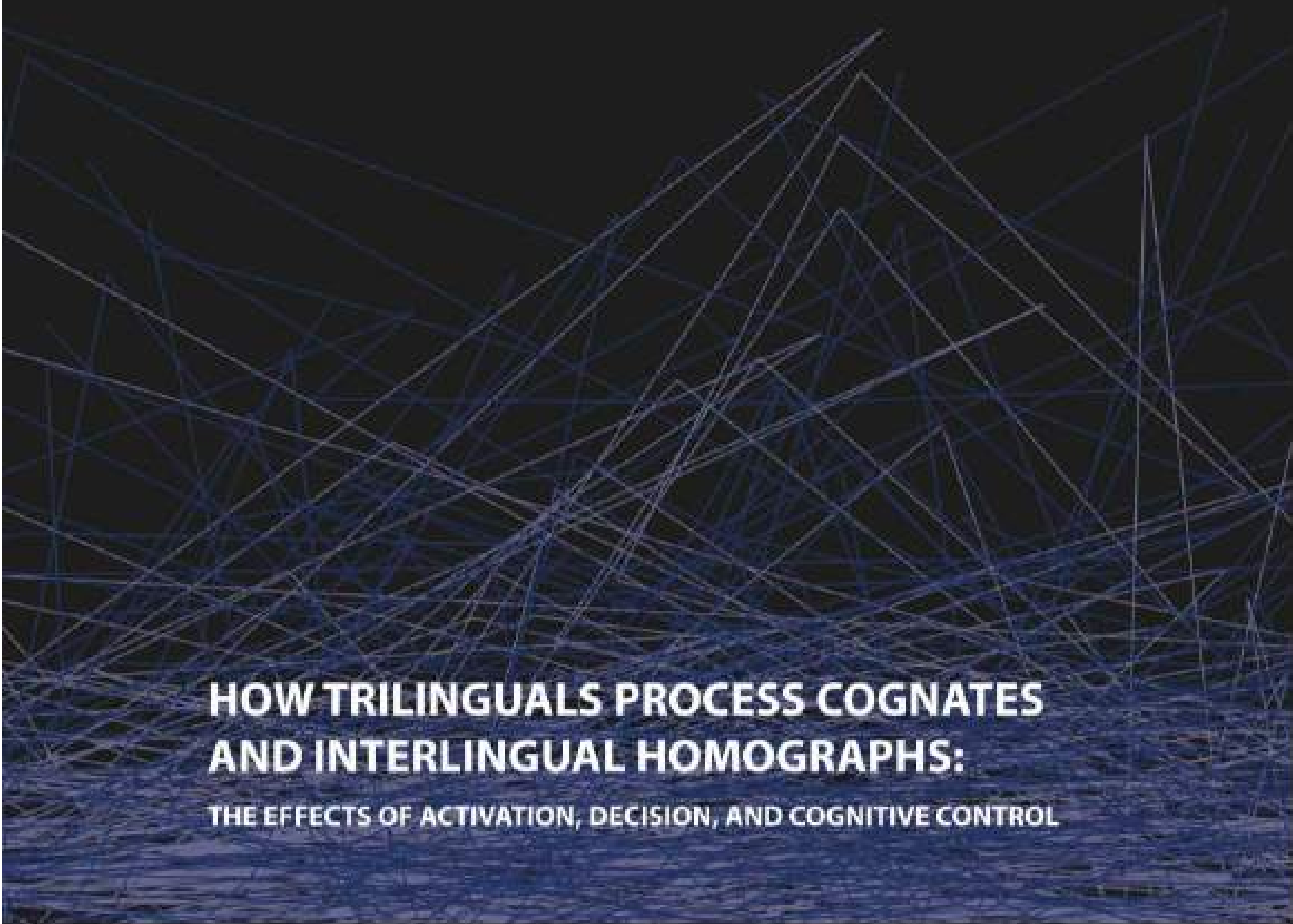
Faculteit Letteren en Wijsbegeerte
Departement Taalkunde

Proefschrift voorgelegd tot het behalen van de graad van doctor
in de Taalkunde aan de Universiteit Antwerpen



HOW TRILINGUALS PROCESS COGNATES AND INTERLINGUAL HOMOGRAPHS:

THE EFFECTS OF ACTIVATION, DECISION, AND COGNITIVE CONTROL

University of Antwerp

CLiPS
Centrum voor Computerlinguïstiek & Psycholinguïstiek
Universiteit Antwerpen

fwo
Fonds Wetenschappelijk Onderzoek
Vlaanderen
Opening new horizons

IHOR BORISOVITCH BILOUSHCHENKO

Promotoren:
Prof dr. Dominiek Sandra
Prof dr. Ton Dijkstra

Antwerpen, 2017

Universiteit Antwerpen
Faculteit Letteren en Wijsbegeerte
Departement Taalkunde

HOW TRILINGUALS PROCESS COGNATES AND INTERLINGUAL HOMOGRAPHS: THE EFFECTS OF ACTIVATION, DECISION, AND COGNITIVE CONTROL

HOE TRILINGUALEN COGNATEN EN INTERLINGUALE HOMOGRAFEN VERWERKEN: DE EFFECTEN VAN ACTIVATIE, DECISIE EN COGNITIEVE CONTROLE

Proefschrift voorgelegd tot het behalen van de graad van doctor
in de Taalkunde aan de Universiteit Antwerpen

te verdedigen door

Ihor Borisovitch Biloushchenko

Promotoren:
Prof dr. Dominiek Sandra
Prof dr. Ton Dijkstra

Antwerpen, 2017

Dedicated to my mother,
who believed in me the most.

Preface

How do you get so far?

Already while at school, I was curious about how brain processes languages and deals with the variety in different languages. But at that time, I was thinking in primitive words and didn't even know that it was psycholinguistics that I was interested in. Through my studies in the Ukraine in languages, pedagogy and psychology I started out on my journey which took 7 years until I finally found my place in the science. Unfortunately, my initial studies didn't give me the possibility to combine different fields into one research, and hence that is why I was always looking out for new challenges. Sitting on the plane to Europe, I was holding some copies taken from the handbooks in psychology that described the research of Vygotsky about Thought and Language. It was all I had access to at that time, but it was my dream to do research in psycholinguistics.

After a hard struggle to win a place at a Belgian university and finishing a Master study in Psycholinguistics, I got lucky to finally receive FWO-funding to research visual recognition in multilinguals.

It is worth mentioning that in my opinion there were different factors that shaped my scientific view through life, but three of them were especially prominent. The first was being raised in a bilingual environment. The constant contact with two languages made me understand that one can utter the same thing differently depending on the environment and/or person. The second was my relationship to reading, not that I do or don't like to read but it costs me more time than the usual person as I believe to think. This is because of my excitement and deeper focus on the words that sit on the pages. Often I read sentences multiple times to enjoy the word use and interesting combinations between them. The third was my life's passion with creativity. It helped me to dare in science, to try novel and risky experimental settings. All these three factors supported me during my seven years of research.

You don't do it alone

Of course, it wouldn't be possible, or at least it would have taken me much longer to finish the research without great support. Firstly, the enthusiasm of Dominiek Sandra gave me confidence and support with his passion for psycholinguistics. The experimental results were not always satisfying and clear, but Dominiek was always supportive in all the ups and downs through the years. Although officially Ton Dijkstra came later, he was permanently present in this research from those early days when only some gloomy items were selected from the pool of the overlapping lexicons. In flesh, he came also at the right time as when I started to doubt, who needs this research and whether I don't just lose my life in doing it, he renewed the passion by his enthusiasm for the results and the experimental design. Great discussions with both supervisors helped me to grow in my scientific expertise.

Also the excellent statistical support made this research much stronger in its message. In particular, Sven De Maeyer, Jarl Kampen and Ardi Roelofs shaped my view on statistical analyses and the choice of suitable methods. Also, Wim Van der Elst was very helpful in all the difficult situations when I came to analysing my data.

Of course I'm grateful to my friends, relatives and colleagues who were always curious about my development in this research, and who unintentionally helped me to organise my findings. For them I learned to formulate the difficult theories in accessible words. Very special thanks go to Duncan who got lucky to be born in an English speaking country and was very kind to read through the whole text to find the inconsistencies in the language.

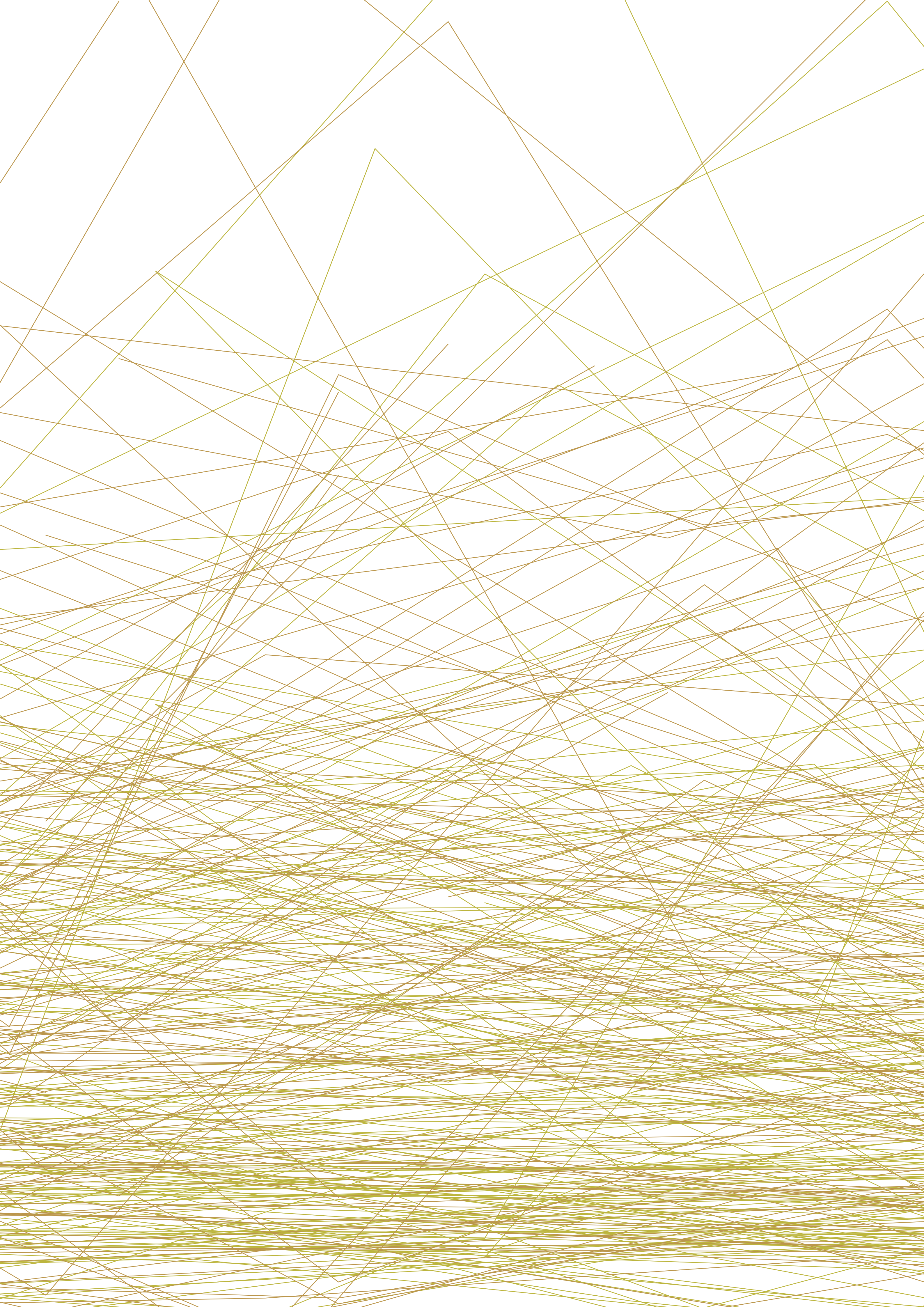
Also, I'm grateful to my jury (Sarah Bernolet, Steven Gillis, Marc Brysbaert and Harald Baayen) who were always a great example to me of passion for the science. They are very worthy of my special appreciation for finding the time to read and judge this dissertation and to be present at the defence.

And last but not least, people (sometimes even friends) who doubt my capacity to finish a PhD based on decent research. This disbelief motivated me to write my dissertation as good as I can to prove my expertise and confidence that all the years were not lost for nothing.

Finally

Finally, here is the physical evidence of my dissertation that once seemed so long away in my dreams and that I can now hold in my hands. This final version went through so many different revisions accompanied by many enriching discussions. Although this is the final point in my PhD story, it proposes new ideas and questions for new research in the deep ocean of psycholinguistics.

CONTENTS



CHAPTER 1
General Introduction

1.1 Basic questions in the investigation of bilingual lexical processing

Todays' growing international connections and the importance of communication with people across the globe since the rise of the electronic media has resulted in large numbers of bilinguals and multilinguals[1]. Not all these people have been raised in two languages from birth (so-called balanced bilinguals), but use one or several non-native languages in their study or job context (so-called unbalanced bilinguals). These large numbers of bilinguals are important from different perspectives. Bilingualism is crucial for economic relations, for the facilitation of social contacts, for language teaching, for debates about the importance of English as a teaching language for certain university courses, and for many other topics. At the same time, it raises fundamental questions about the way the human mind works when a person knows more than one language.

One of the most intriguing observations is that language users do not continually mix their languages. For instance, it seldom happens that bilinguals experience cross-language intrusion errors while speaking in one of their languages. As a matter of fact, rather than randomly mixing words from different languages, they are quite aware of word-finding difficulties in a non-native language. On such occasions, they will often explicitly describe the concept and ask which word they should use. Note that the notion of mixing up words does not refer to the well-known phenomenon of code-switching, i.e., our ability to switch from one language (e.g., Dutch) to another one (e.g., English) when somebody who does not understand the current language joins the group. Just like word-finding difficulties are a conscious experience when speaking a non-native language, language switching is a conscious, deliberate act. In contrast, mixing up words from different languages would result from our inability to unconsciously suppress irrelevant lexical information, which would lead to unintended errors. Also note that language intermixing does not mean that language users never allow words from the other language 'seep through' in their speech. For instance, Dutch-speaking students in informatics and statistics like to use words like 'random', even in ordinary contexts. In contrast to unintended second language intrusions such words result from a conscious choice to use a 'fancy' word that occurs with a high frequency in their courses, and also acts as a verbal signal that they belong to a prestige group. In contrast to the well-known deliberate choice to switch to a different language or the occasional use a second language word as a substitute for a native language word, the evidence confirms our intuition that there are indeed very few uncontrollable intrusions from a second language when people use a foreign language. Gollan, Sandoval, and Salmon (2011) asked younger and older Spanish-English bilinguals to generate as many category members (e.g., fruit or words beginning with an **s**) in a minute and concluded that "Cross-language intrusions were the least common error type. Even older bilinguals […] experienced failure of language control only about 1% of the time on average." (p. 1163).

[1] For the sake of simplicity, I will henceforth use the term 'bilingual' to refer to someone who speaks more than one language, irrespective of the number of foreign languages.

Of course, mixing up languages all the time would make it very difficult (and sometimes impossible) to communicate. For this simple reason, it is necessary that language users develop a mechanism for suppressing the language(s) that is (are) not the target language. However, the fact that language users can do this, raises an important questions for psycholinguists. Which process makes it possible to avoid lexical intrusions from another language? Upon closer inspection, the answer to this question depends on the answer to two other questions, which pertain to the way in which words from different language are organized in memory (the representational question) and the way in which these representations are accessed (the access question). Suppose the vocabularies of different languages are kept separate in memory, such that they can be selectively 'opened' or 'closed'. Such a storage design would readily explain why lexical cross-language intrusions are virtually non-existent. The communicative context would selectively activate the relevant language and suppress the other(s), which would include selective access to the language's mental lexicon. The question is whether it is indeed the case that the words of a second language are stored in a separate memory store, i.e., a second mental lexicon besides the mental lexicon for the native language. And whether a new mental lexicon is added each time we learn a novel language? Without giving this question much thought, the intuitively most appealing answer is that the words from different languages are indeed stored separately, such that the vocabulary of each language can be 'switched on' or 'switched off', as a function of the communicative needs. This view also entails an answer to the question how bilinguals achieve access to the words of one of their languages and make so few cross-language lexical intrusion errors. As all irrelevant languages are in a 'closed' state, the access process would not be able to activate lexical representations stored in these mental vocabularies. Such a representational system would enable language users to selectively access the words in one language without interference from erroneously activations in the other mental lexicons.

Obviously, scientific questions cannot be answered on the basis of intuition or introspection[2]. Hence, an alternative view must also be considered, even though it seems at odds with our conscious experience. The choice of the term 'conscious' is deliberate. As mental representations of words and access to them are (by definition) impermeable to consciousness, consciousness and the intuition derived from it is by definition a guide that should not be trusted in this debate. This alternative view rejects the notion of separate lexicons and the idea that they may be 'sealed off' for access when they are irrelevant in the language context. Rather, words from non-native languages are simply added to the mental lexicon of the native language, in the same way that novel native words are added. According to this view, the brain treats all words in the same way and stores them in a single memory store, housing the lexical representations of the words in all languages known by a language user. Obviously access to such an integrated, language-independent mental lexicon can only be language non-selective, as the integrated lexicon is always in an 'open' state. Hence, lexical access must be a stimulus-driven, bottom-up process, such that all representations that fully or closely match the orthographic pattern of the stimulus are activated, irrespective of whether they belong to the relevant or irrelevant language. Obviously, activated representations that are irrelevant in the language context will be inhibited at a post-lexical level, but initially they will become active.

---

[2]Actually, the history of science, especially the big discoveries, prove that our intuition is almost always wrong. Our intuition would not lead to the idea of a round earth, of an earth that revolves around the sun instead of the opposite, of the intimate bond between time and space, of the way biological evolution works, of the possibility that pure chance may lead to the emergence of intricate structure like the eye, etcetera.

Whereas the hypothesis about separate lexicons predicts that a Dutch-English bilingual who reads the word ***room*** in England will only access the English representation of this word, the hypothesis that posits a single lexicon predicts that the orthographic pattern of the stimulus, being compatible with both an English and Dutch orthographic representation will access both representations. The bottom-up access process in the latter hypothesis is only sensitive to characteristics of the stimulus and is blind to any contextual information, in the broadest sense of the word (e.g., the fact that one is in England, which makes it unlikely to come across a Dutch word). If this is how the system is designed, the answer to the question how language users can, seemingly effortlessly, suppress irrelevant vocabularies would considerably differ from the one in terms of the switchboard metaphor implied by the separate lexicons hypothesis. When there is only one mental lexicon, the vocabularies of irrelevant languages could not be switched off by a top-down suppression process before the onset of the lexical access process. Rather the lexicon of irrelevant languages might be inhibited because another lexicon is more active or the inhibition might be restricted to lexical representations that have been erroneously activated by a bottom-up process, as they are irrelevant in the communicative context. Note that the views of separate lexicons versus an integrated lexicon are both compatible with the idea that language users may differ in their ability to suppress the irrelevant information. Even if most suppression attempts are successful, as cross-language lexical intrusions seldom occur, some language users may be faster than others. This is the issue of cognitive control in the context of bilingual language use, i.e., the ability to suppress conflicting lexical activation. The process responsible for this may be part of the language user's executive control system.

The question about cognitive control will only be touched upon in the final chapter (Chapter 5) of this thesis. The majority of the experiments that will be reported in Chapters 2, 3 and 4 concern the representational and access questions mentioned above, which are so fundamental that they should be answered before the issue of language control can be addressed in a satisfactory way.

1.2 The dawn of lexical processing research in bilinguals

Many psycholinguistic experiments have been designed to address these questions. However, contrary to what one might expect, considering the many people in the world who speak at least two languages (certainly if one includes the native dialect of many people as a form of 'soft bilingualism'), it took quite some time before psycholinguists realized that the bilingual brain is the norm and that monolinguals are rather the exception (at least in the Western world). Before the nineties of the previous century, mainstream psycholinguistics focused almost exclusively on the native language of the experimental participants, thus effectively treating them as monolinguals. With the benefit of hindsight this is understandable. Psycholinguistics, as we conceive of it today, only emerged in the late 1960s (e.g., John Morton's, 1969, logogen model) and many of the first psycholinguists set themselves the goal of testing the 'psychological reality' of Chomskyan linguistics, which was an entirely misguided enterprise (Sandra, 2009)[3]. It was only after psycholinguists had learnt to define their own agenda and to

[3]It is ironic to see that Chomsky himself has always warned against such an interpretation of his theory (Chomsky, 1965, p. 9). His focus was on the boundary conditions that make language possible (so-called competence) whereas the use of this (purportedly) innate knowledge in real-life contexts (so-called performance) is what did not interest him but was, by its very nature, the study object of psycholinguists. Consequently, all experiments attempting to prove the correctness of the Chomskyan claims were doomed to fail, as such tests should not have concerned the mental processes during language use.

successfully apply several experimental paradigms to a variety of lexical processing issues in monolingual situations that they realized the importance of bilingual language processing. The real start of the 'bilingual era' in mental lexicon research is situated in the 1990s, although there had been some attempts to study bilinguals before that time (e.g., Gerard & Scarborough, 1989; Grosjean, 1982; Paradis, 1978)[4].

Suddenly, questions concerning the representation of the lexicons of different languages and the nature of the associated access process sparked the interest of many researchers and produced a lot of experiments. Most of these experiments have been carried out in the domains of speaking and reading (e.g., Finkbeiner, Almeida, Janssen, & Caramazza, 2006; Finkbeiner, Gollan, & Caramazza, 2006; Ivanova & Costa, 2008, in the domain of speaking; e.g., Brysbaert & Duyck, 2010; Dijkstra, Timmermans, & Schriefers, 2000; Dijkstra & van Heuven, 2002; Duyck, Assche, Drieghe, & Hartsuiker, 2007; Kroll & Stewart, 1994; Libben & Titone, 2009; Van Wijnendaele & Brysbaert, 2002, in the domain of reading). All experiments in this thesis will deal with the representation and access of words during the process of visual word recognition.

As mentioned above, the large majority of studies address whether the words of different languages that a language user knows are stored in separate mental lexicons or in an integrated mental lexicon. Researchers attempt to answer this question by investigating which information in the language system is accessed by the lexical input, more particularly, which lexical representations are accessed by it. At the level of production (e.g., speaking) the question is whether a concept will only activate the lemma in the target language or all lemmas representing this concept in the diversity of languages that the speaker masters. At the level of perception (e.g., reading) the question is whether the stimulus will access the word in the target language only or trigger a language non-selective access procedure, which will result in the activation of all lexical representations that (partially) match the spoken or written stimulus, independently of the language they belong to.

As mentioned earlier, the language experience of most native speakers of Dutch makes it hard to believe that they would activate the 'cream' meaning of the Dutch word ***room*** when reading that word in London. At the same time, such interferences occasionally occur. As Dijkstra, Timmermans, and Schriefers (2000) write, many people in the Netherlands and Flanders misinterpreted the title of Steven Spielberg's brilliant movie ***Schindler's list.*** They thought the title referred to the stratagem that Schindler used to save thousands of Jews from the gas chambers in the second World War. The Dutch word ***'list'*** indeed refers to a stratagem, an interpretation that will also have been reinforced by the fact that Schindler in fact did deceive the Germans. Obviously, it would be a big leap to infer from such occasional observations that the process of lexical access in written word recognition is always language non-selective. The following paragraphs will discuss some key results in research on bilingual lexical processing, to find out what psycholinguists have already discovered about the representational and access issues.

[4]Somewhat later, researchers also started to study syntactic processing in bilinguals, mostly relying on the syntactic priming task (e.g., Bernolet, Hartsuiker, & Pickering, 2007; Loebell & Bock, 2003; Schoonbaert, Hartsuiker, & Pickering, 2007).

1.3 Insights from the processing of cognates and interlingual homographs

When studying how bilinguals access their mental lexicon(s) when they recognize a written word, the ideal stimulus is a word whose spelling pattern occurs in the different languages that are familiar to them. If bilinguals develop a distinct mental lexicon for each language they know, the processing of such words will not differ from monolingual control words, whose spelling pattern only occurs in the target language of the experiment (everything else being equal). However, if bilinguals store all words they know in a single mental lexicon, with a single access process, words with identical spellings in two (or more) languages will access their representation in each language. This will set them apart from matched monolingual words, which will access only a single word representation. Hence, a comparison between monolingual and 'bilingual words'[5] in experimental tasks that shed light on lexical access, should make it possible to decide which of the two above hypotheses is correct.

1.3.1 Why interlingual homographs and cognates are so popular

This simple rationale explains why researchers have abundantly used two types of bilingual words: cognates and interlingual homographs (IH). Cognates are words that do not only share their spelling in two or more languages but also their meaning (e.g., ***man, film, hotel, sport***, and ***water***). IHs share their spelling pattern but have different meanings in the languages in which they occur. Examples of Dutch-English IHs are ***room*** ('cream'), ***list*** ('strategem'), ***fee*** ('fairy'), and ***spot*** ('mockery'). Examples of English-French IHs are ***pain*** ('bread'), ***fin*** ('end'), ***coin*** ('corner'), ***four*** ('oven'), and ***sale*** ('dirty'). Language teachers often use the term 'false friends' to refer to IHs, because these words mislead language learners into believing that these non-native words and their meaning are already familiar, whereas only their form is familiar and has to be associated with a novel meaning.

Many studies have also used so-called near-cognates and non-identical IHs (e.g., Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010, for near-cognates; Dijkstra, Grainger, & van Heuven, 1999, for near-homographs). A near-cognate is a word whose orthography is very close to the spelling of the word's translation in another language but differs in one or more graphemes. The overlap between the two orthographic forms is usually large enough to make it easy for language users to capitalize on their knowledge of the form-meaning association in their native language. Examples of Dutch-English near-cognates are ***donder*** ('thunder'), ***groen*** ('green'), ***maan*** ('moon'), ***nacht*** ('night'), ***ster*** ('star'), ***zeven*** ('seven'), and many others. Examples of near-homographs are lief ('leaf'), leek ('lake'), steel ('stale'), koor ('core'). In several experiments, identical and non-identical cognates and IHs were mixed, either with the explicit purpose of studying the impact of orthographic overlap (e.g., Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Duyck et al., 2007; Peeters, Dijkstra, & Grainger, 2013; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011) or not (Dijkstra et al., 1999; Van Hell & Dijkstra, 2002).

Our intuition that there are separate mental lexicons for the vocabularies of different languages is disconfirmed by the large majority of studies with cognates and IHs. The results

[5] I will use the terms 'monolingual words' and 'bilingual words' to refer to words that occur respectively in one and two (or more) languages. Hence, at the item level I will make the same distinction as at the level of participants: units that can be assigned to one language only will be called 'monolingual' and units that can be assigned to two (or more) languages will be called 'bilingual'.

of many experiments, using different tasks, languages, and research techniques converge on the conclusion that bilinguals (a) have an integrated mental lexicon, without any boundaries between the vocabularies of different languages, and (b) rely on a single, language non-selective access procedure to this lexicon. The access procedure initiates a process that ultimately leads to the activation of the lexical representations of a cognate (near-cognate) or an IH (near-homograph) in all languages known by the language user.

In paragraphs 3.2 I will discuss some key findings that have been obtained with IHs. It will become clear that the only way to make sense of these findings is to distinguish between a process of lexical activation, which is blind to the target language, and a post-lexical stage, at which a task-dependent decision is based on the 'output' of the bilingual lexicon. Paragraph 3.3 will be devoted to cognates, which represent the second pillar supporting the language non-selective view on bilingual lexical access. In paragraph 3.4 it will become clear that this view is also based on behavioral experiments with words that do not occur in bilinguals' two languages. It is also interesting to observe (see 3.5) that, since the turn of the millennium, the conclusions with respect to IHs and cognates that are based on many behavioral experiments are corroborated by the data that have been collected with the new neuroscientific methods, more particularly, the registration of ERP and fMRI data. Finally, in paragraph 3.6 the question will be raised whether the conclusion that bilinguals rely on a process of language non-selective access is restricted to the context of isolated word recognition or is also valid when IHs and cognates are encountered during sentence reading.

1.3.2. The processing of interlingual homographs

Let us first consider what is known about the processing of IHs. These words have either been presented in pure lists (mostly only L2 words) or in mixed lists, i.e., where they appear together with monolingual L1 and L2 words. Many more experiments of the second type have appeared in the literature. However, I will first discuss the first type of experiments.

Measuring activation in bilinguals' mental lexicon: IHs in pure language lists
Gerard and Scarborough (1989) presented Spanish-English bilinguals with IHs in an L2 (English) lexical decision task in which the experimental list contained only English words and English-derived non-words. Homographs that were high-frequency words in English were low-frequency words in Spanish and vice versa. Bilinguals' lexical decision latencies on the homographs were only affected by the target word's frequency in the (non-native) target language of the experiment. Moreover, response speed did not differ between the bilinguals and a group of monolingual English participants. Not surprisingly, Gerard and Scarborough presented this outcome as evidence in favor of a language-selective access process to the bilingual lexicon.

About a decade later Dijkstra, Van Jaarsveld, and Ten Brinke (1998) reported a similar finding. They asked Dutch-English bilinguals to perform an English (L2) lexical decision task. The critical items were a set of IHs and a set of cognates and their matched English controls. As in the Gerard and Scarborough study all non-critical words in the experimental list were monolingual English words and all non-words had been derived from such words. Dutch and English frequency were orthogonally manipulated in the set of homographs, giving rise to four subsets. Like Gerard and Scarborough they obtained a null effect obtained for the IHs. Additionally, response latencies in the set of IHs were only affected by English frequency;

Dutch frequency and its interaction with English frequency were non-significant. At face value this seems to provide strong evidence in favor of language non-selective lexical access. However, at the same time a significant effect of cognate facilitation emerged. This effect will be discussed later. However, this seems to be a reliable index of language non-selective lexical access. In this light, the absence of an effect for the IHs is a puzzling outcome, as it suggests that the non-target L1 reading was not accessed. It is strange that even homographs with a high-frequency L1 reading had no effect on response times (and errors). A process of language non-selective lexical access would at least predict that high-frequency L1 readings are accessed faster than low-frequency L2 readings (and probably also high-frequency L2 readings).

De Groot, Delmaar, and Lupker (2000, Experiment 2) added even more evidence of the same type. Dutch-English participants had to make language-specific lexical decisions to a set of IHs and their controls, either in their L1 or in their L2. The previous experiments had only investigated the effect when the target language was the participants' L2. These authors also manipulated the frequency of the two languages: about half of the homographs had a higher frequency in L1 than in L2, whereas the reverse frequency relationship obtained for the other half. No homograph effect appeared in the L2 lexical decision times or error rates, not even when the L1 reading had a higher frequency than the target reading. This is yet another replication of the null effect for IHs in Gerard and Scarborough and in Dijkstra et al. However, when L1 was the target language an effect of homograph inhibition emerged. A significant inhibition effect was observed, both in response times and errors, but only when the L1 target reading occurred was a low-frequency word. This outcome was puzzling too: why would the weaker L2 cause more interference on decision times to L1 words than the stronger L1 on decision times to L2 words?

