

Cumulative Word Occurrence and Cognate Facilitation in L2 Sentence Reading

Ruben John Doorenbos, s1025103

MA Language Variation and Multilingualism

Faculty of Arts, Radboud University

LET-TWM400 Master's Thesis

Dr Sybrine Bultena and Dr Gerrit Jan Kootstra

3 July 2023

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

Signed: Ruben John Doorenbos

Word count: 18750 words

Acknowledgements

This is the second thesis I have written in the number of years, and I am proud of how well I was able to use the lessons from my BA thesis during this project. There were certainly difficult moments, especially since I was simultaneously doing an internship, but I am happy with the final product. I can look back at these five months with pride!

I would not have been able to do get here without help, so I want to thank the people who supported me while I wrote my thesis. I want to thank Anne Dien, Marjolein, and Samuel; it really helped to have friends to complain to who were also working on their thesis. Thank you to my family and friends, especially my mother, Merel, and Jente, for lending an ear when I needed it. I want to thank Marieke Hoetjes and Ferdy Hubers from Statistical Support for helping me figure out how to approach the data analysis and showing me how to run the analyses; I would not have been able to do so otherwise. Finally, thank you Sybrine for supervising this project. I really enjoyed working with you on this and I cannot thank you enough for your support!

Abstract

The cognate facilitation effect is an established effect that occurs in bilingual language processing. Research has found that it is modulated by several factors, including language proficiency, but there are only a few studies looking into the effects of cumulative word frequency (i.e., total exposure to a word in L1 and L2), which is a unique characteristic of form-overlapping words such as cognates. This study examined the effects of cumulative word frequency on identical cognates in L2 sentence context. Forty-five Dutch-English bilinguals took part in an online self-paced reading task in which they read English sentences with an English-Dutch cognate or an English control word. The word frequency for these target words was manipulated to investigate the effect of cumulative word frequency. The findings revealed that the participants deviated from the usual findings as word frequency effects were absent and the cognate facilitation effect was smaller than normal. English word prevalence, an alternative measure of word occurrence, was the best measure to predict reading times. There was also no significant effect of cumulative word frequency on the average reading times. These and previous findings show that more research is needed to elucidate the role of cumulative word frequency on the processing of cognates during bilingual sentence reading.

Keywords: cognates; word occurrence; cumulative word frequency; word prevalence; sentence processing

Contents

Acknowledgements.....	i
Abstract.....	ii
Introduction.....	1
Background.....	1
Cognates in Bilingual Processing	6
Language Proficiency	7
Word Frequency.....	8
Measures of Word Occurrence	11
Word Frequency in Corpora	12
Word Prevalence	13
Present Study	14
Method.....	16
Participants.....	16
Materials	17
Procedure	22
Design and Statistical Analysis.....	23
Results.....	24
Cognate Facilitation, Word Occurrence, and Language Proficiency	26
Word Occurrence Effects on Cognates.....	29
Cumulative Frequency Effects on Cognates.....	29
Discussion.....	31
Cognate Status, Word Frequency, and Language Proficiency	31
Word Frequency and Word Prevalence	34
Cumulative Word Frequency	36
Limitations	38
Future Research	40
Conclusion	42
References.....	44
Appendix A: Cognates.....	50
Appendix B: Control Words	52
Appendix C: Stimulus Sentences and Comprehension Questions.....	54

Introduction

If you were to ask ten people for directions in the Netherlands while speaking English, odds are that nine of these Dutch native speakers will be able to help as the Eurobarometer from 2005 (as cited in Gerritsen et al., 2007) revealed that almost 90% of the Dutch population believed they could hold a conversation in English. The Netherlands has also been ranked as the country with the highest English proficiency skills of the more than one hundred countries on the EF English Proficiency Index three years in a row (EF Education First, 2022). These numbers support the claim that bi- and multilingualism is becoming more commonplace because of globalisation and other changes to the world in the recent decades. The increasingly prominent position of English in Dutch society – whether it be in education or advertising – is not without controversy which is evident from the *Glossary Unnecessary English* (Koops et al., 2009) and the numerous columns arguing fervently against the spread of English and its influence on the Dutch language.

Simultaneously, globalisation has created a platitude of new issues and topics to study. Linguists and psycholinguists, for example, have been working to investigate the effects of bilingualism in several different areas. Many researchers have focussed on how additional languages could be stored in our mental lexicon and how this could affect the language production and comprehension of bilinguals compared to monolinguals (for an overview, see Bultena & Dijkstra, 2023). Studies within the field have looked at word frequency effects (e.g., Brysbaert et al., 2018) and the processing of words like cognates and interlingual homographs that occur in multiple languages (e.g., Kroll et al., 2013). Much has been uncovered about how bilinguals are affected by word frequency and how they deal with cognates during processing. However, few studies have attempted to combine these two topics to investigate how the processing of shared forms, which is typically explained by activation of both languages, is affected by cumulative word frequency (i.e., the total amount bilinguals have been exposed to a word in both their languages). In order to fill this knowledge gap, this study explored whether and how cognate facilitation effects are modulated by two measures of cumulative word occurrence when Dutch-English bilinguals read English sentences.

Background

Grosjean (1989), in what has become one of the most iconic quotes in linguistic research, once said, “Neurolinguists, beware! The bilingual is not two monolinguals in one person.” Essentially, he meant that bilingual language systems will differ from their

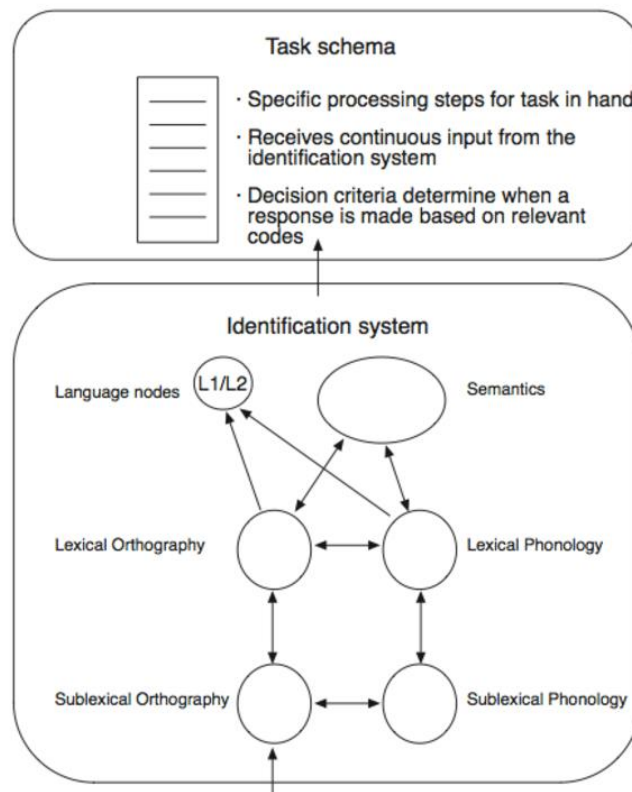
monolingual counterparts, even when it concerns the bilinguals' native language. Findings from different fields, including studies that investigated language processing (for an overview, see Kroll et al., 2013), have confirmed Grosjean's warning. These studies explored whether bilinguals have separate or joint language systems and whether they use language selective or non-selective access during processing. If a bilingual's language systems were language selective, their languages would not influence each other during language processing. Non-selective access to a bilingual's language system, on the other hand, implies that bilinguals access both language systems regardless of the language that they use, even in a completely monolingual setting. Studies found evidence for language non-selective access during isolated visual word recognition (e.g., Lemhöfer et al., 2008; Van Hell & Dijkstra, 2002). Although these findings have been replicated in sentence context (e.g., Dijkstra et al., 2015; Duyck et al., 2007), some studies have demonstrated that language access could become selective due to factors associated with natural reading (Hoversten & Traxler, 2015). Words that occur in multiple languages, such as cognates and interlingual homographs, are key in investigating the nature of bilingual processing. Cognates are translation equivalents that overlap in form, meaning, and pronunciation. These words can either be identical or non-identical, depending on the degree of orthographic, semantic, and phonological overlap. Examples of cognates include the identical pair *hotel* – *hotel* and the non-identical pair *appel* – *apple* in Dutch and English. Bilinguals have been found to process faster to cognates compared to monolingual words (e.g., Dijkstra et al., 2015), proving that non-target language systems can influence processing in the target language. This so-called cognate facilitation effect is influenced by multiple variables such as degree of orthographic and phonological overlap (Arana et al., 2022; Duyck et al., 2007). Next, the relevant models of bilingual language processing and cognate representations will be discussed before examining the cognate facilitation effect and word occurrence measures in more detail.

Research into the mechanisms of language processing is commonplace and it stands to reason that several language models have been developed that attempt to explain the inner workings of our language system (e.g., Levelt, 1989 on word production). After studies had proven that bilingual language processing can diverge from monolingual processing (Kroll et al., 2012), similar models were created to specifically explain how the language system of bilinguals operate (Desmet & Duyck, 2007). One of the most-frequently cited models is the Bilingual Interactive Access Plus model by Dijkstra and Van Heuven (2002; see Figure 1) which describes word recognition by bilingual speakers. The BIA+ explains which operations

take place during word recognition while trying to account for outside factors that can influence the process. According to the model, word recognition relies on lexical and phonological information in a bottom-up fashion: individual characters and phonemes are activated before the lexeme and the pronunciation itself. When the lexical orthography and phonology of a word is active, the lexical representation is connected to the relevant semantics and associated language nodes. The placement of the language nodes after the activation of the lexical information consequently means language membership does not affect a large part of the word recognition process (i.e., this model predicts non-selective language access). In fact, Dijkstra and Van Heuven (2002) updated the initial BIA model (Dijkstra & Van Heuven, 1998) and changed the role of languages nodes to representations of language membership rather than language filters that can exert influence on word activation processes. This is visualised by the feedforward arrows (going upwards) in Figure 1 from lexical orthography and phonology to the language nodes compared to both feedforward and feedback arrows used in much of the BIA+ like between the sublexical and lexical levels. The model is supplemented by a task/decision system that is influenced by non-linguistic factors such as task demands in, for example, a lexical decision task or a picture naming task. On the other hand, linguistic factors

Figure 1

The Bilingual Interactive Activation Plus (BIA+) Model (Dijkstra & Van Heuven, 2002)



are thought to directly influence word recognition through interactions between different language systems: the word identification system and other aspects of the BIA+ model communicate with the sentence parsing system for example, which leads to adjustments in the word recognition process. This aspect of the BIA+ model can be used to explain (some) effects of sentence context on language processing compared to processing in isolation.

One of the key differences between monolingual processing models and bilingual models is that the latter should account for the processing of the form-identical words appearing in multiple languages. Bilingual processing models must account for the interaction between two language systems within an individual and explain how words that occur in both systems are processed. The BIA+ model argues for language non-selectivity with a single integrated lexicon that contains both languages. This position is supported by evidence that form-identical words are processed differently from monolingual words (e.g., Duyck et al., 2007). The activation of interlingual homographs – form-identical words with different meanings, such as *peer* in Dutch (a fruit) and English (someone of equal status) – in the lexicon is slower because the bilingual has to process the conflicting information (i.e., multiple meanings) they receive from both languages (e.g., Libben & Titone, 2009). Conversely, cognates are processed faster because both languages deliver the same information (i.e. overlap in form, pronunciation, and meaning), speeding up word recognition (e.g., Lemhöfer & Dijkstra, 2004).

While the application of an integrated lexicon explains *why* form-identical words are processed differently, it does not immediately explain *how* these words are represented in the mental lexicon. Dijkstra and Van Heuven (2002) included two representations for interlingual homographs in the BIA+ because evidence showed that these words were sensitive to the word frequency in one language rather than cumulative word frequencies (Dijkstra et al., 1998). There was too little information on the representations of cognates at that time for Dijkstra and Van Heuven (2002) to comment on this in the BIA+ model, but Peeters et al. (2013) built on this line of thought and proposed the *two-morpheme view* to explain the processing of identical cognates. The *two-morpheme view* posits that cognates share an orthographic representation but have two underlying language-dependent morphemes. Cumulative word frequency (i.e., the total exposure in L1 and L2) should therefore not affect cognate facilitation effects because one of the two morphemic representations will be prioritised during activation based on the language context; the frequency associated with said morpheme, rather than the combined frequencies of both languages, will exert the most influence on language processing and the facilitation effects.

However, there are also studies that found evidence pointing in the other direction. Peeters et al. (2013) united these findings in two alternative explanations. First, the *form-overlap view* proposes that the identical form of the cognate in both languages means that only one orthographic representation is required in the lexicon. This shared representation is thought to be stronger than monolingual representations because activation of the cognate in both languages adds to that single representation, leading to an effect of cumulative frequency on cognate facilitation. The *shared-morpheme view* predicts the same outcome but gives a different explanation for the cumulative frequency effect. Davis et al. (2009) proposed an additional level in language processing models concerned with morphology to account for evidence that morphology is key to the organisation of the lexicon. This morphological level would, like the orthographic level in the *form-overlap view*, lead to stronger activation for identical cognates due to higher encounter rates for the underlying morpheme.

Peeters et al. (2013) also used this additional level in processing models, in combination with the BIA+ model, to explain the *two-morpheme view*. The important distinction between the *shared-* and *two-morpheme view* is that the latter proposes two morphemes that each hold language-specific information like inflections and frequencies. It is unclear how the *shared-morpheme view* explains the use of language-specific morphology when their position only allows for a single morpheme for identical cognates. Regardless, the juxtaposition between these three theoretical positions is clear: the *form-overlap* and *shared-morpheme view* argue that cumulative frequency can affect cognate facilitation due to joined representations at a level of the word recognition process. The *two-morpheme view*, on the other hand, assumes that both languages receive their own representation at the morphemic level of processing, meaning that facilitation effects are not sensitive to cumulative frequency since only the word frequency for the target morpheme will be activated. In short, the positions in favour of cumulative word frequency effects believe that L1 frequency will be a stronger predictor for processing than L2 frequency if the other frequency is low (i.e., RTs will be faster for cognates with high L1 – low L2 frequency than cognates with high L2 – low L1 frequency). If cumulative word frequency does not affect processing, then the expectation is that the word frequency in the target language affects processing the most. Having gained a better understanding of the different ideas on the mental representation of identical cognates, the next section will discuss cognate facilitation effects and factors that influence the processing of cognates.

Cognates in Bilingual Processing

Research into bilingual processing, and the processing of cognates specifically, has come a long way since linguists started to look into the workings of the bilingual language system. Word recognition was initially investigated in isolation and studies employed methods such as lexical decision (e.g., Van Hell & Dijkstra, 2002) and priming paradigms (e.g., Voga & Grainger, 2007). Research looked at interlingual homographs, for which interference effects were found (e.g., Dijkstra et al., 1998), and cognates which often facilitated language processing (e.g., Lemhöfer & Dijkstra, 2004). This work provided convincing evidence for language non-selectivity, which was bolstered by the findings in Van Hell and Dijkstra (2002). Their paper was one of the first to report evidence for the influence of the second language (L2) on the first language (L1) rather than the other way around. The results from their lexical decision task showed that native language processing is also subject to influences from a second language; in other words, non-target language influences during bilingual processing do not only originate from the L1, but also from the L2.

