence context. Sentences were presented through serial visual presentation, in which each word of the sentence was shown one by one. They found that cognates were recognised significantly faster than control words. In their second experiment, participants read English sentences with cognates or English control words in mid-sentence position in low semantic constraint sentences while their eye movements were recorded (i.e. an eye-tracking paradigm). Now, they found that only orthographically identical cognates showed a cognate facilitation effect in early reading time measures.

Libben and Titone (2009) further extended Duyck et al.'s (2007) study by investigating the recognition of interlingual homographs, in addition to cognates, in L2 sentence context. In the study by Libben and Titone, French-English bilinguals were presented with both low and high semantic constraint sentences for the target-language meanings containing interlingual homographs, cognates, and English control words. The experiment used an eye-tracking paradigm, measuring participants' eye movements as they read the sentences. For the low semantic constraint sentences, it was found that cognates showed significantly shorter reading times, i.e. facilitation, compared to English control words, and interlingual homographs showed significantly longer reading times, i.e. interference, compared to English control words. This applied to both early and late stage reading measures. For the high semantic constraint sentences, however, it was found that cognates only showed facilitation and interlingual homographs only showed interference in early reading time measures.

Hoversten and Traxler (2016) similarly examined the recognition of interlingual homographs in L2 sentences using an eye-tracking paradigm. Spanish-English bilinguals were presented with English sentences that contained interlingual homographs. The sentence context, however, biased either the English or the Spanish meaning of the interlingual homograph. For example, in the sentence '*While eating dessert, the diner crushed his pie accidentally with his elbow*' the English meaning of the interlingual homograph is supported, which is congruent with the sentence context. In contrast, in the sentence '*While carrying bricks, the mason crushed his pie accidentally with the load*' the Spanish meaning of the interlingual homograph ('*foot*') is supported, which is incongruent with the sentence context. It was found that interlingual homographs did not immediately show an interference effect. In fact, interference was only observed on the words after the interlingual homographs, the so-called post-target region, when the English meaning was incongruent with the sentence context. The studies by Duyck et al. (2007), Libben and Titone (2009) and Hoversten and Traxler (2016) thus showed that cognate facilitation and interlingual homograph interference are modulated in L2 sentence context compared to isolated word recognition. Cognate and interlingual homograph effects are smaller, influenced by semantic constraint, delayed, or not present. These findings suggest that cross-linguistic activation is still present in sentence context, but that its effects are more subtle.

Bilingual word recognition in sentence context has not only been investigated in L2 but also in L1 sentences. Van Assche et al. (2009), for instance, examined Dutch-English bilinguals' processing of cognates in sentences in their L1. Participants read Dutch sentences of low semantic constraint while their eye movements were recorded in an eye-tracking paradigm. The sentences contained a cognate or a Dutch control word. It was found that cognates were read significantly faster than Dutch control words, showing facilitation. However, this effect was only found in early reading time measures.

Titone et al. (2011) in parallel investigated bilingual word recognition in L1 sentence context. Orthographically identical cognates and interlingual homographs were used in low and high semantic constraint sentences that were presented to English-French bilinguals in an eye-tracking paradigm. Cognate facilitation was found for early reading time measures in both low and high semantic constraint sentences, but the size of the effect was larger in low semantic constraint sentences than in high semantic constraint sentences. In a second experiment, L2 sentences were included as filler sentences to make the L2 more salient to increase cognate facilitation and interlingual homograph interference. Now, cognate facilitation occurred in both early and late reading time measures. In both experiments, there was interlingual homograph interference for total reading times only, although the size of the effect was much smaller than in the L2 (Libben & Titone, 2009).

Van Assche et al. (2009) and Titone et al. (2009) thus demonstrated that in L1 sentence context, cognate facilitation and interlingual homograph interference are even more subtle than in L2. Cognate facilitation was limited to early reading time measures and was more strongly influenced by semantic constraint. In addition, interlingual homograph interference was only observed in total reading times and the size of this effect was reduced compared to the size of this effect L2.

1.2.5 Sentence reading methodology

Most of the studies on sentence context discussed above used an eye-tracking paradigm to investigate bilingual word recognition. Another method to investigate sentence reading is a self-paced reading paradigm (Jegerski, 2013). In this paradigm, participants can read through a sentence word by word themselves by pressing a button. There are different variants of presenting sentences in a self-paced reading paradigm. In a cumulative display, each word or group of words of the sentence stays visible as the next word (group) appears, so that the sentence gradually builds up on the screen until, at the end, the whole sentence is visible. In contrast, a non-cumulative or moving-windows display only shows one word or group of words at a time, and each new word (group) replaces the previous one, which disappears again. The layout can also differ in positioning: in a centred display, each word or group of words appear in the centre of the screen, replacing the previous one. In a linear display, however, words or group of words are shown one after another from left to right, without overlapping, which is similar to natural reading.

However, when different methods are used, measures and potential effects will also manifest differently (Duyck et al., 2007). The effects of eye-tracking and self-paced reading can be compared using Bultena et al.'s (2014) study with Dutch-English bilinguals. They conducted both a self-paced reading and an eye-tracking study using L2 sentences that contained cognates or non-cognates. Results showed that the size of the cognate facilitation effect was modulated by task type, meaning that cognate facilitation was stronger in self-paced reading due to slower reading times compared to eye-tracking, allowing more time for a cognate effect to occur. These findings thus demonstrated that cognate facilitation can be observed across both methods, though the exact size of the effect may differ across task types.

1.3 The current study

In conclusion, bilingual word recognition is non-selective with respect to language, modelled in the BIA+ model, meaning that two languages can affect each other in processing (Dijkstra & Van Heuven, 2002). This interaction between languages can be investigated using words that contain cross-linguistic overlap, i.e. interlingual homographs and cognates, which is why they have often been used to investigate bilingual word recognition (e.g. Dijkstra et al., 1998). The BIA+ model suggests that both orthography, semantics, and phonology contribute to bilingual word recognition, but the role of phonology is not yet clear (e.g. Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004). Research on the role of phonology has been conducted in isolated word recognition studies (Frances et al., 2021; Kerkemeijer, 2024; Tiffin-Richards, 2024). However, it remains unclear how phonological overlap influences bilingual word recognition in sentence context, in which cross-linguistic activation has been observed in both L2 and L1 sentences, though often modulated or reduced (e.g. Hoversten & Traxler, 2016; Libben & Titone, 2009; Titone et al., 2011). This leads to the following research question:

How does full phonological overlap in interlingual homographs influence bilingual word recognition in L1 sentence context?

This research question will be investigated using a self-paced reading experiment (Jegerski, 2013), in which Dutch-English bilinguals read sentences in their L1 Dutch. Each sentence contains either an interlingual homograph or a Dutch control word. The interlingual homographs are phonologically identical or phonologically non-identical between Dutch and English. Reading times are measured for the interlingual homographs and control words, as well as for the word that followed. This is done because phonological and semantic representations are activated a later stage in processing, particularly for L2 representations (Dijkstra & Van Heuven, 2002). As a result, potential effects may appear with some delay (Hoversten & Traxler, 2016). The reading times are then compared to examine how full phonological overlap in interlingual homographs affects bilingual word recognition in L1 sentence context.

Interlingual homographs are expected to cause interference compared to control words (e.g. Dijkstra et al., 1998; Libben & Titone, 2009; Titone et al., 2011). Within this interference effect, it is expected that phonologically non-identical interlingual homographs will show stronger interference (Dijkstra et al., 1999; Frances et al., 2021; Tiffin-Richards, 2024) than phonologically identical interlingual homographs (Kerkemeijer, 2024). However, when this is combined with sentence contexts, where effects tend to be modulated or reduced compared to word recognition in isolation (Libben & Titone, 2009; Titone et al., 2011), it is expected that only phonologically non-identical interlingual homographs will show interference, while the interference from phonologically identical interlingual homographs will nullify.

2 Method

2.1 Participants

A total of 226 individuals attempted to participate in the self-paced reading experiment, however, 144 dropped out early. Upon inspection, this large number of dropouts appeared to be primarily due to individuals who did not give consent to participate. Although they did not proceed to the experiment, they were still recorded as participants simply by accessing the experiment link. Consequently, a total of 82 participants successfully participated in the self-paced reading experiment. Prerequisites for participating in this experiment were (a) Dutch (dialect) as the first language, English as the second language, and age of 16 years or older. Three participants indicated not to be a native speaker of (a) Dutch (dialect); they were removed from the dataset. The rest of this section reports information of the remaining 79 participants. These 79 Dutch-English bilinguals (58 female; 20 male, 1 non-binary) were aged between 17 and 79, with an average age of 32.91 (*SD* = 17.21). Of these participants, 74 participants reported Dutch as their first language, while five participants reported speaking a Dutch dialect. Participants were recruited through various social media platforms and through personal invitations.

Participants were asked to provide information about their experience with and proficiency in English to get an impression of their language background, inspired by the Language Experience and Proficiency Questionnaire (Marian et al., 2007) and the Language History Questionnaire (Li et al., 2020). However, this questionnaire was administered in Dutch. On average, the participants started acquiring English as a second language at 9.78 years old (*SD* = 2.48; min = 0; max = 14). They had an average of 23.13 years of experience in English (*SD* = 16.40). Participants were asked to self-assess their general proficiency in English on a scale of 0 to 100, where 0 refers to beginner and 100 to fluent (*M* = 73.76; *SD* = 20.41). In addition, they were asked to assess their proficiency in the skills reading, listening, writing, and speaking on a scale from 1 to 7, where 1 refers to not proficient and 7 refers to fully proficient. Table 1 shows the self-assessed proficiency scores in English in each of the four skills.

Table 1

Means and standard deviations of self-assessed proficiency scores in English.

	M	*SD*
Reading	5.95	1.09
Listening	5.71	1.23
Writing	5.30	1.27
Speaking	5.27	1.29

Note. Language proficiency indicated on a seven-point scale: 1 = not proficient; 7 = fully proficient.

Participants were also asked to specify whether they spoke German, French or Spanish, and up to a maximum of three other languages besides English at an intermediate, advanced or a fluent level. Eighty-five participants indicated to speak another language. Among these participants, German was mentioned by 50 participants; French was mentioned 22 times and Spanish was mentioned by three participants. Other languages mentioned were Danish (N = 4), Italian (N = 3), Frisian, Greek, Korean, Norwegian, Polish, Swedish, and Tajik (all N = 1).