In contrast to Cortese and Scarborough, both Dijkstra et al. and De Groot et al. interpreted their findings (across a set of experiments, see below) in favor of language non-specific lexical access. Whereas Dijkstra et al. suggested that the use of only L2 words might make the activity in the L2 lexicon (not to be interpreted as a separate memory store) much stronger than the activity in the L1 lexicon, such that access to the L1 reading could not be detected. De Groot et al., who had only found a null effect in L2, proposed that participants in a single-language experiment could perform a language-specific lexical decision task as a language-neutral (i.e., generalized) lexical decision task. Responses on which a decision was based on the first available reading would have caused facilitation, nullifying the inhibition on responses in which the response was made by relying on the target language. To explain the divergence between a null effect in L2 and an inhibition effect in L1, the proposed that the harder L2 task caused more participants to respond in a language-neutral way.

The experiments reported by Dijkstra, Grainger, and van Heuven (1999) brought more insight into the determinants of IH effects. These authors systematically manipulated the orthographic, phonological and semantic overlap between IHs and cognates. They presented these word types and their controls in a progressive demasking task[6] and a L2 lexical decision task. Interestingly, they found significant homograph facilitation when the homographs only overlapped at the orthographic level (O items, e.g., brand [brænd] – brand [brɑnt],

[6] In progressive demasking the same succession of a word and a visual noise pattern is continuously projected on the screen, while keeping the total duration of the two stimuli constant. Because the ratio between the two stimulus durations changes in a step-wise factor in favor of the word, the later gradually 'emerges' from the noise. Participant identify the word as fast as possible.

'fire'). However, a null effect was obtained when the homographs overlapped at both the orthographic and phonological levels (OP items, e.g., pet [pɛt] – pet [pɛt], 'cap'). Both tasks revealed the same picture: semantic overlap (as in cognates) and orthographic overlap cause facilitation, whereas phonological overlap causes inhibition. For both experiments the authors derived a regression equation from their data, which predicted response times that accounted for 86% and 98% of the observations for progressive demaskling and lexical decision, respectively. Apparently, each of these three lexical dimensions has a linear effect on the response times. None of these effects were found in a control experiment, in which monolingual speakers of English made lexical decisions on the same items, which rules out the possibility that the effects with bilinguals were due to idiosyncratic stimulus properties. The authors argue that a mix of O and OP homographs will often result in a null effect. It also follows that a set of O homographs will yield facilitation, whereas OP homographs will give rise to inhibition (as in the experiments reported in this study). The Dijkstra et al. (1999) study is important because it tears apart item determinants that determine the net effect of IHs in a pure language list. This insight demonstrated beyond doubt that that, even in a pure language list, IHs access their lexical representation in both languages, i.e., that lexical access to the bilingual lexicon is language non-selective.

Studies using other paradigms confirmed that IHs blindly activate their representations in both languages. Semantic priming is one of these paradigms. For instance, Beauvillain and Grainger (1987) asked French-English bilinguals to make lexical decisions on English (L2) targets that were always preceded by French (L1) primes. On critical trials, an IH like ***coin*** (which means 'corner' in French) was followed by an L2 associate of the prime's L2 meaning (e.g., ***money***). A significant priming effect was obtained at a short SOA (150 ms) but not at a long one (750 ms). Hence, even though participants were instructed that all primes would be native language words, they were unable to suppress the fast activation of the word's L1 representation. If they could have 'blocked' access to their English lexicon and 'decided' to read all primes in French only, as instructed, no L2 priming effect would have been found. A second experiment strengthened this conclusion. Using a SOA of 150 ms the authors showed that the frequencies of the IH primes in the two languages were the major determinant of associative priming effects. The higher-frequency reading of the IH primes always facilitated target responses, irrespective of its language (L1 or L2) and irrespective of the instruction to read the primes in a particular language (L1 or L2). For instance, the higher-frequency English (L2) reading of ***four*** ('oven' in French) primed the L2 target five, whether the primes had to be read in L1 or L2. In the same vein, the higher-frequency French (L1) reading of ***pain*** ('bread') primed the L1 target ***beurre*** ('butter') in both instruction conditions. The lower-frequency readings did not cause priming effects. Beauvillain and Grainger's experiments show that bilinguals have no conscious control over which lexicon is accessed. Lexical access is not controlled by the language user but by the IHs themselves: the frequencies of the two readings of IHs determine which language is activated first. Taken together, these results argue against language-selective lexical access and in favor of a process of blind, bottom-up access to an integrated bilingual mental lexicon.

De Moor (1998) also reported an associative priming experiment, using an even more elegant procedure. She demonstrated that, although IHs may yield a null effect in a L2 lexical decision task, they do access their representation in the non-target language. In a first experiment she used the same items as Dijkstra et al. (1998) and replicated their null effect. She used these IHs in a second experiment but presented them as associative primes (presented as ordinary trials) for the next word in a L2 lexical decision task. The target was the English translation of the IH's L1 reading (e.g., ***brand***, i.e., Dutch 'fire', was followed by ***fire***). Like Beauvillain and Grainger she found a reliable associative priming effect. What makes her study important is that she used the same set of IHs and the same task (L2 lexical decision) to show that the null effect in the experiment without associative priming does not necessarily imply that the lexical access process is language non-selective. Recall that Dijkstra et al. (1999) also demonstrated that a null effect for IHs may be the result of a mixture of IHs of the O and OP types. De Moor's finding is especially interesting because she embedded her primes as ordinary trials in the stimulus list, to avoid that the visual presentation of the prime-target pairs would highlight the possibility of a relationship between the two words.

The interim conclusion on the processing of IHs in a pure language list is that these words automatically activate both their L1 and L2 reading. Language users have no control over the process of lexical access to the bilingual lexicon. They cannot decide to 'switch off' lexical access to a non-native lexicon, just as they cannot refuse to recognize a word in their native language (as demonstrated by the well-known Stroop effect, cf. Stroop, 1935). It is the written word itself that has 'control over' the lexical access process. As Beauvillain and Grainger (1998) demonstrated, the frequency relationship between the L1 and L2 readings of an IH determine which reading is accessed first. Whether or not access to the irrelevant reading in the experiment will manifest itself in the response times (and/or errors) depends on other properties of the IHs in the experiment. When most words have a strong orthographic overlap but a weak phonological overlap in the two languages, a facilitation effect will emerge. When most IHs have both a strong orthographic and phonological overlap, an inhibition effect will be found. A mixture of these two types of IHs will yield a null effect, even though access to the lexical representation of the irrelevant language will have occurred (provided that the frequency relationship between both readings makes this possible).

The need for a decision level: IHs in mixed language lists

Many experiments have been reported in which IHs were presented in a stimulus list containing both monolingual words from both L1 and L2. It is a consistent finding in the literature that the effect for IHs is considerably affected by mixing the two languages in which they occur. Dijkstra et al. (1998) added a set of monolingual Dutch words to the same set of IHs that had yielded a null effect in their pure English (L2) list[7]. Recall that the IHs in their study were selected according to an orthogonal combination of English (L2) and Dutch (L1) high-frequency and low-frequency words. Participants performed a L2 lexical decision task in which the monolingual Dutch words had to be rejected as 'not English', i.e., they had to be treated in the same way as the English-like nonwords (the monolingual Dutch words represented 25% of all 'no'-responses). Responses to IHs were both slower and less accurate than responses to their monolingual controls. In addition to this inhibition effect, a regular effect of English frequency was found, i.e., IHs with a high frequency were responded to faster, whereas a reverse frequency effect for Dutch was obtained, i.e., IHs with a high frequency in Dutch were responded to more slowly. Clearly, adding pure L1 words to the stimulus list significantly slowed down L2 responses to the IHs. The frequency effects for the two languages of the IHs indicate that lexical access to the bilingual lexicon is frequency-sensitive, i.e., high-frequency lexical representations are accessed before lower-frequency ones, irrespective of the language of the IH's reading. The reverse frequency effect for L1 shows that, when the L1 reading is accessed first this non-target language makes it difficult for participants to respond. When they respond too fast, they make an error and respond 'no' to an IH. As all monolingual L1

[7] A handful of items that caused to many errors were replaced.

words in the experiment require a 'no'-response, they have learnt to associate the 'no'-response with the Dutch language. However, if they attempt to avoid errors by temporarily postponing the response to check whether the word also occurs in L2 they lose time. In addition, the discovery that the IH is indeed an L2 word as well and, hence, requires a 'yes'-response creates a response conflict, which must be resolved before the correct response can be made. The result is an inhibition on correct responses to IHs.

De Groot et al. (2000) made the same observation. Recall that they, too, obtained a null effect for IHs in a pure L2 list. However, in their L2 list they replaced one third of the English-like nonwords with monolingual Dutch (L1) words and in their L1 list they replace the same proportion of Dutch-like non-words by monolingual L2 words. Recall that about half of the IHs had a higher frequency in L1 than in L2 and vice versa. This change in list composition had the same impact as in Dijkstra et al.'s (1998) study. When monolingual L1 items were added as 'no'-response trials in a L2 lexical decision task, a strong inhibition effect was obtained for the IHs. The same outcome was obtained when monolingual L2 items were added as 'no'-response trials in a L1 lexical decision task. Importantly, this effect only occurred when a 'yes'-response was required to the low-frequency reading of the IH. The findings for the L2 list replicate the results obtained by Dijkstra et al. (1998). The monolingual L1 items cause an association between the native language and the 'no'-response, which causes problems when a 'yes'-response must be made on the low-frequency L2 reading of an IH. The higher-frequency L1 reading becomes more quickly available and either causes errors or delays the response until the L2 reading is accessed and the ensuing response conflict is resolved. Interestingly, the same occurs when participants must respond to the L1 readings of the same IHs. This finding extended Dijkstra et al.'s findings to the context of an L1 list. Apparently, IHs whose L2 frequency is sufficiently high can cause an inhibition effect when responses must be made to the L1 reading with a low frequency.

Dijkstra, Timmermans, and Schriefers (2000) used a different experimental task to study IHs: the go/no-go task. Whereas the lexical decision task requires that a response be made on each item, either a 'yes'-response or a 'no'-response, this task only requires an explicit response to a predefined target category. This task characteristic makes the go/no-go task a promising tool for investigating the access process to the bilingual lexicon. If it is possible to achieve selective access to the lexicon of a single language and 'shut off' access to all languages that are irrelevant for task execution, this task makes it possible (in theory) to exclusively focus on the words of a target language and ignore the words from all other languages. As a result of language blocking, participants would be temporarily unable to recognize the words in these languages (till the end of the experiment). As in previous experiments, Dijkstra et al. manipulated the frequency relationship between the L1 and L2 readings of their IHs. In one experiment the 'go'-response was associated with L2, in another it was associated with L1. The same picture emerged as in the lexical decision tasks described above: a frequency-sensitive inhibition effect was obtained for IHs. The inhibition was strongest when the 'go'-response had to be made to the lower-frequency reading of the IH, independent of target language (L1 or L2). Hence, even in a task that creates optimal conditions for language non-selective access, the results indicated that language users are unable to decide not to recognize words in task-irrelevant languages. As in the lexical decision tasks reported by Dijkstra et al. (1998) and De Groot et al. (2000) the higher-frequency reading was automatically activated first, irrespective of the language to which it belonged. If this reading did not belong to the target language and participants were able to suppress a response ('no-go'), the competition between the two IH's two lexical representations will have delayed access to the lower-frequency reading (compared to a monolingual control word). However, quite probably (see later), upon eventual access to the target reading, a competition between the initial decision not to respond and the novel information that a 'go'-response should be made further delayed the response. In both experiments, a very large inhibition effect (over 200 ms) was found when the IHs had a low frequency in the target language (L1 or L2) but a high frequency in the other language.

Language mixing experiments like these made researchers realize that the story behind language non-selective access to the bilingual lexicon could not only be told in terms of the activation of lexical representations in L1 and L2. Clearly, the findings that IHs activate their L1 and L2 readings in a frequency-sensitive way and that overlap between the orthographic and phonological representations is an essential determinant of their processing time cannot explain why adding L1 words to the stimulus list can change a null effect into an inhibition effect (using the same IHs). Another observation led to the same conclusion.

After having noticed that 'no'-responses to monolingual L1 words caused inhibition on IHs, Dijkstra et al. (1998) decided to use a mixed language list (using the same IHs) in a different task. They used a generalized lexical-decision task, in which a 'yes'-response has to be given to any word (from L1 or L2) and a 'no'-response to a nonword. This time a significant facilitation effect was found for the IHs. Moreover, the data revealed that participants responded to the fastest available IH reading. As responses based on L1 or L2 were both correct a normal frequency effect was found for both languages (in contrast to the reverse frequency effect for L1 in the L2 lexical decision task).

Taken together, the effects of language mixing (pure L2 list vs. mixed language list) and experimental task (L2 lexical decision vs. generalized lexical decision) call for the need of an explanatory level besides lexical activation. Dijkstra and van Heuven (2002) proposed that the effects of list composition and experimental task could only be explained by taking a decision level into account. At this level, the output of the bilingual lexicon is compared to the task demands, a comparison that obviously takes time and that, most importantly, can delay or speed up the decision. A first argument is the finding that the same set of IHs yielded a null effect in a L2 lexical decision task, but produced an inhibition effect in the same task when a set of monolingual L1 words was added (De Groot et al., 2000; Dijkstra et al., 1998). This indicates that the inhibition effect is not due to competition within the mental lexicon but to competition between participants' tendency to make a 'no'-response to IHs (as they readily recognize their native language reading) and the 'yes'-response they eventually must make to comply with task demands. IHs thus confront participants with a response conflict in the decision stage, which causes an inhibition effect. A second argument is the finding that, using the same item types but changing a L2 lexical decision task to a generalized lexical decision task, the inhibition effect turned into a facilitation effect. As the same IHs and the same item types were used, this change in the IH effect could not be attributed to intralexical competition either. Again, this observation can be explained in terms of a decision level. Unlike what was the case in the L2 lexical decision task, participants did not learn to associate L1 with the 'no'-response but with the 'yes'-response, i.e., the same response as for L2 words. Accordingly, any output from the mental lexicon, the L1 or L2 reading of an IH (whichever achieved access first), licensed a 'yes'-response was. The main effects of L1 and L2 frequency for IHs substantiate the claim that the fastest available word was used for responding. Even though the facilitation effect is due to a decision level effect, the faster availability of the higher-frequency reading of an IH is due to intralexical competition and is a major piece of evidence in support of a language non-selective access process to bilinguals' mental lexicon.

To finish this section, one more study is worthwhile considering, if only to demonstrate that a mixed language list does not necessarily affect the size and (even) direction of the IH effect. Language mixing by itself does not affect the decision stage. As demonstrated by the contrasting outcomes of Dijkstra et al.'s (1998) L2 lexical decision task and generalized lexical decision task, the impact on the decision level is jointly determined by list composition and task demands. Kennette and Van Havermaet's (2012) experiment corroborates this view, even though its method considerably differs from the typical research paradigms. IHs were presented in a mixed language list but participants made no responses to the words. They just had to read words from three types on a computer monitor: monolingual French words, monolingual English words and IHs. By definition, the absence of a decision task could neither induce a conflict between response candidates nor the absence of such a conflict. The four word sets (including the L1 and L2 readings of the IHs) were matched on frequency. After a distraction task a set of pictures was shown, and participants had to decide whether the picture name had been presented in Phase 1. Response times on pictures representing the IHs' L1 and L2 readings did not differ from each other, whereas both were significantly longer than those for the monolingual L1 and L2 words. The same pattern was obtained for the errors. Even though this experiment does not tap into the online processing of IHs – it measures the effect of this processing on the creation of episodic memory traces – it is consistent with the experiments reported above. The absence of a difference between response times and errors on pictures representing the two IH readings indicates that both readings had been accessed when reading the word. Moreover, the fact that performance was worse on the IHs (in terms of response speed and accuracy) than on the monolingual French and English controls indicates that there was a conflict between these two readings during word recognition. Importantly, this cannot have been a response conflict as no response was required. It is likely that the competition between the L1 and L2 representations in the bilingual lexicon (and perhaps also their associated conceptual representations) made it more difficult to create episodic memory traces. Unfortunately, the study does not address the question whether the inhibitory effect on IHs is stronger in a mixed list than in a pure L2 (or L1) list.

1.3.3 The processing of cognates

In contrast to IHs, which have produced null effects, facilitation, and inhibition effects (but see 3.2. for the effect of word-specific properties and task-dependency), cognates have a more stable fingerprint. The large majority of studies have shown that cognates cause facilitation effect in behavioral experiments.

The effect of L1 on L2 processing

Back in 1979 Caramazza and Brones already demonstrated that Spanish-English bilinguals make faster lexical decision in their L2 on identical cognates (e.g., ***piano***) and near-identical cognates (e.g., ***eco***, 'echo'). The combination of orthographic identity or near-identity and semantic identity considerably clearly facilitated the lexical decision process. Dijkstra et al. (1998) reported the same finding when they presented both IHs and cognates in a list of otherwise pure L2 words. Although the lexical decision data of their Dutch-English participants revealed a null effect for IHs, suggesting selective access to the bilingual lexicon (but see above), they yielded a significant facilitation effect for the cognates. A cognate facilitation effect indicates that the words of a bilingual's two languages are not stored in separate 'bins' of long-term memory but are interconnected in a single mental lexicon. The nature of these cross-language connections cannot be inferred from these findings. However, the fact that cognates have the same spelling or have a strong orthographic overlap and are associated to the same concept suggests that this form-meaning overlap is the major determinant. After all, the (accidental) connection between a form and a meaning is the very definition of a word. Apparently, whenever two words from L1 and L2 are (nearly) the same, this is reflected 'somehow' in the way they are stored. This general characterization is obviously not sufficiently explicit, but as the storage issue will be the focus of the first experimental chapter (Chapter 2), it will suffice for now.

As discussed in 3.2, Dijkstra et al. (1999) systematically studied the impact of form overlap (at the levels of both orthography and phonology) and meaning overlap on the processing of interlingual words by presenting IHs and cognates in a L2 lexical decision task and a progressive demasking task. Their conclusion that orthographic overlap causes facilitation, whereas phonological overlap causes inhibition (see above), was also based on their findings with cognates. They included three cognates types in their stimulus list, based on a systematic manipulation of the L1 and L2 words' orthographic, phonological, and semantic overlap. SOP cognates, whose semantics, orthography, and phonology overlap (e.g., ***film, lip***) and SO cognates, whose pronunciations are too different (e.g., ***fruit***, [fru:t]-[frœyt]; ***rat***, [ræt]-[rɑt]), gave rise to significant facilitation. In contrast, SP cognates, whose semantics and phonology largely overlap but whose spelling patterns differ (e.g., ***clock***-'klok', ***soup***-'soep'), yielded a null effect. The authors concluded that semantic overlap and orthographic overlap are two independent sources of facilitation, whereas phonological overlap causes inhibition. Recall that the conclusions with respect to the effects of orthographic and phonological overlap were also supported by their results for the three types of IHs (see 3.2). Hence, facilitation was found for SO cognates but not for SP cognates. In the latter word type the semantically induced facilitation component is neutralized by the phonological component. This does not happen in SOP cognates, as the phonological inhibition is not strong enough to cancel the facilitation that is caused by two facilitation components, one due to semantic overlap and one due to orthographic overlap.

Lemhöfer and Dijkstra (2004) replicated the Dijkstra et al. (1999) study, one reason being that Jared and Kroll (2001) had pointed out problems with the P items, i.e., the interlingual homophones. Since monolinguals made more errors on these homophones, bilinguals' difficulties with these items might have had nothing to do with their shared phonology. Lemhöfer and Dijkstra presented the three types of IHs and cognates in separate experiments, thus avoiding that they could affect each others' processing, and presented them in a L2 and a generalized lexical decision task. The results for the cognates in the L2 lexical decision task were a perfect replication of the results in Dijkstra et al. (1999): SOP and SO cognates were recognized significantly faster than their L2 controls whereas a null effect was found for SP cognates. In the generalized lexical decision task the former two item sets were responded to significantly faster than both their L2 (English) and L1 (Dutch) controls, whereas the homophonic cognates (SP items) were recognized as fast as their English controls.

Taken together, these results confirm Dijkstra et al.'s (1999) conclusion that interlingual words whose meaning and spelling considerably overlap are recognized faster than L2 controls, whereas interlingual homophones with the same meaning but a different spelling do not facilitate word recognition. This corroborates Dijkstra et al.'s earlier claim shared semantics and shared orthography are sources of facilitation, whereas shared phonology is a source of inhibition. However, in the generalized lexical decision task both SOP and SO cognates were not only recognized faster than their L2 controls but also faster than their L1 controls. At first sight, this seems to be due to the nature of the generalized lexical decision task, which makes it possible to respond to the first reading that becomes available (often the L1 meaning). Hence, in this task, participants can respond to cognates' L1 readings, which explains why faster responses were made on these cognates compared to the responses on L2 controls. However, it does not explain faster responses compared to the L1 controls. Equally fast responses as to the L1 controls were obtained for IHs of the O and OP type. In the generalized lexical decision task, participants clearly responded to the Dutch reading of these items, which made their responses to these item types faster than to the matched L2 items but not faster than to the L1 controls. The fact that responses to SOP and SO cognates were even faster than to their L1 controls suggests that the explanation is more complex than the principle "respond to the first available response". In addition to this principle, some kind of interaction between the orthographic and semantic representations of cognates must be involved. One possibility is that a cognate's shared meaning causes top down activation feedback to the L1 and L2 orthographic representations. As its semantic representation is activated by these two representations, its activation level will increase faster compared to a matched L1 control, such that this top down process will start sooner. The result will be a faster accumulation of activation at the orthographic level, which will make it possible to respond sooner than to an L1 control (but see Chapter 2, which is devoted to this issue, and considers different accounts of cognate representation).

Whereas Dijkstra et al. (1998) and Lemhöfer and Dijkstra (2004) attempted to unravel the separate contributions of the different levels of cognate representation, Dijkstra, Miwa, Brummelhuis, Sappelli, and Baayen (2010) focused on cognates' cross-language orthographic similarity and their task dependence. In a L2 lexical decision task without monolingual Dutch words, response times linearly decreased (facilitation) as orthographic similarity increased, but in a language decision task they linearly increased (inhibition). It seems that the more a cognate's L2 spelling resembles that of its L1 translation the more top down feedback both lexical representations receive from the shared semantic representation. This will help in making a L2 lexical decision but make it difficult to make a language decision. At the transition between near-identical cognates and identical cognates there was a sudden discontinuity in the evolution of the effect: it became much larger than would be expected from its linear evolution in the near-cognates. Apparently, non-identical orthographic representations compete with each other, whereas no such competition occurs for identical orthographic representations (or a single representation that is shared between L1 and L2). Interestingly, phonological overlap caused a further decrease in response times, but only for identical cognates in the L2 lexical decision task. Cognates with the same spelling and meaning were recognized even faster as they became more similar phonologically as well (until they were identical, e.g., ***film***). This effect seems at odds with the claim that phonological overlap causes inhibition (Dijktra et al., 1999; Lemhöfer & Dijkstra, 2004). However, recall that Lemhöfer and Dijkstra argued that more research on the impact of phonological overlap is needed.

Lemhöfer et al. (2008) approached the issue from a different angle. A group of Dutch, German, and French bilinguals participated in a progressive demasking experiment in which they had to identify 1,025 English words over three sessions. A regression design was chosen because this makes it possible to assess the importance of a variable outside the context of an experiment in which the words have been carefully selected to test the hypothesis. Baayen (2010) argues that, often, a regression design is the best choice to test a hypothesis.

The regression analysis of the entire list revealed a cognate facilitation effect. An inhibition effect for IHs also emerged, but only when the small subgroup of these words (around 60 in each language group) was analyzed separately, which revealed an inhibitory effect of the words' L1 frequency. The authors formulated two important conclusions: (a) the effect of L1 is strikingly similar across different types of L1 (Dutch, German, and French mainly differ in terms of orthographic depth, the consistency in grapheme-phoneme mappings) and (b) the role of L1 in recognizing L2 words is relatively small against the background of all variables that affect word recognition. The latter conclusion should not be interpreted against the view of an integrated bilingual lexicon with a language non-selective access process.

Are cognates in L3 affected by their translations in L1 and L2?

All cognate effects discussed in the previous section were effects of L1 on L2. Could these findings be extrapolated to L3? First, are L3 words affected by identical cognates in L1? Second, are they also affected by identical cognates in a less dominant language than their native language, i.e., their L2? To investigate these questions Lemhöfer, Dijkstra, and Michel (2004) presented identical cognates and their controls to a group of Dutch-English-German trilinguals. They used a L3 (German) lexical decision task and a list in which no monolingual L1 or L2 words appeared. They compared 'double cognates', which were identical in L1 and L3 (e.g., ***kunst***[8], 'art') and 'triple cognates', which were identical in all three language (e.g., ***echo***). Intriguingly, lexical decisions were significantly faster on the cognate triplets than on the (matched) double cognates, which in turn were faster than on the controls. In a control experiment they found no differences among these word types in a group of monolinguals in L3. This study replicates the familiar finding that cognates are processed faster in a non-native language when they also occur in the native language (the double cognate effect). However, it also demonstrates that the cognate facilitation effect does not only depend on the presence of an identical cognate in the native language. Rather, the effect is a cumulative one. Apparently, when target words appear as identical cognates in more than one non-target language, each more familiar language adds a facilitation component to the decision latencies (hence, the stronger effect for the trilingual cognates). As no such effects were found in the monolingual group, they were due to the way the mental lexicon is organized in trilinguals. This outcome can only be explained by assuming a fully integrated mental lexicon, whose access process does not take language differences into account.

The effect of L2 (and L3) on L1 processing

All cognate effects discussed thus far were effects of L1 on L2 and of L1 and L2 on L3. However, the reverse effect has also been found. In a L1 lexical decision task, participants responded faster to words with a near-cognate in L2 than to their matched controls. For instance, Van Hell and Dijkstra (2002) used trilinguals with Dutch as their L1, English as their L2, and French as their L3. In an L1 lexical decision task, decision latencies were shorter on L1 words with a near-cognate in English (e.g, ***appel***, 'apple', but 'pomme' in French; ***bakker***, 'baker', but 'boulanger' in French) than on their controls. When participants were sufficiently proficient in L3 they also responded faster on L1 words with a near-cognate in French, compared to controls (e.g., ***meubel***, 'meuble', but 'piece of furniture in English). It seems that any non-native language in which participants are sufficiently proficient can affect lexical processing of the ***native language***. This means that the cognate facilitation effect does not depend on the existence of a stronger lexical representation in the non-target language (e.g., L1 is stronger than L2, which

[8] Note that the words were presented in capitals, to avoid that differences in spelling conventions might interfere. In German, the first letter of each noun is a capital letter, which is not the case in Dutch and English.

is stronger than L3). Once more, this finding suggests that near-cognates form an integrated representational structure, in which their two orthographic representations are interconnected but also connected to a shared semantic representation, which is more strongly activated compared monolingual words and can, hence, send more top down activation to the orthographic level.

1.4. Effects of cross-language orthographic neighbors and morphological family size

From paragraph 1.3 the reader might infer that all evidence in favor of an integrated mental lexicon and a lexical access procedure that is blind to language distinctions has been built on results with IHs and cognates. However, other factors have been manipulated to investigate the representational and access questions. Such studies have strengthened the conclusions from experiments with interlingual words.

One of these factors is the number of orthographic neighbors, i.e., words with the same length that differ from each other in only one letter position (e.g., ***deal, meal, dear***; see Coltheart, Davelaar, Jonasson, & Besner, 1977). Whereas Coltheart et al. initially claimed that only nonword rejection was inhibited by a large neighborhood size (N), nonwords with many neighbors being more 'word-like', it was later demonstrated that a high N facilitates word recognition in the native language, but only for short low-frequency words (Andrews, 1989, 1992). Wee Hun Lim (2016) arrived at a similar conclusion, but derived it from distributional analyses and an arguably more accurate definition of orthographic neighbors, the so-called Orthographic Levenstein distance[9]. Grainger and Jacobs (1996) argued that homograph facilitation might result from global activation in the mental lexicon, which might either facilitate or inhibit the response process (depending on the task).

Van Heuven, Dijkstra, and Grainger (1998) took the neighborhood effect as their point of departure and showed that Dutch-English bilinguals who had to identify English (L2) words in a progressive demasking task or perform an English lexical decision task made slower responses as the number of orthographic neighbors in the irrelevant language (their L1) increased, resulting in global activation in the wrong language. Monolingual English participants were not affected by Dutch neighbors, i.e., the bilingual effect could not be attributed to uncontrolled item characteristics. Hence, even monolingual words, i.e., words that are neither cognates nor IHs, trigger activation in the non-target lexicon, based on the degree of orthographic overlap with the written stimulus. This finding strongly suggests that the activation in the bilingual lexicon is a stimulus-driven, i.e., bottom-up, process, which is ignorant with respect to the target language and can, hence, not be brought under conscious control by language users, i.e., they cannot decide to 'shutt off' all irrelevant lexicons.

The same rationale was followed when researchers decided to test the effect of cross-language morphological family size (MFS). In their seminal study Schreuder and Baayen (1997) were the first to discover the importance of this variable. A word's MFS is the number of derivations and compound words in which they appear as a constituent (e.g., the morphological family of a word like ***man*** consists of words like ***manful, manhood, manlike***,

[9] The Orthographic Levenstein Distance between two words is the number of transformations (defined as substitutions, deletions, and insertions) needed to turn one word into the other.

manly, mankind, manhole, man-hour, …)[10]. Schreuder and Baayen demonstrated that simple monomorphemic words with a large MFS are recognized faster than words with a small MFS, everything else being equal. The effect of MFS has been shown to be quite robust (e.g., De Jong, Schreuder, & Baayen, 2000; Krott, Baayen & Schreuder, 2001; Moscoso del Prado Martín et al., 2005). Ever since, many studies have matched critical and control words on MFS.

Dijkstra, Moscoso del Prado Martín, Schulpen, Schreuder, and Baayen (2005) found that Dutch-English bilinguals' response times on IHs in a L2 (English) lexical decision task were simultaneously affected by the MFS of their L1 and L2 readings. The MFS in L2 (target language) caused facilitation whereas the MFS in the irrelevant L1 caused inhibition (the same applied to L1 and L2 word frequencies). The (logarithmic) value for the MFS for L2 minus the value for L1 was a significant predictor of lexical decision latencies. When the same IHs were presented in a L1 (Dutch) lexical decision task the reverse pattern emerged: response speed was facilitated by the MFS for L1 but inhibited by the MFS for L2. The difference between the MFS value for L1 minus the MFS value for L2 significantly predicted decision latencies. In contrast, both MFS values caused a facilitation effect in a generalized lexical decision task. In this task, both languages require a 'yes'-response, whereas a language-specific lexical decision task assigns a 'no'-response to one language. When words from L1 and L2 must both be given a 'yes'-response, the MFS of each language will support this response and cause facilitation.