The next major development in the field was to move away from researching cognate facilitation effects in isolation and to look at bilingual processing in sentence context. Duyck et al. (2007) investigated how bilingual process identical and non-identical cognates in different tasks. They asked Dutch-English bilinguals to complete a lexical decision task in isolation or in sentence context, or to participate in an eye-tracking studies. Cognate facilitation effects were present for identical and non-identical cognates in both lexical decision tasks, although the effect was stronger for identical cognates. In the eye-tracking study, however, there was no facilitation effect for the non-identical cognates. Their study made it clear that language non-selectivity and the presence of cognate facilitation could not simply be generalised to sentence comprehension, but that studies also needed to examine the effects in reading tasks. Later studies confirmed their findings regarding identical and non-identical cognates (e.g., Bosma & Nota, 2020) as well as the effects of task demands on cognate facilitation (e.g., Bultena et al., 2014). Researchers have continued to investigate bilingual word recognition in sentences (for an overview, see Van Assche et al., 2012) in an attempt to understand the nuances involved in bilingual sentence comprehension and cognate facilitation effects. Studies have looked at the difference between noun and verbs cognates and found less prominent cognate facilitation effects for verbs (Bultena et al., 2014; Van Assche et al., 2013). Others have replicated the evidence on cognate facilitation effects in sentence context for bilingual children (Bosma & Nota, 2020). Finally, the field has continued to strive for higher ecological validity of previous

findings by investigating cognate processing in larger texts rather than single sentences. Cop et al. (2016) used an entire novel as stimulus material during an eye-tracking study and found evidence for cross-linguistic activation in the L1 and the L2. Balling (2013) also investigated how bilinguals read authentic L2 texts and discovered cognate facilitation only for morphologically simple cognates; the size of this effect was dependent on the context as facilitation was greater if the cognate was used in contexts that matches the L1 definition of the word. Furthermore, the effect of language proficiency and word frequency have been topics of interest in the field. These two influences on the cognate facilitation effect are discussed in more detail below.

Language Proficiency

Another variable that is known to affect cognate facilitation is language proficiency, which is an important variable in much of the research on bilingualism (for an overview, see Van Hell & Tanner, 2012). In the context of cognates and bilingual processing, proficiency effects have been observed both in the L1 and the L2. The same study that proved that the L2 can influence processing in the L1 also showed that certain proficiency thresholds must be reached before this can occur (Van Hell & Dijkstra, 2002), which has been replicated by other studies as well (e.g., Titone et al., 2011). On the other hand, bilinguals have been found to benefit less from cross-linguistic activation in their L2 when their proficiency increases. Libben and Titone (2009) conducted an eye-tracking study in which French-English bilinguals read sentences in their L2. The semantic constraint of the sentences (i.e., how much semantic information does the sentence provide to help the reader predict the next word) was the main manipulation. Cognate facilitation effects were found in both type of sentences, but it disappeared during late reading measures if the sentence had high semantic constraint. Additionally, they observed that the size of the cognate facilitation effect decreased for more proficient bilinguals during their eye-tracking experiment; this effect did not interact with sentence type. Pivneva et al. (2014) and Bultena et al. (2014) reported similar outcomes in their eye-tracking studies in which cognate facilitation effects decreased for bilinguals with higher L2 proficiency. Allen et al. (2020) found that higher proficiency resulted in less facilitation for Japanese-English bilinguals using a self-paced reading task, which shows that this effect also occurs when languages have different scripts. A plausible explanation for the effect is that bilinguals with lower L2 proficiency rely on their first language more, which leads to more noticeable effects due to cross-linguistic activation. Highly proficient L2 speakers, on the other hand, will rely less on their first language because their non-native language system

is better equipped to process the L2; in other words, the L1 contributes less to the word recognition and activation when proficiency grows, resulting in smaller facilitation effects.

Even though these results seem clear, studies have often used different measures of L2 proficiency. This extensive variability meant that Lauro and Schwartz (2017) were unable to take the variable into consideration in their meta-analysis on non-selective lexical access in sentence context. Their appeal for researchers to start using similar proficiency measures when investigating cognate facilitation effects is understandable, but the problem lies in defining the construct of (second) language proficiency. It is unclear which aspects of language should be taken into account when measuring proficiency and, as noted by Bultena et al. (2014), most researchers use a measure that is relevant to their own project. Studies often use self-ratings to measure proficiency (e.g., Pivneva et al., 2014), but it is not uncommon to measure reading speed (e.g., Bultena et al., 2014; Libben & Titone, 2009) or to test vocabulary (e.g., Allen et al., 2020). The use of these different measures – some of which are subjective rather than objective – makes it hard to generalise the effects of proficiency and fully specify how proficiency interacts with cross-linguistic activation. Furthermore, L1 proficiency is often not taken into account during analysis in these types of studies, even though research has shown that L1 proficiency is variable itself (Pakulak & Neville, 2010); the relevance of L1 proficiency during processing will be discussed in more detail in the section on word occurrence. Despite these shortcomings, the importance of language proficiency as a factor in bilingual processing should not be underestimated.

Word Frequency

Research into word frequency has led to the discovery of the word frequency effect (FE) and its central role in language processing. People react faster to words with higher frequency compared to words with lower frequencies. The effect has been found in numerous tasks which explains why word frequency is often controlled for during studies, since it affects various aspects of language processing (Duyck et al., 2008). The FE was first found in monolingual settings and has been the topic of multiple studies since then. However, it took more than thirty years after its discovery before the FE was investigated in the context of bilingual visual word recognition. The study by Duyck et al. (2008) analysed the reaction times from unbalanced Dutch-English bilinguals in a L1 and L2 decision task. The expected effects – faster reaction times in the L1 and for high-frequency words – were found, but the authors also discovered that the FE was significantly larger in the L2. They compared the bilinguals to a group of monolingual English speakers and found that FEs in the native language were similar in size

for both groups. Interestingly, the FEs in English for the bilinguals were significantly larger than the FEs of the monolingual English speakers. This finding, like the interaction between L2 proficiency and the size of cognate facilitation, suggests that speakers with relatively lower proficiency benefit more from facilitatory effects caused by word frequency than speakers with higher proficiency.

Bilingual FEs have been the topic of various other studies by now and the locus of word frequency research in bilinguals has moved towards methodologies that investigate the effects in sentence context, such as eye-tracking. Whitford and Titone (2012) examined FEs during L1 and L2 paragraph reading to test how the L1 and L2, as well as language proficiency, influenced word frequency effects. They found that L2 FEs were larger than L1 FEs, which were modulated by L2 exposure. Similarly, the difference between L2 and L1 FEs became smaller when L2 proficiency increased which matched the predictions made by the BIA+ model (Dijkstra & Van Heuven, 2002) about word frequency. Contrarily, another eye-tracking study explained their results through L1 proficiency (Cop et al., 2015). Eye-movements were measured for mono- and bilinguals while they read a novel and L1 FEs were similar for both groups, while L2 FEs were larger. L1 proficiency modulated these effects as higher proficiency meant less benefit while reading high-frequency words. While it is difficult to properly compare both studies due to methodological differences, it seems logical to conclude that language exposure is an important factor in explaining bilingual FEs. Whitford and Titone (2012) reasoned that higher L2 exposure leads to less L1 exposure, while Cop et al. (2015) argued that L2 proficiency does not necessarily impact L1 proficiency, and that individual overall language exposure is a more reliable predictor. Future research should aim to identify the exact interaction between word frequency and language exposure. This will not only teach us more about how word frequency affects bilingual language processing in general, but it could also help us understand the connection between word frequency and cognate facilitation effects better.

Word frequency is often one of the variables that is controlled for during cognate studies. If this were not the case, it would be difficult to ascertain whether significant findings were caused by the manipulation of cognate status and whether the size of the facilitation effect was influenced by differences in word frequency. However, this does not mean that word frequency does not affect cognate facilitation and that it is not a relevant topic for investigation. In fact, the question arises if the findings about FEs that are described above also apply to cognates. Additionally, the fact that cognates occur in multiple languages might mean they are

affected by cumulative word frequency during processing. In other words, the shared representations of cognates receive more exposure than average because they appear in two languages, which could in turn result in faster processing (Davis et al., 2009). Some studies have mentioned that cognate facilitation might be modulated by cumulative word frequency (e.g., Lemhöfer et al., 2008; Otwinowska & Szewczyk, 2017), but only two studies have investigated this possibility in more detail. These studies tried to find evidence in favour of one of the three theoretical positions discussed earlier: the *form-overlap view* (cognates share a single orthographic representation for both languages), the *shared-morpheme view* (cognates share a single morphemic representation for both languages), and the *two-morpheme view* (cognates have separate morphemes for each language). The first two positions argue in favour of cumulative frequency effects on cognate processing due to shared representations, while the latter believes cumulative word frequency has no effect on cognates as only the relevant morpheme, and thus its word frequency, is activated during processing.

Peeters et al. (2013) conducted a lexical decision task with French-English bilinguals while recording reaction times to investigate these positions. They manipulated the frequency of their cognates in French and English to create four categories: low frequency in both languages, high frequency in both languages, or low and high frequency in one language respectively. They formulated two hypotheses based on the positions above. If cognates are represented by two morphemes in the mental lexicon, then cumulative frequency should not affect cognate facilitation. On the other hand, if cumulative frequency does affect cognate facilitation, it would point towards the *shared-morpheme* or *form-overlap view*. They found the expected L2 FEs, i.e. words with higher frequency in English are reacted to faster than words with lower frequency, and a cognate facilitation effect. Closer examination revealed that cognates with high frequency in English and low frequency in French elicited faster reaction times than cognates with low frequency in English and high frequency in French. These results point towards the *two-morpheme view* because L2 frequency led to faster reading times than L1 frequency if the frequency in the other language was lower, even though exposure in L1 is generally higher. This suggests that bilinguals access the morpheme corresponding to the target language as cognates with high frequency in English and low frequency in French were reacted to faster in an English task than cognates with high French and low English frequency. The preliminary conclusion is thus that cognate facilitation effects are not affected by cumulative frequency, even if the study used isolated lexical decision and a design that reduced word frequency to a binary variable.

The second study that investigated cumulative frequency effects is by Winther et al. (2021) who took a different approach: they used computational modelling to create different models based on data from two previous cognate studies. They found that word frequency, rather than cognate status, was the best predictor for surprisal (a measure reflecting the cognitive effect necessary to process a word). The findings of this study are thus juxtaposed to those of Peeters et al. (2013) as the models that found facilitation effects did so without discriminating between cognate and non-cognates. However, no models were tested that assigned a special status to cognates and the conclusions were drawn based on computational models, which raised the question how well they can mimic human processing. A counterargument in favour of this method is that the frequency measures are likely to be more accurate, since the models were trained from scratch compared to different subjective language exposure that affects processing with human participants. Additionally, studies often use frequency data collected from monolingual corpora which do not represent the amount of exposure late L2 learner receive (Kuperman & Van Dyke, 2011; see below). Combined with the assumption that FEs are dependent on individual differences (Cop et al., 2015), it is plausible that frequency counts from corpora do not provide accurate representations of L2 exposure. The computational models, on the other hand, allow researchers to control and accurately calculate exposure so subjective frequencies can be used during analyses. Winther et al. (2021) correctly point out that their findings show that cumulative frequency could affect and explain cognate processing, although they lacked the data to provide a more definitive answer. It is thus pertinent to carry out further research to investigate this topic by improving on the design by Peeters et al. (2013) and by collecting human data in order to try and verify the conclusions drawn by Winther et al. (2021).

Measures of Word Occurrence

It is important to realise that investigating the effect of word occurrence is not as simple as it might seem, especially because it is often included in studies as the focus or as a control variable. The difficulty lies in defining the construct of word occurrence and how to express the variable in a reliable and valid manner during research. The discussion about FEs in bilingual processing mentioned language exposure, i.e. how often someone is exposed to a word will determine how quickly they process it. It is impractical and perhaps also impossible to calculate language exposure for every participant, which is why research often uses objective word occurrence measures, such as word frequency (how often a word appears in a corpus).

The most important aspects of word frequency are briefly discussed below as well as the lesser-known measure of word occurrence called word prevalence.

Word Frequency in Corpora

In lieu of using subjective measures of word occurrence, research has often relied on objective measures which indicate how often a word occurs within a collection of texts called a corpus. Determining word frequency became much easier with the rise of technology and digitally-available corpora (Brysbaert et al., 2018). This did not only lead to more corpora being available per language (as it was very time-consuming to create corpora by hand), but it also functioned as a catalyst in the field to re-examine the validity of older lists and compare them to the frequencies in the new online corpora. These validity studies showed that word frequency in newer corpora often predicted reaction times better, but also revealed that the corpus needs to contain language to which the participant group has likely been exposed. In other words, the older lists based on newspapers, such as Kucera and Francis (1967), might be appropriate for older participants, but the newer databases created with subtitles explain the variance in students' reaction times better. The difference in variance explained is caused by the contents of the corpus: the older lists were often based on nonfiction materials, while students are exposed to language that is more akin to popular media than to research articles and magazines. Several studies have compared the different corpora to find the best frequency measures for linguistic research by reviewing how well the paper-based lists, older digital corpora like CELEX (Baayen et al., 1995), and newer digital corpora like SUBTLEX (available in multiple languages) explained the variance in data. The SUBTLEX, comprised of television and film subtitles, outperformed the other options in multiple languages, such as Dutch (Keuleers et al., 2010), English (Brysbaert & New, 2009), and German (Brysbaert et al., 2011). Given that these databases have been found to work especially well for students, who often participate in studies, it is not a surprise that the SUBTLEX is used more and more often in research.

Another important development in the field is the move towards using the Zipf scale as the standardised measure for word frequency. The size and contents of the corpora used to calculate word frequencies varies, which also means that the word frequency for one word will differ across corpora. The use of a standardised measure solves this problem by expressing word frequency in a unit that is consistent across corpora. Frequency per million words (fpm) was often used for this purpose, but this measure is not without problems as explained by Brysbaert et al. (2018). Firstly, there are a lot of words with a very small fpm in larger corpora and secondly, fpm is often expressed as a logarithmic value which creates interpretability

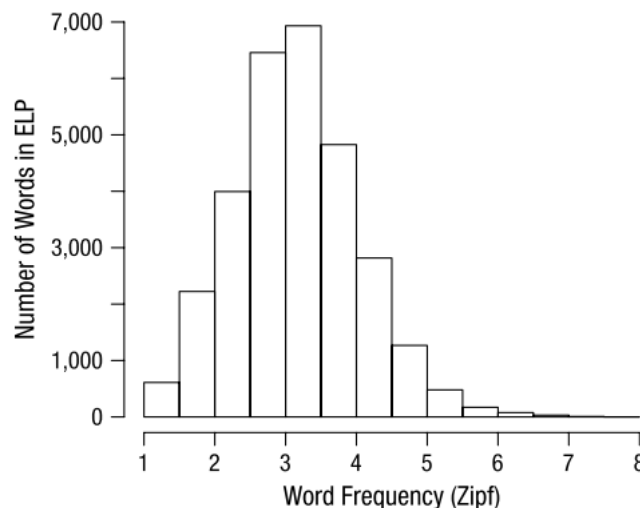
issues: the effect on reaction times is similar for differences between 1 and 2 fpm, and 10 and 20 fpm due to the nature of logarithmic curves. Van Heuven et al. (2014) proposed the Zipf scale as an alternative measure, which has a more straightforward interpretation (low frequency words are located between 1 and 3, and high frequency words between 4 and 7) and a better distribution across the scale (few words at the tails of the scale). The latter benefit especially is illustrated well in Figure 2, which shows the distribution of the words used in the English Lexicon Project carried out by Balota et al. (2007). The first advantage, however, can also lead to a more binary interpretation of word frequency: words with a Zipf score of 3.9 and 4.1 are categorised in low and high frequency respectively, even though the difference between those words is smaller than the difference between words with a score of 2.5 and 5.5. While the use of the Zipf scale as a word frequency measure is becoming more wide-spread, it is important to note that older and even some more recent studies, such as Peeters et al. (2013), use fpm. The comparison between studies with different frequency measures requires caution on behalf of the researcher to ensure that the effects are properly interpreted.