2.2 Materials

Thirty-two Dutch-English interlingual homographs were selected from a list of form-identical words compiled by Sybrine Bultena, which largely matched the interlingual homographs used in the study of Kerkemeijer (2024). Of these 32 interlingual homographs, 16 were phonologically identical and 16 were phonologically non-identical between Dutch and English. Whether the interlingual homographs were phonologically identical or not was determined by comparing their IPA transcriptions in Dutch and English. Based on these transcriptions, two words from the identical interlingual homograph category (*pop* and *spot*) should strictly have fallen into the non-identical category. However, since these words were also used in Kerkemeijer's study and did not appear to affect the observed effect, and in order to allow for close comparison of the results, it was decided to retain these interlingual homographs in the identical category.

Interlingual homographs were selected so that their average word frequencies in Dutch and English matched, to control for potential effects caused by frequency differences between the two languages (see Table 2). Due to limited availability of interlingual homographs, specifically phonologically identical interlingual homographs, word frequencies in Dutch and English were matched across the entire set of interlingual homographs rather than within each interlingual homograph individually. Word frequencies were drawn from the SUBTLEX-DU database (Keuleers et al., 2010) for the Dutch word frequencies and from the SUBTLEX-EN database (Brysbaert et al., 2012) for the English word frequencies, and were expressed as the logarithm (log10) of the absolute word frequency. No *t*-tests were performed to ensure that the Dutch and English word frequencies were successfully matched. This is because null hypothesis significance testing does not allow for the acceptance of the null hypothesis (i.e. there is no difference between two groups); it only allows to reject or fail to reject the null hypothesis. As a result, it is not possible to conclude that the groups come from populations with the same mean. Therefore, only descriptive statistics were looked at during matching (Kover & Atwood, 2013). The means showed that the word frequencies in Dutch and English did not differ strongly across all interlingual homographs (Dutch: 2.90 vs English: 2.98). Further comparisons showed that the difference in word frequencies between Dutch and English for both the phonologically identical (Dutch: 2.79 vs English: 2.85) and the phonologically non-identical (Dutch: 3.01 vs English: 3.11) interlingual homographs was not particularly large either. However, comparison of word frequencies between phonologically identical and non-identical interlingual homographs showed a larger difference: phonologically identical interlingual homographs had somewhat lower word frequencies in both Dutch and English compared to phonologically non-identical interlingual homographs (identical: 2.79 and 2.85 vs non-identical: 3.01 and 3.11). In addition to word frequency, phonologically identical and non-identical interlingual homographs were matched on word length. Interlingual homographs were between three and six letters long, and word length was comparable between phonologically identical and non-identical interlingual homographs (identical: 4.00 vs non-identical: 4.13).

Table 2

Means and standard deviations (in brackets) of word length and word frequency.

Word type	Overlap	Word length	Word frequency	
			Dutch	English
Interlingual homograph		4.06 (0.80)	2.90 (0.72)	2.98 (0.63)
	Identical	4.00 (0.82)	2.79 (0.70)	2.85 (0.64)
	Non-identical	4.13 (0.81)	3.01 (0.75)	3.11 (0.62)
Control word		4.06 (0.80)	2.89 (0.71)	
	Identical	4.00 (0.82)	2.81 (0.69)	
	Non-identical	4.13 (0.81)	2.98 (0.75)	

Each interlingual homograph was then individually matched to a monolingual Dutch control word based on their word frequency in Dutch and their word length, in order to control for potential effects of these factors (see Table 2). To achieve this, the SUBTLEX-DU database (Keuleers et al., 2010) was searched using functions of the LexOPS package (Taylor et al., 2020) in the software RStudio (R Core Team, 2025). For each interlingual homograph, a set of words was generated with the exact same word length, word class, and a word frequency within a range of -0.25 and +0.25 of that of the interlingual homograph to obtain a reasonable number of words to choose from. Although word class was not a matching criterion in the same way as word length and word frequency, it was nonetheless included in the LexOPS search function, as the interlingual homograph and Dutch control word needed to be able to fit in the same sentence position. From the generated set of control words, the word that best fitted the same sentence context as the interlingual homograph was chosen. Comparisons of the means revealed that the Dutch word frequencies of interlingual homographs and control words were highly comparable (2.90 vs 2.89). This applied to both phonologically identical (2.79 vs 2.81) and phonologically non-identical (3.01 vs 2.98) interlingual homographs and their control words. The list of interlingual homographs and their matched Dutch control words can be found in Appendix 1.

For each interlingual homograph, a Dutch sentence was constructed, resulting in 16 sentences containing a phonologically identical homograph and 16 containing a phonologically non-identical interlingual homograph. A corresponding version of each sentence was created by replacing the interlingual homograph with their matched Dutch control word, ensuring that the two versions of each sentence were identical except for the target word (i.e. the interlingual homograph and control word), see Table 3, for examples of experimental sentences. The sentences were intended to be of low semantic constraint; however, as this was not empirically tested in a separate study, the actual degree of semantic constraint was not established. Sentences were between nine and 14 words long (*M* = 11.19, *SD* = 1.31) and the target word was placed between the third and the 11^{th} sentence position (*M* = 7.00, *SD* = 2.21). The initial sentence position was excluded to minimise any potential findings that are related to sentence reading onset. The final sentence position was excluded because reading times were measured not only for the target word but also for the word immediately following the target word (i.e. the post-target word). To ensure that any effects on the post-target word could emerge, this word was required to consist of at least three letters, matching the minimum length of the target words (*M* = 5.63, *SD* = 2.39). Comprehension questions were constructed for 16 of the 32 sentences: eight questions corresponding to sentences in the phonologically identical condition and eight to those in the phonologically non-identical condition. Within each condition, half of the questions were paired with the interlingual homograph version and the other half with the control version of the sentence. To eight of the questions, the answer was '*yes*' and to the other eight questions the answer was '*no*'. The list of sentences and comprehension questions can be found in Appendix 2.

Table 3

Examples of sentences and comprehension questions across conditions.

Overlap	Sentence with **interlingual homograph/control word** and question
Identical	De val had veel schade aan haar **PINK/ADER** veroorzaakt.
	Q: Had ze schade aan haar duim? (no)
Non-identical	Martin vond oude **DOZEN/LADEN** tussen de troep in de kelder.
	Q: Vond Martin dozen/laden in de kelder? (yes)

In addition to the 32 experimental sentences, 32 filler sentences were selected from Doornbos' (2022) study. These were Dutch sentences containing an English loanword. Some sentences were slightly modified to match the length of the experimental sentences. Filler sentences were included for three reasons. First, they served to distract participants from the actual manipulation of the experiment. Second, the presence of English loanwords was intended to increase the salience of English for the participants, thereby potentially increasing the likelihood of an effect on the interlingual homographs in the experimental sentences.

Third, they served to reduce the impact of potential distractions caused by the comprehension questions, which followed directly after the related experimental sentence. The number of filler sentences was set at 32 to ensure a fifty-fifty ratio between experimental and filler sentences, a number that is considered the minimum acceptable amount of filler sentences (Havik et al., 2009). This ratio allowed a balanced alternation between experimental and filler sentences. Ideally, more filler sentences would have been included. However, this would have lengthened the experiment, something that was considered undesirable for the participants, who voluntarily participated in the experiment.

Two stimulus lists were created of all 32 experimental sentences, 32 filler sentences and 16 comprehension questions. The interlingual homograph and control versions of the sentences were counterbalanced across the two lists. Each list thus contained 16 interlingual homograph sentences and 16 control sentences, half of which belonged to the phonologically identical condition and the other half to the phonologically non-identical condition. The order of the two stimulus lists was randomised using the software Mix (Van Casteren & Davies, 2006) to ensure that any potential findings did not arise from the order in which the sentences were presented. The randomisation took into account that each experimental sentence was followed by a filler sentence. As a result of the randomisation, the comprehension questions occurred at random intervals. The two resulting stimulus lists were uploaded to Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), and these were presented randomly and balanced among participants; 41 participants saw list one and 41 participants saw list two. The three participants who were removed from the dataset all saw list one, resulting in a final distribution of 38 participants in list one and 41 in list two.

2.3 Procedure

Participants completed the experiment online in an uncontrolled environment via the Gorilla platform (Anwyl-Irvine et al., 2020), which could only be accessed on a computer. After clicking the link to the experiment, participants were first shown an information letter in Dutch and asked to provide explicit informed consent to participate in the experiment. By doing so, they confirmed their voluntary participation, declared they were at least 16 years old and acknowledged how their data would be used. This procedure ensured compliance with the ethical guidelines for linguistic research set by the Ethics Assessment Committee Humanities of Radboud University, see Appendix 3. Only participants who gave consent proceeded to the next stage, which contained a demographic and language background questionnaire in Dutch.

This gathered information on age, gender, age of English acquisition, and English proficiency. Following the questionnaire, participants received instructions for the self-paced reading task. They were told to read at their own pace by pressing the spacebar to reveal the sentence one word at a time, answer comprehension questions and minimise distractions during the task. After two practice trials and one practice comprehension question, the main task began, consisting of 64 sentences and 16 comprehension questions.

In the non-cumulative linear self-paced reading task, sentences were presented word-by-word in the vertical centre of the screen, using a 24-point monospaced black font on a white background. The use of a monospaced font ensured equal character width, so that words with the same number of letters appeared equally wide. Most, if not all, sentences fitted on a single line, depending on the screen size of the device on which the experiment was performed. Each trial began with a fixation cross shown for 1000 milliseconds (ms), followed by a 500ms pause, after which the first word of the sentence appeared. Pressing the spacebar revealed the next word while the previous word disappeared, with words presented from left to right. A minimum display time was set at 50ms, but there was no maximum display time, allowing participants to read at their own pace. By only presenting one word at a time, a normal reading pace was maintained as distractions from previous words in the sentence were avoided.

2.4 Design & analysis

The experiment consisted of a 2 x 2 within-subjects design with the independent variables being '*Word Type'* (interlingual homograph vs control word) and '*Overlap*' (identical vs non-identical). This resulted in four experimental conditions: identical interlingual homographs, non-identical interlingual homographs, control words that were matched with the identical interlingual homographs, and control words that were matched with the non-identical interlingual homographs. It is important to stress that control words themselves were not phonologically identical or non-identical between Dutch and English; rather, they served as controls for phonologically identical or non-identical interlingual homographs. The dependent variables were the reading times (RTs) of both the target word and the post-target word. As these were continuous dependent variables, a linear mixed effects model was used for statistical analysis. This type of model is advantageous because it can account for variability at both the participant and item level, reflecting the fact that the samples are drawn from a larger population of potential participants and items.