This study supports two conclusions: (a) lexical access to the bilingual mental lexicon is language non-selective, i.e., 'blind' to language distinctions, as even the irrelevant morphological family in the other language is activated, and (b) there is a decision level at the post-lexical level where task demands affect how the lexically activated information is translated into a response. Whatever is activated in the mental lexicon must 'pass through' this decision level, which acts as a 'filter' that determines how the activated information will surface in the data: as a facilitation or inhibition effect (possibly, a null effect) and as a small or large effect.

Whereas Dijkstra et al. (2005) studied the effect of MFS on the recognition of IHs, Mulder, Dijkstra, and Baayen (2015) investigated the role of MFS on cognate recognition. More particularly, they wanted to know whether cross-language MFS affects the processing of identical cognates like ***tent*** and non-identical cognates like ***pill/pil*** differently. If the morphological family of a word is activated by its morphemic representation (Schreuder & Baayen, 1997), the two cognate types should be affected equally by their cross-language MFS. A second question was whether the MFS effect on cognates is task-dependent. In this study, the cross-language MFS caused faster decisions for both identical and non-identical cognates in an English (L2) lexical decision task with Dutch-English bilinguals, and no interaction between cognate type and MFS was found. In a language decision task, which induces a bilingual context, both the L1 and the L2 MFS caused an inhibition effect on the decision times for cognates. The inhibitory effect was stronger for identical cognates than for non-identical cognates. This suggests that the activation of a morphological family is triggered by the spelling of the word, resulting in stronger effects for identical cognates. It also indicates that, when cognates are presented in a task that induces a response conflict, the activated morphological families act to enhance this response competition, which is reflected in a large response delay. Hence, this study, too, supports the view that access to the bilingual lexicon is language non-selective. Furthermore, it also emphasizes the task-dependent nature of

[10]Note that inflected forms are excluded from the morphological family. These are taken up in the variable Lemma Frequency.

effects on bilingual words, and, hence, the importance of a decision level. Task-dependency has mostly been demonstrated with IHs. These experiments demonstrate that the same phenomenon occurs with cognates.

1.5 Neuroscientific evidence

As all experiments in this doctoral dissertation will use behavioral methods we will not attempt to summarize the many experimental findings that have been obtained with the techniques of neuroscience. The evidence that is presented in this paragraph only serves to illustrate that neuroscientific techniques have also been used to study bilingual lexical processing (with a steady increase since the turn of the millennium). A review on the use of neuroscientific methods in the study of bilingual language processing can be found in van Heuven and Dijkstra (2010).

Midgley, Holcomb, van Heuven, and Grainger (2008) used ERP recordings to study the effect of cross-language neighbors. French-English bilinguals read pure language lists, in L1 or L2, in which half of the critical items had many cross-language orthographic neighbors and the other half had few such neighbors. The word list also contained a small number (about 1 in 5) of words that referred to body parts. Participants had to push a button whenever such a word appeared (go/no-go semantic classification), i.e., they gave no response while reading the critical items. Words with many cross-language neighbors produced a more negative-going waveform in the N400 region. This negativity appeared earlier for L2 words. This pattern was not found in a group of monolingual participants, who only read the English list. This indicates that the effects were not due to uncontrolled item characteristics but to the influence of the irrelevant language. These results offer strong evidence in support of language non-selective access to an integrated bilingual lexicon: even though the conditions were optimal for selective access to the French (L1) or English (L2) lexicon, the ERP data clearly show that participants were unable to block access to a language. Note that these findings are consistent with van Heuven et al.'s (1998) findings in behavioral tasks. Midgley et al. interpret the cross-language neighborhood effect as the result of problems in settling on the correct form-meaning association, which would explain why the effect is situated in the region of the N400 (an ERP component that is usually associated with semantic integration).

Kerkhofs, Dijkstra, Chwilla, and de Bruijn (2006) asked Dutch-English bilinguals to perform a L2 lexical decision task. Targets were preceded by a L2 prime. The critical items were IHs (e.g., ***stem***, the Dutch word for 'voice'), which were preceded by a related (root) or unrelated (fool) L2 word. During target processing ERP recordings were made as well. In the behavioral data a (regular) frequency effect was obtained for the target language (L2) in the unrelated condition whereas a reverse frequency effect was obtained for L1. This finding indicates language non-selective access. Even though no monolingual L1 words appeared in the list and the local context (i.e., the prime) further hihglighted that the experiment was about L2 words, participants automatically activated the target's L1 representation. When the unrelated and related conditions were analyzed together, an effect of semantic relatedness was obtained, a (regular) effect of L2 frequency, and an interaction of the latter effect with L1 frequency. Responses were faster when they had a high L2 frequency but slower when they were high-frequency words in L1. However, these effects were not additive, as very slow responses were made when the targets were low-frequency words in both L1 and L2. The fact that L1 frequency had a reverse effect on response times in an experiment on semantic priming in L2, without any monolingual L1 words, is strong evidence that access to the bilingual mental lexicon is indifferent to language distinctions. The ERP data converged with the behavioral data. A regular semantic priming effect was observed in the N400. The amplitude of the N400 was affected by the frequency of the target language (L2) but also by the frequency of the irrelevant language (L1). However, the opposite effects for L1 that were found in the behavioral data were observed here as well: whereas larger N400 amplitudes were found for targets with a low L2 frequency (the typical signature of a low-frequency words in monolingual experiments), larger amplitudes were found when the IH target had a high frequency in L1. Taken together, these findings offer strong evidence that bilinguals cannot ignore their L1, even in a pure L2 context. Apparently, an IH automatically activates its L1 representation, no matter how strong the experimental conditions bias participants towards processing words only in L2.

Besides ERP data, researchers have also tested language users in the MRI scanner to collect so-called fMRI data. These images make it possible to see which brain regions are most active, i.e., use high oxygen levels, while performing a task (e.g., word recognition). Van Heuven, Schriefers, Dijkstra, and Hagoort (2008) made fMRI scans while their Dutch-English participants performed a behavioral task in the scanner: a L2 lexical decision task or a generalized lexical decision task. The critical items were IHs. Their decision to use the same set of IHs in these two tasks was based on a clever rationale. As IHs have the same orthographic form but are semantically unrelated, their lexical representations will compete during lexical processing (on the assumption that there is an integrated lexicon). This will cause problems at the decision level if the experimental task requires participants to choose between these two IH readings (in the L2 lexical decision task). As only one candidate is the correct one, a response conflict will emerge, i.e., the same item will suggest a 'yes'-response and a 'no'-response. In contrast, when the experimental task does not require participants to make a choice (because both support the same response, as in the generalized lexical decision task), no response conflict will emerge. The authors argue that this makes it possible to distinguish between a stimulus-based language conflict, which occurs within the mental lexicon, and a response-based language conflict, which occurs at the decision level. In the response times for the L2 lexical decision task, they found significantly slower responses for the IHs compared to their controls. In contrast, no reliable differences were found between the two item types in the generalized lexical decision task. Dijkstra et al. (1998) also obtained an inhibition effect in the former task but a facilitation effect (rather than a null effect) in the latter task. However, what matters is that the L2 lexical decision task clearly caused a response conflict whereas the generalized lexical decision task did not. The fact that a control group of monolinguals did not show an inhibition effect in the L2 lexical decision task confirms that the effect in the bilingual group was due to their knowledge of the two languages, rather than being the result of uncontrolled item properties.

The imaging data demonstrated that the combination of behavioral and neuroscientific methods can lead to novel insights. Some brain regions were activated by both tasks (sharing the conflict within the mental lexicon), whereas other brain regions were only activated by the L2 lexical decision task (the only task causing a response conflict). Whereas one of the former regions has been shown to be responsible for controlled semantic retrieval, one of the latter regions has been shown to be involved in decision and selection processes. The former function can be linked to the stimulus-related language conflict, whereas the latter can be linked to the response-based language conflict, i.e., the selection difficulty at the decision level (e.g., Nachev, Wydell, O'Neill, Husain, & Kennard, 2007). At any rate, van Heuven et al.'s experimental design made it possible to isolate two distinct conflicts, one of which is induced by the IHs' competition in the mental lexicon and one of which is due to the nature of the experimental task. By correlating the most active brain regions with these two types of conflict, it becomes possible to achieve a deeper understanding of the processing of IHs. This methodology makes it possible to find out which brain regions are involved in either lexical processing or in task execution, and to find out whether the functions that are subserved by each of these regions (as identified in earlier research) fit the nature of the cognitive process that the fMRI technique has associated it with.

1.6 Processing bilingual words in monolingual sentence contexts

All experiments in this dissertation concern the activation and decision aspects of words that are presented in isolation. For that reason, I will not give an extensive review of the literature on the processing of cognates and IHs in sentences. Obviously, this is an important line of research for testing the ecological validity of the effects that have been measured in experiments like the ones discussed above. Indeed, reading words in isolation is the exception, not the default. Words appear in sentences, and even larger linguistic (and non-linguistic) contexts. Hence, it is important to study how sentence contexts interact with, for instance, the processing of cognates and homographs. Top down processes, based on the semantic and/or syntactic properties of the preceding words and/or their language membership, might quickly suppress the rise in the activation level of the non-target representation of a cognate or IH. Note that this would still be in line with the general idea of language non-selective access. A more dramatic finding would be that the preceding sentence context can achieve what cannot be achieved in isolated word lists: that access to all irrelevant languages is blocked, except the one to which all preceding words in the sentence belong, i.e., that bilinguals are able to rely on a process of language selective access in sentence contexts.

It is comforting to know, though, that the experiments that have been done in this line of research do not suggest such a radical conclusion. On the contrary, they support the view of a fully integrated mental lexicon and a strong bottom up access process, which is blind to the language membership of words' orthographic representations (for a review, see Van Assche et al., 2011). For instance, cognates that yielded a typical facilitation effect in an isolated L2 lexical decision task also revealed facilitation in the early stages of lexical processing, as observed in an eye monitoring experiment (Duyck et al., 2007). Van Assche et al., (2011) reported that this is even found in early measures of eye monitoring in high-constraint sentences, in which the cognate is a high-probability word. Interestingly, they observed that cognate facilitation became larger as the orthographic overlap between the L1 and L2 words became larger, a finding that has also been obtained in isolated word experiments (Dijkstra et al., 2010, discussed above). Even more surprisingly, Van Assche, Duyck, Hartsuiker, and Diependaele (2009) demonstrated that eye monitoring Dutch-English bilinguals while they were reading sentences in their native language showed that cognates were processed faster than controls (and replicated this finding with a different set of cognates). Their conclusion is a nice formulation of the view that bilinguals achieve lexical access in a way that is language non-selective: "Becoming a bilingual […] changes one of people's seemingly most automatic skills, namely, reading in one's native language." (p. 926).

Cop, Dirix, Van Assche, Drieghe, & Duyck (2016) went even one step further and monitored their Dutch-English participants' eye movements while they were reading one half of a book in L1 and the other half in L2. The authors observed effects of cognate facilitation, both in L1 and L2. In other words, even the rich (monolingual) context of a story is apparently unable to wipe out cognate facilitation effects.

1.7 An explanatory model: BIA+

The dominant model on bilingual language representation and processing is the BIA+ model (Dijkstra & van Heuven, 2002). This is also the model that I will use in the experimental chapters to interpret the experimental results. The model is a refined version of its predecessor, the BIA (Bilingual Interactive Activation) model. Basically, the BIA model is an adaptation of the Interactive Activation (IA) model, which was developed by McClelland and Rumelhart (1981) and Rumelhart and McClelland (1982) to model monolingual word recognition. To account for bilingual lexical processing the BIA model was developed. First I will briefly describe the BIA model and then specify what has been changed when developing the BIA+ model, which has become the most popular model in the literature.

Like any IA model the BIA model (Dijkstra, van Heuven, & Grainger, 1998; see Figure 1.1) consists of several hierarchically ordered representational layers: a visual feature layer, a letter layer, and a word layer. In addition, its top layer represents language nodes. At each layer, information is represented in the form of nodes. Whenever a node is activated (e.g., a node representing a visual feature that occurs in a letter of the visual stimulus), it transmits this activation to all nodes on the next higher level. When the information represented at a receiving node is compatible with the information, it receives an excitatory input (which inceases its activation level). In the other case, it receives an inhibitory input (which decreases its activation level). This simple principle regulates the communication between any two successive levels. Within the word

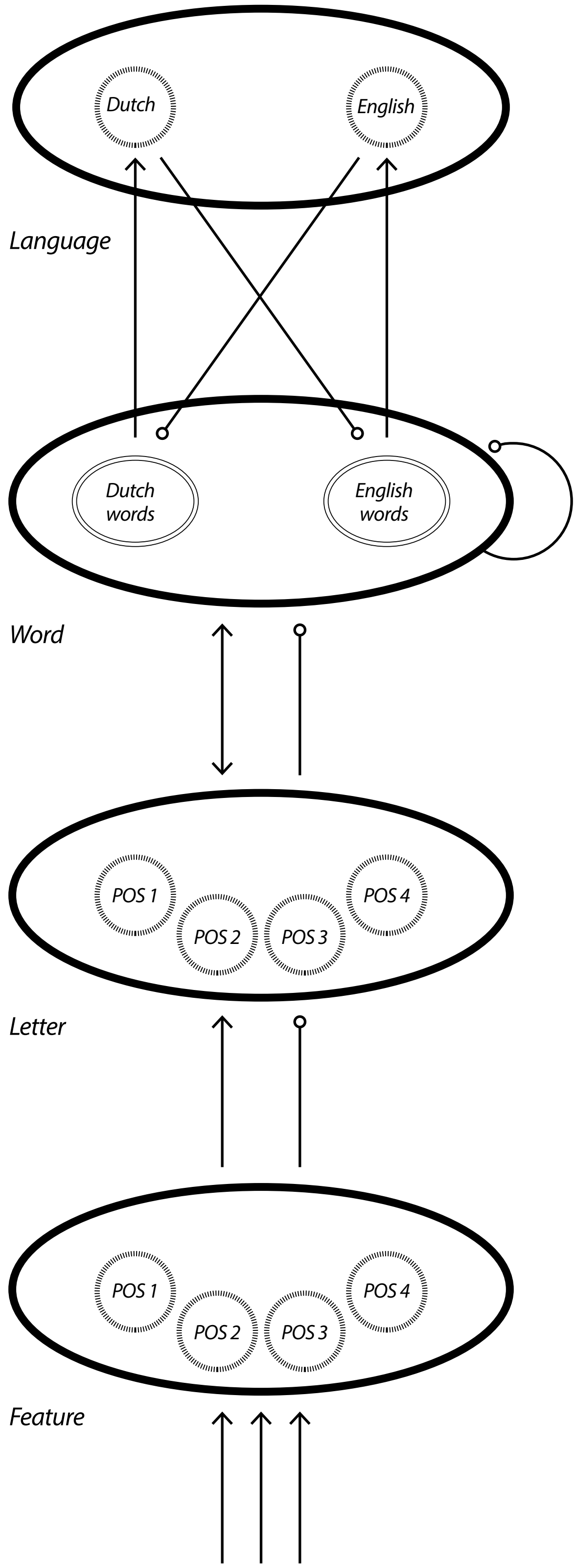


Figure 1.1.
The BIA model. The visual word causes activation at the feature level, which is relayed to the letter level, which is in turn transmitted to the word level, where the activated lexical representations activate the language nodes. Pointed arrows represent excitatory connections, whereas arrows with a black circle represent inhibitory connections. The 'self reference' to the word level indicates that active lexical representations send inhibitory signals to other active lexical representations.

layer, all words are connected to each other with inhibitory links, such that activation at a lexical node sends out inhibitory input to all other words. As a result, the activation level in the receiving lexical nodes is decreased, at least if these nodes have received excitatory input from the letter layer (e.g., because the words they represent are orthographic neighbors of the input word). This inhibitory cross-talk among the lexical representations is a highly dynamic process, which eventually results in threshold activation at one node. This is the moment when the word is recognized.

An important property of IA models is that they are fully interactive, i.e., whenever a node at a higher level is activated, it sends back activation to the lower-level nodes that are compatible with the information stored at the higher-level node. For instance, when the word cold is activated, it will send activation feedback to the letter representations for c, o, l, and d. At the same time these letter nodes will also receive feedback from other partially activated words containing these letters[11], e.g., ***bold, calm, cord, lord***, etc. (a smaller amount of activation, as these words do not receive input from the four input letters of the visual stimulus ***cold***). The combination of these feedforward and feedback processes soon results in a strong, self-reinforcing 'activation loop' between the four target letters and the target word, which makes for a fast rise in the activation level of the target's orthographic representation. This highly interactive process obtains at all representational levels. As this description makes clear, IA models have been named after their two key ingredients: 'interactivity' and 'activation'.

The BIA model has adopted this basic architecture and extended the model in two ways: (a) all words from L2 are integrated in the same lexicon, which is reflected in the form of full interconnectivity among the words of L1 and L2, i.e., L1 and L2 words can inhibit each other's activation level in the same way that L1 representations can inhibit each other in an IA model for a single language and (b) a level of language nodes has been added above the lexical level, one node for each language. All words belonging to a language are connected to the language node representing this language and, when their lexical representation is (partially) activated, they make its activation level rise. The most strongly activated language node does not send activation feedback to the lexical representations from which it receives excitatory input. However, it sends inhibitory feedback to the lexical representations that are linked to the other language node. Hence, the language nodes serve a double purpose: they make it possible to identify a word as a member of a language and they can neutralize the lexical representations in the language that is least activated. Note that a language can thus be selectively blocked. Note that this does not amount to selective lexical access. Access to the BIA model is a non-selective access process. The language suppression that is caused by the top down inhibitory effect of a language node is the final processing stage in the lexical processing course.

Dijkstra and van Heuven (2002) improved the BIA model by making the following changes. First, as the BIA model only incorporated orthographic representations (like other IA models) the BIA+ model is enriched with phonological and semantic representations. Second, they eliminated the top down inhibition between the language nodes and the word level. In the BIA+ model, activation in the mental lexicon is only the result of a bottom up process and cannot be affected by top down knowledge about the probability that the stimulus belongs to a certain language. This turns the original model into a model whose lexical activation process is entirely stimulus-driven. Finally, a decision level is added, which has access to the

[11] In IA models letter representations are position-specific, e.g., there is a representation for the letter 'c' in positions 1, 2, 3, and 4. Note that IA models are generally restricted to the processing of four-letter words.

most activated lexical representations and their language membership, and evaluates this information in the light of external factors, like task demands and list composition (e.g., Dijkstra et al., 1998). This eventually leads to a response, which is then sent to a response buffer for execution. A decision level was added because many experiments had revealed that the same items give rise to different effects (both in terms of size and direction), depending on task demands. With the benefit of hindsight, the inclusion of a decision level is not surprising. Behavioral experiments can only be performed when a task-dependent decision is made. In such experiments, there cannot be a direct 'route' from the mental lexicon to the response. In section 1.5, it has become clear that neuroscientific methods do make it possible to study what happens in the brain when participants perform a task that requires no response on the critical items (e.g., Midgley et al., 2008). However, behavioral tasks require participants to make a conscious decision about the outcome of the lexical activation process, and map this decision onto a response, which is then sent to an output buffer for execution. This makes it impossible for these tasks to directly observe the patterns of lexical activation[12].

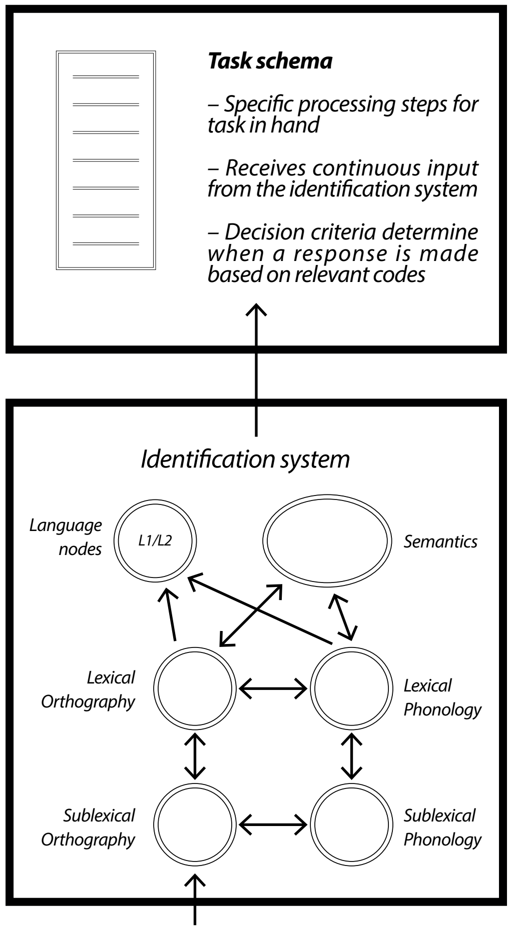


Figure 1.2.
The BIA+ model. Compared to the BIA model this improved version includes phonological and semantic representations. The model is fully interactive (see double-pointed arrows). Only the language nodes are unidirectional: they are activated by lexical activation but cannot suppress activation in the lexicon (as in the BIA model), making the activation process entirely bottom up and autonomous. Finally, a decision system is added, where the output of the mental lexicon is translated into a response.

The BIA+ model (Figure 1.2) has been particularly successful in the literature. Many researchers who find evidence in favor of language-independent access to a fully integrated lexicon attempt to account for their data in terms of the BIA+ framework. This does not by itself mean that the model cannot be improved in the future or has no problems accounting for some findings. However, it is very successful in explaining the large variety of experimental results across a variety of factors: word type (cognates vs. IHs), various types of co-activated information (cross-language neighborhood size, cross-language morphological family size), list composition (pure vs. mixed language lists), task demands (language-specific lexical decision, generalized lexical decision, language decision, go/no-go task), etc.

Note that our commitment to the BIA+ model does not imply that no other models for bilingual processing have been developed. For instance, the Bilingual Single Network Model has been developed in the tradition of connectionist models, in which there are no localist representations (as in all IA models, including BIA and BIA+), but where words are encoded in the form of distributed activation patterns across the network, as the result of a process that

[12] It is comforting to see that there is a strong convergence between the effects that have been observed in behavioral tasks and findings in neuroscientific tasks. This shows that cleverly designed behavioral experiments can inform us a lot on what is not directly observable due to the intervention of a decision stage.

optimizes the mapping of the orthographic input features to a set of semantic features over a layer of hidden units (Thomas, 2002). Attempts have also been made to develop Self-Organizing Models for characterizing bilingual processing. Such models attempt to represent the vocabulary of two languages in a single distributed network and try to find out which information source(s) the model needs to achieve this goal, while producing the standard findings in the literature (e.g., SOMBIP, Li and Farkas, 2002).

1.8 The present dissertation

In this dissertation, I will study lexical processing in Dutch-English-French bilinguals. My focus will be on cognates and IHs. Sometimes, I will address variables that have already been investigated in previous studies but will focus on issues that have not been addressed yet, i.e., by adding conditions or contrasts that have not been studied before. I will also introduce entirely novel factors. Finally, I will address the issue of cognitive control in bilingual lexical processing, but in a context where this question has not been raised before.

Chapter 2 addresses the question how identical cognates are represented in the mental lexicon. Two cognate types will be used for that purpose: L1L2 cognates and L2L3 cognates. They will be presented in a generalized lexical decision task and a go/no-go task. The question is whether identical cognates will only yield faster responses when their two languages match the languages in the list or whether such response facilitation will also occur when there is a match between only one of the cognates' languages and one of the two languages in the list. To investigate this question, the stimulus list composition will be manipulated. Both cognate types will be presented in a pure English (L2) list, a Dutch-English (L1L2) list, and an English-French (L2L3) list. Based on the literature, response facilitation is expected when the L1L2 cognates are presented in the L1L2 list and when the L2L3 cognates are presented in the L2L3 list. However, will such facilitation also occur when the L1L2 cognates are presented in the L2L3 list and the L2L3 cognates in the L1L2 list? Will the mere presence of a second language in the list, which licenses a 'yes'-response, make it easier to respond to any identical cognate that activates two languages? Or will participants learn that the 'yes'-response is associated with the two languages that appear in the list and be hesitant to respond when a cognate activates a language that is absent in the two sets of monolingual words in the list? This issue will be addressed in two experimental tasks: the generalized lexical decision task and the go/no-go task. The major goal behind the use of these two cognate types and this experimental design is to investigate the nature of identical cognates' lexical representation. Several possibilities have been advanced in the literature, but it is only recently that one has begun to experimentally investigate this question (e.g., Peeters, Dijkstra, & Grainger, 2013). The six experiments reported in Chapter 2 are intended to contribute to this debate.

Chapter 3 focuses on IHs. The same three types of stimulus lists will be used as in Chapter 2: a pure L2 list, a L1L2 list, and a L2L3 list. The List factor will be orthogonally crossed with the factor Task. This time, the language-specific L2 lexical decision task and the go/no-go task will be used. As there are insufficient English-French (L2L3) IHs, only IHs of the L1L2 type will be studied. Based on earlier research with these IHs, inhibition is expected in the L1L2 list. The question is what will happen in the L2L3 list, where there are no monolingual words in the list that match one of the homographs' languages (i.e., L1). The presence of any non-target language (L3 in the monolingual words but also L1 in the L1L2 homographs) may trigger a 'no'-response, which will cause a response conflict in the case of L1L2 homographs, and, hence, a delay in response times. Alternatively, participants may learn during the experiment that L2 words require a 'yes'-

response and L3 words require a 'no'-response, and experience no response conflict when L1L2 homographs are presented because they have not learnt to map L1 words to a response. In this chapter, I will make an in-depth analysis of the rich dataset that is obtained as the result of using the same set of IHs and their controls in all six cells of the design. More particularly, the statistical analyses will not be restricted to (generalized) linear mixed models. I will also present the results of quantile analyses to study the evolution of the effect across the distribution of response latencies. Based on these analyses I will derive delta plots and calculate whether the slope of the effect remains constant throughout the distribution of response times or shows sudden 'jumps'. Finally, ex-Gaussian analyses will be reported to separate the normal and exponential distributions that underlie the response time distribution. This will make it possible to find out whether a homograph effect is reflected in the mean of the normal distribution and/or in the region of extreme response times (i.e., the right tail of the distribution). The combination of these analyses makes it possible to not only compare the means of the IHs and their controls but to study where an effect is situated in the distribution of response times and how it evolves as a function of response speed (see e.g., Balota, Yap, Cortese, & Watson, 2008).

Chapter 4 will address a question that has not been investigated thus far. What is the effect of the proportion of monolingual words that belong to the non-target language of the IHs? Again IHs of the L1L2 type will be used. It is well known since Dijkstra et al. (1998) that there is a huge difference in the homograph effects that are observed in stimulus lists that contain 0% vs. 50% monolingual words from the non-target language of the IHs. We replicated these effects and compared them to the effect in a list that contained only 10% of monolingual words from the non-target language. As we used L1L2 homographs and, hence, 10% of monolingual native language words, the addition of a small percentage of these words (L1 being the dominant and, hence, most salient language) may have a considerable impact on the homograph effect. If a significant effect is obtained that is smaller than the effect in the 50% condition, this will imply that the presence of monolingual words from the non-target language of the IHs does not cause an all-or-none effect. Instead, this finding would indicate that the effect depends on the proportion of non-target language words in the list, and, for that reason, gradually increases as the proportion of these words increases.

In the final experimental chapter, Chapter 5, I will address the question whether there is a relationship between the inhibitory effect on a set of IHs of the L1L2 type and two types of cognitive control tasks. The inhibition effect was obtained in a L2 lexical decision task with a list that contained 50% of monolingual L1 words. Bialystok has argued in several papers that bilinguals develop an advantage over monolinguals at the level of cognitive control because their frequent need to keep control over the language they want to use and the language(s) they have to ignore is a permanent 'training' in cognitive control (e.g., Emmorey, Luk, Pyers, & Bialystok, 2008). In her research, Bialystok typically compares bilinguals with monolinguals on their performance in cognitive control tasks. Due to the lack of monolinguals in Flanders (at least in the population used for the experiments in this thesis) I decided to compare the size of bilinguals' inhibition effect on IHs in a L2 lexical decision task with the size of their inhibition effect in two cognitive control tasks. I opted for an experimental task that has been shown to cause a strong inhibition effect, as the result of a response conflict (also in my own experiments in Chapter 3). The cognitive tasks were the Simon Task and the AX-CPT (Continuous Performance Task), two tasks that considerably tax a participant's cognitive control capacity.

Chapter 6 gives an overview of the general conclusions that can be drawn from these experiments. After summarizing the most important findings, I will comment on their significance for research on the bilingual lexicon.

CHAPTER 2
Representation of identical cognates: the role of task demands and stimulus list composition

2.1 Introduction

Consider a series of English words: ***bed, film, hand, lamp, water***, and ***weekend***. What is special about these words is that a monolingual speaker of Dutch will also understand them when seeing them in print. That is because they also occur in Dutch, with exactly the same spelling and meaning, although their pronunciations can differ. Such words are called cognates, more particularly, identical cognates, in contrast to cognates whose orthographic forms are highly similar but not identical, so-called near-cognates (e.g., ***green-groen, seven-zeven***). This special type of words raise the question whether bilinguals, like monolinguals, also access their L1 when reading an identical cognate that needs to be processed in L2 but also exists in L1. A more general question is whether bilinguals access ***all*** languages in which the cognate occurs. If so, which kind of representation(s) mediates this access to multiple languages? This is the basic issue of the current paper.

2.1.1 The representation of identical cognates

A large number of studies on bilingual word recognition have considered whether an input letter string activates word candidates from different languages, or only from the language that is task relevant. Said differently, the question is whether lexical access is language selective or non-selective. Since 1990, more and more evidence has accrued in favor of the language non-selective view. This is the theoretical position that all word representations from both languages that are identical (or similar[13]) to the input are initially co-activated. Many studies (e.g., Dijkstra, Grainger, & van Heuven, 1999; Duyck, 2005; Van Hell & Dijkstra, 2002) have shown that even if bilinguals are processing words in one language, words in the other language become active as well. This language non-selective access view on processing is often combined with the theoretical view that the mental lexicon is shared and integrated across the two languages (van Heuven, Dijkstra, & Grainger, 1998).