Word Prevalence

The Zipf scale seems to be the best measure of word frequency currently available, but that does not mean that word frequency itself does not have any inherent problems. The most glaring issue is perhaps the fact that it is an objective measure of word occurrence, which cannot completely capture the individual differences in language exposure that occur between

Figure 2

Distribution of English Words on the Zipf Scale (Brysbaert et al., 2018)



individuals, and between native and second language speakers (Brysbaert et al., 2018). “Bluetooth”, for example, has a Zipf frequency of 1.292, even though it is a well-known word, especially among younger generations. This example highlights the difference between word frequency and word knowledge: a word might be very well-known but simply not used as often (in a specific corpus) resulting in low frequency. Keuleers et al. (2015) attempted to solve this issue by introducing word prevalence, which could be described as an objective measure of subjective word occurrence or language exposure. Word prevalence is “the proportion of a population knowing a particular word” (Keuleers et al., 2015, p. 1669) and was developed as a complimentary measure to word frequency. As the example above showed, word frequency can be uninformative in the lower regions. Word prevalence should be able to highlight the words that are scarcely used in corpora but that are often known in a population, while word frequency can help differentiate between words with a high prevalence (i.e. words that are part of the core vocabulary). The data necessary to calculate prevalence is collected through a lexical decision task and requires many responses per word, which is why studies collecting prevalence data up until now have all been megastudies. Prevalence can be expressed as the percentage of the population that knows the word or as a z -score (scale of -2.5 to 2.5), which is the standardised version of the data to account for the fact that a large majority of the words is known by most participants (Brysbaert et al., 2019).

It is yet unclear how useful this measure is exactly due to its novelty, but a few studies have collected prevalence data and used it to reanalyse reaction times. The initial study by Keuleers et al. (2015) showed that prevalence was not correlated with word frequency (thus being a different variable) and that both measures explained more than 50% of the variance, confirming that prevalence and frequency are complimentary: whereas Bluetooth is a very low frequency word (Zipf = .292), its prevalence z -score of 2.438 shows it is widely known. Applying word prevalence to more recent Dutch data validated these findings (Brysbaert et al., 2016) and an initial investigation of the effect of word prevalence in English also proved fruitful (Brysbaert et al., 2019). More research is necessary to further validate prevalence as a measure of word occurrence and to explore its use in different languages. Regardless, the variable seems to have useful applications, including in studies that look at cumulative frequency effects in the context of cognate processing.

Present Study

Research into the cognate facilitation effect has yet to pay much attention to the cumulative word frequency and the potential effect it could have on the cognate facilitation

effect. Considering the studies that have shown that word frequency effects occur during bilingual language processing (e.g., Cop et al., 2015; Whitford & Titone, 2012), the question probably is not whether frequency plays a role during cognate processing, but in what way it affects facilitation. Only two studies have dived into this problem, but Peeters et al. (2013) and Winther et al. (2021) found conflicting results regarding the role of cumulative frequency in explaining cognate facilitation effects with different methodological approaches, as described above. This study therefore aimed to expand on these previous studies and contribute to the discussion to get another step closer to discovering whether cumulative frequency modulates the cognate facilitation effects.

In order to achieve this goal, the present study focussed on three factors to add to the existing literature. First, a self-paced reading paradigm (SPR) was used to simulate natural reading and add ecological validity to the findings, but also to expand the methodologies used in previous studies. Previous research into cognate processing has shown that task demands affect the presence and size of cognate facilitation effects (Bultena et al., 2014; Duyck et al., 2007), which means that it is important to investigate the effects of cumulative frequency in sentence context as well. Second, this study investigated how well (cumulative) word prevalence, a new measure for word occurrence, predicts cognate facilitation effects and how it compared to classic corpus frequencies. Third, a linear mixed effects model was used for the primary data analysis to safeguard the continuous nature of word occurrence variables. Studies investigating word frequency effects often use this type of analysis, but the previous literature on cumulative frequency effects have operationalised word frequency as a dichotomous variable: words were either classified as having low or high frequency.

While this study is exploratory in nature since it investigates cumulative frequency effects in sentence context while simultaneously examining the usefulness of word prevalence, four research questions were formulated to serve as a guide during the study. First, do well-established effects, namely the cognate facilitation effect and the effect of English word occurrence (frequency and prevalence), occur in the data? Second, does first and second language proficiency modulate those effects? Third, which measure of word occurrence in Dutch and English is the best predictor for the reading times? Fourth, are there significant differences in reading times when cognates are divided into four different categories based on cumulative word frequency ($L_{E}L_{D}$, $H_{E}H_{D}$, $L_{E}H_{D}$, and $H_{E}L_{D}$; explained below)? Based on previous findings, the presence of a cognate facilitation effect (e.g., Allen et al., 2020; Bultena et al., 2014; Duyck et al., 2007) and a FE is expected (e.g., Cop et al., 2015; Duyck et al., 2008;

Whitford & Titone, 2012). Second language proficiency will likely affect cognate facilitation effects (e.g. Bultena et al., 2014; Libben & Titone, 2009), while both L1 and L2 proficiency is expected to modulate the size of the FE (Cop et al., 2015; Whitford & Titone, 2012). Word frequency and word prevalence, based on previous studies (Brysbaert et al., 2019; Brysbaert et al., 2016; Keuleers et al., 2015), are likely to be complimentary to each other and neither is expected to predict reading times (in general or for cognates) much better than the other. There is little literature to support predictions about the effect of cumulative frequency since studies reported contradictory findings (i.e., Peeters et al., 2013; Winther et al., 2021). However, given the similarity in the design between this study and the study by Peeters et al. (2013) and the fact that both used behavioural data, it seems more likely that there will be no effect of cumulative word frequency on the reading times of cognates.

Method

Forty-five Dutch–English bilinguals took part in a self-paced reading task in which they read English sentences containing an English-Dutch cognate or a monolingual English control word. Reading times were recorded and the reading times (RTs) for the cognates and control words were analysed to investigate whether cumulative word frequency modulated the RTs of these target words during sentence reading. Detailed information on the participants, stimulus materials, procedure, and experimental design can be found below.

Participants

Fifty-one Dutch–English bilinguals completed the experiment, although six participants ended the experiment prematurely; the data of these participants was discarded. The rest of this section reports the demographic information of the participants who completed all the tasks ($N = 45$). Twenty-nine participants were drawn from the student body at Radboud University Nijmegen via the Radboud SONA system. Prerequisites for participation were normal or corrected-to-normal vision, no dyslexia or other reading problems, age between 18 and 35, and Dutch as the native language. Sixteen additional participants were recruited outside the SONA system to gather as much data as possible. These participants had to meet the same prerequisites as the SONA participants; additional questions at the start of the experiment were used to check whether they did.

The 45 participants (14 males; 30 females; 1 transgender) were aged between 18 and 54 ($M = 21.84$; $SD = 5.4$). One participant (age = 54) outside the required age range completed the experiment; they were not excluded after inspection of their RTs as they did not appear

significantly slower than the rest of the sample. Participants were all native speakers of Dutch who acquired English as a second language during their childhood ($M = 9.16$; $SD = 2.77$; range 0-13). Most participants started their acquisition during primary school (between ages 4 and 12), while a few started during secondary school (around ages 12 and 13). One participant indicated age of acquisition as zero, which could mean that they were raised as a balanced bilingual. Inspection of their RTs as well as L1 and L2 proficiency scores showed no deviations from the rest of the sample and the data for this participant was therefore not excluded.

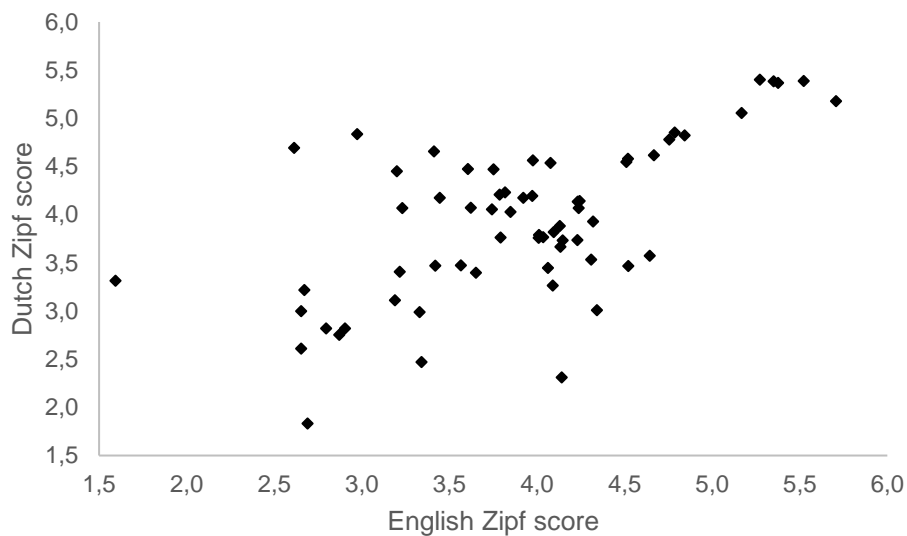
Materials

Sixty-four English-Dutch cognates (see Appendix A) were selected from a list of form-identical English and Dutch words. This list of more than 2000 words was created using a VLOOKUP comparison in Excel to determine which words in the SUBTLEX-US (Brysbaert et al., 2012) overlapped perfectly with words in the SUBTLEX-NL (Keuleers et al., 2010). These shared forms included cognates and interlingual homographs, but also loanwords and coincidentally shared forms (e.g., *job* which is a noun in English but a proper name in Dutch). The list was manually coded for word type and then filtered to create a list of identical noun cognates. The cognates were selected on their word frequency expressed as Zipf scores in English and in Dutch, retrieved from the SUBTLEX-US and SUBTLEX-NL respectively.

Each cognate was assigned to one of the four cumulative frequency conditions: low frequency ($L_E L_D$) or high frequency ($H_E H_D$) in both languages, low frequency in English and high frequency in Dutch ($L_E H_D$), or high frequency in English and low frequency in Dutch ($H_E L_D$). The cognates were evenly spread with sixteen cognates per condition. The design of this study is based on Peeters et al. (2013) who also used these conditions and analysed them with ANOVAs. However, a major downside of the ANOVA is that continuous independent variables, such as word frequency and language proficiency, must be recoded into categorical variables to fit the analysis. The natural distribution of the Zipf scores was therefore also considered during stimulus selection. This ensured that an alternative approach to data analysis, such as linear mixed effects models, would also be an option, preserving the continuous nature of word frequency and other independent variables. The resulting distribution of Dutch and English Zipf scores for the sixty-four cognates are depicted in Figure 3. Any words with Zipf scores below 4.0 are usually considered low frequency and words above 4.0 are considered as high frequency (Van Heuven et al., 2014).

Figure 3

Distribution of English and Dutch Word Frequency (Zipf) for the Cognates (N = 64)



In order to ensure that the cognates in each condition matched the descriptor of said condition, paired-sample t-tests were run to compare the English and Dutch word frequency *within* each condition. Paired-sample t-tests showed that the frequency in English and Dutch was not significantly different for cognates in the $L_E L_D$ ($t(15) = -0.05$, $p = .965$) and $H_E H_D$ ($t(15) = 0.51$, $p = .617$) conditions; the difference in word frequency between languages for the $L_E H_D$ ($t(15) = -5.45$, $p < .001$) and $H_E L_D$ ($t(15) = 5.69$, $p < .001$) conditions was significantly different. This means that matching within conditions was successful as the cognates in the congruent conditions had roughly the same frequency in both languages. In other words, the cognates in the $L_E L_D$ and $H_E H_D$ conditions had equally high or low frequency in both languages, while the differences in word frequency for the cognates in the incongruent conditions ($L_E H_D$ and $H_E L_D$) were significant. See Table 1 for means and standard deviations; the paired-sample t-tests made horizontal comparisons.

Ideally, cognates with high frequency in English would have the same average Zipf score, regardless of whether they are in the $H_E H_D$ or $H_E L_D$ condition. One-way ANOVAs were used to investigate how word frequency differed *across* conditions (i.e., vertical comparison of the word frequency means for the cognates in Table 1). There was a main effect of condition on word frequency in English (one-way ANOVA; $F(3, 60) = 53.24$, $p < .001$) and in Dutch (one-way ANOVA; $F(3, 60) = 59.89$, $p < .001$). Cognates in the four conditions differed significantly in English word frequency as well as Dutch word frequency ($ps \leq .012$, Bonferroni-correction). This means that the cognates with the same frequency in one language

Table 1. Overview of the different measures of word occurrence per condition ($n = 16$)

		<i>Frequency</i>		<i>Prevalence</i>	
		<i>English</i>	<i>Dutch</i>	<i>English</i>	<i>Dutch</i>
<i>L_EL_D</i>					
	<i>cognate</i>	3.02 (0.54)	3.03 (0.48)	2.00 (0.48)	1.69 (0.35)
	<i>control</i>	3.10 (0.26)		2.17 (0.30)	
<i>L_EH_D</i>					
	<i>cognate</i>	3.56 (0.39)	4.34 (0.26)	2.11 (0.33)	1.81 (0.12)
	<i>control</i>	3.43 (0.26)		2.28 (0.17)	
<i>H_EH_D</i>					
	<i>cognate</i>	4.83 (0.52)	4.80 (0.46)	2.37 (0.14)	1.81 (0.21)
	<i>control</i>	4.71 (0.25)		2.41 (0.11)	
<i>H_EL_D</i>					
	<i>cognate</i>	4.20 (0.18)	3.54 (0.41)	2.39 (0.21)	1.73 (0.17)
	<i>control</i>	4.16 (0.12)		2.37 (0.15)	

Note. Frequency is measured as a Zipf-score (1-7) while prevalence is expressed as a z -score (-2.5 – 2.5).

according to their condition did not have the same average word frequency. For example, the average Zipf score in English for the $L_{E}L_{D}$ ($M = 3.02$) and $L_{E}H_{D}$ ($M = 3.56$) conditions was significantly differed, even though the label of those conditions suggests that the average word frequency would be the same. This discrepancy also occurred for all other vertical comparisons in Table 1. It would have been best to match the word frequency per language across conditions for a more optimal comparison and replication of Peeters et al. (2013), but the limited options available during stimulus selection and the goal to have a representative distribution of word frequency in the sample meant that this was not possible.