Prior to statistical analysis, the data were processed using RStudio (R Core Team, 2025) and the ‘tidyverse’ package (Wickham et al., 2019) to prepare them for analysis. Since the dataset provided only cumulative RTs per word within each sentence, individual RTs for the target and post-target words were derived by subtracting the cumulative RT of the following word from that of the current word. Accuracy on comprehension questions was calculated, ranging from 75% and 100% (M = 95.03, SD = 6.00), corresponding to 12 to 16 correct answers per participant. No participants were excluded based on their comprehension performance. Boxplot inspection of the target word RTs revealed a considerable number of outliers at the upper end of the distribution. Although Jegerski (2013) recommended removing only RTs above 3000ms, outlier removal was considered necessary based on the observed skewness and to improve the normality of the data. Two outlier removal methods were therefore explored. First, participant-specific cutoffs were applied, removing RTs exceeding 2.5 standard deviation above each participant’s mean. This method did not sufficiently reduce the number of extreme values, as the distribution remained skewed. Consequently, this method was ultimately not applied. Second, an outlier removal procedure based on the interquartile range (IQR) was performed on all datapoints, removing RTs 1.5 times the IQR above the third quartile, corresponding to a cutoff of 706.75ms. This removed 78 trials (3% of the data) and resulted in a more normally distributed dataset, further improved by a log transformation. For this reason, this method of outlier removal was ultimately selected. A similar inspection of post-target word RTs again revealed extreme values at the upper end of the distribution. However, since these words were not matched on word length and word frequency, applying the same outlier removal method was considered inappropriate. Instead, only post-target RTs corresponding to excluded target RTs were removed. This approach, combined with log transformation of RTs, also led to improved normality. As a result, log transformed RTs were used for both target and post-target RTs in the statistical analysis.

Linear mixed effects regression analyses were conducted in RStudio (R Core Team, 2025), using the ‘lme4’ package (Bates et al., 2015). Two separate models were built: one for target RTs and one for post-target RTs. For the target RTs, the model included fixed effects for Word Type and Overlap (both sum coded), as well as their interaction, along with random intercepts for participants and items. Another coding scheme, Helmert coding, was considered as an alternative to sum coding, using a single three-level categorical variable (control words, identical interlingual homographs, non-identical interlingual homographs). Helmert coding compares each level to the mean of the subsequent levels, thus comparing control words to interlingual homographs and comparing identical interlingual homographs to non-identical interlingual homographs. However, this was rejected, as it compares identical and non-identical interlingual homographs directly, which was undesirable given their slight difference in word frequency. Dutch word frequency (centred) was added to the model as a fixed effect to control for the small differences in word frequency. Additional co-variate fixed effects were evaluated stepwise using likelihood ratio tests (ANOVA), with model fit assessed via the AIC-value. For English proficiency, both overall proficiency and reading proficiency improved model fit, however, the improvement for reading proficiency (centred) was greater, resulting in its inclusion as a fixed effect. Random slopes were added following a principled approach between maximal structures (Barr et al., 2013) and only random intercepts (Baayen et al., 2008). Slopes were included only if they improved model fit without causing convergence errors. Only a slope for word type by item met these criteria, justified by the design where each item appeared in two word type conditions. The same procedure was applied to the post-target RT model, with one key difference: word length (centred) was added as a fixed effect instead of Dutch word frequency. As with the target model, reading proficiency and the random slope for word type by item improved model fit. The model fit of the final models, which met the assumptions of normality, homogeneity, and multicollinearity, was calculated using the 'MuMIn' package (Bartoń, 2022). The analysis code can be found in Appendix 4.

3 Results

After data processing, the final dataset consisted of 2450 trials for target words and for post-target words. This reflects the number of RTs of 32 (post-)target words across 79 participants, with 78 trials removed. Descriptive statistics for the mean RTs and standard deviations across conditions are presented in Table 4. The mean RTs are also visualised in Figure 4 for target words and in Figure 5 for post-target words.

Table 4

Means and standard deviations (in brackets) of target and post-target RTs in ms.

	Identical		Non-identical	
	IH	Control	IH	Control
Target	361 (123)	367 (121)	358 (118)	361 (116)
Post-target	396 (164)	396 (161)	392 (170)	385 (162)

Note. IH stands for interlingual homograph.

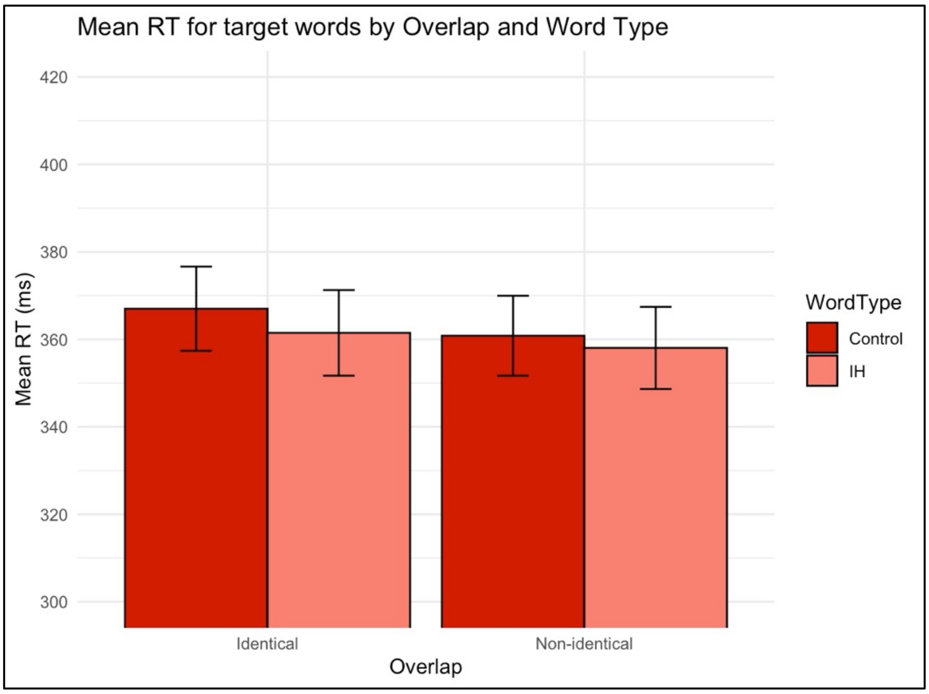

Figure 4. *Visualisation of mean RTs for target words.*

Note. Error bars represent 95% confidence intervals.

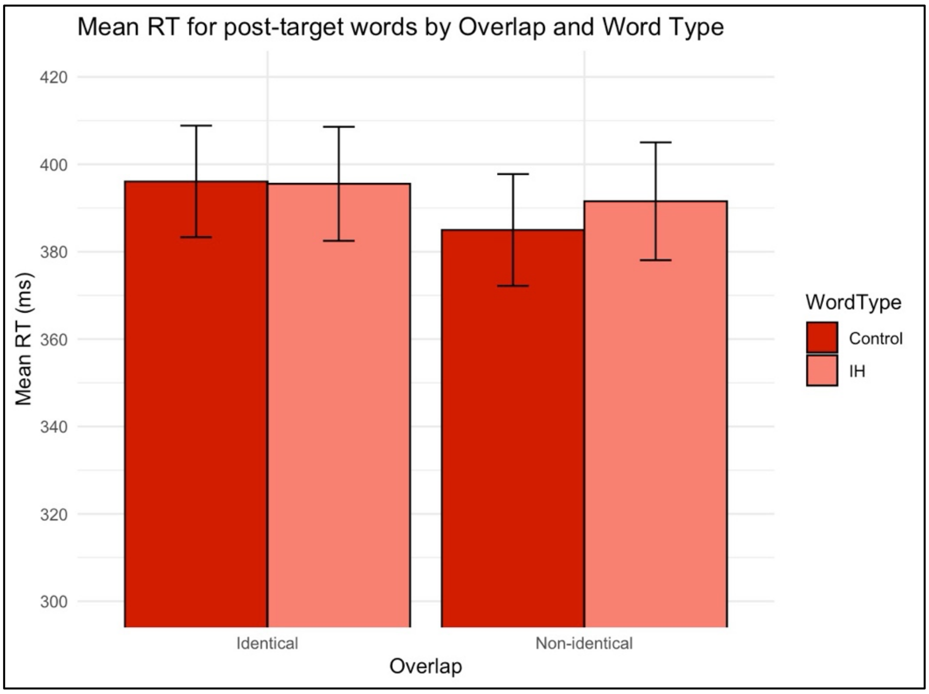

Figure 5. *Visualisation of mean RTs for post-target words.*

Note. Error bars represent 95% confidence intervals.

3.1 Target word analysis

To test the effect of full phonological overlap on bilingual word recognition, two linear mixed effects models were used. The first model had the log transformed RTs of target words as dependent variable and the following predictors as fixed effects: Overlap and Word Type (both sum coded) in interaction, and Dutch word frequency and English reading proficiency (both centred). In addition, the model included random intercepts for participants and items, as well as a slope of Word Type by items. The model fit of the full model (conditional R-squared) was 0.72 and the part of the fixed effects only (marginal R-squared) was 0.14. The intercept corresponds to the predicted log transformed RT for a target word at the mean of the two levels of Overlap and Word Type, at average Dutch word frequency and at average English proficiency. The model outcomes can be found in Table 5.

Table 5

The estimates of the linear regression model[1] for log transformed target RTs.

Coefficient	β	*SE*(β)	*df*	*t*	*p*
Intercept	2.538	0.013	93.997	195.105	<.001
Overlap[2]	-0.002	0.005	31.234	0.507	.616
WordType[3]	-0.003	0.005	30.965	0.591	.559
Freq_DU[4]	-0.002	0.006	31.278	-0.363	.719
ReadingProficiency[4]	-0.052	0.011	78.704	-4.649	<.001
Overlap:WordType	0.001	0.005	30.981	0.248	.806

[1]Formula: RT_Target_Log ~ Overlap * WordType + Freq_DU_c + Reading_c + (1|ParticipantID) + (1+WordType|Item)

[2]Overlap: -1 = Identical, 1 = Non-identical.

[3]WordType: -1 = Control word, 1 = Interlingual homograph.

[4]Centred.

The results of the target word model showed no statistically significant main effect of the predictor Word Type, indicating that interlingual homographs were not read significantly faster than control words. Although the main effect of Word Type was non-significant, estimated marginal means (emmeans) contrasts were conducted to inspect the effect of this predictor across the levels of the other predictor, with Overlap being the other predictor in this case. This showed that non-identical interlingual homographs were not read significantly faster than non-identical control words (β = -0.003, *SE* = 0.013, *df* = 33.2, *t*-ratio = 0.235, *p* =

.995), and that identical interlingual homographs were not read significantly faster than identical control words (β = -0.008, *SE* = 0.014, *df* = 33.2, *t*-ratio = 0.573, *p* = .939). For the predictor Overlap, the model showed no statistically significant main effect: non-identical words were not read significantly faster than identical words. Emmeans contrasts revealed that non-identical interlingual homographs were not read significantly faster than identical interlingual homographs (β = -0.002, *SE* = 0.014, *df* = 33.2, *t*-ratio = 0.172, *p* = .998), and that non-identical control words were not read significantly faster than identical control words (β = -0.007, *SE* = 0.013, *df* = 33.1, *t*-ratio = 0.526, *p* = .952). Dutch word frequency did not significantly predict RTs, suggesting that words with higher word frequency in Dutch were not read faster. In contrast, English reading proficiency significantly predicted RTs on a Dutch target word: higher proficiency was associated with significantly faster RTs. Finally, the interaction between Overlap and Word Type was not significant, indicating that the effect of Word Type did not vary as a function of Overlap.