The issue of bilingual lexical access has been considered by studying the processing of words that have identical or similar orthographic representations across languages. This is where identical cognates come in: translation equivalents that have an identical spelling pattern across two or more languages. Interlingual homographs also lend themselves to this purpose, i.e., words that overlap in orthographic form but differ in meaning across the languages (e.g., room meaning 'cream' in Dutch), which is why they are often referred to as 'false friends'. When these two word types are processed differently from one-language control words, this is evidence in favor of language non-specific access. Overall, cognates are

---

[13] The current paper is concerned with identical cognates only. However, it has also been found that near-cognates (e.g., green-groen) also activate the non-target language representation.

in fact often processed faster than controls, whereas interlingual homographs are sometimes processed slower than controls (Dijkstra, Van Jaarsveld, & Brinke, 1998; Von Studnitz & Green, 2002). Both findings converge on the conclusion that these words access their reading in each language in which they occur, i.e., that lexical access is language-nonselective. Whereas the processing speed of these words sheds light on the nature of lexical access, it does not inform us on the nature of their lexical representation. In this paper, we will focus on this representational issue, in particular with respect to identical cognates. The large degree of cross-language overlap for these words might result in shared orthographic and semantic representations in the bilingual lexicon. In the next paragraph, we will review studies that address this issue.

Recent findings suggest that identical cognates are a special type of cognate, with possible consequences at the representational level. For instance, stronger effects have been observed for identical cognates than for non-identical cognates, both in terms of facilitation and inhibition (Cop et al., 2016; Dijkstra et al., 2010; Font & Lavaur, 2004; Mulder, Dijkstra, Schreuder, & Baayen, 2014; Peeters et al., 2013). Dijkstra et al. (2010) studied how Dutch-English translation equivalents (***lied*** - 'song'), near-identical cognates (***tomaat*** - 'tomato'), and identical cognates (***lamp*** - 'lamp') are processed by Dutch-English bilinguals. They found a significant increase of facilitation in a visual lexical decision task (LDT) and inhibition in a language decision task as a function of increasing orthographic similarity. Unexpectedly, however, the facilitation effect was disproportionally larger for identical cognates than for nearly identical cognates. This finding suggests that identical and non-identical cognates have qualitatively different representations. In this vein, Dijkstra et al. proposed that nearly identical cognates have two orthographic representations, each of which is linked to a shared semantic representation. For these words, activation between the orthographic and semantic representations result in resonance that induces a cognate facilitation effect. At the same time, the two orthographic representations also compete for identification, due to differences in bottom-up activation and a process of mutual inhibition or bottom-up mismatch (i.e., lateral inhibition in an interactive-activation framework). In contrast, identical cognates might display a considerably larger facilitation effect than near-cognates because they have only one orthographic representation. As a consequence, the effect of semantic resonance would not be diminished by lateral inhibition, as in the case of non-identical cognates.

However, alternative representations for identical cognates can be considered. Peeters et al. (2013) proposed three different accounts. All of these hold that a single semantic representation is shared across languages, but the accounts differ in whether their orthographic and/or morphemic representations are shared or not. One view holds that identical cognates share an orthographic representation across languages. The second view posits the shared representation at the morphological level. As this level is believed to capture the relation between the forms and meanings of several words in a single language (Bybee, 1985), a cross-language morphological representation could capture the form-meaning correspondences of identical cognates in two or more languages. Finally, a third position holds that identical cognates have two language-specific morphemes. The first position (Midgley, Holcomb, & Grainger, 2012; Voga & Grainger, 2007) focuses on the cognate advantage in terms of cumulative frequency, which makes its orthographic representation more accessible than that of a matched control. The second view (Davis et al., 2010; Lalor & Kirsner, 2000) attributes the cognate advantage to the increased accessibility of the shared morphemic representation, which is also a cumulative frequency effect. In contrast, the third position (Dijkstra et al., 2010;

Van Hell & Dijkstra, 2002) explains cognate effects in terms of a resonance between a shared semantic representation and two morphemic representations. One theoretical argument that orthographically identical cognates are characterized by two morphemic representations rather than one, is that language-specific plural markers, gender, diminutives, determiners, and so on are specified at the morphological level. For example, the identical Dutch-English cognate '***hand***' has different plural forms in Dutch ('***handen***') and English ('***hands***'). Peeters et al. (2013) present both behavioral and ERP evidence that English-French bilinguals indeed possess different morphological representations in French and English. Their finding that a cumulative frequency effect across L1 and L2 was lacking and that L2 frequency had a stronger effect on low-frequency L1 words was incompatible with the notions of a single orthographic or morphemic representation (Mulder et al., 2014).

2.1.2 The effect of a dominant and a non-dominant language in trilinguals' L2 processing

In the current study, we examined the representation of identical cognates in a series of LDT and go/no-go experiments. We tested trilinguals (Dutch-English-French) rather than bilinguals, because this allowed us to manipulate the dominance relationship between target and non-target language. Most studies (Brenders, van Hell, & Dijkstra, 2011; Bultena, Dijkstra, & van Hell, 2014; Lemhöfer et al., 2008) have studied the effect of the more dominant language (i.e., L1) on cognate processing in the less dominant language (i.e., L2). We wanted to find out if the less dominant L3 also affects the processing of identical cognates in L2. As we explain below, this manipulation provides a window on how identical cognates are lexically represented. To study the effects of the more and less dominant languages separately, we studied identical cognates that only exist in two languages of our trilinguals, i.e., L1 and L2 (henceforth: L1-L2-cognates, e.g., ***hand***) or L2 and L3 (henceforth: L2-L3-cognates, e.g., ***art***).

The cognate recognition process in the different languages of trilingual participants has been studied before (Lemhoefer, Dijkstra, & Michel, 2004; Van Hell & Dijkstra, 2002). Van Hell and Dijkstra (2002) considered the effect of a foreign language on the native language in a word association task and LDT in Dutch, which was the L1 of Dutch-English-French trilinguals. In this experiment, the majority of the cognates were non-identical (e.g., ***bakker-baker*** in Dutch-English). They observed a significant cognate facilitation effect for Dutch-English (L1-L2) and a similar effect for Dutch-French (L1-L3) cognates with participants with a high (but not a low) French (L3) proficiency. For these non-identical cognates, the results were interpreted as a consequence of the co-activation of two different representations.

More relevant for our purposes is a study on the processing of identical cognates in trilinguals. Lemhöfer et al. (2004) studied Dutch-English-German trilinguals in L3 (German) LDT and found faster reaction times for "double" cognates (Dutch-German, e.g., ***kunst*** meaning 'art') and even faster responses for "triple" cognates (Dutch-English-German, e.g., ***wind***). These findings suggest that all languages of the multilingual can simultaneously become active. However, the findings of double and triple cognate effects on their own do not allow us to draw conclusions with respect to the lexical representation of identical cognates. The findings are compatible with a single or multiple orthographic or morphemic representations for identical cognates, because the facilitation effects for cognates could either be the cumulative effect (on a shared orthographic or morphemic representation) of the frequency of cognate usage in the different languages or the effect of the resonance between a single, shared semantic representation and distinct morphemic or orthographic representations in each language.

As mentioned above, we examined this issue by manipulating the dominance relationship between the target and the non-target languages, i.e., by comparing L1-L2 cognates and L2-L3 cognates. If processing proceeds via a shared orthographic or morphological representation (views 1 and 2 above), the representation in L2 should be affected by the ***cumulative frequency*** of all relevant languages (the sum of the cognates frequencies in L1 and L2, or in L2 and L3). Hence, both L1-L2-cognates and L2-L3-cognates should be recognized faster in L2 than matched L2 control words. In contrast, when there are different representations for the cognates in the three languages (view 3 above), the weaker L3 is less likely to affect the processing of the more accessible L2 representation than the stronger L1 representation[14].

2.1.3 Stimulus list context and experimental task

The use of two cognate types allowed us to investigate the background effect of an absent language in the stimulus list. We made the prediction that the cognate facilitation effect for L1-L2-cognates processed in L2 would increase when the list contained a large number of L1 words. However, it is not known whether a cognate facilitation effect arises when the list contains words from a language that is not represented in the list, i.e., from neither the language of the cognates' target reading nor that of their non-target reading (e.g., the effect of L1-L2-cognates in a list consisting of only L2 and L3 words). To study this issue, L1-L2-cognates and L2-L3-cognates were presented in three contexts: pure English, mixed English - Dutch, and mixed English - French. This design allows a study of the effect of the second language in the stimulus list and of the language that does not appear in the list (henceforth: the absent language). For instance, we can test whether a list containing English (L2) and Dutch (L1) words, increases the facilitation effect (relative to the pure English condition) for both Dutch-English (L1-L2) cognates (due to the presence of L1) and English-French (L2-L3) cognates (due to language intermixing, despite the absence of L3). The same question can be raised when the list contains English and French words and the two types of cognate words. In both cases, the question is whether the absent third language will still have an effect, and, if so, whether it is as strong as that by the language that is present in the list.

Finally, we manipulated the experimental task of our study. Besides stimulus list context, properties of the experimental task also affect the obtained patterns and directions of cross-linguistic effects (Bultena et al., 2014; Dijkstra et al., 2010). In our experiments, we used a (generalized) LDT (GLDT) and an English go/no-go task (GNGT). The LDT was only "generalized" in the English-Dutch and English-French context; in the pure English context, it was similar to English LDT, because there were no words from any other languages. The English GNGT is to some extent similar to the Language Decision task, but has hardly been applied with cognates. In the GNGT, the cognate's non-target language, which is associated with the 'no-go'-response, is expected to cause inhibition due to the response conflict with the 'go'-response which is associated with the target language. Hence, L1-L2-cognates in a list that also contains pure L1

[14] Note that the subjective frequency of L2 and L3 words in the unbalanced trilinguals must be lower than that found in word counts for those languages for L1 users.

and L2 words are expected to produce stronger facilitation effects in GLDT than in the GNGT. The same reasoning applies for L2-L3 words in a list containing pure L2 and L3 words.

2.1.4 Contrasting predictions of models proposing shared or multiple representations

An account in which identical cognates share an orthographic and/or morphological representation makes different predictions than one in which there are multiple representations (one for each language) at both the orthographic and morphological levels. In a single representation view, the processing of an identical cognate will necessarily benefit from the word's presence in another language, even when that language is less dominant (L3). Indeed, here the accessibility of the shared representation is determined by the cumulative frequency of the word across the languages in which it occurs. In contrast, the multiple representation view makes it possible that, despite the language non-selective nature of the access process, the orthographic and/or morphological representation in the target language is accessed before or after the representation in the non-target language.

Hence, models with a shared orthographic or morphological representation (the first two views above) predict facilitation in a GLDT relative to one-language controls, for both L1-L2-cognates and L2-L3-cognates. Moreover, they predict this effect in all three list contexts, irrespective of whether the list contains monolingual words from the cognates' non-target language (e.g., L1-words in the case of L1-L2-cognates) or not (e.g., L1-words in the case of L2-L3-cognates or a pure English context). In contrast, the single representation view predicts inhibition in the GNGT, at least when the task requires the 'no-go'-response to be associated with the cognate's non-target language. If cognate processing involves a shared orthographic or morphemic representation, participants will retrieve the information that the word belongs to two languages simultaneously. Thus, they will find themselves confronted with a response conflict, because one language is associated with the 'yes'-response and the other with the 'no'-response. This is the case for L1-L2-cognates in the English-Dutch stimulus list and for L2-L3-cognates in the English-French stimulus list. It is unclear what happens to cognates for which non-target reading belongs to a language that does not occur in the list (e.g., L2-L3 cognates in the English-Dutch list). During the task, the cognates' non-target language should not become associated with the 'no-go'-response, and no response conflict and consequent response delay are expected. However, because the 'go'-response is assigned to only one language, participants might immediately classify any non-target language of a cognate in the 'no-go'-category. This would result in a response conflict and delay in response time. Thus, inhibition might emerge for both cognate types.

An account that postulates different morpheme representations for cognates does not necessarily predict a facilitation effect in the GLDT. Such an effect may depend on the relative accessibility of the cognate readings. According to this kind of account, responses to cognates in this task will be driven by the representation that is most accessible. Facilitation is expected for L1-L2 cognates, because the L1 reading is more accessible than the L2 reading. However, no facilitation is expected for L2-L3 cognates, because the L2 reading is more accessible than the weaker L3 reading. Differences in the accessibility of orthographic and/or morphological representations will also play a role in the go/no-go task when L2 is the target language. Inhibition is expected for L1-L2 cognates: The faster access to the L1 representation will trigger a 'no-go'-response. This will cause a response conflict with the 'go'-response that is triggered by the more slowly available L2 representation. However, no inhibition is expected for L2-L3 cognates. The faster access to the L2 representation will trigger a 'yes'-response. This response can be executed before the representation in the weak L3 is accessed.

The shared representation and multiple representation accounts make the same predictions with respect to the list composition effect for L1-L2 cognates, but not for L2-L3 cognates. When the list contains many words from the cognates' non-target language, the association between that language and the response will be stronger compared to the situation in a pure L2 list. When L1-L2 cognates are presented in a GLDT in a list incorporating many L1 and L2 monolingual words, the L1 words will cause a strong association with the 'yes'-response. According to the shared-representation view, identical cognates result in simultaneous activation of both readings. As a result, the strong language-to-response association for L1 will activate the 'yes'-response more strongly than in a pure L2 list, resulting in a larger facilitation effect. According to the multiple representations view, L1-L2 cognates will activate L1 more quickly than L2 and trigger a stronger association between L1 and the 'yes'-response than in a pure L2 list. Hence, both accounts predict stronger facilitation for L1-L2 cognates when the list contains many L1 monolingual words. The two accounts also predict stronger inhibition for L1-L2-cognates in the GNGT.

Our account for GNGT is almost the same as that for the GLDT, the only difference being that the non-target language L1 is associated with the 'no'-response. As this response is stronger in a list with many monolingual words from the non-target language than in a pure L2 list, a stronger inhibitory effect is predicted. However, the accounts differ with respect to their predictions for the L2-L3 cognates. In a list context where many monolingual L2 and L3 words are presented, a strong connection is established between L3 and the 'yes'-response (GLDT) or the 'no'-response (GNGT). According to the shared-representation view, this stronger association will be triggered simultaneously with the other language-to-response association, i.e., between L2 and the 'yes'-response. This is because the cognates share a representation, which triggers both languages simultaneously. As a result, this account predicts the same effects of stimulus list composition for L2-L3 cognates as for L1-L2 cognates. In contrast, the multiple representations view predicts no effect of stimulus context for L2-L3 cognates. Because the different representations differ in their accessibility, list composition will only have an effect when the strong connection between the non-target language and the response can be activated, i.e., when the non-target language is faster available than the target language. This is the case for L1-L2 cognates but not for L2-L3 cognates. Hence, the multiple representation view predicts that a list including many monolingual words from the non-target language will strengthen the facilitation (GLDT) and inhibition (GNGT) effects for L1-L2 cognates but not for L2-L3 cognates.

To test the predictions of these accounts and answer the research questions formulated above, we will now present a series of six experiments with trilinguals (see Table 2.1) in which

Table 2.1:
Experiments of the current study

		CONTEXT		
		English	English-French	English-Dutch
TASK	Generalized LDT	EXP 1	EXP 2	EXP 3
	English GNGT	EXP 4	EXP 5	EXP 6

three factors are orthogonally manipulated: Cognate Type (Dutch-English and English-French), Context (pure English, English-French, and English-Dutch), and Task (GLDT vs. GNGT).

2.2 Experiment 1: GLDT with L1-L2 and L2-L3 cognates in pure English context

2.2.1 Method

Participants
Twenty Linguistics students of the University of Antwerp were tested (18 women, 2 men; range: 19-28 years of age; mean age: 20.6 years). Nineteen were right-handed and one left-handed. They all had normal or corrected-to-normal vision. All participants were unbalanced Dutch-English-French trilinguals, i.e., native speakers of Dutch, who were proficient in English (L2), and had a good knowledge of French (L3). They had studied English as a foreign language at secondary school for at least six years (mean experience: 9.7 years) and used this L2 on a regular basis in the context of their studies and their spare time (when watching movies, reading on the internet). They had studied French at secondary school for at least six years as well. They were much less familiar with this foreign language, and most of them used it only occasionally in informal situations outside the university.

Materials and Design
The critical items were two sets of identical cognates: 20 Dutch-English ones (L1-L2) and 20 English-French (L2-L3) ones. All cognates were in the same frequency range in all three languages, however we realize that the subjective frequency was higher in the more dominant language i.e., Dutch and English, respectively. When selecting these cognates, we used the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995) for the Dutch-English items and WordGen (Duyck, Desmet, Verbeke, & Brysbaert, 2004) for the English-French ones. We compared these frequencies to the ones that were listed in the SUBTLEX databases, once the latter became accessible (Keuleers, Brysbaert, & New, 2010; New, Brysbaert, Veronis, & Pallier, 2007; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The frequency range remained similar for all items.

We asked 10 randomly selected individuals, drawn from the same population as our participants, to decide for each cognate on (a) the similarity of its meanings in the two languages, (b) the similarity of its pronunciations in these languages, and (c) the familiarity of the word in the participants' L2. Based on these ratings (1 = "none"; 5 = "very strong") the 40 words with the highest mean scores across these three dimensions were selected: 20 Dutch-English cognates and 20 English-French cognates. We excluded words like film, which occur in all three languages of our trilingual participants.

All 20 Dutch-English cognates had similar mean frequency in both languages (Dutch: 4.65, SD = 0.63, range: 3.48-6.00; English: 4.66, SD = 0.62, range: 3.25-5.55). They were all three to five letters long and were matched to a monolingual English control word on an item-by-item basis, both on letter length (mean: 3.85 letters for both cognates and controls) and on the Zipf value[15] (mean: 4.66 for cognates and 4.35 for controls). A matched-pairs t-test showed no significant difference between cognates and controls on these frequencies ($t = 1.86$; $p > 0.05$).

[15] Zipf value = log10(frequency per million*1000), based on (Van Heuven et al., 2014)

All 20 English-French cognates had similar frequency in both languages (English frequency: 4.53, SD = 0.61, range: 3.10-5.60; French: 4.65, SD = 0.43, range: 4.00-5.39). Note that these frequencies apply to native speakers of French, and are unlikely to correspond to our participants' (subjective) frequencies. However, by selecting cognates with a higher frequency in L3 than in L2 we maximized the chance to measure an effect of L3 on the target language (L2) in the experiment. All cognates were three to six letters long and were matched to a monolingual English control word on an item-by-item basis, both on letter length (mean: 5.05 letters for both cognates and controls) and Zipf value (mean: 4.53 and 4.56 for cognates and controls, respectively). A matched-pairs t-test on these frequencies showed no significant difference between the frequency of the cognates and their English controls ($t = 0.23$; $p > 0.1$).

Note that the Dutch-English and English-French cognates had virtually the same mean frequencies and standard deviations, to ensure that different results for the two word types could not be ascribed to differences in the familiarity with the items.

Finally, we selected an extra 20 English filler words from the same frequency range as the other English words in the experiment, such that the item list consisted of 100 words. We added 100 English-like non-words, i.e., items derived from English words, by means of the Wuggy program (Keuleers & Brysbaert, 2010). These non-words were matched item-by-item on letter length with English words.

Procedure
The experiment was run in a dimly lit room on DELL computers (type Optiplex 380) that were connected to a 15" DELL monitor. The stimuli were presented in the center of the screen. Reaction times (RTs) were registered with the DMDX program (Forster & Forster, 2003). A game controller (type Logitech Wingman Precision) was used for responding (using the left and right front buttons) and for initiating a new item block (using the START button).

Participants received written instructions in English, with the purpose of bringing them in an English language mode. They were instructed to decide as quickly and accurately as possible whether the letter string on the screen was a word in any language they knew or not with 'yes'-responses being made with the dominant hand. Which languages occurred in the experiment was not specified, to avoid that expectations might affect the way in which the words were processed. This was also a methodological decision, as we wanted to keep the instructions constant across the first three experiments, which differed only in the composition of the stimulus list (pure English, mixed English-French, and mixed English-Dutch lists). In the current experiment only English words were used, and all non-words were derived from English words, making it a classical lexical decision task in L2. The implicit suggestion in the instructions that different languages might occur in the list did not confuse participants. In addition to the use of English instructions all communication between the experimenter and the participants was in English.

A set of 20 practice items preceded the experimental list. No cognates appeared in this practice session. The experiment itself was divided into five blocks of 40 items each (20 words and 20 non-words). The two sets of cognates and their controls were evenly divided over these blocks, 4 Dutch-English, 4 English-French cognates and their controls in each block. Each item was preceded by a fixation (+) sign, which appeared in the middle of the screen for 500 ms. The subsequent letter string stayed in view until a response had been made or until a time-out of 2,500 ms had passed. The next trial was initiated 500 ms after the response or time-out. The experiment took about 20 minutes.

2.2.2 Results and discussion

We used the R statistical software package (R Core Team, 2014) and the (generalized) linear mixed model in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016), which includes fixed and random effects parameters. The data were analyzed with a linear mixed effects model, using participants and items as crossed random effects (for more details see Baayen, Davidson, & Bates, 2008; Baayen & Milin, 2010; Bates, Maeckler, & Dai, 2007). Response times were analyzed with the lmer function and error rates with the glmer function.

During model-building we started with the model that only included the two random-effect predictors, i.e., participants and items, and added one fixed-effect parameter at a time. To compare two models that differed only in the presence or absence of a single fixed-effect parameter we used likelihood ratio tests with $\alpha = .05$ as the level of significance. If the model with an extra fixed parameter was significantly better than the simpler model, we proceeded with the more complex model and followed the same procedure for the next fixed effect. If no significant difference was found, the simpler model was adopted. The factor Word Type, which is the theoretically important factor in our analyses, was always included. The two types of cognates were analyzed separately.

It is important to be aware of the benefits offered by linear mixed effects models. They make it possible to fit a model to the data that predicts expected values on the dependent variable based on two types of factors: (a) fixed factors, which are generally one or more factors defining the experimental design (and their interactions) and (b) random factors, which introduce random variance in the data because they have been randomly selected from their population, i.e., the units on these variables have not been selected with a theoretical purpose in mind (in contrast to the design factors). This random variance has caused major problems in the analysis of psycholinguistic data ever since Clark's (1973) seminal publication on how to deal with random variance in such data (see also Forster & Dickinson, 1976; Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmer, 1999). The problem is that psycholinguists compare two or more experimental conditions by presenting a random sample of words for each condition to a random sample of participants. Thus, their data are affected by two sources of random variation. Furthermore, the words are generally nested within participants, which means that the two random effects are crossed. Traditional ANOVAs cannot solve this problem, which makes it necessary to perform two separate analyses: one based on participant means (averaging across item data) and one based on item means (averaging across participant data). The problem with such an approach is that information inevitably gets lost: the inter-item variability in the responses of each participant vanishes in her/his participant mean and the inter-participant variability in the responses to each item disappears in the item means. The novel statistical technique that Baayen et al. (2008) introduced in the field can solve this problem by using linear mixed effects models and generalized linear mixed effects models, which can deal with multiple sources of random variation (even when one random variable is nested within the other). Generalized linear mixed effects models are an extension of linear mixed effects models, such that the technique can be applied to other distributions than the normal one, for instance to binary data. This property makes them suitable for the error analyses in psycholinguistic experiments.

The move towards mixed effects models was a huge step forwards, and one that was highly needed, as these models had already been used for about a decade in many areas of science, medicine, etc. by the time Baayen et al.'s paper was published (p. 391). Crucially, these models do not have the weakness that characterizes ANOVAs: unlike ANOVAs they can disentangle the effects of several random factors in a single analysis.

As the same cognates and controls were used in all three generalized lexical decision experiments, list composition being the only difference, the data for these experiments were analyzed together. There are two reasons for performing an omnibus analysis. First, an analysis on the combined data of all six experiments benefits considerably from the choice to use a single set of cognate-control pairs in all design cells. Such an analysis increases the power for detecting a cognate effect and its interactions with the two design factors (List Composition and Experimental Task). Second, an omnibus analysis also avoids the problem that is associated with performing too many statistical comparisons on a single data set, i.e., an increased risk of obtaining Type I errors. The results of the (generalized) linear mixed model that will be reported makes it possible to assess the importance of all theoretically relevant contrasts by fitting a single model to the data.

In the final model-building stage the interaction between Word Type and List Composition was added. If the model including this interaction accounted for significantly more variance in the data, this model was retained. If not, the factor List Composition was included as a fixed factor in the model but its interaction with Word Type was left out. Analyzing the data of the three experiments together increased the power of the statistical analyses.

Table 2.2 shows the mean RTs and error rates for the two cognate types, together with the size and direction of the cognate effects and their associated significance levels. Inspection of the RT distribution revealed non-normality. The inverse transformation was applied to achieve more normally distributed data, using the following function: RTi = -1000/RT[16].

Table 2.2.
Mean Reaction Time (RT) and Error Percentage (ERR) for the two cognate types and their English controls (Experiment 1)

	COGNATE TYPE			
	Dutch-English		English-French	
Word Type	RT	ERR	RT	ERR
cognates	567	0.8	625	5.5
controls	616	7.6	602	4.5
effect	-49***	-6.8**	23	1.0

Note: ** p ≤ .01 *** p≤.001

Dutch-English cognates

One Dutch-English cognate (i.e., elf) and its control (i.e., oar) were removed because the control word produced more than 50% of error responses (n = 40, i.e., 5%). In addition, all

[16] The inverse RT was multiplied with the constant -1000, the inverse function reverses the sign of the beta coefficients and decreases their magnitude. Note that the RT transformation was applied to the RTs in all reported experiments, as it made the data approximate the normal distribution in all cases.

error responses were excluded from the RT analyses (n = 32, i.e., 4%). Finally, all responses outside the RT range between 300 ms and 1500 ms were removed from both the RT and error analyses (n = 3, i.e., 0.4%). This left us with 725 RTs for the two cognate types and their controls (i.e., 95.4% of the 760 remaining data points after removing one cognate-control pair).

The final model included the significant Word Type by List Composition interaction as a fixed-effect predictor. The model's reference level (intercept) represented performance on the identical cognates in a pure English context. Table 2.3 shows the predictors that are relevant for the current experiment and interactions with other experiments that will be discussed later. The effect of Word Type was highly significant (ß = 0.15, t = 5.50, $p < 0.001$), indicating that, in a pure English context, responses to Dutch-English cognates were significantly faster than to their English controls (567 vs. 616 ms). The effect of the previous response time was also highly significant (ß = 0.0003, t = 10.80, $p < 0.001$), indicating that fast (slow) responses tend to be followed by fast (slow) responses.

Table 2.3.
Estimates for the fixed-effect predictors in the RT analysis of the Dutch-English cognates in pure English context (Experiment 1) and interaction with the English-French and English-Dutch context.

	Estimate ß	SE ß	df	t	p
Intercept	-2.0940	0.05094	128.60	-41.10	< 0.001
Word Type (English list)	0.1465	0.02665	2078.0	5.50	< 0.001
Word Type x English-French list	0.0037	0.03623	2078.0	0.10	> 0. 1
Word Type x English-Dutch list	0.0795	0.03681	2078.0	2.16	< 0.05
Previous RT	0.0003	0.00002	2146.00	10.80	< 0.001

In the analysis of the errors neither the interaction between Word Type and List Composition nor the effect of List Composition were included in the final generalized linear mixed effects model. Hence, the three experiments in this series were analyzed together, using Word Type as the only fixed-effect predictor. As can be seen in Table 2.4 participants made significantly fewer errors on the Dutch-English cognates than on their English controls (ß=2.47, z=8.05, $p < 0.001$; cognates: 0.8%; controls: 7.6%).

Table 2.4.
Estimates for the fixed-effect predictors in the error analysis of the Dutch-English cognates

	Estimate ß	SE ß	z	p
Intercept	-5.6189	0.5510	-10.20	< 0.001
Word Type	2.4720	0.3072	8.05	< 0.001

English-French cognates

None of the cognate-control pairs caused errors systematically. Other error responses were excluded from the RT analyses (n = 40, i.e., 5%). Finally, all responses outside the RT range between 300 ms and 1500 ms were removed from both the RT and error analyses (n = 5, i.e., 0.6%). This left us with 755 RTs for these cognates and their controls (i.e., 94.4% of the 800 data points).

The final model for the RT analysis included Word Type, List Composition, and Previous RT as fixed-effect predictors. As the interaction between Word Type and List Composition did not improve the model's explanatory power, it was removed from the final model. Table 2.5 shows the effects of the predictors. As the non-significant effect of Word Type ($t < 1$) did not depend on the stimulus context in which the cognates appeared, this null effect obtained across all three experiments. In the current experiment the response times were even numerically larger for the English-French cognates than for their English controls (625 ms vs. 602 ms, respectively). The effect of Previous RT was significant ($t = 7.59$, $p < .001$), indicating a positive correlation between the RTs on successive trials.

Table 2.5.
Estimates for the fixed-effect predictors in the RT analysis of the English-French cognates (Experiment 1)

	Estimate ß	SE ß	df	T	p
Intercept	-1.8430	0.04846	120.1	-38.04	< 0.001
Word Type	-0.0090	0.01290	2178	-0.70	> 0.10
Previous RT	0.0002	0.00003	2240	7.59	< 0.001

In the error analysis, the model with the interaction term accounted for significantly more variance than the model without this term. Table 2.6 shows the effects of the predictors and interactions. The effect of Word Type was not significant ($z < 1$), i.e., the error rates on cognates and their controls did not differ significantly from each other (5.5% and 4.5%, respectively).

Table 2.6.
Estimates for the fixed-effect predictors in the error analysis of the English-French cognates (Experiment 1)

	Estimate ß	SE ß	z	p
Intercept	-3.5446	0.3807	-9.31	< 0.001
Word type (English list)	-0.2342	0.3278	-0.72	> 0.1
Word type x English-French list	0.5570	0.4756	1.17	> 0.1
Word type x English-Dutch list	-0.4664	0.4321	-1.08	> 0.1

2.3 Experiment 2: GLDT with L1-L2 and L2-L3 cognates in English-French context

2.3.1 Method

Participants
Twenty-five students from the same population as in Experiment 1 were tested (21 women, 4 men; range: 19-31 years of age, mean age: 21.7 years). Twenty-four were right-handed, one left-handed. All participants had normal or corrected-to-normal vision. They had the same

experience with English and French as the participants in Experiment 1. Their mean experience with English was 9.8 years.

Materials and Design
The same items were used as in Experiment 1. The critical items were 20 Dutch-English and 20 English-French cognates and their monolingual English controls. To create an English-French context we added 100 monolingual French words. Hence, the words in the list consisted of monolingual French words (n=100, 50%), monolingual English words (n=60, 30%), Dutch-English cognates (n=20, 10%), and English-French cognates (n=20, 10%). The monolingual French words were matched on an item-by-item basis to the 100 English words, both on letter length and word frequency. Frequency matching with respect to the cognates occurred based on their English readings (Dutch-English cognates) or French readings (English-French cognates). It allowed us to select 20 French controls that were matched on an item-by-item basis, both on the letter length (mean 3.85) and on the Zipf-value (mean 4.70, SD = 0.50, range: 3.60-5.63). A matched paired t-test with the French readings of the cognates showed no significant difference between cognates and controls (t=0.30, p>0.1)

Additionally, 100 French-like non-words (derived from different French words than those in the experiment) were generated by means of the Wuggy program (Keuleers & Brysbaert, 2010). These non-words were matched on letter length to the French words. Note that none of the French words and French-like non-words had language specific graphemes (e.g., é, è, ç etc.), because in this way the decision can be made only on basis of the orthographical and not lexical features.