Word prevalence in both English and in Dutch, expressed as z -scores, was obtained for each cognate from Brysbaert et al. (2019) and Keuleers et al. (2015) respectively. Word prevalence, the percentage of a population that is familiar with that word, is a relatively new measure of word occurrence and has thus not been used in cognate studies so far. It was difficult to find suitable cognates matched on both Zipf scores and word prevalence in both languages. Prevalence was therefore not included as a matching variable but as a factor in the analysis to explore its effect on reading times and its usefulness in this context. Paired-samples t -test

showed that the prevalence scores in English and Dutch respectively were significantly different *within* all conditions of cumulative word frequency ($ps \leq .005$) unlike the Zipf scores. In other words, the word prevalence in English for cognates in one condition was significantly different than the Dutch word prevalence for those same cognates. A one-way ANOVA indicated that Dutch word prevalence was not significantly different across all conditions ($F(3, 60) = 1.15, p = .336$). The average Dutch word prevalence for cognates in the L_EL_D condition was thus the same as the score for the cognates in the other three conditions of cumulative frequency. There was a significant difference between the four conditions for English word prevalence ($F(3, 60) = 5.75, p = .002$) as word prevalence for the L_EL_D condition was significantly lower than for the H_EH_D ($p = .012$, Bonferroni-correction) and H_EL_D ($p = .007$, Bonferroni-correction) conditions. These inconsistencies can be attributed to the lack of matching on word prevalence. The means and standard deviations are listed in Table 1.

After the cognates had been selected, each was matched to a monolingual English control word (see Appendix B). Interlingual homophones, non-identical cognates, and loanwords were not considered to ensure that the selected words were as far removed from Dutch as possible. Each pair was matched on word length ($M = 6.63, SD = 1.82$; range: 4 – 10) and number of syllables ($M = 2.31, SD = 0.68$; range: 1 – 4); this information was retrieved from CELEX (Baayen et al., 1995). The pairs were also matched on English Zipf scores, but they were not matched to the cognates on word prevalence (see Table 1). Paired-samples t-tests revealed no significant differences between cognates and their matched control word for any of the variables above within all four conditions ($ps > .05$). Word length and number of syllables were not significantly different across conditions for both cognates and control words (one-way ANOVAs; $ps > .05$). Zipf scores (one-way ANOVA; $F(3, 60) = 160.15, p < .001$) and word prevalence (one-way ANOVA; $F(3, 60) = 4.79, p = .005$) were significantly different across conditions for the control words. Control words in all four conditions differed significantly from each other on word frequency ($ps < .001$, Bonferroni-correction), while the word prevalence scores for the L_EL_D condition was significantly lower than for the H_EH_D ($p = .005$, Bonferroni-correction) and H_EL_D ($p = .033$, Bonferroni-correction) conditions. These patterns are identical to those described above for the cognates.

Sixty-four low-constraint sentences, each with a cognate and a control version, were created as stimulus materials; there were sixteen sentences per condition (see Table 2). The only difference between the two versions of a sentence was whether the cognate or the control word was inserted, and the preceding article was adjusted to the target word when necessary.

The sentences were between 10 and 14 words long ($M = 11.95$; $SD = 0.93$) and the target word was placed between the seventh and tenth position in the sentence ($M = 8.19$; $SD = 0.83$). The sentence continued for at least another three words after the target with a maximum of six words for longer sentences. The sentences were written to be low-constraint, but there was no time to conduct a cloze probability study. Consequently, there is no empirical evidence that the sentence context allowed for a large range of possible words in the target position. Comprehension questions were constructed for 16 of the 64 sentences: four per condition (also see Table 2). The questions did not include direct references to the target words to keep the questions equal across lists and had an unambiguous “yes” or “no” response; the answer to eight questions was “yes”, while the answer to the other half was “no”. All stimulus sentences and comprehension questions can be found in Appendix C.

Two experimental lists were created with cognate and control words counterbalanced across the lists. Each list thus contained 32 cognate sentences and 32 control sentences, plus the 16 comprehension questions; the number of cognate and control sentences per condition in each list was equal. The order of the lists was randomised to ensure that any findings did not result from the order in which the sentences appeared on the screen. Randomisation was done according to two principles: sentences from the same condition could not appear consecutively,

Table 2. Example sentences and comprehension questions per condition

<i>Condition</i>	<i>Sentence with cognate/control target word and comprehension question</i>
<i>L_EL_D</i>	The young socialite was wondering when the IMPERIUM/TAPESTRY would be hers. Was the socialite young? (yes)
<i>L_EH_D</i>	The firefighters were arguing about the INTERNET/EQUATION and its usefulness. Did the firefighters agree about the usefulness? (no)
<i>H_EH_D</i>	The bartender forgot to write down the WEEK/HOUR of her next appointment. Did the bartender write down the details of their appointment? (no)
<i>H_EL_D</i>	The wizard denied knowing about the TRAUMA/PILLOW when asked by his apprentice. Did the wizard deny knowing about it? (yes)

and the comprehension questions appeared at random intervals. In the end, 22 participants completed the experiment with list one and 23 participants saw list two.

Procedure

The sentences in the self-paced reading task were centred in the middle of the screen; depending on the size of the computer screen, most if not all sentences fitted on a single line. The sentences were presented black 14-point Verdana font to a white background. The task used a non-cumulative linear paradigm (Jegerski, 2013). In other words, participants saw dashed lines instead of letters when a new sentence was presented on the screen. Pressing the spacebar replaced the dashed lines of the first word with the actual letters; another press and the first word reverted to dashes and the next word became visible. This paradigm preserves the number of letters and words in a sentence to help participants maintain a normal reading pattern. Reading times were recorded per individual word from the moment it appeared on the screen until the participant pressed the spacebar. There was a 1000 ms pause after each sentence before the next sentence started. This pause was 2000 ms long after a comprehension question to avoid that the subsequent reading times would be affected. Participants used a mouse to click on “yes” or “no” to answer the comprehension questions.

Participants were tested online outside the lab in an uncontrolled environment. The experiment was conducted with the Radboud Online Linguistic Experiment Generator (ROLEG), which is an application developed at Radboud University to create and carry out online (linguistic) experiments, including self-paced reading tasks. Participants recruited via the SONA system could start the experiment via a link provided by the platform, while the link was included in the recruitment message for the participants recruited outside the SONA system. They first read the relevant study information (e.g., duration, tasks, ethical approval) before they were asked to give their consent. If they agreed, the experiment began with demographic questions. Participants were asked their gender, age, and age of acquisition of English. SONA-participants were asked to give their SONA number to ensure course credit could be awarded afterwards, while participants from outside SONA were asked to confirm they spoke Dutch as a native language, they had normal or corrected-to-normal vision, and that they had no dyslexia; these prerequisites were included in SONA pre-screening and were thus not included in the SONA version of the experiment. The participants recruited outside of SONA did not receive any form of compensation.

Next, participants were instructed on what they would do during the experiment and how they should behave, i.e., read at a normal pace and avoid small breaks in the middle of a sentence. The instructions were followed by four practice sentence and a practice question, before the self-paced reading task began during which participants read 64 English sentences and answered 16 comprehension questions. Subsequently, the participants completed the Dutch and English version of the LexTALE test to measure their L1 and L2 proficiency. The LexTALE is an un-speeded lexical decision task that consists of 60 items in the target language (Lemhöfer & Broersma, 2012) and is often used in research to gain a measure of general proficiency. Each version was prefaced with instructions provided in the respective language and three practice items. Finally, participants were asked if they noticed anything during the experiment or if they had something else on their mind about the experiment. Then, they were either send back to SONA to receive their credit or they ended the experiment by closing the window, depending on recruitment method.

Design and Statistical Analysis

In line with Peeters et al. (2013), this study has a 2 x 4 design with within-subject factors word type (cognate or control) and cumulative frequency ($L_E L_D$, $H_E H_D$, $L_E H_D$ and $H_E L_D$) so ANOVAs could be used for data analysis. However, ANOVAs cannot be used to analyse continuous variables, such as word frequency, without recoding these into categorical variables; they also do not consider how variation in reading times can be caused by individual differences (i.e. RTs are dependent on participant characteristics). Van Hell and Tanner (2012) recommend the use of linear mixed effects models (LMEs) to address these shortcomings when analysing language proficiency, and their advice also applies to studies interested in word frequency effects. The stimulus materials, as explained before, were therefore also selected to allow for the use of LMEs *and* ANOVAs during analysis.

Inspection of the data showed that it was possible to use LMEs to answer the research questions. Specifically, variance inflation factor (VIF) was calculated for the measures of word occurrence and language proficiency to make sure none of the variables were correlated too strongly. The VIF for English and Dutch Zipf scores, English and Dutch word prevalence, and L1 and L2 LexTALE scores were all below 2; these variables were significantly correlated ($p < .001$), but the correlations were low to moderately positive ($r < .70$). These outcomes indicate that there is no problem of multicollinearity that could affect the models. Given the advantages of this method (Van Hell & Tanner, 2012) and that it has not been used before to investigate

the effects of cumulative word frequency on cognate processing, the decision was made to mainly use LMEs, followed by ANOVAs for more specific analyses regarding cumulative word frequency. The latter also created the opportunity to compare the results of this study more closely with those in Peeters et al. (2013). After the data was prepared for analysis (see Results), various models were built to investigate the effects of cognate status and word occurrence on the reading times of the target words (i.e. cognate and control words). First, effects of cognate status on the reading times were analysed and the effect of L1 and L2 proficiency were investigated. Next, models were built to examine how various measures of word occurrence (Zipf scores and word prevalence in English and Dutch) affected the reading times of the cognates and whether an effect of cumulative frequency could be observed. All analyses were conducted in SPSS version 27.0.

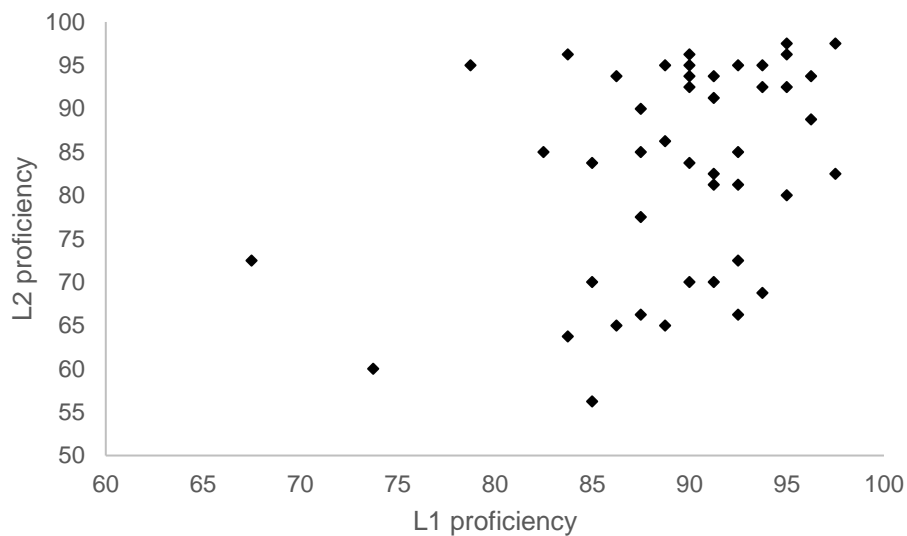
Results

Prior to the analysis of the reading times, the accuracy of the comprehension questions and the L1 and L2 proficiency scores were calculated per participant and then analysed. Participants performed well on the comprehension questions ($M = 87.64$; $SD = 10.53$) with an average of 14 correct answers and ranging from 62.5% to 100% (i.e., 10 to 16 correct answers). Performance on the Dutch ($M = 89.36$; $SD = 5.87$) and English ($M = 83.14$; $SD = 12.04$) LexTALE revealed a broad range of proficiencies and different proficiency combinations (see Figure 4). The majority ($n = 27$) scored better in Dutch than in English, a large minority ($n = 16$) scored better in the L2 than in the L1, and two participants scored the same in both languages. Most participants ($n = 31$) were advanced learners of English (score above 80%), while the rest were high intermediate ($n = 13$; score between 60-80%) or low intermediate ($n = 1$) learners (Lemhöfer & Broersma, 2012). Sadly, the conversion from LexTALE scores to proficiency indicators has not been validated for the Dutch version of the test, but an overwhelming majority ($n = 42$) scored above 80% on the test and the other three participants did not score lower than 67.5%. The proficiency scores in both languages were centred around their respective mean for analysis. Similarly, the Zipf scores were standardised to allow for easier interpretation of the outcomes since word prevalence was also expressed as z -scores.

No participants were excluded based solely on their accuracy or proficiency scores, but the RTs of participants with lower accuracy or proficiency scores in either language (lower than 75%) were inspected individually. This led to the removal of two participants. The data from the first participant was removed due to low accuracy on the comprehension questions

Figure 4

Distribution of L1 and L2 LexTALE Scores for All Participants (N = 45)



and a lower L2 proficiency (68.75% and 65% respectively) in combination with very fast reading times ($M = 174.02$; $SD = 108.98$) compared to the rest of the sample before outliers were removed ($M = 318.93$; $SD = 144.69$). The data of the second participant was removed because their RTs appeared to be affected by a system glitch: all recorded response times for this participant, also for non-reading inputs, were rounded to hundreds (e.g., 200 ms and 400 ms instead of 205 ms and 436 ms). Both cases were removed to err on the side of caution and prevent contamination of the data. These participants saw different experimental lists, so their removal did not affect the distribution as described in the Method.

Next, three different methods for outlier removal were applied to the data. The goal was to arrive at a dataset in which the data was (close to) normally distributed and in which as little data as possible was removed. All three methods were applied to RTs in ms as well as log transformed RT values. The first method removed outliers through absolute cut-offs (RTs below 100 ms and above 1000 ms), while the second and third method did so by establishing upper and lower bounds. The second method used the overall mean of the dataset and set the bounds at 2.5 standard deviations below and above this mean; the third method calculated the mean and standard deviation per participants and used these in a similar way to remove outliers in the RTs per participant. The normal distribution of the data for each method was checked with the Shapiro-Wilk test. Neither the original RTs nor the log transformed values were normally distributed after applying any of the three methods. Visual inspection of the accompanying Q-Q plots identified two possible options that would lead to a distribution

closest to normal: trimming the log RTs with absolute cut-offs (100 ms – 1000 ms) or with bounds calculated per participant. The latter would have led to the removal of 156 (5.7%) of the data, while 52 cases (1.9%) needed to be deleted for the former method. Outliers were therefore removed based on absolute cut-offs to preserve as much data as possible, resulting in a dataset with 2700 datapoints.