3.2 Post-target word analysis

The second linear mixed effects model had the log transformed RTs of post-target words as dependent variable and the following predictors as fixed effects: Overlap and Word Type (both sum coded) in interaction, and word length and English proficiency (both centred). The model similarly included random intercepts for participants and items, as well as a slope of Word Type by items. The model fit of the full model (conditional R-squared) is 0.64 and the part of the fixed effects only (marginal R-squared) is 0.12. The intercept corresponds to the predicted log transformed RT for a post-target word at the mean of the two levels of Overlap and Word Type, at average word length and at average English proficiency. The model outcomes can be found in Table 6.

Table 6

The estimates of the linear regression model[1] for log transformed post-target RTs.

Coefficient	β	*SE*(β)	*df*	*t*	*p*
Intercept	2.566	0.014	97.547	181.320	<.001
Overlap[2]	-0.003	0.006	31.827	0.438	.664
WordType[3]	0.0002	0.005	31.511	-0.037	.971
WordLength_PostTarget[4]	0.004	0.002	30.764	1.908	.066
ReadingProficiency[4]	-0.053	0.012	78.691	-4.411	<.001
Overlap:WordType	0.002	0.005	31.517	0.343	.734

[1]Formula: RT_PostTarget_Log ~ Overlap * WordType + WordLength_PostTarget_c + Reading_c + (1|ParticipantID) + (1+WordType|Item)

[2]Overlap: -1 = Identical, 1 = Non-identical.

[3]WordType: -1 = Control word, 1 = Interlingual homograph.

[4]Centred.

The results of the post-target word model likewise revealed no statistically significant main effect of the predictor Word Type, suggesting that interlingual homographs were not read significantly slower than control words. Emmeans contrasts showed that non-identical interlingual homographs were not read significantly slower than non-identical control words (β = 0.004, *SE* = 0.016, *df* = 33.2, *t*-ratio = -0.260, *p* = .993), and that identical interlingual homographs were not read significantly faster than identical control words (β =-0.003, *SE* = 0.016, *df* = 33.3, *t*-ratio = 0.209, *p* = .997). Similarly, for the predictor Overlap, the model showed no statistically significant main effect: non-identical words were not read significantly faster than identical words. Emmeans contrasts revealed that non-identical interlingual homographs were not read significantly faster than identical interlingual homographs (β = -0.001, *SE* = 0.020, *df* = 33.7, *t*-ratio = 0.067, *p* = .999), and that non-identical control words were not read significantly faster than identical control words (β = -0.008, *SE* = 0.011, *df* = 32.9, *t*-ratio = 0.775, *p* = .865). The predictor Word Length did not reach significance, though this non-significant trend suggested that longer words may be associated with slightly slower RTs. As with the target word model, English reading proficiency significantly predicted RTs on a Dutch post-target word: higher proficiency was linked to significantly faster RTs. The interaction between Overlap and Word Type was not significant, indicating that the difference in RTs between control words and interlingual homographs was similar across the identical and non-identical conditions.

4 Discussion

The aim of the current study was to investigate how full phonological overlap in interlingual homographs influences bilingual word recognition in L1 sentence context. This was examined using a self-paced reading experiment in which Dutch-English bilinguals read Dutch sentences containing an interlingual homograph or a Dutch control word. The interlingual homographs were phonologically identical or phonologically non-identical between Dutch and English. Reading times were measured for both the target and the post-target word and analysed using linear mixed effects models. The results for both the target and post-target word showed no significant main effects of word type (interlingual homograph vs control word) or overlap (identical vs non-identical), nor a significant interaction between the two. This indicates that neither the reading times for interlingual homographs and control words, nor the difference in reading times between these word types across the identical and non-identical conditions, differed significantly. However, a significant main effect of English proficiency was found, with higher English proficiency associated with faster reading times.

The current findings do not seem to provide evidence that full phonological overlap in interlingual homographs influences bilingual word recognition in L1 sentence context. It was hypothesised that interlingual homographs would show interference, but that this interference effect would be attenuated for phonologically identical interlingual homographs compared to phonologically non-identical interlingual homographs. Consequently, in L1 sentence context, interference effects of phonologically non-identical interlingual homographs were expected to persist, whereas interference effects of phonologically identical interlingual homographs were expected to nullify. However, based on the current findings, this hypothesis cannot be confirmed. The current study demonstrated that no cross-linguistic activation emerged in L1 sentence context. This suggests that cross-linguistic activation does not occur in every context, indicating that there are limits to the contexts in which such activation can emerge. It therefore seems that the experimental context of the current study was not conducive to cross-linguistic activation, preventing interference effects from occurring. Although the hypothesis was based on previous literature reporting evidence for cross-linguistic activation, the absence of such effects in the current study was not entirely unexpected. This will be further explained in the remainder of this discussion, after which limitations of the current study will be considered and suggestions for future research will be offered.

4.1 Word type, overlap, and proficiency

The current study did not show any statistically significant main effects of word type in either the target or post-target word analyses, meaning that interlingual homographs and their control words did not differ significantly in reading times, leading to the conclusion that no cross-linguistic activation emerged. Previous studies, in contrast, did find effects of cross-linguistic activation in sentence context (e.g. Hoversten & Traxler, 2016; Libben & Titone, 2009; Titone et al., 2011). However, several differences between these studies and the current study may account for the absence of cross-linguistic activation, in the form of interlingual homograph interference, making the lack of a main effect of word type less surprising.

Libben and Titone (2009), for instance, found interlingual homograph interference in low-constraint sentences, which were the same type of sentences that were also used in the current study, in both early and late reading time measures. This interlingual homograph interference effect was modulated in high-constraint sentences, where interference had only been found in the early reading time measures, but sentences of this type were not present in the current study. A difference, however, is that participants in Libben and Titone's study read sentences in their L2, while participants in the current study read sentences in their L1. This difference may have led to a different degree of cross-linguistic activation in the experiments. Indeed, the L1 has a stronger influence on the L2 than the L2 has on the L1 (e.g. Duyck, 2005; Jared & Szucs, 2002), which can be explained by the higher resting activation of L1 representations due to higher frequency in that language (Dijkstra & Van Heuven, 2002). This explains why interlingual homograph interference was found in Libben and Titone's study and not in the current study; the L1 context of the current experiment may have led to faster activation of L1 representations compared to L2 representations, thereby limiting L2 influence and preventing cross-linguistic activation.

This explanation partially coincides with studies that have investigated the influence of the L2 on L1. These studies suggest that a minimum level of proficiency in the L2 is required in order for effects of the L2 to become noticeable in the L1 (e.g. Grainger et al., 2010; Van Hell & Dijkstra, 2002). It appears that bilinguals need a reasonable level of L2 proficiency before effects in the L1 can emerge. In the current study, participants' mean self-assessed overall proficiency was 73.76 on a scale from 0-100 and for reading proficiency this was 5.95 on a scale from 1-7, which seems to be a reasonable level of proficiency. However, there was some variation among participants in terms of proficiency; for instance, the standard deviation of 20.41 indicates that 68% of the participants rated their overall proficiency between 53.35 and 94.17. This variation could mean that the level in the L2 of the participants was not consistently high enough to observe an effect of the L2 on the L1, and thus cross-linguistic activation.

Hoversten and Traxler (2016), in addition, found interlingual homograph interference in L2 sentences in the region after the interlingual homograph in early reading time measures, but only when the meaning of the interlingual homograph was incongruent with the sentence context. This suggested cross-linguistic activation, as the correct meaning of the homograph was activated at a later stage in sentence reading. However, this was the only condition showing interlingual homograph interference, while this condition was not present in the current study. The sentences in which the meaning of the interlingual homograph was congruent with the sentence context, as was also the case in the current study, showed no interlingual homograph interference in early and late reading time measures at all. This is therefore consistent with what was found in the current study. This thus shows that interlingual homographs, even in L2 sentences, do not even consistently show an interference effect. Since these effects should be stronger than in L1 (Dijkstra & Van Heuven, 2002), and even these are not consistently present, it is not entirely unexpected that the current study using L1 sentences did not find interlingual homograph interference either.

Yet, Titone et al. (2011) found interlingual homograph interference in L1 low-constraint sentences, which are the same type of sentences used in the current study. This effect was only found for late reading time measures, particularly total reading times. Since the current study also included filler sentences in which the L2 was made salient by using English loanwords, that experiment was the most comparable to the current study. However, in both experiments performed by Titone et al. the interlingual homograph interference effect was reduced: only 22 milliseconds in the experiment without L2 filler sentences and 66 milliseconds in the experiment with L2 filler sentences. This is in contrast to the interlingual homograph interference effect observed in, for example, Libben and Titone's (2009) study, in which the interference effect on total reading times was 117 ms. The effects in the L1 thus appear to be very reduced, which shows that cross-linguistic activation is weaker in L1 sentence context, possibly also explaining why no interference was found in the current study.

Besides, there was another overarching difference between the studies of Libben and Titone (2009), Hoversten and Traxler (2016) and Titone et al. (2011) in comparison to the current study, namely that those studies used an eye-tracking paradigm to register reading times. With this method it is possible to record very subtle eye movements for specific measures. For example, Libben and Titone found interference effects both in the early and late reading time measures, Hoversten and Traxler found interference effects only in the early reading time measures, and Titone et al. found only an interference effect on total reading times. Unlike eye-tracking, self-paced reading does not provide detailed information on specific eye movements. Therefore, if the previously reported effects were limited to a particular eye-tracking measure, it may have gone undetected in the current self-paced reading experiment. However, Bultena et al. (2014) did find effects in both eye-tracking and self-paced reading measures. A difference between their study and the current study, however, is that they investigated cognates, whereas the current study focused on interlingual homographs. As such, these differences in experimental context could explain the discrepancy in findings between the current study and the studies discussed above.

In addition, an interaction with phonological overlap was examined within the broader effect of interlingual homograph interference. It was hypothesised that interlingual homographs with identical phonological overlap would not show interference in L1 sentence context, whereas interference was expected for interlingual homographs with non-identical phonological overlap. However, no significant interaction was found between word type and overlap for both the target and post-target words, suggesting that phonology did not influence word recognition at this point, contrary to what might be expected based on prior research.