Procedure
The same procedure was used as in Experiment 1.

2.3.2 Results and discussion

Dutch-English cognates
The Dutch-English cognate-control pair elf-oar was excluded in this experiment as well because the error rates on the control word was high (n=50, i.e., 5%). All error responses on the critical items (n=50, i.e., 5%) and RTs falling outside the range between 300 and 1,500 ms (n=10, i.e., 1 %) were excluded. This left us with 890 observations for these cognates and their controls (i.e., 93.7 % of the 950 remaining data points after removal of one cognate-control pair).

Table 2.7.
Mean reaction time (RT) and error percentage (ERR) for the two cognate types and their English controls (Experiment 2)

	COGNATE TYPE			
	Dutch-English		English-French	
Word Type	RT	ERR	RT	ERR
Cognates	580	1.30	635	3.00
Controls	643	9.30	624	4.00
Effect	-63***	-8.00**	-11	1.00

Note: ** p ≤ .01 *** p≤.001

Mean response times (RT) and errors (ERR) for the Dutch-English cognates are presented in Table 2.7, together with the size and direction of the cognate effects and their associated significance levels.

Recall that the cognate facilitation effect in the pure English context of Experiment 1 was significant. The absence of an interaction between Word Type and List Composition at the level of the English-French context ($t < 1$, see Table 2.3) reveals an equally strong cognate facilitation effect in the English-French list (63 ms) as in the pure English list (49 ms, reference level). The cognate facilitation effect in this experiment was also significant by itself. When the RTs for the cognates in the English-French context were used as the model's reference level the effect of Word Type was significant ($ß = 0.15$, $t = 6.12$, $p < 0.001$, data not shown).

In the error analysis, the final model did not include List Composition nor its interaction with Word Type. As mentioned in Experiment 1, the effect of Word Type was significant in the combined analysis of the three experiments in this series, indicating that significantly fewer errors were made on Dutch-English cognates than on their English controls. The absence of an interaction with List Composition indicates that the magnitude of this facilitation effect was not affected by the languages that were represented in the list.

English-French cognates

All error responses on the critical items (n=35, i.e., 3,5 %) and RTs falling outside the range between 300 ms and 1,500 ms (n=8, i.e., 0.8%) were excluded. This left us with 957 observations for these cognates and their controls (i.e., 95.7 % of the 1,000 data points).

The final model for the RT analysis included Word Type, List Composition, and Previous RT as fixed-effect predictors. As mentioned in Experiment 1, the effect of Word Type was not significant in the combined RT analysis of the three experiments in this series. The interaction between Word Type and List Composition at the level of the English-French context was not significant either (reference level: pure English context), indicating that the effect in the current experiment did not differ from the null effect in Experiment 1. The effect of Word Type was not significant when the cognates in the English-French context were used as the model's reference level either ($t < 1$). Hence, no cognate facilitation was obtained on the RTs for the English-French cognates in the English-French list.

In the error analysis (Table 2.6), the final model included the interaction between Word Type and List Composition. The interaction term between Word Type and List Composition at the level of the English-French context (i.e., pure English list vs. English-French list) was not significant ($z = 1.17$, $p > 0.1$). Hence, the effect of Word Type in the English-French list did not differ from the null effect in Experiment 1. When the data for these cognates in the English-French context were used as the model's reference level the effect of Word Type was not significant either ($ß = 0.32$, $z = 0.94$, $p>0.1$, data not shown). Hence, no cognate facilitation was obtained on the errors for the English-French cognates in the English-French list. It means, that comparable error rates were observed for the English-French cognates and their English controls in Experiments 1 and 2.

We also used a (generalized) linear mixed model to compare the English-French cognates with their matched French controls. The French controls were associated with significantly longer RTs than the cognates ($ß = 0.20$, $t = 7.37$, $p < 0.001$; 635 vs. 730 ms, respectively) and with significantly higher error rates (ß = 3.16, z = 4.74, p < .001; 3.00% vs. 28.8%, respectively).

2.4 Experiment 3: GLDT with L1-L2 and L2-L3 cognates in English-Dutch context

2.4.1 Method

Participants
Twenty-two students, drawn from the same population as the participants in the previous experiments, were tested (20 women, 2 men; age range: 19-31, mean age: 21.7 years). Seventeen were right-handed and 5 left-handed. They all had normal or corrected-to-normal vision. They had the same experience with English and French as the participants in the previous experiments. Their mean experience with English was 9.7 years.

Materials and Design
We used the same items as in Experiment 2, with the exception that the French words and French-like non-words were replaced by Dutch words and Dutch-like non-words. The Dutch words were matched on an item-by-item basis on letter length to the 100 English words. Frequency-matching was also performed on an item-by-item basis, more particularly, on the Dutch readings of the 20 Dutch-English cognates and on the English readings of the 20 English-French cognates and the 60 monolingual English words. It allowed us to use 20 Dutch words as additional L1 control items for Dutch-English cognates, all were matched on the letter length (mean 3.85) and on the Zipf-value (mean 4.40, SD = 0.79, range: 2.91-6.03). A matched paired t-test with the Dutch readings of the cognates showed no significant difference between cognates and controls (t=1.13, p>0.1)

Dutch-like non-words (derived from other Dutch words than the ones used as filler words or cognates in the experiment) were generated with the Wuggy program (Keuleers & Brysbaert, 2010). These non-words were matched item by item on letter length with the 100 monolingual Dutch words.

Procedure
The same procedure was used as in Experiment 1.

2.4.2 Results and discussion

Dutch-English cognates
As in the previous experiments, the Dutch-English cognate-control pair elf-oar was excluded because of the high error rates on the control word (n=44, i.e., 5%). All error responses were removed from the RT data (n = 44, i.e., 5%), as were all RTs that fell outside the range between 300 and 1,500 ms (n = 11, i.e., 1.4 %) This left us with 781 observations (i.e., 93.42 %, given the 836 data points).

Table 2.8 presents the mean RTs and error percentages for the two cognate types. In the final model for the RT analysis the interaction between Word Type and List Composition was included. The significant interaction between Word Type and List Composition at the level of the English-Dutch context (pure English vs. Dutch-English; ß = 0.08, t = 2.16, $p < .05$) indicates that the cognate facilitation effect was larger in the English-Dutch list than in the pure English list (103 ms vs. 49 ms, respectively). When the data for the cognates in the English-Dutch context were used as the reference level, the model's output revealed that the cognate facilitation effect in this context was also significantly larger than in the English-French context (ß = -0.08, t = -2.15, $p < .05$, data not shown). The latter model also revealed that the effect in the English-Dutch context was significant by itself, i.e., the effect of Word Type was highly significant (ß = 0.23, t = 8.90, $p < 0.001$, data not shown).

Table 2.8.
Mean reaction time (RT) and error percentage (ERR) for the two cognate types and their English controls (Experiment 3)

	COGNATE TYPE			
	Dutch-English		English-French	
Word Type	RT	ERR	RT	ERR
Cognates	589	1.00	695	8.90
Controls	692	9.60	701	5.00
Effect	-103***	-8.60**	-6	3.90

Note: ** $p \leq .01$ *** $p \leq .001$

In the error analysis, the final model did not include List Composition nor its interaction with Word Type. Hence, Word Type was the only fixed-effect predictor in the combined analysis on the error data of the three experiments in this series. As mentioned in Experiment 1, the significant effect of Word Type indicated that fewer errors were made on the Dutch-English cognates than on their English controls. The absence of an interaction with List Composition indicates that the size of this cognate facilitation effect was independent of the presence of other languages in the list.

Finally, we compared the Dutch-English cognates to their Dutch controls. This comparison revealed significant effect of Word Type in the analysis of response times (cognates: 589 ms, controls: 611 ms; ß = -0.05, t = -2.15, $p < 0.05$). However, no significant effect of Word Type appeared in the analysis of the accuracy data (cognates: 1.0%, controls: 0.7%; ß = 1.40, t = 1.24, $p > 0.1$).

English-French cognates
All error responses on the critical items (n=61, i.e., 7 %) and RTs falling outside the range between 300 ms and 1500 ms (n=9, i.e., 1.1%) were excluded. This left us with 810 observations for these cognates and their controls (i.e., 92.04 % of the 880 data points).

Recall that the final model for the RT analysis did not include the Word Type by List Composition interaction and that the effect of Word Type was not significant in the combined RT analysis of the three experiments in this series. The effect of Word Type was not significant (see Table 2.4).

In the analysis of the error data, the interaction between Word Type and List Composition (pure English vs. English-Dutch list) was not significant (ß = -0.47, z = -1.08, $p > 0.1$). As there was no significant cognate effect in the error analysis of Experiment 1, this finding indicates the absence of a cognate effect in the error data of the Dutch-English context as well. The effect of Word Type was significant by itself, i.e., when the English-Dutch list was adopted as the reference level ($ß = -0.70$, $z = -2.49$, $p < 0.05$, data not shown), but the effect of List Composition was not significant, i.e., comparable error rates were observed for the English-French cognates and their controls in an English-Dutch list as in a pure English list.

2.5 Experiment 4: GNGT with L1-L2 and L2-L3 cognates in pure English context

2.5.1 Method

Participants
Twenty students from the same population as in the previous experiments were tested (16 women, 4 men; range 19-26 years of age, mean age 20.5 years). Sixteen of them were right-handed and 4 left-handed, had normal or corrected-to-normal vision. They had the same experience with English and French as participants in the previous experiments. Their mean experience with English was 9.7 years.

Materials and Design
The same stimuli from the Experiment 1 were used.

Procedure
The experiment was run in a dimly lit room on the same computers as previous experiments using the DMDX program (Forster & Forster, 2003). A different written instruction in English was given, where participants were instructed to decide as quickly as possible and as accurately as possible, whether the letter string presented on the screen was an English word ("go" button). If not, participants were instructed to wait till the next word appeared (no-go). The button on the side of the dominant hand was assigned to the "go" button. Practice and experimental trials were presented in the same order as in the Experiment 1. The item stayed in view until a response had been made or until a time out of 2,500 ms had passed. The next trial was initiated 500 ms after the response or time out.

2.5.2 Results and discussion

As we also used the same cognates in the GNGT, we analyzed all three experiments together. We followed the same model-building steps in (generalized) linear mixed model (see more

Table 2.9.
Mean Reaction Time (RT) and Error Percentage (ERR) for the two cognate types and their English controls (Experiment 4)

	COGNATE TYPE			
	Dutch-English		English-French	
Word Type	RT	ERR	RT	ERR
Cognates	603	1.6	634	2.8
Controls	628	8.7	630	3.0
Effect	-25	-7.1*	4	-0.2

Note: ** $p \leq .01$ *** $p \leq .001$

detailed description in Experiment 1). Table 2.9 shows the mean RTs and error rates for the two cognate types, together with the size and direction of the cognate effects and their associated significance levels.

Dutch-English cognates

Similarly to the previous experiments, the Dutch-English cognate-control pair elf-oar was excluded because of the high error rates on the control word (n=40, i.e., 5%). All error responses were removed from the RT data (n = 34, i.e., 4.25%), as were all RTs that fell outside the range between 300 ms and 1,500 ms (n = 16, i.e., 2 %). This left us with 710 observations (i.e., 93.42 %, given the 760 data points).

Word Type and List Composition interaction was significant as a fixed-effect predictor and was included in the final model. Table 2.10 shows the predictors and interactions for the pure English context as reference level (intercept). The effect of word type was marginally significant (ß = 0.07, t = 1.98, p = 0.05), however analyzing this experiment and cognate type separately showed non-significant results (ß = 0.08, t = 1.90, p > 0.05, data not shown). It indicates, that cognate effect is very weak and, although participants tend to react faster to the cognates (603 ms) than to the pure English controls (628 ms), the effect is not strong enough to reach significance. The effect of the previous response time was highly significant (ß = 0.0003, t = 6.38, p < 0.001), indicating that fast (slow) responses tend to be followed by fast (slow) responses.

Table 2.10.
Estimates for the fixed-effect predictors in the RT analysis of the Dutch-English cognates in pure English context (Experiment 4) and interaction with the English-French and English-Dutch context.

	Estimate ß	SE ß	df	T	p
Intercept	-2.0050	0.06381	140.70	-31.43	< 0.001
Word Type (English list)	0.0773	0.03877	1166.0	1.98	= 0.05
Word Type x English-French list	-0.0600	0.05083	1165.0	-1.18	> 0. 1
Word Type x English-Dutch list	-0.3526	0.05165	1167.0	-6.83	< 0.001
Previous RT	0.0003	0.00004	1220.0	6.38	< 0.001

Table 2.11.
Estimates for the fixed-effect predictors in the error analysis of the Dutch-English cognates (Experiment 4)

	Estimate ß	SE ß	z	p
Intercept	-5.1908	0.6004	-8.65	< 0.001
Word type (English list)	1.9809	0.4707	4.21	< 0.001
Word type x English-French list	-2.0924	0.6670	-3.14	< 0.01
Word type x English-Dutch list	-3.6393	0.6263	-5.81	< 0.001

Also in the analysis of the errors Word Type and List Composition interaction was significant and was included in the model. Table 2.11 shows the effects of the predictors and interactions. The effect of Word Type was significant in the pure English context (ß = 1.98, z = 4.21, p < 0.001) indicating less errors for cognates than for English controls (1.6% and 8.7%, respectively).

English-French cognates

All error responses on the critical items (n=19, i.e., 2.37 %) and RTs falling outside the range between 300 ms and 1,500 ms (n=12, i.e., 1.5%) were excluded. This left us with 769 observations for these cognates and their controls (i.e., 96.12% of the 800 data points).

The final model for the RT analysis included Word Type, List Composition, and Previous RT. The interaction between Word Type and List Composition improved the model. Table 2.12 shows the effects of the predictors and interactions. The effect for the reference level (English) was not significant (ß = -0.03, t = -0.88, p > 0.1). There was no difference in the RTs on the English-French controls and pure English controls (634 and 630, respectively).

Table 2.12.
Estimates for the fixed-effect predictors in the RT analysis of the English-French cognates in pure English context (Experiment 4) and interaction with the English-French and English-Dutch context.

	Estimate ß	SE ß	df	T	p
Intercept	-1.8390	0.05800	148.70	-31.70	< 0.001
Word Type (English list)	-0.0292	0.03323	1348.0	-0.88	> 0.1
Word Type x English-Dutch list	-0.0216	0.04422	1351.0	-0.49	> 0. 1
Word Type x English-French list	-0.1576	0.04421	1355.0	-3.57	< 0.001
Previous RT	0.0002	0.00004	1411.0	5.31	< 0.001

In the error analysis, the model with interaction between the Word Type and List Composition was also significantly better. Table 2.13 shows the effects of the predictors and interactions. The Word Type was not significant for the current reference level (ß = 0.10, t = 0.22, p > 0.1).

Table 2.13.
Estimates for the fixed-effect predictors in the error analysis of the English-French cognates (Experiment 4)

	Estimate ß	SE ß	z	p
Intercept	-4.9786	0.6080	-8.19	< 0.001
Word type (English list)	0.1000	0.4474	0.22	> 0.1
Word type x English-Dutch list	-1.5563	0.9215	-1.69	> 0.05
Word type x English-French list	-2.1699	0.5941	-3.65	< 0.001

2.6 Experiment 5: GNGT with L1-L2 and L2-L3 cognates in English-French context

2.6.1 Method

Participants
Thirty students from the same population as in the previous experiments were tested (25 women, 5 men; range 20-28 years of age, mean age 22.4 years). Twenty-six of them were right-handed and 4 left-handed. All participants had normal or corrected-to-normal vision. They had also the same experience with English and French as participants of the previous experiments. Their mean experience with English was 10.5 years.

Materials and Design
The same stimuli from the Experiment 2 except non-words were used. Excluding non-words created more conflict between the cognate readings of the presented languages.

Procedure
The same procedure from Experiment 4 was used.

2.6.2 Results and discussion

Dutch-English cognates
As in the previous experiments, the Dutch-English cognate-control pair elf-oar was excluded because of the high error rates on the control word (n=60, i.e., 5%). Two participants were excluded from the analyses because of high error rate (>15%); they either didn't understand the task correctly or pressed wrong buttons (n=80, i.e., 6.67%). All error responses were removed from the RT data (n = 14, i.e., 11.7%), as were all RTs that fell outside the range between 300 and 1,500 ms (n = 19, i.e., 15.8%) This left us with 1031 observations (i.e., 96.9 %, given the 1,064 data points).

Table 2.14.
Mean reaMean Reaction Time (RT) and Error Percentage (ERR) for the two cognate types and their English controls (Experiment 5)

	COGNATE TYPE			
	Dutch-English		English-French	
Word Type	RT	ERR	RT	ERR
Cognates	619	1.9	741	8.9
Controls	617	1.7	618	1.6
Effect	2	0.2	123***	7.3*

Note: ** p ≤ .01 *** p≤.001

Mean response times (RT) and errors (ERR) for Dutch-English cognates are presented in Table 2.14. As mentioned in the Experiment 4, the interaction of the Word Type and List Context was significant for RTs and ERRs and was included in the model. The cognate effect, however, at current reference level (English-French context) was not significant (ß = 0.02, t = 0.53, p > 0.1, data not shown). The interaction between the English-French and English was also not significantly different (ß = -0.06, t = -1.18, p > 0.1). It indicates that the recognition of the Dutch-English cognates is similar in the pure English context and English-French context.

The effect of the Word Type was not significant by itself (ß = -0.11, t = -0.24, p > 0.1, data not shown), but the effect of Word Composition was significant (ß = -2.09, t = -3.14, p < 0.01). It indicates, that participants made significantly more errors in the pure English context.

English-French cognates
The same two participants were excluded because of too many errors (n=80, i.e., 6.67%). All error responses on the critical items (n=47, i.e., 4 %) and RTs falling outside the range between 300 ms and 1500 ms (n=32, i.e., 2.7%) were excluded. This left us with 1,041 observations for these cognates and their controls (i.e., 92.9 % of the 1,120 data points).

The model for the RT and ERR analysis contained the interaction between the Word Type and List Context. The cognate inhibition effect (123 ms) was highly significant in the English-French list (ß = -0.19, t = -6.42, p < 0.001, data not shown). The interaction with the English context was also significant (ß = -0.15, t = -3.57, p < 0.001). The presence of French words in the stimulus list increased the conflict between the cognate readings and resulted in the significant inhibition that was significantly stronger than the situation (no effect) observed in the pure English context.

The error analysis showed a similar situation. The cognate inhibition effect was significant for the English-French reference level (ß = -2.07, t = -5.30, p < 0.001, data not shown) indicating more errors for cognates than for the control words (8.9% and 1.6%, respectively). The interaction with pure English context shows significantly stronger cognate effect in the English-French context (ß = -2.17, t = -3.65, p < 0.001)

2.7 Experiment 6: GNGT with L1-L2 and L2-L3 cognates in English-Dutch context

2.7.1 Method

Participants
Thirty students from the same population as in the previous experiments were tested (27 women, 3 men; range 20-24 years of age, mean age 21.6 years). Twenty-six of them were right-handed and 4 left-handed. All participants had normal or corrected-to-normal vision. They had the same experience with English and French as the participants in previous experiments. Their mean experience with English was 10.0 years.

Materials and Design
The same stimuli from the Experiment 3 were used except for the English-like and Dutch-like non-words.

Procedure
The same procedure from Experiment 4 was used.

2.7.2 Results and discussion

Dutch-English cognates

As in the previous experiments, the Dutch-English cognate-control pair elf-oar was excluded because of the high error rates on the control word (n=62, i.e., 5%). Four participants were excluded from the analyses because of high error rate (>15%); they either didn't understand the task correctly or pressed wrong buttons (n=152, i.e., 12.26%). All error responses were removed from the RT data (n = 35, i.e., 2.8%), as were all RTs that fell outside the range between 300 and 1,500 ms (n = 25, i.e., 2%) This left us with 966 observations (i.e., 94.15 %, given the 1,026 data points).

Table 2.15 presents the mean RTs and error percentage for the two cognate types. As mentioned in the previous experiments the interaction between Word Type and List Composition was significant and included in the final model for RTs and errors. The cognate effect showed significant inhibition for English-Dutch context (ß = -0.28, t = -8.08, p < 0.001, data not shown). The significant interaction between English-Dutch list and pure English list (Table 2.10, ß = -0.35, t = -6.83, p < 0.001) and English-Dutch list and English-French list (ß = 0.29, t = 6.18, p < 0.001, data not shown) indicates that the cognate effect was strongest in the list with Dutch words. In other words, including the language of the non-target reading of the cognate increases the conflict between the readings and results in the significant stronger effect (here inhibition) than in other language lists.

Table 2.15.
Mean Reaction Time (RT) and Error Percentage (ERR) for the two cognate types and their English controls (Experiment 6)

	COGNATE TYPE			
	Dutch-English		English-French	
Word Type	RT	ERR	RT	ERR
Cognates	730	7.1	627	1.5
Controls	618	1.6	593	0.4
Effect	112***	5.5*	34	1.1

Note: ** p ≤ .01 *** p≤.001

Similar situation was observed in the error rates analysis. The cognate effect was significant by itself for the English-Dutch context (ß = -1.66, z = -4.04, p < 0.001, data not shown), i.e. significantly more errors for cognates (7.1%) than for controls (1.6%). The interaction with pure English context (ß = -3.64, z = -5.81, p < 0.001) and English-French context (ß = 1.55, z = 2.47, p < 0.05) indicates significantly stronger cognate effects in English-Dutch context.

English-French cognates

The same four participants were excluded because of too many errors (n=160, i.e., 12.9%). All error responses on the critical items (n=9, i.e., 0.7 %) and RTs falling outside the range between 300 ms and 1500 ms (n=13, i.e., 1.05%) were excluded. This left us with 1058 observations for these cognates and their controls (i.e., 98% of the 1080 data points).

As mentioned in the previous experiments, the interaction between Word Type and List Composition was significant and included in the model for the analysis of RTs and error rates. The effect of Word Type was not significant by itself for English-Dutch list as reference level (ß = -0.05, t = -1.75, p > 0.05, data not shown). The interaction with English-French context was significant (ß = 0.14, t = 3.30, p < 0.001), but not significant with pure English context (ß = -0.02, t = -0.49, p > 0. 1), i.e. the RTs on English-French cognates are not different in the contexts where non-target reading is not presented and are not significantly different with the RTs on controls (627 and 593, respectively).

For the error analysis, Word Type was not significant by itself for English-Dutch context as reference level (ß = -1.45, z = -1.81, p < 0.05, data not shown). There was no significant interaction with pure English context (ß = -1.55, z = -1.69, p > 0.05), and no significant interaction with English-French context (ß = -0.61, z = -0.69, p > 0. 1). It indicates, that participants similarly to the English-French context tend to make more errors on English-French cognates than on controls in English-Dutch context (1.5% and 0.4%, respectively), however this difference is not significant similarly to the not significant differences in the pure English list.

2.8 General Discussion

The goal of this study was to investigate how identical cognates are represented and processed by trilinguals. In particular, we examined whether in L2 processing such cognates are affected by both the more dominant native language (L1), and the less dominant foreign language (L3). To be able to manipulate Cognate Type within participants, we asked unbalanced Dutch-English-French participants to process Dutch-English (L1-L2) and English-French (L2-L3) identical cognates. We further studied to what extent task demands and stimulus list composition modulate the observed effects of the non-target language (Dutch or French).

The design underlying the experiments involved a full orthogonal crossing of three variables: cognate type, stimulus list composition, and experimental task. In three generalized lexical decision tasks and three English (L2) go/no-go tasks, we varied stimulus list context for one set of Dutch-English (L1-L2) and one set of English-French (L2-L3) identical cognates. Objective word frequency based on SUBTLEX databases (Keuleers et al., 2010; New et al., 2007; van Heuven et al., 2014) was similar across the readings of the cognates in all three languages. However, because participants were unbalanced trilinguals, subjective word frequency was actually different depending on the participant's proficiency in each language (resulting in the following language order, from strong to weak: L1 – L2 – L3). The cognates were presented in three contexts: a pure L2 context, a context including both L1 and L2 words (i.e., the languages of the L1-L2 cognate readings), and a context including both L2 and L3 words (i.e., the languages of the L2-L3 cognate readings).

Across the set of experiments, cognate effects varied from facilitation to null effects and even to inhibition effects, depending on the type of cognate, stimulus list composition, and task. In the generalized lexical decision task, we found no significant effects for English-French cognates, but a significant facilitation effect for Dutch-English cognates across the three stimulus lists. The latter effect did not differ between the pure English and English-French contexts, but was significantly larger in the Dutch-English context, where both languages represented by the cognates were present. In the go/no-go task, a cognate inhibition effect appeared, but only when other words belonging to the language of the cognates' non-target reading were also part of the experiment. Hence, in an English-French context, inhibition was found for English-French cognates, but not for Dutch-English cognates. Conversely, in a Dutch-

English context, significant inhibition arose for Dutch-English cognates, but not for English-French cognates. Finally, in a pure English context, neither English-French, nor Dutch-English cognates showed any significant effects.

To account for these result patterns, we require only two assumptions, each concerned with a different level of processing. The first assumption is concerned with lexical representations and their activation. It holds that identical cognates are characterized by two morphological representations that compete for selection. The second assumption relates to task and decision processes. It holds that each task and list composition poses its own unique demands on processing and has its own decision mechanism. Figures 2.1 through 2.4 clarify our account by visualizing the processes on the representation and decision levels for both cognate types (compared to monolingual words) in the two experimental tasks. Note that, for simplification, we refer in these figures to the lexical collective of orthographic, semantic, and phonological representations for an item with one term: "Dutch", "English", or "French".

2.8.1 Activation level: multiple representations

The first assumption, that multiple representations exist for identical cognates, is supported by the following observations. In the generalized LDT, participants reacted faster to Dutch-English cognates than to L2 (English) or L1 (Dutch) control words. At first sight, we might consider this as support for the cumulative account and one shared cognate representation. However, participants responded equally fast to English-French cognates and their English controls, and significantly slower to French controls. It is therefore more likely that participants responded predominantly to the dominant reading of the cognates, because in generalized LDT participants could respond to any of the activated reading. In the case of the English-French cognates, responses were made to the English reading (Figure 2.2), as testified by equally fast responses to these cognates and purely English matched controls. Hence, no facilitation was obtained for this cognate type (but no inhibition either). In the case of Dutch-English cognates, the participants apparently reacted to the fastest activated reading, that could be either Dutch (because this is their native language) or English (because their proficiency is high). This resulted in a facilitation effect relative to their matched English or Dutch controls (Figure 2.1). In other words, depending on the familiarity of the cognate in one or another language, participants could always respond to the fastest available reading.

These findings are in line with the existence of a separate representation for identical cognates in each language, the accessibility of which is affected by familiarity with that language (roughly indexed by language dominance). Because L1 is the most dominant language, and L2 is more dominant than L3, Dutch-English (L1-L2) cognates undergo facilitation whereas English-French (L2-L3) cognates do not. Apparently, lexical access in the weak L3 (French) was so much slower than in the L2 (English) that it was not able to affect the response anymore. In sum, the trilingual lexicon is accessed in a language-nonselective way, such that the three different morpheme representations of the cognate triplet are co-activated to an extent that depends on their subjective frequency for participants.

This proposal is in line with the conclusion reached by Peeters et al. (2013). For English-French cognates in an English lexical decision task, these authors failed to find an effect of cumulative frequency (i.e., summing the frequencies of the cognate in the two languages). A cumulative frequency effect would be predicted if cognates had only a single representation that is shared across languages. Instead, the authors found that the cognates' frequency

Figure 2.1. *Dutch-English cognates in generalized LDT*

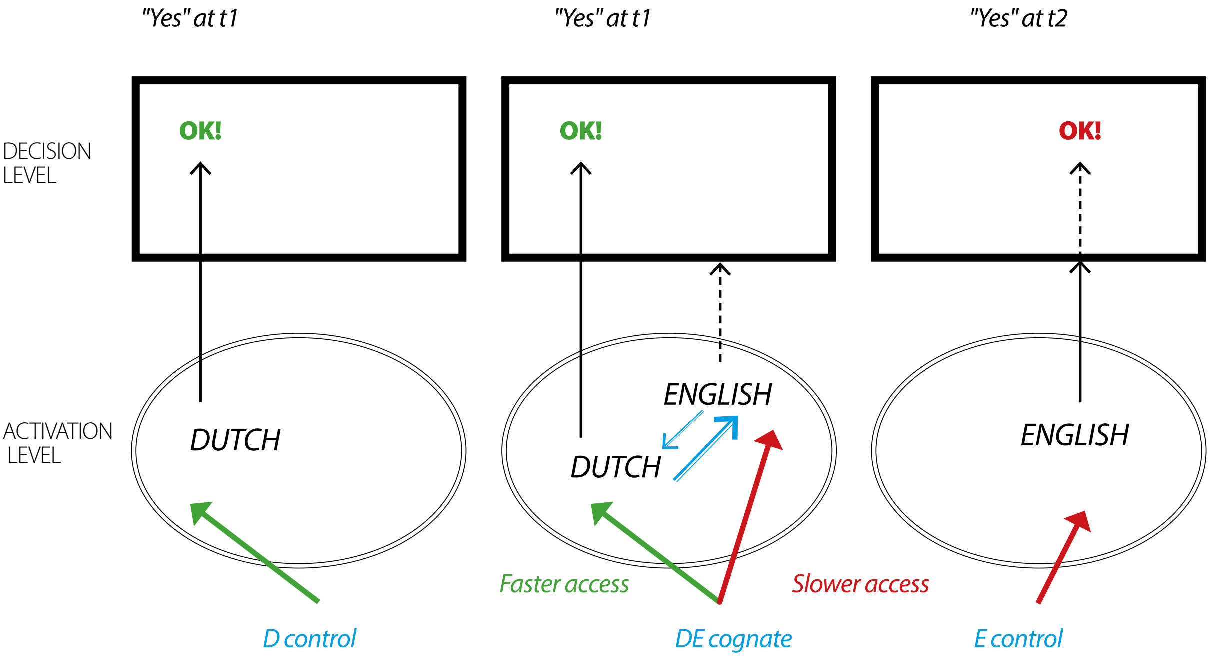


Figure 2.1.
Processing of the Dutch-English (DE) cognates versus Dutch (D) and English (E) control words in a Generalized Lexical Decision task. The middle panel shows the two morphemic representations of the cognate and their interaction, the more familiar reading having a stronger effect on the less familiar reading. The readings are successively accessed as a function of their familiarity. Note that response at t1 < response at t2.