The models below thus used the log transformed RTs of the cognates and control words. The effects of cognate facilitation, word frequency and word prevalence, and language proficiency were investigated in models that included the entire dataset after outliers had been removed. The effects of (cumulative) word frequency and word prevalence on the reading times for cognates were examined in models that did not include the RTs for the control words. The missing Zipf and prevalence scores for the control words meant that these could not be considered in the models and that only cognate RTs were used; Peeters et al. (2013) took the same approach during their analysis. The LME models started simple with a single independent fixed variable and were expanded step-by-step by adding additional variables and by adding random slopes. The AIC-value for models were compared to arrive at the best model for the data. The following sections describe the steps that were taken during the analysis and report on the models that fit the data best.

Cognate Facilitation, Word Occurrence, and Language Proficiency

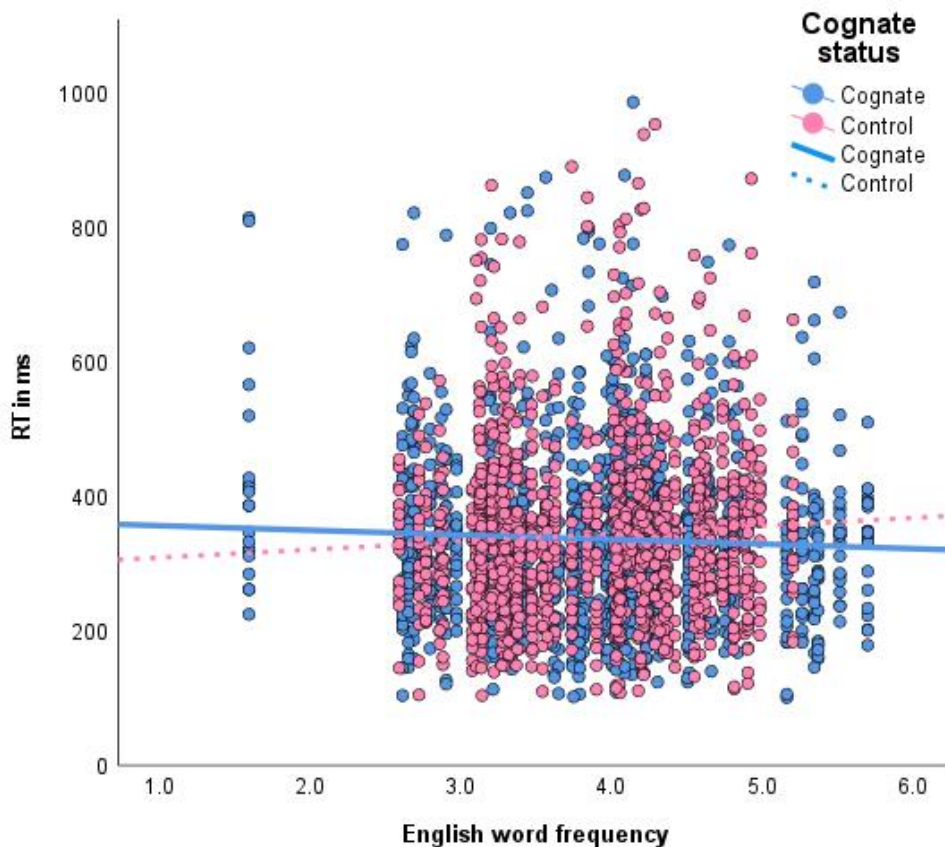
The first step was to investigate whether the data replicated previous findings of cognate facilitation effects, word frequency effects and language proficiency effects. The initial model included participant as a fixed intercept and cognate status as a fixed slope. Next, additional slopes were added one by one, and participant was changed to a random intercept to account for the data structure. The added slopes were L1 and L2 proficiency, English word frequency, and interactions between cognate status and L1 and L2 proficiency, word frequency and L2 proficiency, and cognate status and word frequency. The final model did not include any random slopes because adding any of the predictors above as a random slope and fixed factor led to convergence warning; these models were therefore not taken into consideration. Thus, the final model included a random intercept for participant and the fixed factors listed above; the outcomes of this model can be found in Table 3. RTs showed significant variance across participants ($p < .001$). Additionally, cognate status ($F(1, 2657.00) = 4.33, p = .038$) and the interaction between cognate status and English word frequency ($F(1, 2657.12) = 9.83, p = .002$) significantly predicted the reading times. The main effect of cognate status indicates that

cognates were read slightly faster than control words ($t = 2.08$). The interaction ($t = 3.14$) was investigated further by creating separate models for cognates and control words with the same predictors as the main model; the main effect of cognate status and the interactions involving cognates status were not included as these predictors became irrelevant by running separate models. These analyses showed that English word frequency did not significantly predict the reading times of cognates ($b = -0.003$, $t(1307.52) = -1.17$, $p = .243$), while it did significantly predict the reading times of control words ($b = 0.01$, $t(1307.04) = 3.00$, $p = .003$). Cognates were thus not influenced by word frequency, but control words were read slightly slower with higher word frequency (see Figure 5).

The same predictors were entered into another model in which word frequency was substituted for word prevalence. Reading times varied once again significantly per participant ($p < .001$). Cognate status was not significant for this model ($F(1, 2657.21) = 1.68$, $p = .196$), but there was a main effect of word prevalence ($F(1, 2657.02) = 22.76$, $p < .001$). Reading times decreased when a word was more prevalent ($t = -4.77$). None of the interactions were

Figure 5

Scatterplot of the Reading Times by Cognate Status and English Word Frequency



significant (see Table 4). The AIC for the prevalence-model is worse than the AIC for the frequency-model (-3774.46 compared to 3761.60), which means that the addition of word frequency to the model explains the data better than adding word prevalence. However, both measures of word occurrence influenced reading times to some extent. Given that the cognates also have word occurrence measures for Dutch and that the influence of word frequency differed between cognates and controls (see Figure 5), the following analyses considered only cognate the reading times.

Table 3. Fixed effects of the linear mixed effects model including word frequency for reading times on all target words

<i>Variable</i>	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
(Intercept)	2.50	0.02	44.69	155.30	<.001
Cognate	0.009	0.004	2657.00	2.08	.038
English frequency	-0.004	0.003	2657.28	-1.24	.217
L2 proficiency	0.0004	0.001	44.66	0.27	.787
L1 proficiency	0.0003	0.003	44.86	0.09	.930
Cognate x L2 proficiency	-0.0006	0.0004	2656.98	-1.39	.165
Cognate x L1 proficiency	-0.0002	0.0008	2657.10	-0.19	.848
English frequency x L2 proficiency	7.5276E-5	0.0002	2657.15	0.39	.695
Cognate x English frequency	0.0143	0.005	2657.12	3.14	.002

Table 4. Fixed effects of the linear mixed effects model including word prevalence for reading times on all target words

<i>Variable</i>	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
(Intercept)	2.59	0.03	278.92	101.06	<.001
Cognate	-0.05	0.04	2657.21	-1.29	.196
English prevalence	-0.04	0.009	2657.02	-4.77	<.001
L2 proficiency	-0.0005	0.002	177.12	-0.22	.826
L1 proficiency	0.0003	0.003	44.85	0.09	.926
Cognate x L2 proficiency	-0.0006	0.0004	2656.99	-1.49	.136
Cognate x L1 proficiency	-0.0002	0.0008	2657.10	-0.21	.837
English prevalence x L2 proficiency	0.0004	0.0007	2657.20	0.58	.559
Cognate x English prevalence	0.03	0.02	2657.21	1.61	.107

Word Occurrence Effects on Cognates

The same procedure as described above was used to investigate the effects of word frequency and word prevalence in both languages on the reading times of cognates only; this was done by creating a model with the relevant predictors for word frequency and another model with the predictors for word prevalence (see Table 5). The random intercept for participant was included in both models and showed that reading times varied significantly across participants ($ps < .001$). The model for word frequency was built with the predictors English and Dutch word frequency as fixed factors and random slopes. No interaction was added to the model as it did not improve model with. Neither English ($F(1, 28.67) = 0.07, p = .778$) nor Dutch ($F(1, 31.31) = 0.82, p = .372$) word frequency was significant (see Table 5).

The model for word prevalence included English word prevalence as fixed factor and Dutch prevalence as a fixed factor and a random slope to achieve best model fit; adding both predictors as random slopes led to a non-converging model. There was a main effect of English word prevalence on the reading times ($F(1, 1282.30) = 11.73, p < .001$), but not of Dutch word prevalence ($F(1, 54.67) = 2.68, p = .108$). Cognates with higher prevalence in English were read faster compared to cognates with lower prevalence ($t = -3.43$).

Cumulative Frequency Effects on Cognates

The effect of cumulative frequency was examined with the help of analyses of variance (ANOVA). This analysis was also carried out by Peeters et al. (2013), which made it easier to compare the results of both studies. Additionally, analysing the reading times of the four different cumulative frequency conditions ($L_E L_D, L_E H_D, H_E H_D, \text{ and } H_E L_D$) with an ANOVA

Table 5. Fixed effects of the linear mixed effects model for reading times on the cognates

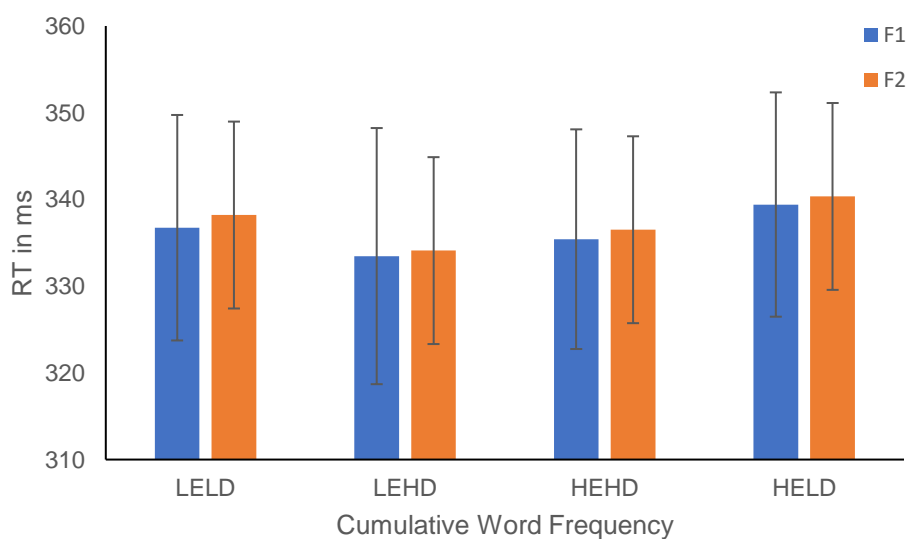
<i>Variable</i>	<i>Estimate</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>p</i>
Word frequency					
(Intercept)	2.50	.016	43.31	154.12	< .001
English frequency	-.001	.004	28.67	-0.27	.788
Dutch frequency	-.004	.004	31.31	-0.91	.372
Word prevalence					
(Intercept)	2.62	0.03	46.55	85.40	< .001
English prevalence	-0.03	0.01	1282.30	-3.43	< .001
Dutch prevalence	-0.03	0.02	54.67	-1.64	.108

created a direct comparison between different frequency combination compared to interaction effects that could have been significant in the LME models. Aggregated means per participant per conditions were calculated using RTs in ms; in other words, each participant had an average reading time per condition. First, a repeated-measures ANOVA with cognate status and English word frequency as independent variables was used to investigate whether this analysis would reveal a cognate facilitation effect and a word frequency effect. There was no main effect of cognate status ($F_1(1, 42) = 0.57, p = .454$; $F_2(1, 124) = 0.67, p = .415$). There was a main effect of English frequency for participants ($F_1(1, 42) = 15.44, p < .001$), but not for sentences ($F_2(1, 124) = 2.08, p = .152$). Participants read low frequency words ($M = 334.13, SD = 80.72$) faster than high frequency words ($M = 345.52, SD = 83.90$); this effect was thus not significant on an item level.

The repeated-measures ANOVA with cumulative frequency as independent variable and L2 proficiency as a covariate analysed the RTs for cognates only and showed no significant effect of cumulative frequency ($F_1(3, 123) = 1.23, p = .302$). There also was no significant interaction between cumulative frequency and L2 proficiency ($F_1(3, 123) = 1.42, p = .239$). A one-way ANOVA was used for the F_2 analysis since the design of the dataset did not facilitate for the use of a repeated-measures ANOVA in this case. This ANOVA also revealed no significant effect of cumulative frequency ($F_2(3, 60) = 0.60, p = .980$). Numerical observation showed that reading times were highest for the $H_{E}L_{D}$ condition, followed by cognates in the $L_{E}L_{D}$ condition. Words with high Dutch frequency were read faster the cognates

Figure 6

Mean Cognate RTs Split by Cumulative Word Frequency and Analysis



with low Dutch word frequency: cognates in the L_EH_D condition were read fastest, followed by the cognates in the H_EH_D condition (see Figure 6).

Discussion

I set out to investigate the potential effects of cumulative word frequency on the cognate facilitation effects during bilingual processing by conducting a self-paced reading task. Dutch-English bilinguals participated in an online study in which they read English sentences which contained an English-Dutch cognate or a monolingual English control word. The word frequency, expressed as a Zipf score, of the cognates was manipulated in both languages to create four distinct cumulative word frequency categories: (1) low frequency in both languages, (2) low frequency in English and high frequency in Dutch, (3) low frequency in both languages, and (4) high frequency in English and low frequency in Dutch. The reading times for the target words were analysed with linear mixed effects models and analyses of variance to answer the research questions. There was a small cognate facilitation effect – although the effect disappeared when the data was analysed with a repeated-measures ANOVA – and an interaction between cognate status and the effect of English word frequency on reading times. Additionally, English word prevalence affected the reading times of cognates as expected, but cognate status was no longer a significant predictor in that model; there also was no interaction between word prevalence and cognate status. Lastly, there were no significant differences between any of the four different cumulative word frequency categories. Surprisingly, most of the hypotheses based on solid research – such as a word frequency effect and an interaction between cognate status and L2 proficiency – were not confirmed by the data or smaller than usual. Possible explanations for these findings are discussed below in the context of the research questions that guided the study, followed by the limitations of this study and suggestions for follow-up research.

Cognate Status, Word Frequency, and Language Proficiency

Studies have found evidence for cognate facilitation effects (e.g., Allen et al., 2020; Bultena et al., 2014) and word frequency effects (e.g., Duyck et al., 2008; Whitford & Titone, 2012) time and again. While cognates in this study were also read faster, the effect appeared to be weaker than what has previously been observed. Participants read identical cognates faster, but the size of the facilitation effect was quite small: the estimated increase in RT from cognates to control words is slightly more than a single millisecond according to the LME model. Other studies have found much larger differences between cognates and control words; for example,

Peeters et al. (2013) observed a facilitation effect of 32 ms. The small size of the facilitation effect is reflected by the low t -value for this effect ($t = 2.08$), which in turn explains why there was no significant effect of cognate status in the ANOVA. t -values must be greater than 2 to be significant, which means that the effect in the LME model was barely significant in the first place. Given that ANOVAs work with aggregated means per participants and thus with fewer data points than LME models, the absence of a significant cognate facilitation effect is not too surprising. There are two points to take away from these findings. First, it is important to critically assess the design of a study and consider the data structure before choosing the statistical analysis to apply to the data. Small but significant effects might remain undiscovered if this is done incorrectly as demonstrated by the comparison of the repeated-measures ANOVA and LME model above. Second, the fact that the cognate facilitation effect was small says something about the sample and their RTs in comparison to other studies. Namely, it seems that the participants in this study behaved differently than expected based on earlier findings, which means that the hypotheses formulated in the Present Study might not bear fruit. The most likely explanation seems to be that something task-related caused a lack of variation in the reading times, resulting in effects being smaller than normal or even disappearing. The exact cause of the lack of variation will be discussed in the Limitations below, while the rest of this section will explore how it affected the outcomes of this study and whether there are other potential explanations for the findings.