Dijkstra et al. (1999), Frances et al. (2021) and Tiffin-Richards (2024) all found an inhibitory role of phonology, suggesting that phonological overlap which was not necessarily identical led to increased interference. However, several differences between these and the current study should be noted. First, in these three studies this result for phonological overlap was based on cognates, whereas the current study focused on interlingual homographs. Second, these studies examined the effect of phonological overlap using lexical decision tasks, while the current study used a self-paced reading experiment. Kerkemeijer (2024) did examine interlingual homographs and found reduced interference of phonologically identical interlingual homographs compared to phonologically non-identical interlingual homographs, but also using a lexical decision task. In contrast, the current study did not find a significant interaction in either the target or post-target analyses that would confirm these previous findings, which may be the result of limited cross-linguistic activation in L1 sentence context, as discussed above. Nonetheless, the reading time pattern on the post-target word did follow the predicted pattern: reading times for phonologically identical interlingual homographs were identical to their control words, whereas phonologically non-identical interlingual homographs showed numerically slower reading times than their control words. Although this interaction was not statistically significant, its direction is consistent with the prediction that no effect would emerge in L1 sentence context for phonologically identical interlingual homographs, whereas phonologically non-identical interlingual homographs were expected to show interference. The pattern in the data thus seems to correspond with the findings by Dijkstra et al. (1999), Frances et al. (2021) and Tiffin-Richards (2024), who found interference for phonological overlap that was not necessarily identical, as well as the findings by Kerkemeijer (2024), who found reduced interference for phonologically identical interlingual homographs. Given that interlingual homograph interference in L1 sentence context was limited, as discussed above, and that phonological overlap was examined within this already subtle effect, the absence of a statistically significant interaction was not unexpected.

Finally, the current study did reveal a significant main effect of proficiency in English, indicating that participants with higher proficiency in English showed faster reading times on both target and post-target words. The measure used for proficiency in English was participants' self-assessed reading proficiency. This indicates that reading both monolingual Dutch words and interlingual homographs was faster with higher proficiency in English. Previous studies did show that L2 knowledge, such as proficiency or age of acquisition, modulated the effects found (Titone et al., 2011; Van Hell & Dijkstra, 2002). However, this was not the case in the current study, as the addition of an interaction with proficiency did not improve the model fit. It is important to note that these previous studies did not specify whether higher L2 proficiency leads to faster L1 reading. A possible explanation could be that reading speeds in the L1 and L2 are correlated, in the sense that participants with higher L2 reading proficiency generally also read faster in Dutch. This is confirmed by a study that investigated correlations between reading speed in the L1 and the L2 (Baró, 2018). This study indeed showed that a faster reading speed in L1 also led to a faster reading speed in L2.

Although the current study did not find statistically significant effects, the alignment in the direction of the data with prior findings, especially for phonological overlap, suggests that the influence of full phonological overlap may still be present, but too subtle to be able to appear or detect in the current study as a result of limited cross-linguistic activation in L1 sentence context. In this light, the BIA+ model (Dijkstra & Van Heuven, 2002) remains a useful model for interpreting the data, and the absence of significance in the current study should not be taken as evidence against the model, as it can potentially explain the findings of the current study.

4.2 Limitations

Apart from the possibility that the experimental context in the current study limited cross-linguistic activation to such an extent that interlingual homograph interference effects and effects of phonological overlap could not emerge, several other limitations may have also influenced the outcomes of the current study.

The first limitation concerns the uncontrolled online testing environment. While this allowed for efficient participant recruitment by removing the need for lab attendance and thus increasing participation, it likely introduced substantial variability in the data. Participants completed the experiment at home, using different devices and setups, leading to differences in attention, external distractions, screen sizes and understanding of the task. Furthermore, feedback from some participants revealed that they misunderstood the instructions or started the task without reading them carefully, potentially resulting in inaccurate or inconsistent reading times. Such variability can be particularly problematic in self-paced reading tasks, where small timing differences are critical. This lack of standardisation of the testing environment could have introduced noise into the reading time data, which could potentially explain why certain data points had to be excluded from analysis. In addition, the length of the sentences was controlled for so that they fitted on one single line, however, this is dependent on the size of the screen on which the experiment is performed.

Another potential limitation of the current study relates to the semantic constraint of the experimental sentences. Prior studies have shown that cognate facilitation and interlingual homograph interference are most reliably observed in sentences of low semantic constraint, in which the context does not strongly predict the target word (e.g. Libben & Titone, 2009). In the current study, no empirical testing of sentence constraint was conducted (e.g. through a cloze probability test), due to time constraints. As a result, it is unclear whether the target words occurred in contexts that limited lexical competition or prediction. If some sentences were more semantically constraining than intended, this may have reduced the likelihood of cross-language activation, weakening potential interference effects and complicating direct comparisons with earlier findings.

The last potential limitation lies in the use of filler sentences containing English loanwords in the Dutch language. These fillers were included to increase the salience of English in the task, to potentially increase cross-language activation. However, these words, although of English origin, may have acquired strong and independent Dutch representations, including Dutch phonological representations, as they are common in Dutch and therefore do not activate any English representations. Examples of such English loanwords in Dutch include ‘*office*’, ‘*loser*’, ‘*spotlight*’, ‘*sweater*’, and ‘*issue*’. Consequently, the presence of such fillers might have failed to sufficiently prime the non-target language (English), leading to weaker cross-language activation, thereby reducing the probability of observing interference effects and effects of phonological overlap on the experimental trials.

4.3 Future research

In light of the findings of the current study and the abovementioned limitations, several suggestions for future research can be made. Future studies could replicate the current study using a more sensitive method such as eye-tracking. Unlike self-paced reading, eye-tracking allows for measuring early and late reading time measures separately. This distinction is crucial because previous research on interlingual homographs and cognates have shown that facilitation and interference effects often emerge only in early reading time measures and/or in specific late reading time measures (e.g. Titone et al., 2011), which cannot be extracted from self-paced reading. Additionally, eye-tracking can reveal whether effects occur on the interlingual homograph itself or on the post-target words. By replicating the current design using eye-tracking, it is possible to determine whether the absence of significant effects in this study reflects a true null result or a methodological limitation in measurement sensitivity.

While the current study focused on L1 sentences, future research could explore the role of phonological overlap in L2 sentence context. The current study focused on L1 Dutch reading, where English, the non-target language, may have been less strongly activated and effects potentially only emerge at a certain level of proficiency (Grainger et al., 2010; Van Hell & Dijkstra, 2002) and consequently effects were small (e.g. Libben & Titone, 2009; Titone et al., 2011). However, according to the BIA+ model, L1 representations are more dominant and accessed earlier than L2 ones due to their higher relative frequency. This suggests that in L2 reading, cross-language interference from the L1 may be more robust. Investigating the influence of full phonological overlap compared to non-identical phonological overlap in L2 sentence contexts, in which cross-linguistic activation is more likely to emerge, could reveal whether identical overlap strengthens or reduces interference.

Future research could manipulate the language context more directly by adding English filler sentences instead of Dutch sentences with English loanwords. This approach used by Titone et al. (2011) has shown that cross-language activation was higher and cognate facilitation and interlingual homograph interference was present in more measures. Including

English filler sentences may increase the salience of English, boosting cross-language activation and potentially the likelihood of effects to emerge. However, it should be carefully considered that there is a potential influence of language switching costs, which themselves can affect reading times. Solutions could be to structure the experiment in language-specific blocks, in which English and Dutch sentences are alternated in blocks to minimise switching costs or to use a predictable switching pattern to minimise these switching costs. By manipulating the language context more directly, future studies could better examine whether increasing non-target language activation modulates bilingual word recognition.

5 Conclusion

This study investigated how interlingual homographs with identical or non-identical phonological overlap influence bilingual word recognition in L1 sentence context. Using a self-paced reading paradigm, no significant evidence of interlingual homograph interference was found, nor of an interaction with phonological overlap. These results indicate that cross-linguistic activation may be subject to certain limitations. Although the results were not statistically significant, the pattern of data seemed to align with previous findings and predictions made by the BIA+ model (Dijkstra & Van Heuven, 2002). Subtle numerical trends in the direction of predicted interference suggest that cross-linguistic activation may still be present at a low level, but is strongly reduced under L1 conditions. Furthermore, the significant role of L2 proficiency in general reading speed shows the relatedness of L1 and L2 reading fluency. Overall, the findings show that cross-linguistic activation cannot just emerge in any situation, but that it depends on the context in which the bilingual is situated.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N.Z., & Evershed, J. K. (2020). Gorilla in our midst: an online behavioural experiment builder. *Behavior Research Methods*, *52*, 388-407. https://doi.org/10.3758/s13428-019-01237-x

Arana, S. L., Oliveira, H. M., Fernandes, A. I., Soares, A. P., & Comesaña, M. (2022). The cognate facilitation effect depends on the presence of identical cognates. *Bilingualism: Language and Cognition*, *25*(4), 660–678. https://doi.org/10.1017/s1366728922000062

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baró, À. L. (2018). Reading in English as a foreign language: Examining differences in reading speed, comprehension, efficacy and L1 cross-linguistic influence across grades. *Investigaciones Sobre Lectura*, 9, 1–49. https://doi.org/10.37132/isl.v0i9.226

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. https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2022). MuMIn: Multi-Model Inference. R package version 1.47.1, https://CRAN.R-project.org/package=MuMIn.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bultena, S., & Dijkstra, T. (2023). How bilinguals read words in their two languages (Lexical access in visual word recognition in second language processing). In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. John Wiley & Sons. https://doi.org/10.1002/9781405198431.wbeal0685.pub2

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. https://doi.org/10.1080/17470218.2013.853090

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*(4), 991–997. https://doi.org/10.3758/s13428-012-0190-4

Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496–518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., Timmermans, M., & Schriefers, H. (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. https://doi.org/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). Lawrence Erlbaum Associates Publishers.

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. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51–66. https://doi.org/10.1017/s1366728998000121

Dijkstra, A., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekke, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition, 22*(04), 657–679. https://doi.org/10.1017/s1366728918000287

Doorenbos, R.J. (2022). *Exploring the lexical processing of English loanwords during L1 sentence reading: Evidence from self-paced reading* [BA thesis]. Radboud University Nijmegen.