Figure 2.2. *French-English cognates in generalized LDT*

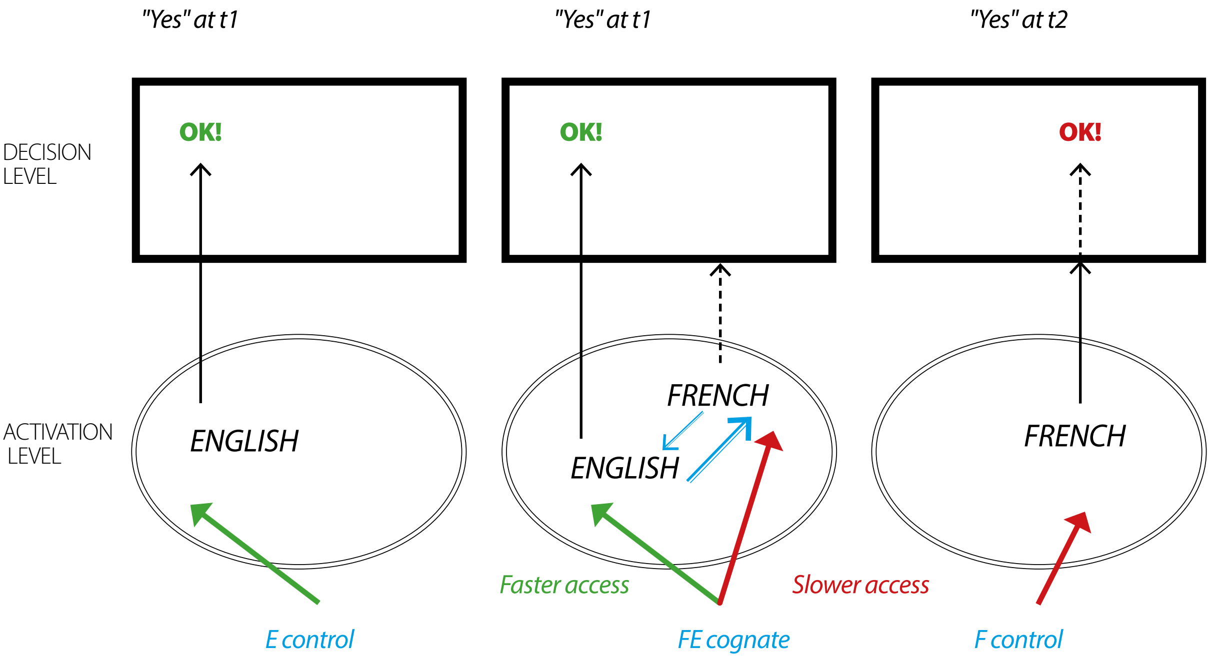


Figure 2.2.
Processing of the English-French (FE) cognates versus Dutch (D) and English (E) control words in Generalized Lexical decision task. The middle panel shows the two morphemic representations of the cognate and their interaction, the more familiar reading having a stronger effect on the less familiar reading. The readings are successively accessed as a function of their familiarity. Note that response at t1 < response at t2.

in the target language in the experiment (the participants' L2) more strongly determined response times than the cognates' native language frequency. For instance, cognates with a high frequency in the target language English and a low frequency in native French were recognized faster than cognates with a low frequency in English and a high frequency in French. The N400 data in their ERP measures confirmed the behavioral effects, i.e., they also followed the target language rather than the native language. Based on these findings, the authors argued that cognates have a separate morphemic representation in each language.

Our findings are compatible with two further studies. First, they support Lemhöfer and Dijkstra (2004), who concluded that Dutch (L1) cognate representations are activated earlier than English ones (L2) and therefore can affect their English readings in an English task. Second, the importance of relative language dominance, i.e., the participant's proficiency level in each language, was emphasized by Van Hell and Dijkstra (2002) for Dutch-English-French trilinguals. When participants were proficient in L3, and only then, the authors obtained a significant effect of L3 on L1 during cognate processing in a lexical decision task. This outcome is in line with the assumptions that cognates have different (morpheme) representations in each language and that the degree of activation of these representations depends on a participant's level of language proficiency. The importance of L3 proficiency suggests that the cognate's representation in a later acquired language can become equally accessible as that in an earlier acquired language if the proficiency in that language is sufficiently high. One might ask why interlingual words with an identical spelling and meaning should have separate morphemic representations. Recent studies (Baayen, 2010; Peeters et al., 2013) assume that two representations arise for orthographically identical cognates, because late bilinguals acquire their languages in different situations and contexts (mostly L1 at home and L2 at school). Moreover, the presence of different morphological properties of a cognate in different languages supports the notion of different morphological representations in different languages, e.g., different plural forms, gender, or syntactic categories across the languages. For instance, the word troll has different determiners and plurals in English, Dutch, and German: the troll, two trolls; de trol, twee trollen; die Trolle, zwei Trolle.

2.8.2 Decision level: flexible response criteria

The second assumption that is needed to account for our result patterns in the lexical decision and go/no-go tasks pertains to the decision level. According to, for instance, the BIA/BIA+ models (Dijkstra & van Heuven, 1998, 2002) and the Inhibitory Control model (Green, 1998), different tasks and list compositions put different demands on the participant, implying that the activation state of the word recognition system can be used (read out) in a different way depending on task and context. For instance, whereas for lexical decision most studies have reported cognate facilitation effects relative to control words (Lemhoefer et al., 2004; Van Hell & Dijkstra, 2002), inhibition effects have been reported for the same items in the language decision task. In this task, cognates confront participants with an active reading in each response language. In the case of identical cognates, this obviously leads to problems in deciding which language they should respond to and this causes inhibition (e.g. Brenders et al., 2011; Dijkstra et al., 2010; Font & Lavaur, 2004; Peeters et al., 2013). An observation that indicates that the direction of the cognate effect (facilitation or inhibition) is task-dependent. Facilitation effects appear to arise in tasks where both readings of the cognate support the same response, while inhibition effects arise in tasks where the two readings are linked up to different responses.

The same line of reasoning can be applied to the go/no-go task in our study. Suppose one processes an identical Dutch-English cognate in the English go/no-go task in a Dutch-

Figure 2.3. *Dutch-English cognates in English-Dutch context in go/no-go task*

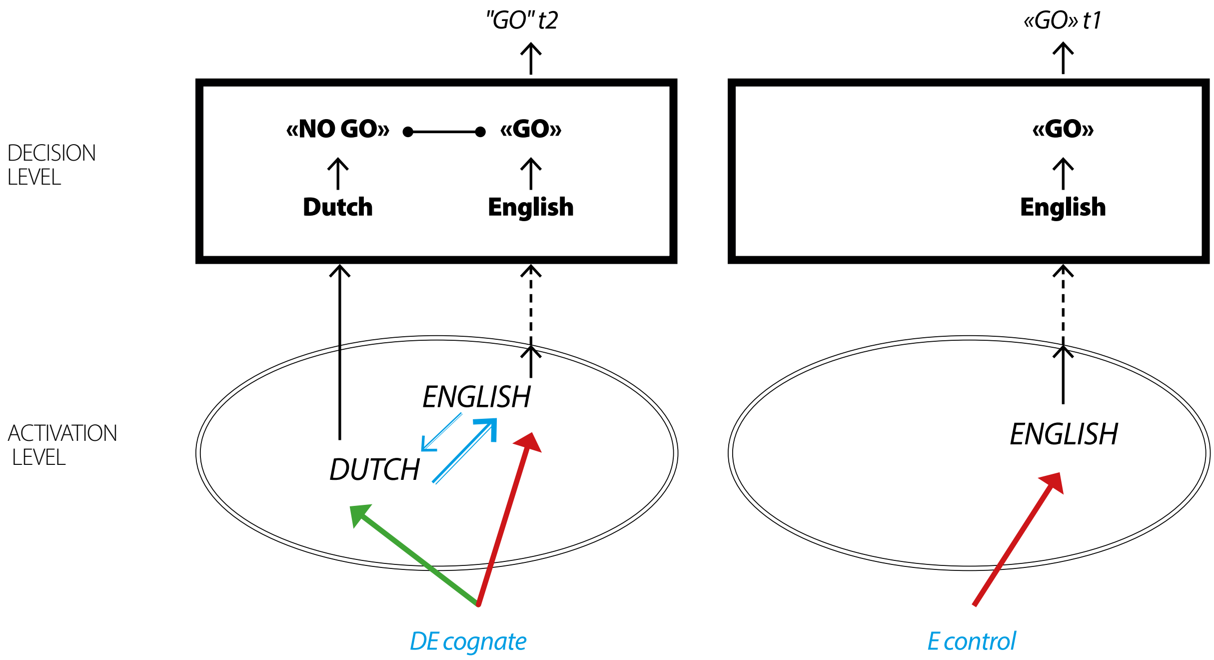


Figure 2.3.
Processing of the Dutch-English (DE) cognates versus English (E) control words in a Go/no-go task. The task allows both readings to be successively accessed. An English-Dutch stimulus list causes a response binding between Dutch and the "no-go"-response, which causes a response conflict for cognates at the decision level (t1 < t2). Such a conflict does not occur in a pure English and English-French context, such that no inhibition effect emerges.

Figure 2.4. *French-English cognates in English-French context in go/no-go task*

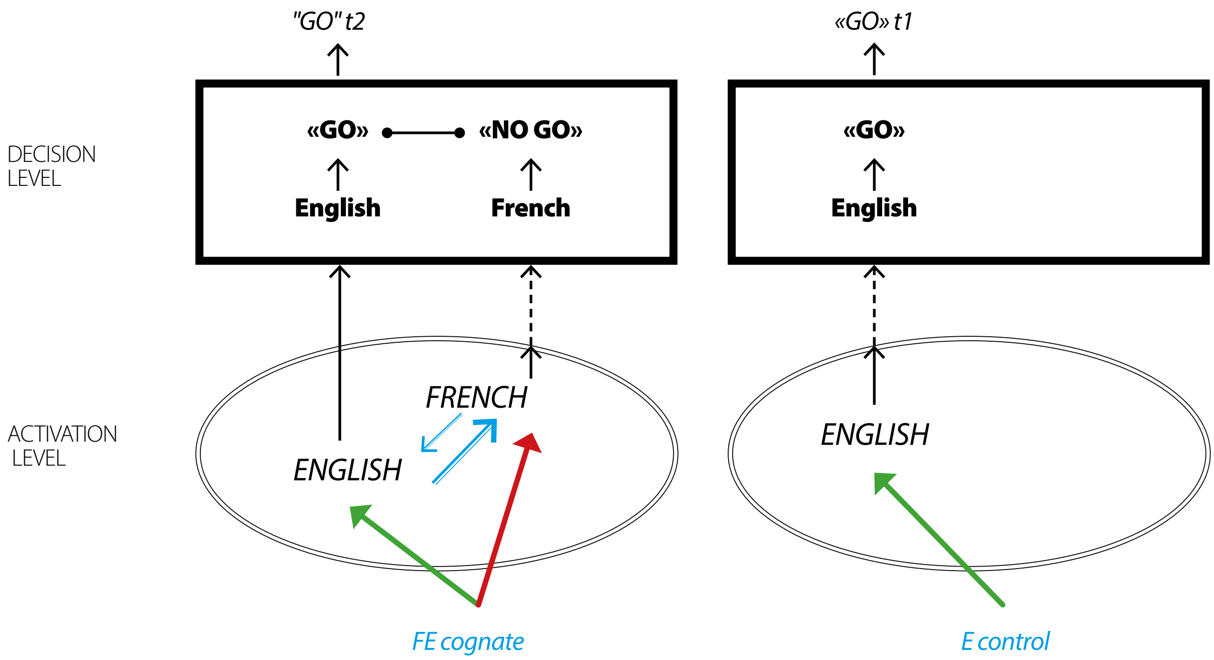


Figure 2.4.
Processing of the English-French (FE) cognates versus English (E) control words in Go/no-go task. The task allows both readings to be successively accessed. An English-French stimulus list causes a response binding between Dutch and the "no-go"-response, which causes a response conflict for cognates at the decision level (t1 < t2). Such a conflict does not occur in a pure English and English-French context, such that no inhibition effect emerges.

English context, i.e., 'go'-responses are given to English words and 'no-go'-responses are given to Dutch words. If identical cognates have multiple morpheme representations, the L1 Dutch reading is most active and linked to the 'no-go'-response, while the L2 English reading is activated somewhat later, but still before response execution, and is linked to the 'go'-response (see Figure 2.3). The ensuing response conflict creates doubt, causing a response delay for correct decisions and more errors. Note that errors will emerge if the English reading becomes relatively active only after the response deadline has been passed.

This account, combined with the above conclusions concerning lexical activation, predicts that in a go/no-go task English-French cognates will not elicit an inhibition effect in an English-French stimulus context. Recall that the null effect for such cognates in the generalized lexical-decision task, even in a stimulus list with 50% monolingual French words, made us conclude that the morphemic representation in L2 was accessed faster than the morphemic representation in L3. A similar activation flow will occur in a L2 go/no-go task as task demands are not expected to affect the activation level during word recognition. As a result, no decision conflict is anticipated, because it should be possible to link the 'go'-response to the L2 reading. Contrary to this prediction, we observed a strong inhibition effect. Apparently, the French reading was also accessed in the English-French stimulus context and linked to the 'no-go'-response, causing a response conflict and an inhibition effect (Figure 2.4).

We argue that our findings point to an essential difference between the two tasks in terms of decision-making. In a generalized lexical decision task, the recognition of a word can immediately give rise to a 'yes'-response. In contrast, in a go/no-go task, the recognized word must still be classified as a member of a particular language before a correct response can be made. This extra task component must require time, which is supported by our observation of much higher mean response times in go/no-go experiments than in generalized lexical decision experiments. Longer processing at the decision level makes it possible that the more slowly accessible French cognate reading still becomes available before response selection and execution. Hence, in the language classification phase of the task, the English reading is linked to the 'go'-response, whereas the French reading becomes active and is linked to the 'no-go'-response, yielding an inhibition effect.

However, such a response conflict did not arise for Dutch-English cognates in the English-French context of the go/no-go task. Note that Dutch was not relevant in this task. This suggests that only the two languages actually present in the experiment are linked to the 'go'- and 'no-go'-responses. Hence, in the case of an English-Dutch cognate in an English-French context, the quickly available Dutch reading is not linked to the 'no-go'-response. No response conflict occurs and, hence, no inhibition is found. In contrast, the later available English reading is linked to the 'yes'-response, at about the same time as its matched monolingual control, and therefore gives rise to a null effect in the data. The same account explains why no inhibition occurs for Dutch-English cognates in a purely English context in a go/no-go task.

The null effects for English-French cognates in the English-Dutch and pure English contexts can be explained in the same vein. Here the irrelevant French cognate reading may not even become active (cf. null effects in generalized lexical decision). Indeed, as the English reading is more quickly available than the French one, a 'go'-response may be made even before the French reading becomes activated to any significant extent. Alternatively, the extra language classification phase in the go/no-go task may provide sufficient time for the French reading to become available (cf. above). However, since the 'no-go'-response is linked to Dutch in an English-Dutch context and to nonwords in a pure English context, an activated French cognate reading will not be linked to the 'no-go'-response. Hence, it will cause no response conflict and no inhibition effect. It follows that there should be no difference in response speed between English-French cognates and their matched English words in these two stimulus contexts. This is indeed what the data show.

In short, cognate inhibition in a go/no-go task was only found when the list context contains pure language words that belong to the same language ('non-language' in the case of nonwords) as the cognates' non-target reading. In our experiments, the list contained 50% of such words from the cognates' non-target language. Further research is needed to find out whether a smaller proportion is sufficient to bring about an inhibition effect. Our current results indicate that the non-target language of identical cognates is not assigned to the 'no-go'-response when it does not also occur in non-cognates in the list; hence, it will not induce any inhibition effect.

Whereas some aspects of the decision process (e.g., response conflict) can affect the direction of the effect, other components of this process can affect its size. Our results suggest that the composition of the stimulus context may affect participants' decision criteria. Participants seem prepared to make faster 'yes'-responses in case the experimental context makes them more confident about their decision. This becomes clear in the effect of stimulus composition on the magnitude of facilitation in the generalized lexical decision task. Although we observed significant facilitation for Dutch/English cognates in pure English and English-French contexts, this facilitation was much stronger in a mixed English-Dutch context. This suggests that the presence of many words (50%) in the stimulus list that also belong to the cognates' non-target language, makes participants feel confident that the quickly available Dutch reading in identical cognates supports a 'yes'-response. In contrast, participants in the experiments without monolingual Dutch words seem to be more cautious when a Dutch reading 'pops up'. We conclude that list composition affects the decision criterion for making a 'yes'-response to the Dutch reading of an identical cognate.

Another finding further clarifies the impact of list composition on cognate facilitation. In contrast to exclusively Dutch words, monolingual French words in the stimulus list did not affect the size of the facilitation effect for Dutch-English cognates in the generalized lexical decision task, compared to a pure English context. Note that this could not have been predicted. The presence of any second language might have raised participants' awareness that the words in the list were drawn from several languages and that any language membership was sufficient for making a 'yes'-response. If this were true, facilitation would be expected for Dutch-English cognates in an English-French context. However, the magnitude of the cognate facilitation effect was about the same in the English-French list as in the pure English list. Thus, the presence of a second language in the experiment does not have a general effect, i.e., participants do not 'set' their decision criterion such that they are prepared to respond 'yes' to any language. Rather, the effect is restricted to the second language that is (strongly) represented in the experiment. Apparently, participants in a generalized lexical decision experiment with a mixed list learn to associate a 'yes'-response only to the languages in the experimental word set. Note that this account parallels our explanation of our go/no-go results, more particularly, of the finding that an inhibitory effect is only caused by a language that is explicitly linked to the 'no-go'-response. We argue that these effects are learning effects that occur during the experimental task and emerge at the level of the task's decision criteria. Note that previous studies (Brenders et al., 2011; T. Dijkstra et al., 2010) also observed differently sized cognate effects depending on context.

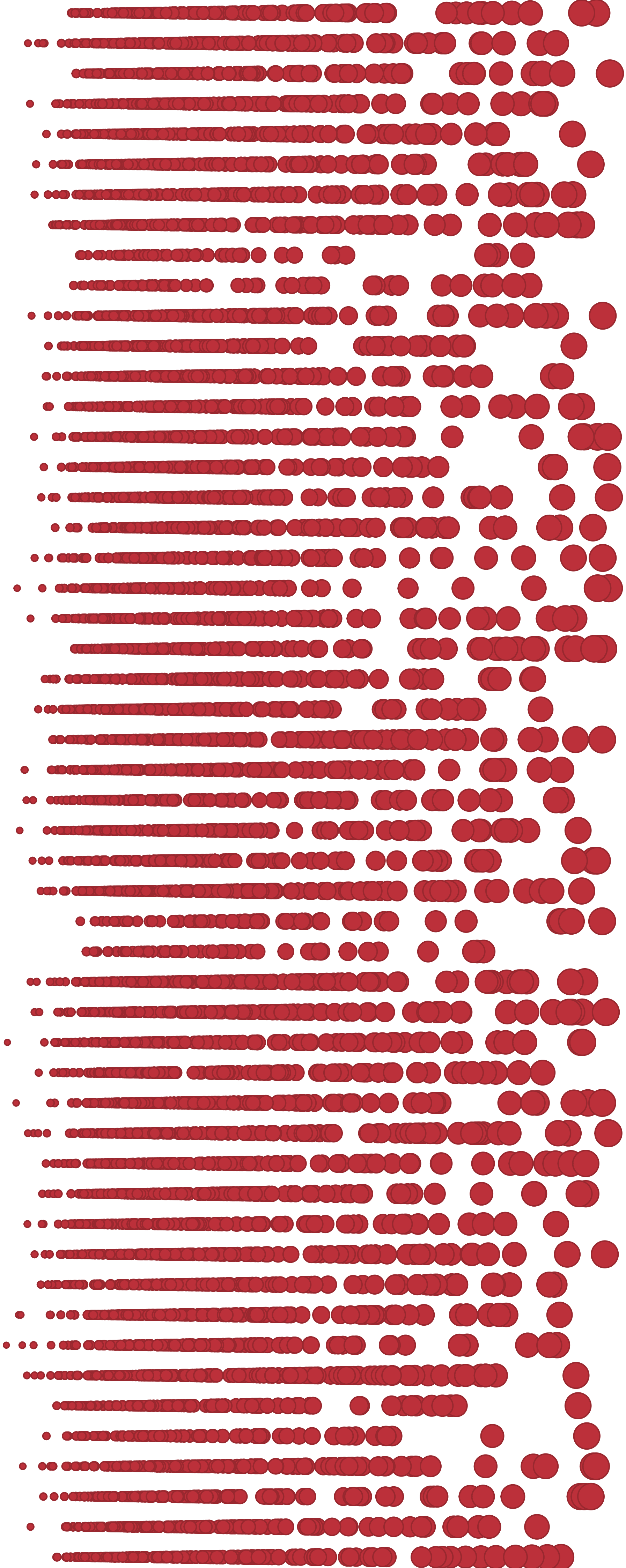
In sum, whereas the initial activation of multiple morpheme representations of a cognate is driven by a non-selective bottom-up process, a task-sensitive and context-sensitive decision component should be considered as well. Representations and processes at both levels jointly determine the direction and the size of the cognate effect. Note that this account is in line with the BIA+ model, in which the activation and decision levels are clearly separated.

2.8.3 Conclusion

The above account makes sense of the diverging result patterns in the generalized lexical decision and go/no-go tasks within a single explanatory framework. The presence or absence of a facilitation effect in generalized lexical decision (L1-L2 cognates vs. L2-L3 cognates) is in line with the notion that identical cognates have multiple morphemic representations, which are accessible as a function of their subjective frequency (which strongly correlates with language dominance). At the same time, the large size of this effect underscores the importance of flexible decision criteria at the postlexical level. Effects in go/no-go also highlight these decision criteria. The presence or absence of an inhibition effect reflects the impact of list composition on decision criteria. In this task, the latter depend on response bindings and the two languages that participants apply during the task.

Our account offers support for the BIA+ model, which focuses on different stages of bilingual language processing: a preconscious lexical activation level, where representations are activated as a function of orthographic similarity and familiarity (subjective frequency), and a decision level, involving conscious, task-induced processes operating on the activated representation(s) and determining the ultimate size and direction of the effects.

To sum up, our study provides a contribution to the ongoing debate on the nature of cognates' representation in the mental lexicon. Our findings support a view in which cognates have different orthographic, semantic, and phonological representations, and, crucial in the context of this paper, different morphemic representations in each language in which they occur. In line with the BIA+ model, the results support the distinction between (a) time-dependent parallel access to these morphemic representations during an early language-nonselective lexical activation process, and (b) postlexical processes operating on active representations at the level of task execution. The latter processes are affected by stimulus list composition and task-specific demands. Together with the events at access level, they co-determine the size of the effects (possibly null effects) and their direction (facilitation or inhibition).



Appendix: Cognates used in this study.

ENGLISH-FRENCH COGNATES			
word	length	English Zipf	French Zipf
aide	4	3.1	5.23
art	3	5.15	4.82
cause	5	4.83	5.33
double	6	5.1	4.46
effort	6	4.76	4.37
empire	6	4.43	4.28
image	5	4.64	4.71
menace	6	3.55	4.32
page	4	4.52	4.40
parent	6	4.21	4.00
phrase	6	4.29	4.19
point	5	5.6	5.27
port	4	4.49	4.47
queue	5	4.08	4.59
regard	6	4.19	4.72
saint	5	4.07	4.09
secret	6	4.93	4.91
six	3	5.47	5.07
trace	5	4.15	4.47
train	5	4.98	5.39
mean:	5.05	4.53	4.65
sd:	1.00	0.61	0.43
range:	3-6	3.10-5.60	4.00-5.60

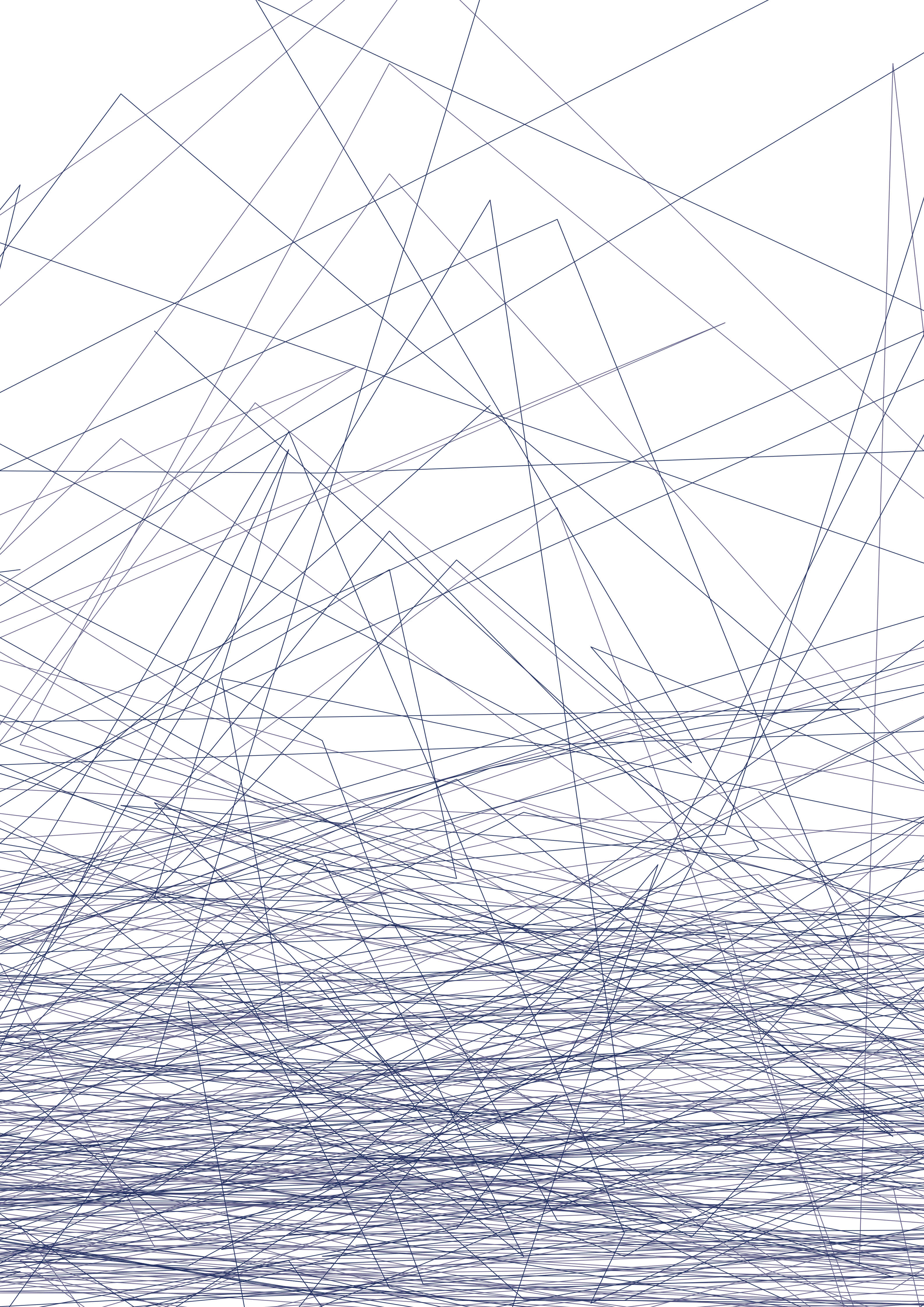
ENGLISH CONTROLS		
word	length	English Zipf
barn	4	4.32
leg	3	4.88
adult	5	4.4
shrimp	6	3.63
future	6	5.3
gossip	6	3.86
board	5	5.1
nephew	6	3.81
seat	4	4.78
reason	6	5.19
search	6	4.81
money	5	5.84
tail	4	4.45
spoon	5	4.29
sleeve	6	3.88
sauce	5	4.79
target	6	4.81
raw	3	4.32
reply	5	3.98
noise	5	4.78
mean:	5.05	4.56
sd:	1.00	0.57
range:	3-6	3.63-5.84

FRENCH CONTROLS		
word	length	French Zipf
vert	4	4.39
ami	3	5.56
vingt	5	4.47
demain	6	5.63
bouche	6	4.94
caisse	6	4.47
odeur	5	4.67
mérite	6	3.84
vide	4	4.22
laisse	6	4.84
proche	6	4.58
lèvre	5	3.60
égal	4	4.44
repos	5	4.63
chacun	6	4.97
effet	5	5.00
vivant	6	4.88
roi	3	5.22
étage	5	4.60
sujet	5	5.03
mean:	5.05	4.70
sd:	1.00	0.50
range:	3-6	3.60-5.63

DUTCH-ENGLISH COGNATES			
word	length	English Zipf	Dutch Zipf
bar	3	4.75	4.73
bed	3	5.12	5.38
elf	3	3.25	4.44
golf	4	4.52	4.24
half	4	5.55	5.02
hand	4	5.44	5.30
land	4	5.18	5.39
lip	3	3.91	3.77
model	5	4.67	4.31
motor	5	4.25	4.63
nest	4	4.4	4.05
net	3	4.53	6.00
norm	4	3.52	3.48
pan	3	4.6	3.97
ring	4	4.92	4.72
warm	4	4.92	4.85
water	5	5.53	5.39
wild	4	4.96	4.41
wind	4	5	4.70
wolf	4	4.17	4.31
mean:	3.85	4.66	4.65
sd:	0.67	0.62	0.63
range:	3-5	3.25-5.55	3.48-6.00

ENGLISH CONTROLS		
word	length	English Zipf
owl	3	4.07
buy	3	5.37
oar	3	3.03
hint	4	3.95
poor	4	5.05
life	4	5.81
food	4	5.45
egg	3	4.8
taste	5	4.95
trunk	5	3.93
worm	4	3.84
oak	3	4.23
howl	4	3.28
cap	3	4.42
crow	4	3.65
nice	4	5.83
growl	5	3.1
fame	4	4.18
horn	4	4.1
clap	4	3.92
mean:	3.85	4.35
sd:	0.67	0.84
range:	3-5	3.03-5.83

DUTCH CONTROLS		
word	length	Dutch Zipf
das	3	4.13
mes	3	4.67
zit	3	6.03
dijk	4	2.91
geit	4	3.91
laan	4	3.09
dief	4	4.48
vis	3	4.70
barst	5	4.07
laars	5	3.62
haak	4	4.10
bek	3	4.71
trui	4	4.07
mol	3	3.88
bril	4	4.39
gauw	4	4.78
beurt	5	4.62
vrij	4	5.37
feit	4	4.63
huis	4	5.91
mean:	3.85	4.40
sd:	0.67	0.79
range:	3-5	2.91-6.03



CHAPTER 3
Effects of language dominance, task demands, and list composition on interlingual homograph processing in trilinguals

3.1 Introduction

One of the famous German beer companies used the slogan ***Die Nacht wird Hell*** ('The night will be bright'), in which the polysemous word ***Hell*** carried overtones of its two meanings, i.e., light and blond, and was also the name of a German pale lager beer. This slogan confused English-German speakers, who linked the word ***Hell*** to its meaning in their native language, and did not understand why an advertisement for beer would say that drinking will cause a bad night. Most multilinguals have experienced similar situations with so-called "false friends" or interlingual homographs (henceforth: IHs), i.e., words that are orthographically identical or similar across languages but have different meanings in these languages. Such experiences suggest that, while reading a homograph in an L2 context, language users may not be able to prevent its L1 reading from becoming active as well. This would mean that bilinguals' lexical access process is language-nonselective, i.e., that the orthographic representation of a written word automatically activates its representation in all languages in which it exists.