The hypothesis that highly frequent words would have been read faster than words that are less frequent was indeed disproven for this study. Neither the model in which both cognate and control words were considered nor the model in which the effect of word frequency on cognate RTs was investigated found a significant effect of English word frequency, although reading times did decrease slightly as frequency increased. The absence of FEs itself was striking enough as this is one of the most well-documented phenomena in language research: word frequency effects have been observed in numerous different contexts with high frequency words being processed faster than low frequency words (Duyck et al., 2008). However, the FE was not only absent in the LME models; it was also reversed – and significantly so – when the data was analysed with an ANOVA to investigate the reading times of all target words. The average reading times for low frequency English words were about 11 ms faster than the reading times for the high frequency English words in this analysis. Given that effect was not present in the LME model, it seems likely that the difference is caused by differences between the two methods, particularly the change from frequency as a continuous variable to a

dichotomy between low and high frequency. The absence of a significant effect in the model could be attributed to the sample and the lack of variation in the RTs: the cognate facilitation effect was smaller than usual, so the expected FE could have shrunk compared to other studies resulting in it no longer being significant.

The previous explanation, however, does not account for the observed interaction between cognate status and English word frequency in the LME. Cognates were read faster when they became more frequent in English, while the opposite was true for control words, as visualised in Figure 5. The FE effect for cognates was as expected, but the inverted FE for the control words was not. Explaining the advantage of low frequency words compared to high frequency words is quite difficult: FEs are well-established, especially since studies started to explore the effects of word frequency during bilingual processing (e.g., Cop et al., 2015; Whitford & Titone, 2012). The most plausible explanation would be that the objective word frequency for the control words in this experiment as collected from the SUBTLEX-US (Brysbaert et al., 2012) did not match the subjective word frequency for the participants, which caused the subversion of the word frequency effect. The mismatch between subjective and objective word frequency could originate from the monolingual nature of the database. Many of the corpora with word occurrence measures are based on L1 sources and thus match the experiences and exposure of native speakers. Research has shown that individual differences such as vocabulary size (Kuperman & Van Dyke, 2011) and second language exposure (Cop et al., 2015) influences word frequency effects. If the word frequency measure does not match these differences (in this case the difference in exposure for second language learners), it is not unexpected that predictions do not bear fruit. After all, the measure used to test the prediction was not fit to do so because it was not calibrated to the target group. The negative FE for the control words could be an additional explanation for the ANOVA results, i.e., a main effect of word frequency in the opposite direction. Sadly, it was not possible to recreate the interaction between cognate status and word frequency in the ANOVA. The dataset was built in such a way that a single datapoint would have counted towards both independent variables, leading to invalid analyses. It was therefore impossible to investigate the issue further.

Lastly, the lack of significant effects of L1 and L2 proficiency was also unexpected. The cognate facilitation effect often interacts with L2 proficiency as more experienced bilinguals rely less on the added benefits of the shared form and meaning compared to beginners (Bultena et al., 2014; Libben & Titone, 2009). Both L1 and L2 proficiency also influence FEs as they represent language exposure: higher proficiencies mean more exposure

and a smaller difference between low and high frequency words, leading to less benefit from FEs (Cop et al., 2015; Whitford & Titone, 2012). Again, the lack of variation in the data is the most likely culprit; since the LexTALE has been extensively validated (Lemhöfer & Broersma, 2012), making it improbable that the measure of language proficiency was the problem. Additionally, Bultena et al. (2014) found effects of L2 proficiency operationalised as self-rated L2 proficiency or as L2 reading speed in a self-paced reading task. These measures correlated significantly with the LexTALE score of participants, making it even more unlikely that the absence of proficiency effects can be explained by the measure itself. Aside from the uniformity in the data obscuring language proficiency effects, it is also possible that language proficiency did affect reading times, but that these effects were not found due to the narrow focus during the analysis. Only the RTs for the target words were included during the analysis to prevent the overcomplication of the LME models. However, this does mean that there are no insights into how the rest of the sentence was read and if language proficiency affected reading in general.

Word Frequency and Word Prevalence

Word prevalence – the percentage of population that knows a word – has recently been introduced as an alternative to word frequency as a measure of word occurrence. Studies have investigated how well this measure explains language processing and how it compares to word frequency. They found that, while both measures do correlate slightly, they explain different aspects of word occurrence, as word prevalence works better for words that are rarely used (Keuleers et al., 2015); word frequency, on the other hand, provides more information when a word is used very often. Brysbaert et al. (2019) describe multiple applications of word prevalence, including preventing experiment bias which might occur when selecting stimulus materials based on Zipf scores only. That could lead to words being labelled as low frequency even though people are very familiar with them, since those words might rarely be used in films and tv-series (e.g., *binocular*), leading to a low Zipf score in corpora based on subtitles. For these reasons, word prevalence was included in this study to test the usefulness of the measure in the context of cognate facilitation effects and to compare it to word frequency.

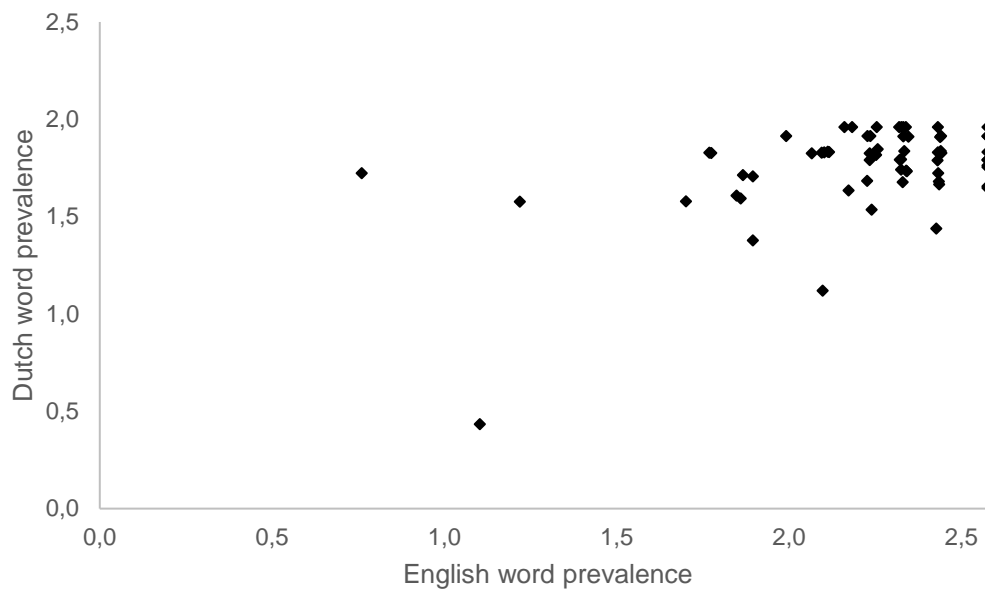
These questions were investigated with two LME models which each had a prevalence variant and a frequency variant. The first model looked at RTs for cognates and controls words to examine effects of cognate status, English word occurrence, and language proficiency (see Tables 3 and 4). As described above, there was no main effect of word frequency, but there

was an interaction between cognate status and word frequency: cognates were read faster if they had higher frequency (although not significantly so), while control words were read slower if they had higher frequency. When word prevalence was entered into the equation instead of word frequency, the main effect of cognate status and the interaction effect were no longer significant. However, there was a significant main effect of word prevalence: more prevalent words were read faster than less prevalent words. The main effect of cognate status in the first model was quite small, so it seems that word prevalence predicted reading times to such an extent that cognate status no longer contributed to explain the variance in the data. These findings show that word prevalence is an important and useful measure of word occurrence, as found by earlier studies (Brysbaert et al., 2019; Brysbaert et al., 2016; Keuleers et al., 2015). However, the absence of an effect of word frequency suggests that word prevalence might work even better than word frequency when trying to account for word occurrence. Further research is necessary to figure out exactly how these two measures relate to each other and whether their usefulness might change depending on the situation in which they are used.

Similarly, word frequency in either language did not predict reading times when considering a model that included only the cognate reading times to analyse the effects of L1 and L2 word occurrence (see Table 5). While Dutch word prevalence also did not significantly affect RTs, an increase in English word prevalence resulted in a significant decrease in RT. It is curious that word prevalence turned out to be the better predictor of the two, given how often FEs have been found when using Zipf scores. The Zipf scores of the cognates used in the self-paced reading task are mostly concentrated between 3 and 5 (see Figure 3), which means that neither word prevalence nor word frequency was expected to outperform the other as there was no preference for rare or well-known words. There can be large discrepancies between Zipf scores and prevalence (see Table 7 in Brysbaert et al., 2019), which could mean that word prevalence was a better measure of word occurrence for this selection of cognates. However, this would mean that many of the chosen cognates are words under- or overused in movies and tv-series, which would have led to a less representative Zipf measure in the SUBTLEX. On the other hand, this outcome could also suggest that word prevalence is a stronger predictor for cognate reading times since the effect remained when the data lacked variation; the well-known FE, on the other, did not withstand the uniformity in the data. Whatever caused the significant effect of word prevalence, it is important to remember that word prevalence was not controlled for or matched on during stimulus selection, while word frequency was. The distribution of

Figure 7

Distribution of English and Dutch Word Prevalence for the Target Cognates (N = 64)



word prevalence, visualised in Figure 7 (also see Table 1 for word prevalence per condition), shows that cognates did not have a z -score higher than 2.0 for Dutch word prevalence while their scores for English were mostly above 2.0. Future studies should focus solely on word prevalence in this context and use stimulus materials which are matched on word prevalence similarly to how word frequency was matched in this study. The findings from those studies would be able to provide more definitive insight into the role of word prevalence during bilingual processing and the processing of cognates specifically.

Cumulative Word Frequency

The findings above seem to suggest that word occurrence in the first language does not affect the reading times of cognates in the second language, contrary to what Winther et al. (2021) and Peeters et al. (2013) found. While these studies did not agree whether cumulative frequency affected RTs, both studies did find that first language exposure played a role in the processing of cognates. In order to investigate the effect of cumulative frequency more closely, the data was analysed with an ANOVA in the same manner as Peeters et al. (2013). The results confirmed the findings from the LME models that there were no significant differences between the four conditions. Considering the lack of or smaller size of certain effects found in previous studies for this study, it seemed wise to look at the means for each condition as seen in Figure 6 to gain more insight into which way the data leaned and to potentially find leads for future research. The fastest cognates were those in the L_EH_D condition, followed by the

cognates in the H_EH_D , L_EL_D , and H_EL_D conditions. Cognates with high frequency in Dutch are thus read faster than those with low word frequency. Furthermore, English word frequency has an interesting influence on the reading times: cognates with low frequency in English are read faster than cognates with high English word frequency when the Dutch frequency is the same (i.e., $L_EH_D > H_EH_D$ and $L_EL_D > H_EL_D$). This order was quite surprising: regardless of the role of cumulative frequency, the cognates with high frequency in both languages would be expected to be read faster and those with low frequency in both languages would be expected to be read slower. According to Peeters et al. (2013), the order of the incongruent conditions provided the clearest evidence for the role of cumulative frequency. They found that the French-English cognates with high frequency in the L2 were read slower than those with high frequency in the L1 if the frequency in the other language was low (i.e., they found the order $H_FH_E > L_FH_E > H_FL_E > L_FL_E$). Here, L1 frequency mainly determined the speed at which the cognates were read, while L2 frequency was less influential. Even more curious, L2 frequency did not behave as expected since low frequency words were read faster than high frequency words when L1 frequency was kept the same (note that this matched the main effect of English word frequency in the ANOVA for overall RTs).

The findings thus align more closely to the conclusions drawn by Winther et al. (2021) who argued, based on computational models which processed cognates and non-cognates similarly, that cumulative frequency could explain the cognate facilitation effect: the only difference between cognates and non-cognates in their models were the amount of exposure to the words in the L1 and the L2, and they still found that cognates were processed faster. This study did not investigate four different frequency combinations as done by Peeters et al. (2013), but their findings can be interpreted in the context of the *form-overlap* and *shared-morpheme view*. The findings in this study, while not significant, add to the findings in Winther et al. (2021) by providing tentative evidence in favour of cumulative frequency based on behavioural data.

The question remains why L2 frequency, which was a secondary predictor of RTs compared to L1 frequency, did not adhere to previously observed FEs in which high frequency words are processed faster than low frequency words. The effect of L2 frequency on cognate reading times was not significant in the LME models in Table 3 and 4, and while it trended in the expected direction, the estimated size of the effect was very small: an increase in English Zipf score would lead to a decrease in RT of around 1 ms. Like the effect of cognate status disappearing in the ANOVA, it could be that the grouping of RTs in the four different

conditions meant that the effect disappeared because it already was very weak in the LME model. While definitive conclusions cannot be drawn due to the absence of significant effects and lack of variation in the data, it seems that the *form-overlap* and *shared-morpheme view* are the most likely explanations for how cognates are represented in the mental lexicon. Both views argue that cognates share crucial representations that make them susceptible to the influence of cumulative word frequency during processing. The *shared-morpheme view* posits that cognates share a single morphological representation which causes effects of cumulative word frequency (Davis et al., 2009), while the *form-overlap view* point towards the single orthographic representation of cognates as the source (Midgley et al., 2011). The *two-morpheme view*, however, does not explain the data, since it argues against an effect of cumulative frequency because cognates would have two separate morphemes with their own word frequency (Dijkstra & Van Heuven, 2002; Van Hell & Dijkstra, 2002).

Limitations

While the findings provided multiple interesting points for discussion and angles for future research, the fact remains that there were only a few significant effects and that those were all fairly small compared to the effects found in previous studies. This was likely caused by the lack of variation in the RT data, as mentioned before. This uniformity is illustrated by the mean reading times for cognates and control words: the mean RT for cognates ($M = 336.35$, $SD = 82.83$) was roughly 7 ms lower than the mean RT for control words ($M = 343.19$, $SD = 81.45$), compared to the 32 ms difference found by Peeters et al. (2013). This discrepancy was likely caused by the self-paced reading task used to gather the data. The main advantage of self-paced reading tasks is that they make it easy to collect data in sentence context without it requiring specialist equipment (e.g., the cameras for eye-tracking tasks) or additional tasks that distract from the reading itself (e.g., a lexical decision task at the end of the sentence). However, participants are required to press a button multiple times per sentence for the entire experiment (here: 64 sentences with an average length of 12 words means approximately 768 button presses during the task). Even though participants are asked to pay attention while reading and to avoid falling into a certain rhythm, it is difficult to ensure this does not happen. When participants do end up losing focus, it could result in less accurate reading times and the lack of variation in the data found in this study.