Duyck, W. (2005). Translation and associative priming with cross-lingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology Learning Memory and Cognition*, *31*(6), 1340–1359. https://doi.org/10.1037/0278-7393.31.6.1340

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

Frances, C., Navarra-Barindelli, E., & Martin, C. D. (2021). Inhibitory and facilitatory effects of phonological and orthographic similarity on L2 word recognition across modalities in bilinguals. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-92259-z

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology Learning Memory and Cognition*, *15*(2), 305–315. https://doi.org/10.1037/0278-7393.15.2.305

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology Section A*, *39*(2), 295–319. https://doi.org/10.1080/14640748708401788

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), Language acquisition across linguistic and cognitive systems (pp. 267–283). https://doi.org/10.1075/lald.52.18gra

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73–112. https://doi.org/10.1111/j.1467-9922.2009.00501.x

Hoversten, L. J., & Traxler, M. J. (2016). A time course analysis of interlingual homograph processing: Evidence from eye movements. *Bilingualism: Language and Cognition*, *19*(2), 347–360. https://doi.org/10.1017/s1366728915000115

Jared, D., & Szucs, C. (2002). Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism: Language and Cognition*, *5*(3), 225–239. https://doi.org/10.1017/s1366728902003024

Jegerski, J. (2013). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Kerkemeijer, A. (2024). *Phonological activation in bilinguals: Interlingual homographs in a native context* [BA thesis]. Radboud University Nijmegen.

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. https://doi.org/10.3758/brm.42.3.643

Kover, S. T., & Atwood, A. K. (2013). Establishing equivalence: Methodological progress in group-matching design and analysis. *Am J Intellect Dev Disabil*, *118*(1), 3–15. https://doi.org/10.1352/1944-7558-118.1.3

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. https://doi.org/10.3758/bf03195845

Li, P., Zhang, F., Yu, A., & Zhao, X. (2000). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism Language and Cognition*, *23*(5), 938–944. https://doi.org/10.1017/s1366728918001153

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. https://doi.org/10.1037/a0014875

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067

Palma, P., & Titone, D. (2020). Bilingual lexical access and reading. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Lexical Ambiguity Resolution* (pp. 159–183). Cambridge University Press. https://doi.org/10.1017/9781316535967.008

R Core Team. (2025). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods*, *52*(6), 2372–2382. https://doi.org/10.3758/s13428-020-01389-1

Tiffin-Richards, S. P. (2024). Cognate facilitation in bilingual reading: The influence of orthographic and phonological similarity on lexical decisions and eye-movements. *Bilingualism: Language and Cognition*, 1–18. https://doi.org/10.1017/s1366728923000949

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology*, *3*, 1–8. https://doi.org/10.3389/fpsyg.2012.00174

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. https://doi.org/10.1111/j.1467- 9280.2009.02389.x

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/bf03193889

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. https://doi.org/10.3758/bf03196335

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *The Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Appendices

Appendix 1: List of interlingual homographs and matched control words

Table 7

Word frequency (log10) and word length of interlingual homographs and their matched control words.

IH	Overlap	Freq DU	Freq EN	Control	Freq	Word length
Pop	Identical	3,019	3,537	Aap	3,097	3
List	Identical	2,218	3,614	Snee	2,305	4
Pink	Identical	2,100	3,162	Ader	2,155	4
Spin	Identical	2,533	2,873	Geit	2,548	4
Spot	Identical	2,520	3,497	Kook	2,517	4
Stem	Identical	3,578	2,061	Kerk	3,541	4
Tips	Identical	2,348	2,670	Zalm	2,364	4
Vent	Identical	3,920	2,354	Hond	3,868	4
Wet	Identical	3,546	3,301	Les	3,463	3
Step	Identical	1,362	3,782	Erwt	1,380	4
Links	Identical	3,477	2,152	Sterk	3,620	5
Slim	Identical	3,688	2,783	Lief	3,698	4
Pet	Identical	2,761	3,013	Das	2,767	3
Lessen	Identical	2,842	1,491	Glazen	2,846	6
Kin	Identical	2,533	2,340	Rok	2,498	3
Strip	Identical	2,241	2,904	Nagel	2,248	5
Rug	Non-identical	3,548	2,726	Tas	3,407	3
Angel	Non-identical	2,668	3,601	Stier	2,623	5
Tube	Non-identical	1,591	2,924	Riet	1,580	4
Pad	Non-identical	3,264	2,619	Bos	3,308	3
Files	Non-identical	1,940	3,133	Meren	1,968	5
Arts	Non-identical	3,157	2,673	Heks	3,068	4
Steel	Non-identical	2,576	2,974	Naald	2,571	5
Room	Non-identical	2,521	4,351	Saus	2,621	4
Brand	Non-identical	3,288	2,853	Kogel	3,282	5

Blank	Non-identical	2,866	2,696	Bruin	2,777	5
Hoop	Non-identical	4,206	2,140	Auto	4,302	4
Leg	Non-identical	3,750	3,460	Sla	3,562	3
Kind	Non-identical	4,164	4,479	Zoon	4,183	4
Dozen	Non-identical	2,658	3,091	Laden	2,670	5
Boot	Non-identical	3,623	2,755	Reis	3,597	4
Map	Non-identical	2,297	3,211	Kas	2,181	3

Appendix 2: List of sentences and comprehension questions

Table 8

Experimental sentences and comprehension questions across conditions.

Item	Overlap	Sentence with **interlingual homograph/control word** and question
1	Identical	Het meisje kan alleen maar naar de **POP/AAP** blijven kijken.
		Q: Keek een meisje naar de pop/aap? (ja)
2	Identical	Hij vertelde dat de **LIST/SNEE** veel pijn heeft gedaan.
3	Identical	De val had veel schade aan haar **PINK/ADER** veroorzaakt.
		Q: Had ze schade aan haar duim? (nee)
4	Identical	Jan was verbaasd toen hij hoorde dat een **SPIN/GEIT** gevaarlijk kan zijn.
5	Identical	Ik heb een goede band met mijn vrienden, dus ik **SPOT/KOOK** graag met ze.
		Q: Heb ik een goede band met mijn vrienden? (ja)
6	Identical	Anna rilde toen ze een **STEM/KERK** opmerkte in het verlaten landschap.
7	Identical	Piet deelt **TIPS/ZALM** uit aan iedereen die op het feest is.
		Q: Deelt Piet complimenten uit? (nee)
8	Identical	Vol trots presenteerde Kim haar nieuwe **VENT/HOND** aan haar beste vriendin.
9	Identical	Voor Eline was het niet duidelijk hoe de **WET/LES** vormgegeven moest worden.
10	Identical	Marie vond nog een **STEP/ERWT** onder de kast tijdens de verbouwing.
		Q: Vond Marie nog iets onder de kast? (ja)
11	Identical	Daan stemt op partijen die **LINKS/STERK** zijn in hun overtuigingen.
12	Identical	Zij vond het heel **SLIM/LIEF** dat ze de sleutel kreeg van haar vriend.
		Q: Kreeg zij de sleutel van haar moeder? (nee)
13	Identical	Henk gaat graag terug naar de winkel waar hij zijn **PET/DAS** gekocht had.
14	Identical	Iedereen vindt het fijn dat bij mij de **LESSEN/GLAZEN** altijd goed gevuld zijn.
		Q: Zijn bij mij de lessen/glazen goed gevuld? (ja)
15	Identical	Na het eten zat haar **KIN/ROK** onder de vlekken van de soep.
16	Identical	Niemand wist meer welke kleur er voor de **STRIP/NAGEL** gebruikt was.
		Q: Wist iemand nog welke kleur was gebruikt? (nee)
17	Non-identical	Na een lange dag ging de vrouw op haar **RUG/TAS** liggen slapen.
		Q: Ging de vrouw op haar rug/tas liggen? (ja)
18	Non-identical	De vrouw schrok toen zij de **ANGEL/STIER** van dichtbij zag.
19	Non-identical	Lisa had niet verwacht dat er nog wat uit de/het **TUBE/RIET** zou komen.

		Q: Had Lisa verwacht dat er iets uit de doos zou komen? (nee)
20	Non-identical	Naarmate we dichter bij het **PAD/BOS** kwamen, werd het steeds donkerder.
21	Non-identical	Op deze kaart staan alle **FILES/MEREN** heel duidelijk aangegeven.
		Q: Stonden er files/meren op de kaart? (ja)
22	Non-identical	Het jongetje was bang geworden voor de **ARTS/HEKS** uit de film.
23	Non-identical	Hij pakte de **STEEL/NAALD** voorzichtig van de grond op.
		Q: Pakte hij het voorwerp van de tafel? (nee)
24	Non-identical	Julia mag de overgebleven **ROOM/SAUS** hebben nadat de lunch is afgelopen.
25	Non-identical	De beelden van de **BRAND/KOGEL** zullen hen voor altijd bijblijven.
26	Non-identical	Het hout was nog helemaal **BLANK/BRUIN** voordat het geschilderd werd.
		Q: Was het hout eerst blank/bruin? (ja)
27	Non-identical	Ze waren erg verdrietig nadat ze de **HOOP/AUTO** waren kwijtgeraakt.
28	Non-identical	De muur is niet meer te redden, dus ik **LEG/SLA** houten platen ertegenaan.
		Q: Was het dak niet meer te redden? (nee)
29	Non-identical	Het verzorgen van een **KIND/ZOON** was een grote droom van hem.
30	Non-identical	Martin vond oude **DOZEN/LADEN** tussen de troep in de kelder.
		Q: Vond Martin dozen/laden in de kelder? (ja)
31	Non-identical	Tim kon eindelijk die dure **BOOT/REIS** betalen nadat hij de loterij won.
32	Non-identical	De onderzoeker keek in de **MAP/KAS** naar de resultaten van het experiment.
		Q: Keek een student naar de resultaten? (nee)

Appendix 3: Ethics assessment checklist

CHECKLIST ETHICS ASSESSMENT THESIS PROJECTS

Name	Huisman, M. (Merit)
Email	
Course Name	LET-TWM400 Master's Thesis
Programme	Master
Thesis Theme	The role of phonology in bilingual word recognition in L1 sentence context
Radboud Supervisor	Sybrine Bultena
Which situation applies to you?	I will collect new data
Is use of the existing data set allowed?	
Will you collect data from social media platforms and/or newspapers/news sites?	No
Will you collect data from participants?	Yes
Do patient/clients of a health care facility (e.g., a hospital or a nursing home) participate in the study?	No
Does the research include medical-scientific research that might carry risks for the participant?	No
Will the research be conducted on vulnerable or non-healthy participants?	No
Will the research be conducted amongst minors (<16 years of age) or (legally) incapable persons?	No
Is participation in the study voluntary?	Yes
Does the study use material/images/information that could be perceived as shocking or offensive?	No
Can participation in the study cause physical and/or mental harm to the participant?	No
Is compensation for participants in line with the EACH guidelines?	Yes
Does deception take place?	No
Do you make audio and/or video recordings (photo/video) of your participants?	No
Are participants recruited via the Radboud Research Participation System (SONA) and/or is the research conducted in the CLS Lab?	No
I declare that I have answered the questions truthfully.	Yes

UITKOMST

CHECKLIST RESULT: Review by the Ethics Assessment Committee Humanities is not necessary. If applicable: the next step is to draft an information document, consent form and Data Management Plan. For further explanation, see https://www.radboudnet.nl/facultyofarts/research/ethics-assessment-committee-humanities/student-thesis-research/	I understand (end of checklist)
CHECKLIST RESULT: Review by the Ethics Assessment Committee Humanities is necessary. Contact your thesis supervisor for the assessment procedure.	
CHECKLIST RESULT: This application may have to be reviewed by a recognised Medical Institutional Review Board, for example the METC Oost-Nederland. Contact your thesis supervisor for the assessment procedure.	