3.1.1 The homograph effect: lexical activation and decision processes

As the spellings of homographs like ***Hell/hel***l are identical or near-identical (e.g., Dutch ***groen***, 'green') in two (or more) languages, they have been a favorite research tool for investigating how bilinguals access their mental lexicon. The majority of experiments converge on the view that lexical access to the bilingual (or multilingual) lexicon is indeed a language-nonselective process. Mechanisms of cognitive control appear unable to block this bottom-up activation process.

Different properties of homographs have been manipulated to arrive at this conclusion: their orthographic and phonological overlap between languages, their frequency relationship between languages, and their morphological family size[17] in both languages. Larger orthographic similarity between homographs causes larger facilitation, whereas larger phonological similarity causes larger inhibition (e.g., Dijkstra et al., 1999; Schwartz, Kroll, & Diaz, 2007). These effects can only be explained if the homograph also activates the lexical representation in the non-target language of the experiment. The frequency relationship between the homograph's readings also supports the non-selectivity view (e.g., Dijkstra, Timmermans, et al., 2000; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998). When the homograph's frequency is higher in L1 than in L2, inhibition is found when 'no'-responses must be made to L1 words (e.g., in a L2 lexical decision task), whereas facilitation occurs when 'yes'-responses are expected for both L1 and L2 (e.g., in a generalized lexical-decision task. The same applies to morphological family size (e.g., Dijkstra et al., 2005). When the morphological family is larger in L1, homograph inhibition occurs when responses must be made in L2, whereas facilitation occurs when responses can be based on either language.

[17] The family size of a monomorphemic word (the set of words relevant to the present study and many studies on homographs) is the set of all derivations and compound words in which this word is a constituent morpheme.

The conclusion that bilinguals' lexical access process is language-nonselective pertains to the ***activation*** of representations in the mental lexicon. However, several studies have shown that the homograph effect is also determined by what happens at a post-lexical processing level, more particularly, the so-called decision level (Dijkstra & van Heuven, 2002; Green, 1998). This level is involved whenever bilinguals have to make conscious decisions in an experiment, as in all behavioral experiments[18]. Whereas word-related variables, like the three factors mentioned above, affect the activation level, word-external factors such as the experimental task and the composition of the stimulus list affect decision criteria and, hence, also influence the homograph effect. The intervention of these factors ultimately determines both the direction and the size of the homograph effect (e.g., De Groot et al., 2000; Dijkstra, De Bruijn, Schriefers, & Ten Brinke, 2000; Dijkstra, Timmermans, et al., 2000; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; Lemhöfer & Dijkstra, 2004; Smits, Martensen, Dijkstra, & Sandra, 2006; Smits, Sandra, Martensen, & Dijkstra, 2009).

Dijkstra, Van Jaarsveld, and Ten Brinke (1998) found clear evidence of effects of task demands. In an English lexical decision task, containing English and Dutch monolingual words (i.e., a 'no'-response was required for Dutch words), a homograph inhibition effect was obtained with Dutch-English bilinguals. In contrast, when the same list was presented in a generalized lexical decision task, in which a 'yes'-response had to be made on words from both languages, a facilitation effect (relative to the English controls) was found. Lemhöfer and Dijkstra (2004) reported that generalized lexical decisions on Dutch-English homographs were even as fast as on Dutch controls, indicating that a decision was made on the fastest available reading (i.e., L1), which explains the facilitation effect with respect to English controls. Dijkstra et al. (1998) also demonstrated an effect of list composition. The homograph inhibition effect in the English lexical decision task with English and Dutch monolingual words turned into a null effect when the list contained only monolingual English words. These experiments show that the observable effects (in response times [RTs] and/or errors) in bilingual (visual) word recognition reflect the effects at two levels: (a) the activation level, where the language-nonselective nature of bilinguals' lexical access process activates a lexical representation in two languages and (b) the decision level, where task demands and stimulus list composition determine how this dual activation will surface in the data. It may give rise to a response conflict or not. In the former case, inhibition will be observed. In the latter case, facilitation will be found if the two representations are linked to the 'yes'-response, i.e., when the fastest available reading can trigger the response. A null effect will be obtained if only one representation is linked to a response (as in a pure L2 list). Hence, decision level phenomena determine both the ultimate direction (inhibition vs. facilitation) and the size of the homograph effect (null effect vs. measurable effect).

The BIA+ model (Dijkstra & van Heuven, 2002) makes an explicit distinction between an activation stage (the lexical level) and a decision stage (the post-lexical level), such that it can account for the effect of word-internal factors, which influence the former level, and word-external factors, which influence the latter level. The crucial notion at the decision level is the task schema, i.e., a set of criteria that can flexibly adapt to the nature of the experimental task and the task environment, i.e., stimulus list composition (also see Green, 1998). Hence, the BIA+ model can account for the effects of these two factors in terms of differences in task schemas. The task effect can be explained in terms of how representations and responses

[18] Unlike the situation in experiments targeting word recognition in a natural context, in which participants do not have to make any decision but only have to read the word (e.g., eye monitoring, ERP studies, fMRI studies, etc.).

are linked. When the list contains monolingual words in both L1 and L2, the nature of these connections depends on the task. In a L2 lexical decision task, the 'yes'-response is linked to L2 and the 'no'-response to L1, whereas in a generalized lexical decision task the 'yes'-response is linked to both languages. Even though the L1 reading is generally faster available than the L2 reading, the presence of the homographs makes participants cautious not to respond too fast to L1 words in the former task. This gives rise to a response conflict and, hence, inhibition. In contrast, the latter task yields facilitation, as L1 words also require a 'yes'-response, i.e., do not alarm participants that a L2 reading might still become available. The effect of list composition, i.e., the presence or absence of monolingual L1 words in a L2 lexical decision task, can be accounted for in a similar way. When the list contains monolingual L1 words, the L1 is linked to the 'no'-response. This does not occur when the list only contains monolingual L2 words. In the former situation, the fast activation of a homograph's L1 reading and the participants' suspicion to respond too fast to it, yield a response conflict (hence, inhibition), but in the latter situation, the L1 reading is not linked to either response (hence, resulting in a null effect).

3.1.2 The current study

In this study, our topic of investigation was how the homograph effect is handled at the decision level. To do so, we manipulated the composition of the stimulus list and the experimental task. Our first goal was to extend previous research by introducing one type of list composition that has not been studied yet. In all earlier experiments, inhibition effects have only been studied in a pure list or in a mixed list containing monolingual words from the two languages in which the homographs occur. The current set of experiments will address the question what happens when the homographs are presented in a mixed list but occur in only one of the two languages in the list, i.e., when there is a mismatch between the languages of the homographs and the languages in the list. This effect can shed further light on how list composition affects the setting of decision criteria. Our second goal was to compare the effect of list composition in different tasks. We chose to apply both a variant of the lexical decision task and a go/no-go task. As will be argued below, the latter task is more sensitive to decision-level processes, as it is cognitively more demanding than tasks where a choice between two overt responses must be made (e.g., Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003). Our third goal was to provide a more fine-grained analysis of the homograph effect. Specifically, besides using linear mixed models, we applied so-far infrequently used statistical techniques to investigate the relationship between the size of the homograph effect and response speed. We set out from the hypothesis that effects at the decision level become more pronounced the longer it takes to respond, i.e., when more time is available for decision level processes to affect RTs. Accordingly, we will perform several analyses that make it possible to study the development of the homograph effect over RTs. Each of these three goals will be described more in detail below.

List Composition

The same homographs and controls are presented in three stimulus lists: (1) ***A pure L2*** list: All 'yes'/'go'-trials for filler items are monolingual English words, whereas all 'no'/'no-go'-trials are nonwords derived from English words; (2) ***An L1-L2*** list: All 'yes'/'go'-trials for filler items are monolingual English words, whereas 'no'/'no-go'-trials are monolingual Dutch words; (3) ***An L2-L3*** list: All 'yes'/'go'-trials for filler items are monolingual English words, whereas 'no'/'no-go'-trials are monolingual French words. Whereas a pure L2 list and an L1-L2 list have been used in earlier research, an L2-L3 list has not been studied before. In this list, L1-L2 (Dutch-English) homographs are presented in an L2-L3 context (English-French), where a 'no'/'no-go'-response has to be made to French words. Will the presence of L3 words affect the processing of L1-L2 homographs? Obviously, French words should have no effect on the activation stage of Dutch-English homographs. If they have any effect at all, it must occur at the decision level. Even though monolingual Dutch filler words in an English lexical decision task are also assumed to affect the decision stage only (see above), there can be no doubt about this when the filler words come from a language that is not one of the languages of the homographs.

The homograph effect in the L2-L3 list might manifest itself in three different ways. A first possibility is that the presence of any non-target language (i.e., here: French L3) makes participants treat all non-target languages equally. In that case, the 'no'/'no-go'-response will not only be linked to the French language but to all non-target languages (here: Dutch L1). This view predicts an inhibition effect for the homographs: Their Dutch reading will be linked to the 'no'/'no-go'-response, whereas their English reading is linked to the 'yes'/'go'-response. This should give rise to a response conflict and, hence, inhibition. If this reasoning is correct, the results for the L2-L3 list will be similar to those for the L1-L2 list. The fact that L3 is a much weaker language than L1 does not discredit this hypothesis. The monolingual L3 words will obviously be processed more slowly than L1 and L2 words. However, what matters is that they will cause all non-native languages to be connected to the 'no'/'no-go'-response. Hence, according to this view, the effect on the processing of L1L2-homographs should be the same as in the L1-L2 list.

A second possibility is that French does not affect how participants deal with the homographs' Dutch reading. During the experiment, they learn to respond 'no'/'no-go' to French words but learn no particular response to Dutch words, such that the 'no'/'no-go'-response is linked to the French language only. If this is true, no response conflict is expected and, hence, no inhibition effect either. In that case, the results for homographs in the L2-L3 list should be similar to those in the pure L2 list.

A third possibility is that the effect in an L2-L3 list will lie somewhere in between the null effect for the pure L2 list and the inhibition effect for the L1-L2 list. That would be the case if the occasional confrontation with a language that is not explicitly linked to any response (i.e., when reading an interlingual homograph) creates temporary decision doubt, persisting beyond the moment at which the English reading becomes available but not as long as the time needed to resolve a response conflict in an L1-L2 list. Hence, this doubt would cause a smaller inhibition effect than in a L1-L2 list. Note that the latter possibility could not be explained in terms of response bindings with languages but in terms of lack of confidence at the moment of decision-making.

As we made clear in the above paragraph, the study of L1-L2 homographs in an L2-L3 list requires that the results be compared with the effects in a pure L2 list and a L1-L2 list. By including the latter two, our list factor contains conditions that are expected to yield minimal and maximum inhibition, respectively. These two benchmarks are necessary to position the homograph effect in the L2-L3 list. Note that the minimal condition has often been found to yield a null effect. In a monolingual context, Gerard and Scarborough (1989) found an effect of a homograph's frequency in the target language but not of its frequency in the non-target language (see also Caramazza & Brones, 1979; Soares & Grosjean, 1984). These authors interpreted this result as evidence supporting language-selective access. Most researchers now reject this hypothesis, although a null effect for homographs in a monolingual list has been replicated in several studies (e.g., Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; De Moor, 1998; De Groot, Delmar, & Lupker, 2000, see also Chapter 1 for a possible explanation of the unstable homograph effect in pure list contexts). Language non-selective access models, such as the BIA+ model, can easily account for this null-effect: Because the quickly available L1 reading is not linked to either response type, participants are not confused when the L2 reading becomes available and incur no response delay. Interestingly, De Moor (1998) found evidence supporting the idea that a homograph's L1 reading becomes activated in a pure L2 (English) list. In such a list, she too observed a null effect when comparing homographs and controls. However, she also found that the Dutch reading of the homograph primed its English translation on the next trial (e.g., ***brand-fire***, the Dutch word 'brand' meaning 'fire' in Dutch), which suggests that the L1 reading was activated but did not interfere with the L2 lexical decision on the homograph itself.

Maximal inhibition for IHs was obtained when L1-L2 homographs were presented in a L1-L2 mixed list, where pure L1 and pure L2 words are intermixed, and when L2 was the target language. This stimulus list caused a strong response conflict in the decision process of homographs and, hence, a sizeable inhibition effect. Importantly, this inhibition arises because participants learn to link the 'no'-response to the L1 words during the experiment, not because they act on expectations. In a study by Dijkstra, De Bruijn, Schiefers, and Ten Brinke (2000), participants were explicitly told that Dutch (L1) words would appear as negative trials in the experiment but they only encountered such words in the second half of the experiment (which they did not know in advance). The authors found no homograph effect in the first half, despite participants' expectations that L1 words would be presented, but a clear effect in the second half. Hence, only the actual ***experience*** that L1 words require a 'no' response causes links between the words in a language and a specific response.

Experimental Task

The three experimental lists were presented in two experimental tasks: a lexical decision task (henceforth: LDT)[19] and a go/no-go task. These tasks have already been used in previous studies (e.g., Dijkstra et al., 2000). However, they have never been crossed with stimulus list composition in an orthogonal design.

Even though the go/no-go task allows participants (at least theoretically) to block lexical access to all words that do not belong to the target language, it is unlikely that this will happen in our experiments. In their go/no-go task, Dijkstra et al. (2000) found that the two readings of IHs had been automatically activated, i.e., that language users cannot control their process of lexical access. Although they did not compare the homograph effects across tasks, the inhibition for the homographs with a higher frequency in L1 (Dutch) was even much larger in the go/no-go task than in their L2 (English) lexical decision task (235 ms vs 54 ms, respectively). This larger effect is likely due to a more difficult response process. Indeed, accurate performance in the go/no-go task comprises (a) a component of conflict monitoring (i.e., a response conflict) and (b) a component of inhibitory control, i.e., the deliberate suppression of a motor response when a negative decision has been made (Nieuwenhuis et al., 2003; Donkers

[19] For the sake of convenience, we will use the term 'lexical decision task', even though the task in two experiments was, strictly speaking, not an LDT. In the pure L2 list, 'no' responses had to be made to nonwords, i.e., participants performed a real LDT. In the mixed lists, 'no' responses had to be made to L1 words (in the L1-L2 list) or to L3 words (in the L2-L3 list), i.e., participants performed a language decision task. However, the instructions made no reference to other-language words. As L2 was the target language in all lists, the tasks were quite similar, i.e., the discrimination of L2 words from items that did not occur in L2. The use of nonwords was the only way to define a benchmark for assessing the minimal effect of homograph inhibition.

& van Boxtel, 2004). Hence, we expected longer RTs in the go/no-go task than in our variant of the lexical decision task and an interaction between the factors Experimental Task and List Composition, i.e., more inhibition in the L1-L2 list (and possibly, the L2-L3 list) in the go/no-go task.

Fine-grained analyses of homograph inhibition

The data of all experiments will be analyzed with linear mixed models, using participants and items nested within participants as random effects. However, an important aim of our study was to investigate the homograph effect in much more detail than in previous studies. In these studies, the authors only made comparisons between the means in the homograph and control conditions, using either ANOVAs or (more recently) mixed-effects analyses. Means comparisons, by definition, shed no light on the entire RT distributions underlying the conditions being compared. Our final goal was to find out whether the homograph inhibition effect is due to a shift of the entire RT distribution for the controls, i.e., is constant across the full range of RTs, or is restricted to a specific region in the RT distribution. As the homograph effect is due to a response conflict, it seems reasonable that it will become larger in the region of longer RTs. It may even be absent in the region of the fastest RTs. Indeed, long RTs may indicate that the participant has a hard time inhibiting the irrelevant response, whereas short RTs may reveal that the participant can quickly select the correct response. Fast responses may also indicate that the L2 reading was faster available than the L1 reading, although this was rather unlikely in our experiments (see Materials). To our knowledge, this is the first attempt to learn more about the evolution of the homograph effect across the range of RTs in the experiment. However, the idea to delve deeper into the data to learn more than what can be achieved with mean comparisons is not new.

Recently, several authors have criticized the practice in psycholinguistic journals to report only statistical analyses that are based on mean comparisons (Balota & Yap, 2011; Balota et al., 2008; Lo & Andrews, 2015). For instance, Balota and colleagues demonstrated that this approach causes researchers to overlook phenomena that occur in the right tail of the RT distribution, i.e., the region of long RTs. RT distributions are typically not normally distributed but have a denser right tail than the normal distribution. The left end of the scale stops at a particular point, lets say around 300 ms, because participants in behavioral experiments are often unable to press a button before that time. However, at the other end of the scale there is no such upper RT limit, although many researchers use a time-out (which can very between 2 s in some experiments and even 5 s in others). The problem is not so much the time-out itself but the fact that there is a broad region beyond individual participants' average RT to still make a response. This means that a lot of RTs will be more 'spread out' towards the right end of the RT scale than towards the left end. These RTs will quite likely be the result of difficult items, temporary lapses of attention, fatigue, etc. As all participants can respond far beyond their own average (perhaps better: median) RT, there will be more RTs in the right tail of the distribution than would be the case in a normal distribution. This makes the right-end tail 'fatter' than normal, i.e., the real distribution of RTs is often right-skewed. This is a problem because the relatively many long RTs in such a skewed distribution make the mean shift to the right, making it no reliable estimate of the central tendency of the distribution.

Ex-Gaussian analyses Rather than normalizing the data by means of data transformation (like taking the logarithm or the inverse of the RTs), to meet the requirements of the parametric test, they argue that it is important to analyze the actual RTs. Interestingly, the typical right-skewed RT distribution in a psycholinguistic experiment can be mathematically decomposed into two distributions: a Gaussian (i.e., normal) distribution and an exponential distribution (see Ratcliff, 1978, 1979, who demonstrated the usefulness of splitting up an asymmetric distribution into these two types of distributions). The exponential distribution is responsible for the denser right tail of the RT distribution. Three parameters are crucial in this decomposition: the mean (mu, symbol: μ) and the standard deviation (sigma, symbol: σ) of the underlying normal distribution component, and the mean and standard deviation of the underlying exponential component (both captured by tau, symbol: τ). An analysis in which these two components of the RT distribution are analyzed separately, a so-called ex-Gaussian analysis, makes it possible to find out whether the effect of an independent variable is constant across the RT range. In that case, the means (μ) of the Gaussian distributions for the experimental and control conditions will be significantly different from each other, i.e., the distribution will simply be shifted along the RT axis. Sometimes, the means will differ but also the standard deviations (σ), which means that the size of the effect varies across participants and/or items. When the effect is restricted to the right tail of the distribution, the means and standard deviations of the Gaussian component will not be statistically different, as the effect will be located in the exponential distribution. In that case, the effect will be reflected in tau (τ). Most psycholinguistic variables cause both a shift of the entire RT distribution and a larger effect in the region of the longest RTs, i.e., the most difficult items (Balota et al., 2008).

Balota et al. (2008) demonstrated the added value of ex-Gaussian analyses in a set of semantic priming experiments. They found that the semantic priming effect only shifts the RT distribution for the unprimed condition when both prime and target are clearly visible. However, when targets are degraded, there is both a shift of the RT distribution and an increase in the number of RTs in its right tail. This is the case when the primes are clearly visible but also when they are masked. Whereas the former result suggests the validity of the headstart metaphor for semantic priming, the latter findings indicate an increased reliance on prime information for the more difficult items, i.e., those taking a long time to recognize. Clearly, the use of ex-Gaussian analyses makes it possible to achieve a better insight into the phenomenon under study. As mentioned earlier, we used this rationale to find out whether the Gaussian and/or the exponential component of the RT distribution in experiments with IHs are effected by the design factors, and what this means for a theory of bilingual lexical processing.

Quantile analyses Along the lines of the same logic, Balota (2008, 2011) and other researchers (Rouder, Lu, Speckman, Sun, & Jiang, 2005) proposed a second type of analysis, using quantiles (they refer to them with the term 'vincentiles'). This makes it possible to achieve an even more detailed view on the temporal evolution of the effect under study. Whereas the separation of the RT distribution into a normal and an exponential distribution separates the effect in the region of the longest RTs from the effect on the RTs in the Gaussian distribution component, it does not yield a very detailed picture of the evolution of the effect across the entire RT range. To study this evolution of the effect, Balota et al. (2008) divided the theoretically relevant RTs into a number of bins, the same number for each condition in the experimental design. Per condition they rank-ordered each participants' RTs from the smaller to the larger ones and then divided this ordered sequence into 10 bins, each containing the same number of RTs (note that the experimenter decides on the optimal number of bins). Mean participant RTs were entered into an ANOVA, for each combination of bin number and condition (primed vs. control), to find out whether the size of the experimental effect varied across bins. No interaction was found in experiments with clearly perceptible visual primes and auditory targets. However, the interaction was significant in the experiments with degraded auditory targets and (unmasked or masked) visual primes: the semantic priming effect became larger in the bins containing the longest RTs. As Balota et al. (2008) achieved a better insight into the mechanism of semantic priming by using quantile analyses, we decided to follow their example with the purpose of finding out how the interlingual homograph effect evolves over response time.

Delta plot analyses Another way to analyze the data in the bins is to make use of a delta plot analysis. Such an analysis is used, for instance, to compare two participant groups differing on a theoretically important proficiency measure (e.g., Roelofs & Piai, 2017; Roelofs, Piai, & Rodriguez, 2011). However, it can also be used with a single set of participants. The purpose of the analysis is to find out whether the difference between the sizes of an experimental effect in two successive bins changes over the range of response speed. Hence, whereas a quantile analysis assesses whether the magnitude of the effect changes across bins, a delta plot analysis compares these changes. More particularly, it assesses whether the change in the magnitude of the effect between successive bins is constant (e.g., there is a constant increase or decrease of the effect from one bin to the next) or is discontinuous (e.g., a larger increase between bins with the longest RTs). Delta plots visualize this temporal development. Analysis of these plots consists in the comparison of two slopes (each connecting successive bins) in the graph plotting the size of the experimental effect per bin. The analysis involves a pairwise comparison of the beta values (one for each participant) defining the steepness of the two slopes. For instance, one can ask whether the slope between the effects in bins 1 and 2 differs from the slope between the effects in bins 2 and 3, etcetera. Delta plot analyses allow one to hypothesize why the size of an effect does not change (flat line in delta plot), why the increase/decrease in an effect is constant (straight line in delta plot), or why the effect suddenly increases, decreases, or stabilizes (discontinuous line in delta plot).

In short, in the statistical analyses of our experiments, we will use linear mixed models in which fixed and random factors are used to predict RTs and error rates. These will give us an idea of the average homograph effect in each experiment. However, to achieve a better understanding of how the size of the effect is related to response speed, and, hence, probably depends on the amount of doubt at the decision level, we will also report ex-Gaussian analyses, quantile analyses, and delta plot analyses.

To summarize, the experiments reported below were designed with three goals in mind. First, we studied whether L1-L2 homographs give rise to an inhibition effect in a L2 decision task when they are presented in a stimulus list that contains many monolingual L1 or L3 words. This effect was then compared to that in a pure L2 list condition, where a null effect was expected, and a L1-L2 list, where a robust effect was expected. This enabled us to make statements with respect to the task schema at the decision level that is used in an L2-L3 list. Second, by comparing variant of the language-decision task to the go/no-go task, which is cognitively more demanding and, hence, expected to cause longer decision latencies, we wanted to find out more about the role of (late) decision level processes as determinants of the size of the homograph effect. Third, we wanted to achieve a more detailed picture of the homograph effect by analyzing (with various techniques) whether the effect is constant across the entire RT range (distributional shift) or becomes larger in the region of the longest RTs. If the effect is not constant, we will use delta analyses to determine whether the size of the effect grows at a constant rate, i.e., is a linear effect, or whether there are discontinuities in the evolution of the effect's magnitude. The cognitively demanding go/no-go task, involving both conflict monitoring and response suppression, may create ideal circumstances for observing steeper slopes in the higher RT regions.

3.2 Experiments

Six experiments will be reported with English as the target language in which trilingual participants (Dutch-English-French) are presented with Dutch-English (L1L2) homographs. Different participants were used in each experiment, but the same set of IHs and their controls was used across experiments. The experiments represent the orthogonal combination of two factors: List Composition and Task. The Task factor represents two levels: a language-specific (i.e., English, L2) variant of the lexical decision task and a go/no-go task. The List Composition factor has three levels, depending on the language(s) from which the monolingual words are drawn: L2 only, both L1 and L2, and both L2 and L3. The latter condition has not been studied thus far.

Considering the many commonalities across the experiments, it is more convenient to first describe what was common to all experiments. Below, the design factors behind the experiments, the materials, the procedure, and the statistical techniques that will be used to analyze the data will be described. Next a (brief) description of the specific aspects of each experiment will be given. Finally, the results will be presented.

3.2.1 Method

Design
List Composition
The Dutch-English IHs were presented in three different experimental lists. In each of these lists, they belonged to the item set whose members had to be classified as L2 words. The other members of this L2 set were the controls for the IHs and a small set of fillers (see below). The other item set in the experiment differed as a function of the condition on the factor List Composition. In the pure L2 list, words from another language were not allowed, i.e., L1 and L3 words. Hence, this list consisted of the L2 set and a set of English-like nonwords, making the discrimination task a classical L2 lexical decision task. In the mixed L1-L2 list, the L2 set was combined with a set of monolingual L1 (Dutch) words. Hence, even though the instructions informed participants to decide whether the target on the screen was a L2 item, this task closely resembles a language decision task. However, there was a crucial difference with respect to a typical language-decision task with IHs: the participants were not free to decide whether IHs were L2 or L1 words (e.g., as in Dijkstra et al., 2000). They had to respond "yes" for each item that was an L2 word. In the mixed L2-L3 list, the L2 set was combined with a set of monolingual L3 (French) words. Here, too, participants were told to make a 'yes'-response on each item that existed in their L2 and a 'no'-response on all other items. Accordingly, in all three experiments, participants were obliged to treat L2 as the positive response category.

Experimental Task
As discussed earlier, two experimental tasks were used. The first task can best be characterized as a variant on the language decision task (henceforth: LDT), where the L2 is explicitly presented as the target language. Given this focus on L2 in all three lists, there are more similarities between the task in Experiment 1, where negative responses are given to

nonwords, and the tasks in Experiments 2 and 3, where negative responses are given to L1 words and L3 words, respectively. The second task is the go/no-go task.

These two experimental tasks will be systematically compared across the three types of list composition described above. Previous studies have studied the effect of list composition (e.g., Dijkstra et al., 1998; De Groot et al., 2000) or experimental task (e.g., Dijkstra et al., 1998, 1999, 2000, 2008), but, to the best of our knowledge, a systematic comparison between tasks with respect to the effect of list composition has never been made.

3.2.2 Materials

The critical items for LDT were 20 Dutch-English (L1-L2) orthographically identical homographs. Originally, we selected words from the CELEX database (Baayen et al., 1995). Later we compared the frequencies with the SUBTLEX databases, once the latter became accessible (Keuleers et al., 2010; van Heuven et al., 2014)[20]. Based on the SUBTLEX databases, the frequency of the Dutch reading for Dutch-English homographs was still higher than of the English reading (see below).

Ten randomly selected students from the same population as the experimental participants were asked to rate (1 = "none"; 5 = "very strong") a set of Dutch-English homographs: (a) Do they have different meanings in the two languages?, (b) Do they have a similar or very close pronunciation?, and (c) Are they familiar to the average university student? Based on these ratings we selected 20 Dutch-English homographs with clearly distinct meanings, a different pronunciation in both languages[21], and a sufficient degree of familiarity to most university students. These IHs can be found in Appendix to this chapter.

All selected homographs had a higher frequency in Dutch (mean: 4.77, SD = 0.53, range: 3.79 – 5.76) than in English (mean: 4.03, SD = 0.54, range: 3.33 – 5.60). Note that the English frequencies are based on a frequency count that gives estimates of word frequencies for native speakers of English. These values obviously overestimate the subjective L2 frequencies in our sample of Dutch-English language users. Accordingly, it is highly likely that the Dutch language reading of the IHs will reach its activation threshold first. For each homograph, an English monolingual control word was selected with the same number of letters (mean: 3.90 for homographs and controls) and a matched frequency (mean: 4.10, SD = 0.62, range: 3.02 – 5.29). A t-test indicated that homographs and controls were accurately matched on frequency ($t = 0.19$; $p > 0.1$). In addition, we added 10 English filler words from the same frequency range as the other English words in the experiment. Additionally, we generated 50 English-like non-words (derived from different English words) using the Wuggy program (Keuleers & Brysbaert, 2010). All non-words were matched on number of letters with English words and homographs.

Most of the IHs would fall in Dijkstra et al.'s (1999) category of O homographs, i.e., words with an identical spelling but different pronunciations and meanings in L1 and L2 (e.g., ***brand***, Dutch for 'fire'). Based on this study, in which the orthographic, phonological, and semantic overlap between L1 and L2 words was systematically manipulated, O homographs are

[20] Note, we used the formula log10(frequency per million*1000) from van Heuven et al., 2014, to calculate Zipf values for Dutch data. It made it easier to compare different data sets.

[21] Dijkstra et al. (1998) reported evidence that phonological overlap causes an inhibition effect. Van Hell and Dijkstra (2004) could not replicate this finding but found that phonological overlap reduced response times to IHs.

expected to yield facilitation, at least in a pure L2 list (inhibition is expected in a mixed L1-L2 list). The authors indeed found that orthographic and semantic overlap causes facilitation, whereas phonological overlap causes inhibition. In their O homographs the net result of these three components was a strong (53 ms) facilitation effect. However, Lemhöfer and Dijkstra (2004) were not able to replicate essential findings in this earlier study, although they used the same IHs. Hence, based on these two studies, it is difficult to predict the effect for the IHs. However, all IHs in the Dijkstra et al. study were equally frequent in both languages, to maximize the possibility that their homograph effects were not due to frequency differences between the two readings but to their overlap or mismatch at the levels of orthography, phonology, and/or semantics. In contrast, in the present study, the two readings of the IHs were deliberately mismatched on frequency (their L1 reading being more frequent, certainly when taking into account that L2 was the weaker language), such that a response conflict at the decision level was almost inevitable. Based on this frequency relationship our IHs are likely to cause an inhibitory effect, even in a pure L2 list, unless participants do not experience a response conflict in the latter list due to the absence of monolingual L1 words (see above).

All selected homographs were three to five letters long (mean length: 3.90 letters). In addition, 10 English filler words were added with similar length and frequency properties as the IHs and their controls (mean length: 4.62 letters; frequency = 41 per million, logarithmic frequency: 1.23). Using the Wuggy program (Keuleers & Brysbaert, 2010), 50 English-like non-words (derived from English words) were generated. Each non-word was matched on an item-by-item basis with an English word on its number of letters.