There are two additional factors besides the inherent characteristics of the self-paced reading tasks that could have contributed to the less-than-optimal data, namely setting and

application. The experiment was conducted online to make it easier to recruit more participants since there seems to be fewer people willing to go to the lab for study participation at the time of writing. The downside of online studies is that it is impossible to control the environment in which participants complete the tasks, besides suggestions and warnings in the instructions. There is simply no way of knowing whether there was something in the room or outside that distracted a participant or if they completed the study while scrolling on their smartphone. Additionally, the choice to use ROLEG also did not help to guarantee the integrity of the data. While the application is easy to use for both experimenter and participant, and includes all the tools necessary for this study, it suffers from limited response time accuracy (ROLEG manual: [Response time accuracy](#)). Similarly, five participants reported at the end of the study that they noticed slight delays while reading. These delays should not have affected measuring the reading times themselves, according to the developers, but noticeable disruptions are likely to catch participants' attention and slow down their reading times when it distracts them. This means that replication of this study with more accurate software, such as PsychoPy, and conducting the experiment in a more controlled setting will be necessary to improve the reliability of the data. Such replication would provide insight into whether the lack of variation in the data was due to the participants, application and setting, or if it was caused by the stimulus material.

This study might also have tried to do too much within the available time and resources. Researching the effect of cumulative frequency on cognate facilitation effects in sentence context itself was already a novel approach to the topic, but language proficiency and word prevalence were also added to the design of the study based on findings from previous studies. Adding more factors means that the experimental design got more complicated, which meant certain compromises had to be made. This, in turn, may have reduced the purpose of certain factors. For instance, word prevalence was added but it was not possible to match the target words on both prevalence and frequency. The findings showed that word prevalence was more influential than word frequency, but it cannot be said with certainty that the target words were a representative sample of a normal distribution for word prevalence. Additionally, the statistical analyses were not decided on beforehand which meant that the stimulus materials and the data might not have been optimally prepared for every analysis. The most prominent example are the four conditions of cumulative word frequency used in the ANOVA. Peeters et al. (2013) made sure that cognates in separate conditions but with the same word frequency (e.g., cognates with high English frequency but low or high French frequency) were matched

on average word frequency. This was not possible for this study as I also tried to achieve a natural distribution of the Zipf scores for the LME models (see Figure 3), but that did mean that the conditions were less optimal for the ANOVA. It might have been better to divide this study into multiple smaller projects to ensure that different aspects of the study did not clash and weaken each other.

Future Research

While the findings in this study were not very strong, they provide enough possibilities for future research. Aside from the suggestion above to reduce the scope of the study to focus on either word frequency or word prevalence in the context of cumulative frequency, the current design could also be kept largely intact while making improvements. These improvements could include the setting and application of the study (i.e., conducting the experiment in a controlled environment with software fit to detect small deviations in reading times). It would furthermore help to either increase the area of interest during analysis beyond the target word or change methodologies from self-paced reading to eye-tracking. Research has shown that the effects of a certain word, such as cognate facilitation effects of interlingual homograph interference, can also affect the next word(s) (e.g., Hoversten & Traxler, 2015). It could be possible that the effects that were small or not significant in the current analysis would have been stronger if this so-called spillover region was also considered. This can be achieved by either adding the RTs of the next two words in the sentence to the analysis or by using chunks instead of individual words as a unit in the self-paced reading task. The latter would have the added benefit that participants would have to press the button less often because a single press of the button would not reveal one word but rather three or four. Alternatively, the experiment could be restructured into an eye-tracking task. This would also expand the area of interest as it becomes easier to see how participant react to cognates of different frequencies (e.g., do they regress less or more often when reading cognates with certain word frequencies?). As added benefit, it would also increase the ecological validity of the study compared to self-paced reading tasks since participants in an eye-tracking experiment can focus purely on reading and not on a secondary task such as pressing a button.

Another venue for investigation is the matter of cumulative frequency with a different set of languages. Dutch and English are both Germanic languages and bear several similarities. The global influence of English on language and culture in general is also felt in the Netherlands, intertwining both languages even further. Children in the Netherlands are exposed

to English at the end of primary school but are likely to encounter it outside of a school setting at an even earlier age. This raises the question to what extent differences in cumulative frequency, if it were to affect cognate facilitation effects, would be visible in Dutch-English bilinguals. The choice to use these languages for the present study was rooted in convenience: the researcher speaks both language at a high level and, more importantly, there are plenty of Dutch-English bilinguals in the country who would be suitable candidates. Schepens et al. (2013) found that cognates in languages that are closely related to each other are used more often than cognates in languages that are less similar. This means that the distinction between low and high frequency for English-Dutch cognates might be less clear than expected. After all, word frequency measures are taken from monolingual corpora that do not account for exposure in both languages. Peeters et al. (2013) investigated cumulative frequency effects with French and English for this reason; replicating their study in sentence context or working with another set of less-related languages could provide further insight into the effects of cumulative frequency (or lack thereof). Afterwards, it would be interesting to revisit language pairs such as Dutch and English to compare findings and see whether degree of relatedness influences the size or presence of the effects.

Finally, the finding that word prevalence ended up being a better predictor of reading times than word frequency bears further investigation. Due to the relative novelty of this word occurrence measure, the first step would be to gain more insight into how different prevalence scores affect language processing. This would be especially helpful to establish boundaries between lower and higher word prevalence. Knowing where to draw the line would allow for the replication of the current design and analysis of the effect of cumulative word prevalence with ANOVAs. If such a study could be designed, it would enable researchers to focus solely on word prevalence and properly match stimulus materials to each other to figure out whether the effects found in this study were affected by the lack of matching (see Figure 7). Another useful approach would be to collect L2 word prevalence data to use in studies on (cumulative) word frequency. The current databases for word frequency and prevalence, like the SUBTLEX, are based on data collected from native speakers and might therefore not reflect foreign language exposure.

Having access to an L2 measure would also help to create a single word occurrence measure. Currently, one of the biggest problems when investigating cumulative word frequency is that there is no measure that has been proven to express the combined frequency of a word in multiple languages. Studies can use Zipf scores or other frequency measures, but

these are based on monolingual input that are unlikely to reflect the bilingual nature of cumulative word frequency. Additionally, there is no evidence that adding the Zipf score in one language to the other properly represents cumulative word frequency. It would therefore be useful to create a single variable that could express these nuances so studies looking into cumulative word frequency would have a stronger foundation. This measure would need to reflect the language exposure of second language learners, which cannot be done with corpora based on sources such as movies. After all, there is no guarantee that L2 learners are exposed to the same sources as native speakers. Word prevalence, on the other hand, would lend itself well to this measure as second language learners can indicate whether they know a word to create a L2 variant of the current data available. The largest downside of word prevalence is that it is less effective for highly frequent words. However, this data could be provided by L1 corpora, since the most frequent words are also the ones to which learners are most likely to be exposed to. Creating such a measure would be a difficult undertaking as the word prevalence data would have to be collected in a megastudy, in addition to other methodological decisions that would have to be made during the process (e.g., how to deal with those highly frequent words). However, if this unified measure of cumulative word occurrence were to exist, it would open new possibilities for research and to gain more insight into the workings of cognates and how they are affected by word frequency effects.

Conclusion

As two of the most well-known facilitatory effects in language processing, word frequency and cognates have been researched in many different contexts with varying methods, providing new insights into bilingual language processing. A relatively unexplored area in these fields is the effect that cumulative word frequency might have on the processing and facilitation of identical cognates, and none of the previous studies examined this topic in sentence context. These words are unique since they exist in multiple languages and share form, meaning, and phonology in both, which means that the effect of word frequency on bilingual processing could come from either language. This study reported on a self-paced reading task in which Dutch-English bilinguals read sentences in their L2 containing cognates or control words. Analysis of the reading times of these target words revealed that the sample in this experiment might have been different from previous studies as familiar effects were either absent or reduced in size, most likely due to a lack of variation in the data; conclusions drawn from the data should therefore be verified through replication studies. Word prevalence turned out to be the better predictor of RTs when compared to word frequency. Additionally, tentative

evidence was found that cumulative frequency influenced the reading times of cognates since high L1 word frequency, rather than high L2 word frequency, led to the fastest reading times. The evidence thus suggests that the cognates are likely represented in the mental lexicon according to the *form-overlap* (Midgley et al., 2011) or *shared-morpheme view* (Davis et al., 2009). This finding contradicts the only other study on cumulative frequency and cognate facilitation working with behavioural data, who found evidence in favour of the *two-morpheme view* of identical cognate representation which argues that cumulative word frequency does not affect processing (Peeters et al., 2013).

Finding a more definitive answer to the question whether cumulative word frequency affects cognates would lead to a clearer idea on how these cognates are presented in the mental lexicon and how they interact with bilingual language processing models such as the BIA+ model (Dijkstra & Van Heuven, 2002). In a world where bi- and multilingualism is becoming the norm in many countries, it is even more relevant to understand the interaction between languages systems and how bilinguals function differently from monolingual speakers. This understanding can be achieved by researching cumulative word frequency effects, not only with cognates but also with interlingual homographs and other form-overlapping words. The rise of bilingualism also calls the validity of current corpora into question and the construction a word occurrence measure (whether that is word frequency, word prevalence or a combination) for second language learners seems crucial for future research.

References

- Allen, D., Conklin, K., & Miwa, K. (2020). Cross-linguistic lexical effects in different-script bilingual reading are modulated by task. *International Journal of Bilingualism*, 25(1), 168-188. <https://doi.org/10.1177/1367006920943974>
- 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. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database *Philadelphia, PA: University of Pennsylvania Linguistic Data Consortium.*
- Balling, L. W. (2013). Reading authentic texts: What counts as cognate? *Bilingualism: Language and Cognition*, 16(3), 637-653. <https://doi.org/10.1017/S1366728911000733>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3), 445-459. <https://doi.org/10.3758/bf03193014>
- Bosma, E., & Nota, N. (2020). Cognate facilitation in Frisian–Dutch bilingual children’s sentence reading: An eye-tracking study. *Journal of Experimental Child Psychology*, 189, 1-18. <https://doi.org/https://doi.org/10.1016/j.jecp.2019.104699>
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bolte, J., & Bohl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412-424. <https://doi.org/10.1027/1618-3169/a000123>
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45-50. <https://doi.org/10.1177/0963721417727521>
- Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, 51(2), 467-479. <https://doi.org/10.3758/s13428-018-1077-9>
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word

- frequency measure for American English. *Behavior Research Methods*, 41(4), 977-990. <https://doi.org/10.3758/BRM.41.4.977>
- 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>
- Brysbaert, M., Stevens, M., Mander, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 441-458. <https://doi.org/10.1037/xhp0000159>
- 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* (pp. 1-8). 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>
- Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2016). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, 20(4), 747-769. <https://doi.org/10.1017/s1366728916000213>
- Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin and Review*, 22(5), 1216-1234. <https://doi.org/10.3758/s13423-015-0819-2>
- Davis, C., Sánchez-Casas, R., García-Albea, J. E., Guasch, M., Molero, M., & Ferré, P. (2009). Masked translation priming: Varying language experience and word type with Spanish–English bilinguals. *Bilingualism: Language and Cognition*, 13(2), 137-155. <https://doi.org/10.1017/s1366728909990393>
- Desmet, T., & Duyck, W. (2007). Bilingual language processing. *Language and Linguistics Compass*, 1(3), 168-194. <https://doi.org/10.1111/j.1749-818X.2007.00008.x>
- Dijkstra, T., Van Hell, J. G., & Brenders, P. (2015). Sentence context effects in bilingual word recognition: Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition*, 18(4), 597-613. <https://doi.org/http://dx.doi.org/10.1017/S1366728914000388>

- 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.
- 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., & Brinke, S. T. (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>
- Duyck, W., Assche, E. V., 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>
- Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, 15(4), 850-855. <https://doi.org/10.3758/pbr.15.4.850>
- EF Education First. (2022). EF English Proficiency Index. <https://www.ef.com/wwen/epi/downloads/>
- Gerritsen, M., Nickerson, C., Van Hooft, A., Van Meurs, F., Nederstigt, U., Starren, M., & Crijns, R. (2007). English in product advertisements in Belgium, France, Germany, the Netherlands and Spain. *World Englishes*, 26(3), 291-315. <https://doi.org/10.1111/j.1467-971X.2007.00510.x>
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), 3-15. [https://doi.org/10.1016/0093-934x\(89\)90048-5](https://doi.org/10.1016/0093-934x(89)90048-5)
- Hoversten, L. J., & Traxler, M. J. (2015). 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>
- Jegerski, J. (2013). Self-paced reading. In *Research methods in second language psycholinguistics* (pp. 36-65). Routledge.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch word based on film subtitles. *Behavior Research Methods*, 42(3), 643-650. <https://doi.org/10.3758/BRM.42.3.643>

- Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, 68(8), 1665-1692. <https://doi.org/10.1080/17470218.2015.1022560>
- Koops, E. J., Slop, P., Uljé, P., Vermeij, K., & van Zijderveld, D. (2009). *Funshoppen in het Nederlands: Woordenlijst onnodig Engels*. Bert Bakker.
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In *Psychology of learning and motivation* (Vol. 56, pp. 229-262). Elsevier. <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-394393-4.00007-8>
- Kroll, J. F., Gullifer, J. W., & Rossi, E. (2013). The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two or more languages. *Annual Review of Applied Linguistics*, 33, 102-127. <https://doi.org/10.1017/s0267190513000111>
- Kucera, H., & Francis, W. (1967). *Computational analysis of present day American English*. Brown University Press.
- Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 65(1), 42-73. <https://doi.org/10.1016/j.jml.2011.03.002>
- Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, 92, 217-233. <https://doi.org/http://dx.doi.org/10.1016/j.jml.2016.06.010>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325-343. <https://doi.org/10.3758/s13428-011-0146-0>
- Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, 32(4), 533-550. <https://doi.org/10.3758/bf03195845>
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 12-31. <https://doi.org/10.1037/0278-7393.34.1.12>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. The MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001>

- 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>
- Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, 23(7), 1634-1647. <https://doi.org/10.1162/jocn.2010.21463>
- Otwinowska, A., & Szewczyk, J. M. (2017). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*, 22(8), 974-991. <https://doi.org/10.1080/13670050.2017.1325834>
- Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, 22(12), 2728-2744. <https://doi.org/10.1162/jocn.2009.21393>
- Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, 68(4), 315-332. <https://doi.org/https://doi.org/10.1016/j.jml.2012.12.003>
- Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 787-796. <https://doi.org/10.1037/a0035583>
- Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLoS One*, 8(5), e63006. <https://doi.org/10.1371/journal.pone.0063006>
- Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1-L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1412-1431. <https://doi.org/10.1037/A0024492>
- Van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts: The effect of cognate status and verb tense. *Studies in Second Language Acquisition*, 35(2), 237-259. <https://doi.org/https://doi.org/10.1017/S0272263112000873>
- Van Assche, E., Duyck, W., & Hartsuiker, R. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology*, 3, 1-8. <https://doi.org/10.3389/fpsyg.2012.00174>