Appendix 4: Data analysis script

```
read.csv("data_tasks_v2.csv", header = T, sep = ",") -> taskdata_v1_raw

# SELECT USEFUL COLUMNS
taskdata_v1_raw[, c("Participant.Private.ID", "Participant.Status", "Current.Spreadsheet",
                "Response.Type", "Response", "Absolute.Reaction.Time",
"Spreadsheet..Answer",
                "Spreadsheet..Item", "Spreadsheet..Condition", "Spreadsheet..Overlap",
                "Spreadsheet..Word", "Spreadsheet..Freq_DU", "Spreadsheet..Freq_EN",
                "Spreadsheet..Length")] -> taskdata_v2_selectedcolumns

# DELETE FILLER AND PRACTICE TRIALS
taskdata_v2_selectedcolumns[!taskdata_v2_selectedcolumns$Spreadsheet..Condition ==
"Filler" & !taskdata_v2_selectedcolumns$Spreadsheet..Condition == "Practice", ] ->
taskdata_v3_onlyitems

# CHANGE COLUMN NAMES
taskdata_v4_columnnames <- taskdata_v3_onlyitems %>%
  rename(ParticipantID = Participant.Private.ID,
         Status = Participant.Status,
         Version = Current.Spreadsheet,
         RT = Absolute.Reaction.Time,
         Answer = Spreadsheet..Answer,
         Item = Spreadsheet..Item,
         WordType = Spreadsheet..Condition,
         Overlap = Spreadsheet..Overlap,
         TargetWord = Spreadsheet..Word,
         Freq_DU = Spreadsheet..Freq_DU,
         Freq_EN = Spreadsheet..Freq_EN,
         WordLength_Target = Spreadsheet..Length)

# ADD ITEM NUMBERS TO ANSWERS
taskdata_v5_additemno <- taskdata_v4_columnnames %>%
  mutate(
    Item = if_else(Answer %in% c("C", "M"), lag(Item), Item)
  )

# CALCULATE RTS FOR EACH WORD
taskdata_v6_RTdiff <- taskdata_v5_additemno %>%
  mutate(
    RT_lead = lead(RT),
    Item_lead = lead(Item),
    RT_diff = if_else(
      Item == Item_lead,
      RT_lead - RT,
      RT
    )
  ) %>%
  select(-RT_lead, -Item_lead)
```

```
taskdata_v7_onlywords <- taskdata_v6_RTdiff %>%
  filter(
    Response.Type == "info" |
      (Response.Type == "response" & Answer != "")
  )

# CHECK WHETHER RESPONSE TO QUESTION IS CORRECT (1) OR INCORRECT (0)
taskdata_v8_correctresponse <- taskdata_v7_onlywords %>%
  mutate(
    Correct = case_when(
      Response == "yes" & Answer == "M" ~ 1,
      Response == "no" & Answer == "C"  ~ 1,
      Response == "yes" & Answer == "C" ~ 0,
      Response == "no" & Answer == "M" ~ 0,
      TRUE ~ NA
    )
  )

# PASTE CORRECT OR INCORRECT RESPONSES TO ROWS WITH THE SAME ITEM
NUMBER
taskdata_v9_correctitem <- taskdata_v8_correctresponse %>%
  group_by(ParticipantID, Item) %>%
  mutate(
    Correct_filled = Correct[Answer != ""][1]
  ) %>%
  ungroup() %>%
  select(-Correct)

# PASTE POST-TARGET RT AND LENGTH IN NEW COLUMN
taskdata_v10_addposttargetrt <- taskdata_v9_correctitem %>%
  mutate(
    RT_PostTarget = lead(RT_diff)
  )
taskdata_v10.1_addposttargetrt <- taskdata_v10_addposttargetrt %>%
  mutate(
    WordLength_PostTarget = lead(nchar(Response)-1)
  )

# CHANGE TO LOWERCASE
taskdata_v11_lowercase <- taskdata_v10.1_addposttargetrt %>%
  mutate(TargetWord = tolower(TargetWord))

# KEEP ONLY TARGET ROWS
taskdata_v12_onlytarget <- taskdata_v11_lowercase[
  !is.na(taskdata_v11_lowercase$Response) &
    !is.na(taskdata_v11_lowercase$TargetWord) &
    trimws(tolower(as.character(taskdata_v11_lowercase$Response))) ==
trimws(tolower(as.character(taskdata_v11_lowercase$TargetWord))),
]
```

```
# READ QUESTIONNAIRE DATA FILE
read.csv("data_questionnaires_v2.csv", header = T, sep = ",") -> questdata_v1_raw

# SELECT USEFUL COLUMNS
questdata_v1_raw[, c("Participant.Private.ID", "Participant.Status",
"First.language.object.13.Response",
                "Gender.object.12.Response", "Age.object.14.Value", "AoA.object.15.Value",
"General.proficiency.object.17.Value",
                "Reading.object.5.Response", "Listening.object.6.Response",
"Writing.object.7.Response", "Speaking.object.8.Response",
                "Other.foreign.languages.object.18.Geen",
"Other.foreign.languages.object.18.Duits", "Other.foreign.languages.object.18.Frans",
"Other.foreign.languages.object.18.Spaans",
                "Other.foreign.languages.object.18.Andere",
"Specification.of.other.languages.object.19.Taal",
"Specification.of.other.languages.object.19.Taal.1",
                "Specification.of.other.languages.object.19.Taal.2")] ->
questdata_v2_selectedcolumns

# CHANGE COLUMN NAMES
questdata_v3_columnnames <- questdata_v2_selectedcolumns %>%
  rename(ParticipantID = Participant.Private.ID,
         Status = Participant.Status,
         FirstLanguage = First.language.object.13.Response,
         Gender = Gender.object.12.Response,
         Age = Age.object.14.Value,
         AoA = AoA.object.15.Value,
         GeneralProficiency = General.proficiency.object.17.Value,
         Reading = Reading.object.5.Response,
         Listening = Listening.object.6.Response,
         Writing = Writing.object.7.Response,
         Speaking = Speaking.object.8.Response,
         NoOtherLanguage = Other.foreign.languages.object.18.Geen,
         German = Other.foreign.languages.object.18.Duits,
         French = Other.foreign.languages.object.18.Frans,
         Spanish = Other.foreign.languages.object.18.Spaans,
         Other = Other.foreign.languages.object.18.Andere,
         OtherLanguage1 = Specification.of.other.languages.object.19.Taal,
         OtherLanguage2 = Specification.of.other.languages.object.19.Taal.1,
         OtherLanguage3 = Specification.of.other.languages.object.19.Taal.2)

# CALCULATE YEARS OF EXPERIENCE
questdata_v4_experience <- questdata_v3_columnnames[-1, ]
questdata_v4_experience <- questdata_v4_experience %>%
  mutate(YearsExperience = as.numeric(Age) - as.numeric(AoA))

# MERGE DATAFRAMES
mergeddata_v1 <- left_join(taskdata_v12_onlytarget, questdata_v4_experience, by =
"ParticipantID")
```

```
# ORGANISE MERGED DATAFRAME
mergeddata_v2_organised <- mergeddata_v1[, c("Version", "ParticipantID", "FirstLanguage",
"Gender", "Age", "AoA", "YearsExperience", "Item", "TargetWord", "WordType", "Overlap",
"RT_diff", "RT_PostTarget", "Correct_filled", "Freq_DU", "Freq_EN",
"WordLength_Target", "WordLength_PostTarget", "GeneralProficiency", "Reading",
"Listening", "Writing", "Speaking", "NoOtherLanguage", "German", "French", "Spanish",
"Other", "OtherLanguage1", "OtherLanguage2", "OtherLanguage3")]

# CHANGE COLUMN NAMES
mergeddata_v3_columnnames <- mergeddata_v2_organised %>%
  rename(
    Correct = Correct_filled,
    RT_Target = RT_diff
  )

# ADD PARTICIPANT NUMBERS AND FIRSTLANGUAGE
mergeddata_v3_columnnames$ParticipantID <-
match(mergeddata_v3_columnnames$ParticipantID,
unique(mergeddata_v3_columnnames$ParticipantID))
mergeddata_v4_firstlang <- mergeddata_v3_columnnames %>%
  mutate(FirstLanguage = case_when(
    FirstLanguage == 1 ~ "Dutch",
    FirstLanguage == 2 ~ "Dutch dialect",
    FirstLanguage == 3 ~ "Not Dutch",
    TRUE ~ as.character(FirstLanguage)
  ))

# ADD LOG TRANSFORM TO RTS
mergeddata_v5_logtarget <- mergeddata_v4_firstlang %>%
  mutate(RT_Target_Log = log10(RT_Target))

mergeddata_v6_logposttarget <- mergeddata_v5_logtarget %>%
  mutate(RT_PostTarget_Log = log10(RT_PostTarget))

# PARTICIPANT ANALYSIS
  # Delete non-native speakers of Dutch
mergeddata_v7_onlynativedutch <- mergeddata_v6_logposttarget %>%
  filter(FirstLanguage != "Not Dutch")

  # Percentage correct answers
resultaat <- mergeddata_v7_onlynativedutch %>%
  group_by(ParticipantID) %>%
  summarise(
    aantal_correct = sum(Correct == 1, na.rm = TRUE),
    totaal = sum(!is.na(Correct)),
    percentage_correct = round((aantal_correct / totaal) * 100, 1)
  )

  # Add percentage correct to dataframe
mergeddata_v8_perccorrect <- mergeddata_v7_onlynativedutch %>%
```

```
  left_join(resultaat, by = "ParticipantID")

# INSPECT DISTRIBUTIION
qqnorm(mergeddata_v8_perccorrect$RT_Target, pch = 1, frame = F)
qqline(mergeddata_v8_perccorrect$RT_Target, col = "steelblue", lwd = 2)

qqnorm(mergeddata_v8_perccorrect$RT_PostTarget, pch = 1, frame = F)
qqline(mergeddata_v8_perccorrect$RT_PostTarget, col = "steelblue", lwd = 2)

# REMOVE OUTLIERS IQR-METHOD
  # Participant SD
mergeddata_v9_SDoutlierremoval <- mergeddata_v8_perccorrect %>%