- 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>
- Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62, 148-171. <https://doi.org/10.1111/j.1467-9922.2012.00710.x>
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176-1190. <https://doi.org/10.1080/17470218.2013.850521>
- Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, 35(5), 938-952. <https://doi.org/10.3758/bf03193467>
- Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, 19(1), 73-80. <https://doi.org/10.3758/s13423-011-0179-5>
- Winther, I. E., Matushevych, Y., & Pickering, M. J. (2021). Cumulative frequency can explain cognate facilitation in language models. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43. <https://doi.org/10.31234/osf.io/e2vft>

Appendix A: Cognates

cognate	condition	length	syllables	word frequency		word prevalence		Zipf difference
				EN	DU	EN	DU	
accountant	1	10	3	3,792	3,762	2,339	1,736	-0,029
continent	1	9	3	3,566	3,473	2,252	1,817	-0,093
coupon	1	6	2	3,341	2,469	2,434	1,681	-0,872
dilemma	1	7	3	3,419	3,469	2,066	1,826	0,051
fiasco	1	6	3	3,216	3,409	1,894	1,707	0,192
imperium	1	8	4	1,593	3,315	0,760	1,724	1,722
jargon	1	6	2	2,797	2,818	1,847	1,608	0,021
kiosk	1	5	2	2,672	3,219	1,866	1,713	0,547
kiwi	1	4	2	2,905	2,818	2,325	1,794	-0,087
lobbyist	1	8	3	2,654	2,611	1,701	1,578	-0,043
toffee	1	6	2	2,654	2,999	2,112	1,834	0,345
marathon	1	8	3	3,652	3,397	2,441	1,914	-0,255
optimist	1	8	3	3,189	3,111	2,331	1,960	-0,078
panda	1	5	2	3,329	2,989	2,257	1,847	-0,340
alumnus	1	7	3	2,690	1,833	1,103	0,433	-0,857
supporter	1	9	3	2,872	2,753	2,322	1,793	-0,118
alibi	2	5	3	3,922	4,175	1,769	1,829	0,253
angst	2	5	1	2,973	4,837	1,774	1,826	1,864
cent	2	4	1	3,977	4,564	2,183	1,960	0,587
chaos	2	5	2	3,973	4,196	2,341	1,733	0,223
chauffeur	2	9	2	3,751	4,470	2,102	1,830	0,719
cola	2	4	2	3,744	4,056	2,326	1,960	0,313
commandant	2	10	3	3,412	4,658	1,219	1,576	1,245
sector	2	6	2	3,849	4,030	2,234	1,791	0,181
informant	2	9	3	3,622	4,072	2,330	1,677	0,449
internet	2	8	3	3,817	4,232	2,432	1,960	0,415
journalist	2	10	3	3,786	4,209	2,576	1,793	0,423
kilo	2	4	2	3,200	4,451	1,991	1,914	1,251
kilometer	2	9	4	2,614	4,692	2,235	1,914	2,078
revolver	2	8	3	3,444	4,176	2,234	1,826	0,732
tempo	2	5	2	3,231	4,068	2,116	1,832	0,837
whisky	2	6	2	3,606	4,475	1,860	1,594	0,869
baby	3	4	2	5,706	5,178	2,254	1,960	-0,529

camera	3	6	3	4,755	4,781	2,576	1,832	0,025
circus	3	6	2	4,232	4,136	2,319	1,960	-0,096
concert	3	7	2	4,244	4,143	2,338	1,960	-0,101
contract	3	8	2	4,510	4,547	2,442	1,826	0,037
incident	3	8	3	4,238	4,068	2,576	1,790	-0,170
moment	3	6	2	5,271	5,401	2,432	1,831	0,130
professor	3	9	3	4,842	4,824	2,239	1,537	-0,018
restaurant	3	10	3	4,667	4,617	2,334	1,837	-0,050
school	3	6	1	5,522	5,389	2,440	1,836	-0,133
senator	3	7	3	4,520	4,582	2,331	1,912	0,061
sheriff	3	7	2	4,785	4,854	2,097	1,119	0,068
basis	3	5	2	4,077	4,537	2,227	1,915	0,460
team	3	4	1	5,169	5,054	2,324	1,742	-0,114
water	3	5	2	5,352	5,385	2,576	1,913	0,033
week	3	4	1	5,377	5,369	2,346	1,911	-0,008
album	4	5	2	4,010	3,760	2,431	1,789	-0,250
ballet	4	6	2	4,011	3,788	2,161	1,960	-0,223
baron	4	5	2	4,129	3,883	2,095	1,828	-0,246
concept	4	7	2	4,035	3,767	2,576	1,912	-0,268
dessert	4	7	2	4,147	3,734	2,227	1,683	-0,413
district	4	8	2	4,321	3,928	2,576	1,755	-0,392
hockey	4	6	2	4,062	3,449	2,576	1,656	-0,613
item	4	4	2	4,090	3,264	2,435	1,666	-0,826
madam	4	5	2	4,643	3,573	1,894	1,377	-1,070
magazine	4	8	3	4,521	3,466	2,433	1,723	-1,055
producer	4	8	3	4,096	3,818	2,576	1,647	-0,278
punk	4	4	1	4,342	3,009	2,172	1,635	-1,333
reporter	4	8	3	4,309	3,534	2,576	1,763	-0,775
tennis	4	6	2	4,134	3,665	2,576	1,960	-0,469
transcript	4	10	2	4,141	2,310	2,427	1,440	-1,831
trauma	4	6	2	4,231	3,736	2,439	1,910	-0,496

Note. Word frequency is expressed as Zipf score (1 – 7 with a score below four for low frequency words and a score above 4 for high frequency words) and word prevalence is expressed as z-scores.

Note 2. The difference in Zipf score is negative if the word is more frequent in English while the difference is positive if the word is more frequent in Dutch.

Appendix B: Control Words

control	length	syllables	word frequency	word prevalence
researcher	10	3	2,883	2,303
physicist	9	3	3,137	2,049
pastry	6	2	3,287	2,200
forgery	7	3	3,296	2,139
idiocy	6	3	2,769	1,833
tapestry	8	3	2,860	2,211
anthem	6	2	3,279	2,576
crypt	5	1	3,143	2,052
lute	4	1	2,593	1,301
reviewer	8	3	2,723	2,224
famine	6	2	3,206	2,105
campfire	8	3	3,105	2,254
outsider	8	3	3,378	2,576
rhino	5	2	3,321	2,436
prodigy	7	3	3,124	2,119
foreigner	9	3	3,471	2,325
relic	5	2	3,131	2,258
agony	5	3	3,575	2,181
cube	4	1	3,477	2,576
havoc	5	2	3,137	2,043
spokesman	9	2	3,077	2,307
tuna	4	2	3,904	2,431
accomplice	10	3	3,628	2,165
runway	6	2	3,739	2,235
caretaker	9	3	3,392	2,441
equation	8	3	3,545	2,119
apprentice	10	3	3,292	2,330
acre	4	2	3,265	2,245
proximity	9	4	3,221	1,938
beverage	8	3	3,371	2,442
shelf	5	1	3,843	2,427
celery	6	3	3,274	2,167
aunt	4	1	4,741	2,431
wallet	6	2	4,358	2,576

estate	6	2	4,324	2,234
airport	7	2	4,580	2,346
marriage	8	2	4,886	2,448
accident	8	3	4,910	2,444
choice	6	1	4,989	2,576
gentleman	9	3	4,621	2,429
government	10	3	4,814	2,576
prison	6	2	4,819	2,325
soldier	7	2	4,590	2,331
grandma	7	2	4,658	2,433
image	5	2	4,354	2,340
army	4	2	4,932	2,425
candy	5	2	4,553	2,246
hour	4	2	5,210	2,440
cabin	5	2	4,293	2,440
soccer	6	2	4,099	2,322
eagle	5	2	4,060	2,324
verdict	7	2	4,039	2,318
pumpkin	7	2	4,035	2,432
entrance	8	2	4,195	2,338
liquor	6	2	4,238	2,244
poem	4	2	4,135	2,419
lover	5	2	4,425	2,428
solution	8	3	4,199	2,431
elephant	8	3	4,056	2,431
guts	4	1	4,375	2,434
customer	8	3	4,182	2,576
parole	6	2	4,019	1,926
confession	10	3	4,217	2,576
pillow	6	2	4,057	2,271

Note. Word frequency is expressed as Zipf score (1 – 7 with a score below four for low frequency words and a score above 4 for high frequency words) and word prevalence is expressed as z-scores.

Appendix C: Stimulus Sentences and Comprehension Questions

Stimulus sentences

- 1 The little boy wished to become a(n) **ACCOUNTANT/RESEARCHER** when he was older.
- 2 The teacher gave a lecture about the **CONTINENT/PHYSICIST** at the beginning of the year.
- 3 The competitor came in last and got a **COUPON/PASTRY** from the baker.
- 4 The newlyweds were left with a **DILEMMA/FORGERY** after their honeymoon ended.
- 5 The failed assassination would go down as the **FIASCO/IDIOCY** of the last century.
- 6 The young socialite was wondering when the **IMPERIUM/TAPESTRY** would be hers.
- 7 The engineer tried to convince his peers that the **JARGON/ANTHEM** sounded completely ridiculous.
- 8 The traveller was gutted that the **KIOSK/CRYPT** had closed for the night.
- 9 The uncle gave his niece a **KIWI/LUTE** for her tenth birthday.
- 10 The company was unable to convince the **LOBBYIST/REVIEWER** of their proposal's worth.
- 11 The lecturer decided to give a talk about **TOFFEE/FAMINE** and its history.
- 12 The French village wanted to organise an annual **MARATHON/CAMPFIRE** to boost tourism.
- 13 The wife told her husband he was the **OPTIMIST/OUTSIDER** in their family.
- 14 The children were excited to see a **PANDA/RHINO** at the zoo.
- 15 Many considered her to be the most successful **ALUMNUS/PRODIGY** in twenty years.
- 16 The shopkeeper decided to help the **SUPPORTER/FOREIGNER** to find their destination.
- 17 The princess doubted the authenticity of the **ALIBI/RELIC** her husband gave her.
- 18 The nanny was surprised by the **ANGST/AGONY** radiating from the teenager.
- 19 The toddler found and returned the **CENT/CUBE** to its rightful owner.
- 20 The social worker was dumbfounded by the **CHAOS/HAVOC** she found in the house.
- 21 The graduate was unsure if he could be the **CHAUFFEUR/SPOKESMAN** the company needed.
- 22 The roommate was upset when there was no more **COLA/TUNA** left in stock.
- 23 The lawyer formally requested that the **COMMANDANT/ACCOMPLICE** was called to testify.
- 24 The spy misremembered the place of the **SECTOR/RUNWAY** at a crucial time.
- 25 The cleaning lady was waiting for the **INFORMANT/CARETAKER** to turn up.
- 26 The firefighters were arguing about the **INTERNET/EQUATION** and its usefulness.
- 27 The new neighbour worked as a(n) **JOURNALIST/APPRENTICE** at the local newspaper.
- 28 The gardener debated the necessity of the **KILO/ACRE** as a unit of measurement.
- 29 The librarian had to check everything within a/the **KILOMETER/PROXIMITY** from the forest.
- 30 The nurse told her friends about the **REVOLVER/BEVERAGE** she had found in the trash.
- 31 The waiters agreed on what the right **TEMPO/SHELF** should be in the end.
- 32 The sophomore swore he would never touch **WHISKY/CELERY** again in his life.
- 33 The boyfriend was excited to meet the **BABY/AUNT** for the first time.
- 34 The guard decided she wanted a new **CAMERA/WALLET** for her birthday.
- 35 The dentist announced that he had bought a(n) **CIRCUS/ESTATE** for his newborn child.
- 36 The army chief ordered to close the **CONCERT/AIRPORT** for the public's safety.
- 37 The pair hesitated to agree to the **CONTRACT/MARRIAGE** his parents proposed.
- 38 The thieves swore to never speak of the **INCIDENT/ACCIDENT** ever again.
- 39 The strategist thought it was the right **MOMENT/CHOICE** to win the war.
- 40 The chancellor clapped and congratulated the **PROFESSOR/GENTLEMAN** with their award.
- 41 The dean did not understand why the **RESTAURANT/GOVERNMENT** decided to shut down.
- 42 The neighbourhood association agreed to build a **SCHOOL/PRISON** for troubled teenagers.
- 43 The majority of the people nominated the **SENATOR/SOLDIER** for the highest honour.
- 44 The biologist was anxious and called the **SHERIFF/GRANDMA** for moral support.
- 45 The screenwriter had to look up the **BASIS/IMAGE** on the internet.
- 46 The translator formally requested that the **TEAM/ARMY** helped with the assignment.
- 47 The dog got yelled at for stealing **WATER/CANDY** off the dinner table.

- 48 The bartender forgot to write down the **WEEK/HOUR** of her next appointment.
 - 49 The salesperson was distraught because he lost the **ALBUM/CABIN** his father had given him.
 - 50 The employee strongly felt that watching **BALLET/SOCCER** was a waste of time.
 - 51 The farmer did not believe that a(n) **BARON/EAGLE** was spotted in their village.
 - 52 The musician failed to fully understand the **CONCEPT/VERDICT** and wept in frustration.
 - 53 The pregnant woman was hoping to buy a **DESSERT/PUMPKIN** to satisfy her cravings.
 - 54 The hunter asked for directions to the **DISTRICT/ENTRANCE** before driving away.
 - 55 The grandpa felt no one should be introduced to **HOCKEY/LIQUOR** before turning thirty.
 - 56 The painter wondered why someone would want a(n) **ITEM/POEM** on banana trees.
 - 57 The novelist had never imagined someone calling her **MADAM/LOVER** to her face.
 - 58 The carpenter only managed to find the **MAGAZINE/SOLUTION** after calling his husband.
 - 59 The cop wanted to take a picture with the **PRODUCER/ELEPHANT** for his nephews.
 - 60 The aunt scolded the preschooler for saying **PUNK/GUTS** to her baby sister.
 - 61 The investigator recorded a conversation with the **REPORTER/CUSTOMER** on the recent events.
 - 62 The queen wanted to learn more about **TENNIS/PAROLE** before the end of the year.
 - 63 The squad confessed that they buried the **TRANSCRIPT/CONFESSION** out of fear.
 - 64 The wizard denied knowing about the **TRAUMA/PILLOW** when asked by his apprentice.
-

Note. The target words are capitalised and in bold; the control word always follows the cognate.

Comprehension questions	Sentence	Answer
1 Did the teacher give a lecture at the end of the year?	2	no
2 Was the assassination a success?	5	no
3 Was the socialite young?	6	yes
4 Was the village French?	12	yes
5 Was the social worker dumbfounded?	20	yes
6 Did the firefighters agree about the usefulness?	26	no
7 Did the gardener debate the existence of the unit of measurement?	28	no
8 Were the waiters in agreement?	31	yes
9 Did the guard make a decision?	34	yes
10 Did the army chief give an order?	36	yes
11 Did everyone support the nomination?	43	no
12 Did the bartender write down the details of their appointment?	48	no
13 Was the salesperson upset?	49	yes
14 Did the hunter already know the way?	54	no
15 Did the aunt praise the preschooler?	60	no
16 Did the wizard deny knowing about it?	64	yes

Note. The order in which the comprehension questions appeared in each list was different and they did not appear in the order in which they are listed here.