  group_by(ParticipantID) %>%
  mutate(
    mean_RT = mean(RT_Target, na.rm = TRUE),
    sd_RT   = sd(RT_Target, na.rm = TRUE),
    z_score = (RT_Target - mean_RT) / sd_RT
  ) %>%
  filter(abs(z_score) <= 2.5) %>%
  select(-mean_RT, -sd_RT, -z_score)

qqnorm(mergeddata_v9_SDoutlierremoval$RT_Target_Log, pch = 1, frame = F)
qqline(mergeddata_v9_SDoutlierremoval$RT_Target_Log, col = "steelblue", lwd = 2)

qqnorm(mergeddata_v9_SDoutlierremoval$RT_PostTarget_Log, pch = 1, frame = F)
qqline(mergeddata_v9_SDoutlierremoval$RT_PostTarget_Log, col = "steelblue", lwd = 2)

  # Interquartile range
boxplot(mergeddata_v8_perccorrect$RT_Target)

mergeddata_v9_IQRoutlierremoval <- mergeddata_v8_perccorrect %>%
  mutate(
    Q1 = quantile(RT_Target, 0.25, na.rm = TRUE),
    Q3 = quantile(RT_Target, 0.75, na.rm = TRUE),
    IQR = Q3 - Q1,
    upper_bound = Q3 + (1.5 * IQR)
  ) %>%
  filter(RT_Target <= upper_bound) %>%
  select(-Q1, -Q3, -IQR, -upper_bound)

boxplot(mergeddata_v9_IQRoutlierremoval$RT_Target)

qqnorm(mergeddata_v9_IQRoutlierremoval$RT_Target, pch = 1, frame = F)
qqline(mergeddata_v9_IQRoutlierremoval$RT_Target, col = "steelblue", lwd = 2)

qqnorm(mergeddata_v9_IQRoutlierremoval$RT_Target_Log, pch = 1, frame = F)
qqline(mergeddata_v9_IQRoutlierremoval$RT_Target_Log, col = "steelblue", lwd = 2)

qqnorm(mergeddata_v9_IQRoutlierremoval$RT_PostTarget_Log, pch = 1, frame = F)
qqline(mergeddata_v9_IQRoutlierremoval$RT_PostTarget_Log, col = "steelblue", lwd = 2)
```

```
# DESCRIPTIVE STATISTICS
  # Participant characteristics
participant_characteristics <- mergeddata_v8_perccorrect %>%
  mutate(Age = as.numeric(Age),
         AoA = as.numeric(AoA),
         FirstLanguage = as.factor(FirstLanguage),
         Gender = as.factor(Gender),
         GeneralProficiency = as.numeric(GeneralProficiency),
         Reading = as.numeric(Reading),
         Listening = as.numeric(Listening),
         Writing = as.numeric(Writing),
         Speaking = as.numeric(Speaking)) %>%
  summarise(Mean_Age = mean(Age), SD_Age = sd(Age), Min_Age = min(Age), Max_Age
= max(Age),
            Mean_AoA = mean(AoA), SD_AoA = sd(AoA), Min_AoA = min(AoA), Max_AoA =
max(AoA),
            Mean_Experience = mean(YearsExperience), SD_Experience = sd(YearsExperience),
Min_Experience = min(YearsExperience), Max_Experience = max(YearsExperience),
            Mean_OverallProficiency = mean(GeneralProficiency), SD_OverallProficiency =
sd(GeneralProficiency),
            Mean_Reading = mean(Reading), SD_Reading = sd(Reading),
            Mean_Listening = mean(Listening), SD_Listening = sd(Listening),
            Mean_Writing = mean(Writing), SD_Writing = sd(Writing),
            Mean_Speaking = mean(Speaking), SD_Speaking = sd(Speaking)
  )

prop.table(table(mergeddata_v8_perccorrect$Gender))
prop.table(table(mergeddata_v8_perccorrect$FirstLanguage))
prop.table(table(mergeddata_v8_perccorrect$AoA))
prop.table(table(mergeddata_v8_perccorrect$German))
prop.table(table(mergeddata_v8_perccorrect$French))
prop.table(table(mergeddata_v8_perccorrect$Spanish))
prop.table(table(mergeddata_v8_perccorrect$Other))
prop.table(table(mergeddata_v8_perccorrect$NoOtherLanguage))
prop.table(table(mergeddata_v8_perccorrect$OtherLanguage1))
prop.table(table(mergeddata_v8_perccorrect$OtherLanguage2))
prop.table(table(mergeddata_v8_perccorrect$OtherLanguage3))

  # M and SD of RTs
as_tibble(mergeddata_v9_IQRoutlierremoval) %>%
  group_by(WordType, Overlap) %>%
  summarise(Gemiddelde = mean(RT_Target), Standaarddeviatie = sd(RT_Target))

as_tibble(mergeddata_v9_IQRoutlierremoval) %>%
  group_by(WordType, Overlap) %>%
  summarise(Gemiddelde = mean(RT_PostTarget), Standaarddeviatie = sd(RT_PostTarget))

  # M and SD of correct answers
mean(mergeddata_v8_perccorrect$percentage_correct)
sd(mergeddata_v8_perccorrect$percentage_correct)
```

```
# TEST MODEL
  # Center freq, length, general proficiency
mergeddata_v9_IQRoutlierremoval <- mergeddata_v9_IQRoutlierremoval %>%
  mutate(ParticipantID = factor(ParticipantID),
         Freq_DU_c = Freq_DU - mean(Freq_DU),
         WordLength_Target_c = WordLength_Target - mean(WordLength_Target),
         Item = as.factor(Item),
         GeneralProficiency = as.numeric(GeneralProficiency),
         GeneralProficiency_c = GeneralProficiency - mean(GeneralProficiency),
         Reading = as.numeric(Reading),
         Reading_c = Reading - mean(Reading),
         WordLength_PostTarget_c = WordLength_PostTarget - mean(WordLength_PostTarget))

  # Explore helmert coding
    # Add helmert coding
mergeddata_v10_helmert <- mergeddata_v9_IQRoutlierremoval %>%
  mutate(
    Condition = case_when(
      WordType == "Control" ~ "Control",
      WordType == "IH" ~ Overlap,
      TRUE ~ NA_character_
    )
  )

CvsIH <- c(-2/3, 1/3, 1/3)
IdvsNonId <- c(0, -1/2, 1/2)

condition.helmert <- cbind(CvsIH, IdvsNonId)
mergeddata_v10_helmert <- mutate(mergeddata_v10_helmert, ConditionH =
as.factor(Condition))

contrasts(mergeddata_v10_helmert$ConditionH) <- condition.helmert

  # Model with helmert coding
target_model_helmert <- lmer(RT_Target_Log ~ ConditionH + Freq_DU_c + Reading_c +
(1|ParticipantID) + (1|Item), data = mergeddata_v10_helmert, REML = F)
summary(target_model_helmert)

posttarget_model_helmert <- lmer(RT_PostTarget_Log ~ ConditionH +
WordLength_PostTarget_c + Reading_c + (1|ParticipantID) + (1|Item), data =
mergeddata_v10_helmert, REML = F)
summary(posttarget_model_helmert)

  # Use sum coding
    # Add sum coding
mergeddata_v9_IQRoutlierremoval$WordType_Sum <-
factor(mergeddata_v9_IQRoutlierremoval$WordType, levels = c("IH", "Control"))
contrasts(mergeddata_v9_IQRoutlierremoval$WordType_Sum) <- contr.sum(2)
contrasts(mergeddata_v9_IQRoutlierremoval$WordType_Sum)
```

```
mergeddata_v9_IQRoutlierremoval$Overlap_Sum <-
factor(mergeddata_v9_IQRoutlierremoval$Overlap, levels = c("Non-identical", "Identical"))
contrasts(mergeddata_v9_IQRoutlierremoval$Overlap_Sum) <- contr.sum(2)
contrasts(mergeddata_v9_IQRoutlierremoval$Overlap_Sum)

  # Model with sum coding
target_model_sum <- lmer(RT_Target_Log ~ Overlap_Sum * WordType_Sum + Freq_DU_c
+ Reading_c + (1|ParticipantID) + (1+WordType_Sum|Item), data =
mergeddata_v9_IQRoutlierremoval, REML = F)
summary(target_model_sum)

r.squaredGLMM(target_model_sum)

posttarget_model_sum <- lmer(RT_PostTarget_Log ~ Overlap_Sum * WordType_Sum +
WordLength_PostTarget_c + Reading_c + (1|ParticipantID) + (1+WordType_Sum|Item), data
= mergeddata_v9_IQRoutlierremoval, REML = F)
summary(posttarget_model_sum)

r.squaredGLMM(posttarget_model_sum)

  # Conduct emmeans
emm <- emmeans(target_model_sum, ~Overlap_Sum*WordType_Sum)
pairs(emm)

emm_post <- emmeans(posttarget_model_sum, ~Overlap_Sum*WordType_Sum)
pairs(emm_post)

# TEST ASSUMPTIONS
res <- residuals(target_model_sum)
plot(fitted(target_model_sum), res)
vif(target_model_sum)

res <- residuals(posttarget_model_sum)
plot(fitted(posttarget_model_sum), res)
vif(posttarget_model_sum)

# PLOT DATA
summary_data <- mergeddata_v9_IQRoutlierremoval %>%
  group_by(WordType, Overlap) %>%
  summarise(
    mean_RT = mean(RT_PostTarget, na.rm = TRUE),
    sd_RT = sd(RT_PostTarget, na.rm = TRUE),
    n = n(),
    se = sd_RT / sqrt(n),
    ci = qt(0.975, df = n - 1) * se
  )

ggplot(summary_data, aes(x = Overlap, y = mean_RT, fill = WordType)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), color = "black") +
  geom_errorbar(aes(ymin = mean_RT - ci, ymax = mean_RT + ci),
```

```
              position = position_dodge(width = 0.9),
              width = 0.2) +
  scale_fill_manual(
    values = c("Control" = "#D30000",
               "IH" = "#FA8072")) +
  coord_cartesian(ylim = c(300, 420)) +
  scale_y_continuous(
    name = "Mean RT (ms)",
    breaks = seq(300, 420, by = 20)) +
  labs(
    x = "Overlap",
    y = "Mean RT (ms)",
    fill = "WordType",
    title = "Mean RT for post-target words by Overlap and Word Type"
  ) +
  theme_minimal()

# SAVE DATA FILE
write.csv(mergeddata_v9_IQRoutlierremoval, file = "Data Thesis Merit.csv", row.names = F)
```