3.2.3 Procedure

Instructions
Written instructions were provided in English. The experimenter welcomed and briefed the participants in English to avoid activation of other languages than the target language in the experiment. The same instructions were used for all six experiments that together made up the experimental design. The instructions informed participants to make 'yes'-responses to L2 words and 'no'-responses to all items that did not belong to L2. Importantly, the instructions made no explicit reference to the nature of the items in the negative response set, i.e., nonwords in Experiment 1 and other-language words in Experiment 2 (L1 words) and Experiment 3 (L3 words).

Stimulus Presentation
The experiment was run on DELL computers of the type "Optiplex 380", connected to a 15" DELL monitor in a dimly illuminated room. The stimuli were presented one by one in the center of the screen. RTs were registered with millisecond accuracy with the DMDX program (Forster & Forster, 2003). Game controllers of the type Logitech 'Wingman precision' were used for responding (with the two front buttons) and for initiating a block of trials (with the 'start' button). Written instructions in English were used. In the L2 lexical decision task, participants were instructed to decide as quickly and accurately as possible whether the letter string on the screen was an English (L2) word or a non-word. They used their dominant hand for 'word'-responses. In the L2 go/no-go task, a response had to be made when the stimulus on the screen was a word in L2. If so, participants had to push a response button ('go'-trials). If not, they did not have to make an overt response ('no go'-trials). The button on the side of the dominant hand was used for 'go'-responses. The instructions did not specify that some words also existed in Dutch, to avoid that participants would expect L1 words as well.

A set of 12 practice trials preceded the experimental list. None of these words and non-words were repeated in the experimental list. Homographs and their controls were evenly spread across the blocks in the experiment. After each block, participants could take a short break. Each trial started with a fixation sign (+), which appeared in the middle of the screen for 500 ms, when it was replaced by a word or non-word. This item stayed in view until a response had been made or until a time out of 2,500 ms had passed. The next trial was initiated 500 ms after the response or time-out.

3.2.4 Data analysis

Several analyses of the data will be reported: (a) (generalized) linear mixed effects analyses (henceforth (G)LMMs), (b) ex-Gaussian analyses, (c) quantile analyses, and (d) delta plot analyses. As mentioned above, analyses (b) to (d) are reported to achieve a better insight into the evolution of the homograph effect across the RT range, which is one of the goals of the current study.

Now that all general aspects of the experiments have been described, the experiment-specific information will be provided. These brief descriptions will be followed by the statistical analyses.

3.2.5 Experiment 1: LDT in pure English context

Participants

Twenty-four Linguistic students of the University of Antwerp participated voluntarily (22 women, 2 men; age range: 19-24, mean age 22.3 years). Twenty-three were right-handed and one left-handed. Each participant had normal or corrected-to-normal vision. They were all unbalanced Dutch-English-French trilinguals: native speakers of Dutch who were proficient in English and had a good knowledge of French. The participants had studied English as a foreign language at secondary school for at least six years and used this language regularly (mean experience: 9.7 years). They had also studied French at secondary school and most of them used it only occasionally in informal situations outside their study (with friends or relatives).

Method

The IHs and their controls were presented in a pure L2 context, i.e., all words (including the IHs) were existing English words and all nonwords were derived from English words. Participants performed a L2 lexical decision task. The complete experimental list consisted of practice block and 4 experimental blocks of 20 trials each.

3.2.6 Experiment 2: LDT in English-Dutch context

Participants

Twenty-four students participated voluntarily (18 women, 6 men; age range: 19-23 years, mean age 20.3 years). They were all Dutch-English-French trilinguals and were drawn from the same population as in Experiment 1. Twenty-two were right-handed and two left-handed. All of them had normal or corrected-to-normal vision. Their mean experience with English was 8.8 years.

Materials and procedure
The IHs and their controls were presented in a mixed L1-L2 context. The items in the experiment were all words: words that exist in English and words that exist in Dutch (the IHs representing the critical items, which exist in both languages). Participants decided whether the item on the screen existed in L2 (hence, there was a strong focus on L2, as in the previous experiment). Given the nature of the items that required a 'no'-response, i.e., monolingual Dutch words, participants performed a variant of the lexical decision task (see above).

The same 20 Dutch-English homographs and their monolingual English controls were used, together with the 10 English filler words. To create an English-Dutch list, the nonwords in Experiment 1 were replaced by a set of 50 monolingual Dutch words. These were matched on an item-by-item basis with the IHs and the monolingual English words, both on number of letters (mean length: 3.90 letters) and word frequency. This frequency-matching was accurate, both with respect to the IHs ($t = 0.22$; $p = 0.83$) and with respect to the English controls ($t = 1.26$; $p = 0.23$). The mean frequency of the Dutch items was 43 per million (mean logfreq = 1.23).

3.2.7 Experiment 3: LDT in English-French context

Participants
Twenty-four students participated voluntarily (17 women, 7 men; age range: 19-23 years of age, mean age: 20.2 years). They were all Dutch-English-French trilinguals and were drawn from the same population as in the previous experiments. Twenty-two were right-handed and two left-handed. All of them had normal or corrected-to-normal vision. Their mean experience with English was 8.7 years.

Materials
The stimulus list represented the critical condition on the variable List Composition. The IHs and their controls were presented in a mixed L2-L3 context. Hence, the non-target language of the IHs (i.e., L1) did not occur in the two sets of monolingual words in the experiment. More particularly, whereas the IHs were words in both English (the L2 target language) and Dutch (L1), all monolingual words in the list were English (L2) and French (L3) words. The question is whether the presence of a non-target language will cause decision problems on the IHs because these words also activate a non-target language. Do participants tend to associate the 'no'-response with all languages that are not the target language in the experiment? Or do they selectively associate this response to the particular non-target language that appears in the experimental list? In the former case, an inhibition effect is expected on IHs of the L1-L2 type in a L2-L3 list. Participants will classify the activated L1 reading as a non-target language and, as they have learnt to respond "no" to L3 (i.e., non-target language) words, they will associate L1 with this response as well. In the latter case, no inhibition effect is expected. Because participants have only made 'no'-responses to L3 words, they will not associate this response to an activated L1 reading. In that case, no response conflict will emerge and, a 'yes'-response will be made as soon as the word's L2 reading becomes availble. Given the nature of the items that required a 'no'-response, i.e., monolingual L3 words, participants performed a variant of the lexical decision task (see above).

The same set of 20 Dutch-English homographs and their monolingual English controls was again used, together with the 10 monolingual English fillers. To create an English-French list, 50 monolingual French fillers were added. These were matched on an item-by-item basis with the 50 English words on letter length (mean length: 3.9 letters). None of the French words contained letters that do not occur in English, such as vowel letters with accents (e.g., é, è, ê, ô, à, ù, œ, …), or consonant letters that only occur in French (e.g., ç). The French words were selected from the same frequency range as the English words (using Lexique2, see New, Pallier, Brysbaert, & Ferrand, 2004). A paired t-test indicated that the set of IHs and their controls did not differ in frequency from the French words they were matched with ($t = 1.41$, $p = .17$). Note, however, that this matching was based on frequencies for native speakers of English and French, respectively. As the participants were more proficient in L2 than in L3, they were likely more familiar with the English words than with the French ones. However, note that this is not a problem for this experiment. The most important issue was that the French words could activate lexical representations in the participants' mental lexicon, such that the language "French" could become associated with the 'no'-response. Even though one cannot be sure that all participants were familiar with all French words in the list, there is no doubt that many of them should be familiar to everyone who has studied French at secondary school (e.g., ***sel***-'salt', ***sud***-'south', ***mer***-'sea').

The stimulus lists that were used in Experiments 1-3, which instantiated the three conditions on the factor List Composition, were also used in Experiments 4-6. The only difference was that a different task was used in the latter experiments, more particularly, the go/no-go task. If participants can selectively attend to the language of the 'go'-responses, and somehow be temporarily 'blind' to the non-target language during the experiment, an inhibition effect in the lexical/language decision task should be seriously reduced or even wiped out in the go/no-go task.

3.2.8 Experiment 4: GNG in pure English context

Participants
Twenty-four students participated voluntarily (20 women, 4 men; age range: 19-31, mean age: 21.6 years). They were all Dutch-English-French trilinguals and were drawn from the same population as in the previous experiments. Twenty-two were right-handed and two left-handed. All of them had normal or corrected-to-normal vision. Their mean experience with English was 9.7 years.

Materials
The same stimulus list was used as in Experiment 1: half of the items were existing English words (including the L1L2 IHs), the other half were English-like nonwords.

3.2.9 Experiment 5: GNG in English-Dutch context

Participants
Thirty-one students participated voluntarily (27 women, 3 men; age range: 19-28 years, mean age: 21.4 years). They were all Dutch-English-French trilinguals and were drawn from the same population as in the previous experiments. Twenty-six were right-handed and four left-handed. All of them had normal or corrected-to-normal vision. Their mean experience with English was 9.7 years.

Materials
The same stimulus list was used as in Experiment 2: half of the items were existing English words (including the L1L2 IHs), the other half were monolingual Dutch (L1) words.

3.2.10 Experiment 6: GNG in English-French context

Participants

Thirty students participated voluntarily (25 women, 5 men; age range: 19-31 years, mean age: 22.4). They were all Dutch-English-French trilinguals and were drawn from the same population as in the previous experiments. Twenty-six were right-handed and four left-handed. All of them had normal or corrected-to-normal vision. Their mean experience with English was 10.5 years.

Materials

The same stimulus list was used as in Experiment 3: half of the items were existing English words (including the L1L2 IHs), the other half were monolingual French (L3) words.

3.3 Results for Experiments 1-6

Participants whose error rates exceeded 15% for the whole stimulus list were excluded from the data in all six experiments (n = 8)[22]. Two IHs (slang, Dutch 'snake', and hoop, Dutch 'hope') and their English controls (birch and fowl) were removed from the data in all experiments because more than 50% errors were made on the controls words in all six experiments. RTs shorter than 300 ms and longer than 1,500 ms (outliers) were excluded (n = 140)[23].

The remaining data (n = 5,364) for the IHs and their controls (the 18 remaining pairs) were used for the RT and error analyses. There were 4,846 (90.34%) correct responses for the RT analysis and 518 error responses (9.66%) for the error analysis.

3.3.1 Analysis 1: (Generalized) Linear mixed effects models

Reaction Times

As mentioned above, the RT data were analyzed with a LMM. Visual inspection of the distribution of the RTs revealed non-normality, such that a data transformation was called for. A comparison of the logarithmic and inverse transformations showed that the latter was more successful in reducing the non-normality. All RTs were transformed with the formula -1000/RT[24].

The R statistical software package (R Core Team, 2014) and the LMM in the lmerTest package (Kuznetsova et al., 2016) were used. The latter include fixed and random effects parameters. We used the forward model selection procedure, in which the simplest model, containing only the random effects parameters, is tested first. Next, theoretically important fixed variables are added one at a time. When an extra factor is added to the model, a likelihood-ratio test is used to compare the simple model to the more complex model, which is nested within the simpler model. The test calculates how many times the likelihood of the

---

[22]In Experiments 1-6 the numbers of removed participants were 1, 0, 0, 0, 5, and 2, respectively.

[23]In Experiments 1-6 the number of outlier RTs were 13, 10, 15, 41, 38, and 23, respectively.

[24]The minus sign is needed to correctly interpret the beta values in the model's output (as the inverse operation reverses the relationship between the original values, e.g., 4 > 2 but ¼ < 1/2). The inverse RT is multiplied by 1000 to increase the beta values, which makes the output easier to read.

observations is higher in the best model compared to the worst model (α=.05 was used as the significance level). If the extra factor leads to a significant improvement of the simpler model, the added factor is kept in the model; otherwise it is removed. Next, another fixed variable is added to the model formula, and the procedure is repeated. The same rationale is used for the introduction of interaction terms. For instance, when fixed factors A and B are already included in the model, the next complex model includes their interaction. When no more fixed factors can improve the model, the best-fitting model has been identified and is used to analyze the data. Table 3.1 presents the mean RTs for the IHs and their controls as a function of both List Composition and Experimental Task.

Table 3.1.
Mean RTs (ms) and error rates (%) on IHs and controls as a function of List Composition and Task

	Pure L2		Mixed L1L2		Mixed L2L3	
Word Type	LDT	GNGT	LDT	GNGT	LDT	GNGT
Controls (L2)	647	670	594	607	648	595
IHs (L1L2)	678	739	702	781	673	666
Effect	31	69	108	174	25	71

Note. LDT refers to the lexical decision task in the Pure L2 condition and to a variant of the lexical decision task in the other two conditions of List Composition. GNG refers to the go/no-go task.

The final model contained three second-order interactions as fixed factors: Type x Context, Type x Task, and Context x Task. Previous RT was also included, as this is a covariate that is often found to account for a significant portion of the variance in the data. As a LMM assumes the observations to be independent from each other, such temporal dependencies between the data must be brought under statistical control, so as not to violate the model's assumptions. Moreover, many analyses demonstrate that the inclusion of this covariate leads to a better fit of the model (Baayen & Milin, 2010). This was also the case in the current analysis. The results of the LMM can be found in Table 3.2.

The beta estimates for each term in the output table and the estimates for each participant and item on the two random factors can be used to calculate the value that the model predicts for each individual response. These predicted values can then, in turn, be used to calculate the model-predicted means in each condition. If one wants to determine the 'pure' effect of a factor, i.e., without the 'contamination' by the random factors, one can use only the beta values for the different conditions on that factor to calculate the predicted means in each condition of that factor. Below we will briefly demonstrate how this can be done.

Each beta estimate must be considered with respect to the intercept. The intercept is the point where the regression line, running through the multidimensional space that is defined by the various factors in the model, crosses the axis that represents the dependent variable. It represents one of the conditions in the design, representing one level on each of the model's fixed effects (these are known as the 'reference levels'). The reference levels in this LMM are the control words on the factor Word Type, the pure L2 list on the factor List Composition, and the LDT on the factor Experimental Task. Taken together, all beta estimates can be used to determine the model-predicted value in each design condition. For instance, to calculate the predicted value for the interaction between homographs and the English-Dutch list, one starts with the value for the intercept, i.e., -1.77, and adds (a) the value for the level

Table 3.2.
Estimates for the fixed-effect predictors in the RT analysis of the Dutch-English homographs

	Estimate β	SE β	df	t	p
Intercept	-1.7700	0.05036	201.8	-35.15	< 0.001
Homograph	0.0550	0.02142	287.4	2.57	<0.05
English-Dutch	-0.1466	0.06273	145.0	-2.34	<0.05
English-French	-0.0082	0.06259	143.8	-0.13	>0.1
go/no-go	0.0185	0.06512	161.4	0.28	>0.1
PreviousRT	0.0002	0.00002	2987	8.63	< 0.001
Homograph:English-Dutch	0.2099	0.02937	2885	7.15	< 0.001
Homograph:English-French	0.0034	0.02805	2880	0.12	>0.1
Homograph:go/no-go	0.0907	0.02465	2891	3.68	< 0.001
English-Dutch:go/no-go	0.0422	0.08839	148.3	0.48	>0.1
English-French:go/no-go	-0.1803	0.08740	146.6	-2.06	<0.05

Note. The intercept refers to control words in the pure English condition in the LDT task.

'homographs' (term 'Homograph'), i.e., 0.0550, (b) the value for the level 'English-Dutch' (term: 'English-Dutch'), i.e., -0.1466, and (c) the value for the interaction between these two levels (term: 'Homograph:English-Dutch'), i.e., 0.2099[25]. The latter is needed to assess whether there is a significant interaction effect. If the change from 'controls' to 'homographs' is the same for the English-Dutch list as for the pure L2 list, i.e., if the value for homographs in the English-Dutch list barely deviates from the linear combination of the effects under (a) and (b) there will be no interaction. However, if the value that results from the linear combinations of these two effects is far removed from the observed value, the model will estimate a value for the interaction effect that reaches significance.

Recall that the RTs have been transformed by means of the formula -1000/RT. Hence, the beta values and the values for the design conditions that can be calculated from them, are expressed on this inverse scale. It is easy to back-transform the model predictions to the original RT scale. If RTi = -1000/RT, then RT = -1000/RTi, where RTi is the inverse of RT. Hence, when the value that has been calculated for the IHs in the English-Dutch condition is back-transformed into a value on the RT scale, one will find that the model predicts a mean RT of 605 ms for that condition (note that the level on the task factor was LDT). This is (obviously) not the same value as the one in Table 3.1, i.e., 702 ms, as no model perfectly fits the data. In this case, there are two reasons for the mismatch. First, the predicted value is only based on the fixed effects and does not take the random effects of participants and items into account (which are, by definition, part of the observed data). Second, the model does not (and cannot) take the unexplained error variance into account, i.e., the difference between the observed values and the ones predicted by the model. As a result, the model-predicted means will almost never coincide with the observed means.

[25] Hence, the predicted value for homographs in the English-Dutch list is calculated as: -1.77 + 0.05498 - 0.1466 + 0.2099 = -1.65172 (note that this value is expressed on the scale of the natural logarithm).

In what follows, we will only discuss what is theoretically important, more particularly, the impact of List Composition and Experiment Task on the size (and, possibly, direction) of the homograph effect. For obvious reasons, the effect of these design factors on the monolingual control words will not be considered. For instance, the effect for the condition labelled 'English-Dutch', i.e., the set of control words in the LDT in the English-Dutch stimulus list, is significant ($t = -2.34$, $p = 0.02$), but this finding is not important for the questions under study.

List Composition and the homograph effect

The IHs caused an inhibition effect on the RTs in the pure L2 list in the LDT. The term Homograph is significant ($t = 2.57$, $p = 0.01$) when compared to the control words at the intercept. Its estimated beta value is positive, i.e., 0.05498, which indicates that this value must be added to the intercept value. Back-transforming these two values (-1.77 and -1.715) yields the predicted values on the original RT scale for the controls and IHs in a pure L2 list, given the LDT: 565 ms and 583 ms, respectively. Hence, in a pure L2 list the IHs cause an inhibition effect of 18 ms.

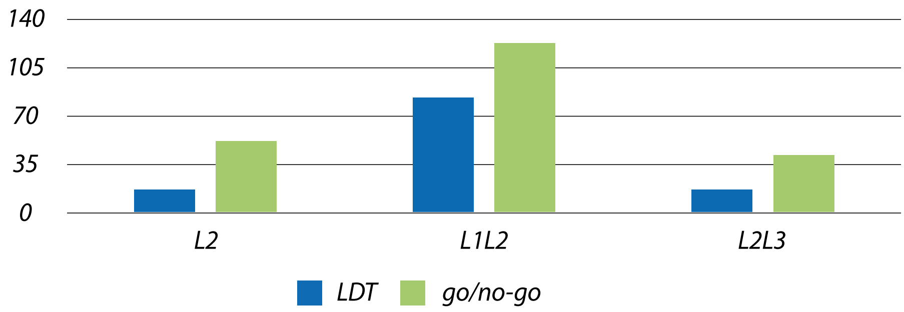
How does the homograph effect interact with the two mixed lists in the LDT? The model's interaction terms Homograph:English-Dutch and Homograph:English-French provide this information. These interaction terms reflect whether the expected values for IHs in the English-Dutch and English-French conditions are the linear combination of the homograph effect in the pure L2 list (in the LDT, i.e., the condition for the task variable in the intercept) and the effect of English-Dutch and English-French, respectively. If so, the beta value for the interaction term will not be significant. However, if the homograph effect in these lists is larger or smaller than would be expected on the basis of the linear combination, the beta value will be associated with a significant effect.

The homograph effect in the English-Dutch (L2L1) list is significantly larger, i.e., even more inhibitory, than in the pure L2 list ($t = 7.15$, $p < .0001$). In contrast, the homograph effect in the English-French (L2L3) list is not statistically different from the effect in the pure L2 list ($t < 1$). The model-predicted values for the English-Dutch condition, after back-transforming the values on the inverse RT scale to the original scale, are 522 ms vs. 605 ms for the controls and IHs, respectively (84 ms inhibition). For the English-French condition, these values are 562 ms and 581 ms (19 ms inhibition, which is virtually identical to the 18 ms inhibition in the pure L2 list).

As there was no third-order interaction between Word Type, List Composition, and Experimental Task, the interaction between Word Type and List Composition did not differ statistically between the two tasks. Hence, in the go/no-go task, the inhibition effect in the pure L2 list was also significantly smaller than in the L2L1 list, but did not differ from the inhibition effect in the L2L3 list.

The predicted values in the go/no-go task can be calculated according to the same rationale as above. Starting from the predicted beta values for the controls and IHs in the three list compositions in the LDT, one adds (a) the beta value for the effect of the go/no-go task in all six cells for the go/no-go task, (b) the beta value for the interaction between the go/no-go task and the two mixed lists in the four cells for the mixed list conditions, and (c) the beta value for the interaction between the go/no-go task and the IHs in the three cells for the IHs. When back-transforming these predicted values to the RT scale, for each list in the go/no-go task, the

Figure 3.1.
Inhibition effects on RTs as a function of list composition and task



following values are obtained for the controls and the IHs, respectively: 571 ms vs. 623 ms (52 ms inhibition) for the pure L2 list, 533 ms vs. 658 ms (125 ms inhibition) for the English-Dutch list, and 511 ms vs. 553 ms (42 ms inhibition) for the English-French list. Figure 3.1 visualizes the homograph effects as a function of list composition and experimental task.

When comparing these homograph effects to those for the LDT, it is clear why a likelihood ratio test indicated that a model with the third-order interaction between Word Type, List Composition, and Experimental Task did not significantly improve a model that only contained the three second-order interactions. The predicted homograph effects for the three lists indeed reveal a highly similar pattern of inhibitory effects in the two tasks. In both the LDT and the go/no-go task, a much larger inhibition effect was obtained in the English-Dutch list than in the other two lists (see below), whose effects were very similar. The difference between the latter two homograph effects was not significant ($t < 1$), as revealed by the interaction term that compares the effect in the English-French condition to that in the pure English condition. Finally, the inhibition effects in the go/no-go task were always larger than in the LDT (see below).

To make sure that the inhibition effect in the English-Dutch list was also significantly larger than in the mixed English-French list, the same model was rerun by replacing the pure L2 condition by the English-Dutch condition in the intercept (using the function relevel in R). In LDT, the interaction term testing whether the homograph effect differed between the English-Dutch list and the pure English list was highly significant (ß = -0.2065, $t = -7.51$, $p < .0001$). As there was a Word Type by Experimental Task interaction but no third-order interaction between the factors Word Type, Experimental Task, and List Composition, the same outcome was obtained when including the go/no-go task in the reference level (ß = -0.2065, $t = -7.51$, $p < .0001$).

Experimental Task and the homograph effect

Even though the experimental task does not affect the relation between the homograph effects in the three list conditions, Figure 3.1 suggests that the homograph inhibition effect is larger in the go/no-go task than in the variant of the lexical decision task. The model output indeed confirms that the interaction term testing whether there is a difference between the homograph effect in the go/no-go task and the LDT is highly significant ($t = 3.68$, $p = .0002$). As expected, the predicted RTs for the controls and IHs in the two tasks reveal a much larger mean inhibition effect in the go/no-go task than in the LDT. Mean predicted RTs for controls and IHs were 565 ms vs. 583 ms (18 ms inhibition) in the LDT and 571 vs. 623 ms (52 ms inhibition) in the go/no-go task.

Finally, note that the effect of Previous RT was indeed highly significantly ($t = 8.63$, $p < .0001$). As mentioned by Baayen and Milin (2010) this is often one of the strongest effects on the RTs in psycholinguistic experiments. As mentioned earlier, it is important to statistically control for this effect because its existence demonstrates that the observations are non-independent, as required by linear mixed models (and many parametric tests).

The homograph effect in each experiment

To investigate whether the homograph effect was significant in each of the six experiments, i.e., all combinations of List Composition and Experimental Task[26], the same model was re-run by changing the reference levels on the factors List Composition and/or Experimental task. The homograph effect was significant in each experiment. For the LDT the following results were obtained: $t = 2.57$, $p = .01$ in the pure L2 list, $t = 11.82$, $p < .0001$ in the L1L2 list, and $t = 2.83$, $p < .005$ in the L2L3 list. For the go/no-go task the outcomes were: $t = 5.14$, $p < .0001$ for the pure L2 list, $t = 13.96$, $p < .0001$ for the L1L2 list, and $t = 6.17$, $p < .0001$ for the L2L3 list.

Error analyses

List Composition and the homograph effect

The same analyses were performed on the error rates. Table 3.3 represents the error percentages for the controls and the IHs, and the resulting homograph effects, as a function of List Composition and Experimental Task.

The process of model selection yielded the same model as for the analysis of the RTs: the three second-order interactions were included in the model but the third-order interaction did not significantly improve the model's fit to the data anymore. The factor Previous RT was not included either. Table 3.4 shows the beta estimates, their standard error, and the associated z and p values for each for each effect in the model's output table.

Table 3.3.
Mean error rates (%) on IHs and controls as a function of List Composition and Task (Experiment 1)

	Pure L2		Mixed L1L2		Mixed L2L3	
Word Type	LDT	GNGT	LDT	GNGT	LDT	GNGT
Controls (L2)	4.1	7.7	3.5	1.3	6.7	1.2
IHs (L1L2)	7.0	12.3	33.1	16.2	14.8	8.5
Effect	2.9	4.6	29.6	14.9	8.1	7.3

Note. LDT refers to the lexical decision task in the Pure L2 condition and to a variant of the lexical decision task in the other two conditions of List Composition. GNG refers to the go/no-go task.

[26]Note that this does not increase the risk of Type I errors, as when a series of pairwise comparisons are performed when, for instance, a significant interaction is found in an ANOVA with a complex design. Post-hoc comparisons do affect the likelihood of a Type I error, which is why corrections are taken into account (e.g., lowering the alpha level, as in the Bonferroni correction). However, changing the reference level of a LMM (or GLMM) does not amount to a novel analysis on the same data set. The only change is that the reference level at the intercept is changed on one or more variables, to be able to assess all contrasts that are theoretically important. A change in reference level does not in the least change the analysis. The same contrasts in two models with different reference levels yield the same beta estimates, although their sign sometimes changes. One looks at the same data from a different 'vantage point', so to speak. When A is the reference level and B has a negative ß value, a change of reference level (B becomes the reference level) will not affect the numbers but the sign will reverse (as A remains the level with higher scores than B).

Table 3.2.
Estimates for the fixed-effect predictors in the RT analysis of the Dutch-English homographs

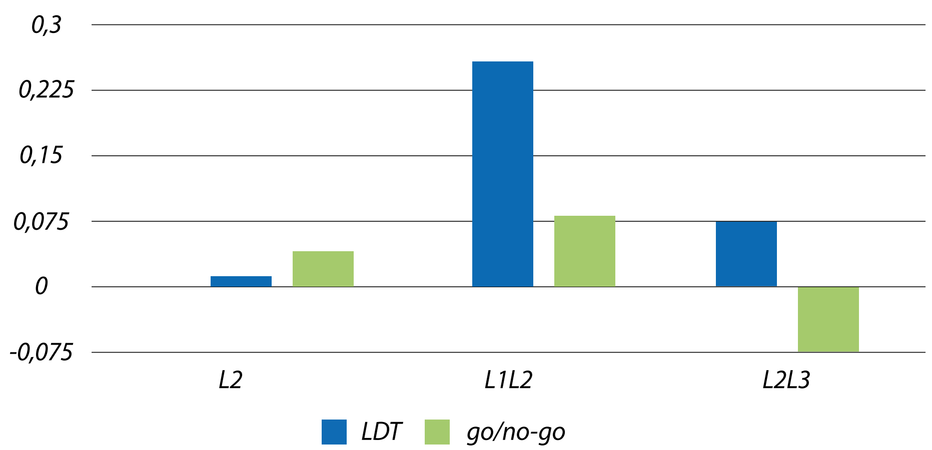
	Estimate β	SE β	z	p
Intercept	-3.6327	0.3515	-10.33	< 0.001
Homograph	0.3763	0.2497	1.507	>0.1
English-Dutch	-0.2115	0.3791	-0.558	>0.1
English-French	0.3124	0.3546	0.881	>0.1
go/no-go	0.5269	0.3439	1.532	>0.1
Homograph:English-Dutch	2.5420	0.3248	7.826	< 0.001
Homograph:English-French	0.8579	0.2920	2.938	< 0.01
Homograph:go/no-go	0.3567	0.2601	1.372	>0.1
English-Dutch:go/no-go	-2.0161	0.4242	-4.752	< 0.001
English-French:go/no-go	-1.8561	0.4301	-4.316	< 0.001

Note. The intercept refers to control words in the pure English condition in the LDT task.

The homograph effect in the pure L2 list was not significant ($t = 1.51$, $p = .13$). The effects in the two mixed lists were significantly more inhibitory ($t = 7.83$, $p < .0001$ and $t = 2.94$, $p = .003$ for the English-Dutch and English-French lists, respectively). Based on the fixed effects in the model, the predicted error probabilities for the pure L2, L1-L2, and L2-L3 lists in the LDT were 0.026, 0.021, and 0.035, respectively, for the controls, and 0.037, 0.284, and 0.110 respectively, for the IHs. The corresponding probabilities in the go/no-go task were 0.043, 0.035, and 0.058 for the controls, and 0.085, 0.113, and 0.045 for the IHs. Figure 3.2 visualizes the inhibition effects (i.e., control minus IH) as a function of list composition and experimental task.

To find out whether the effects for the L1-L2 and L2-L3 conditions differed, the model was run again, using an intercept in which the pure L2 condition had been replaced by the L1-L2 condition. The difference between L1-L2 and L2-L3 was highly significant: many more errors were made in the L1-L2 list than in the L2-L3 list ($ß = -1.68$, $t = -5.27$, $p < .0001$).

Figure 3.2.
Inhibition effects on error proportions as a function of list composition and task



Experimental Task and the homograph effect
The error rates did not differ statistically between the two experimental tasks. There was no effect of Experimental Task on the controls ($t = 1.53$, $p = .13$) nor was there a significant interaction with the IHs ($t = 1.37$, $p = .17$). In other words, the error proportions (across lists) did not differ between the two experimental tasks, neither on the control items nor on the IHs.

The homograph effect in each experiment
By making each of the six combinations between List Composition and Experimental Task part of the intercept (by changing the reference level for the factors List Composition and Experimental Task) and rerunning the model, we assessed the homograph effect in each experiment. Each time an inhibitory effect was found, except in the pure L2 list in the LDT. In LDT we obtained the following results: $t = 1.51$, $p = .13$ in the pure L2 list, $t = 11.47$, $p < .0001$ in the L1L2 list, and $t = 5.53$, $p < .0001$ in the L2-L3 list. In the go/no-go task we obtained the following outcomes: $t = 3.35$, $p = .0008$ in the pure L2 list, $t = 10.48$, $p < .0001$ in the L1-L2 list, and $t = 5.86$, $p < .0001$ in the L2-L3 list.

As discussed above, the purpose of the following analyses is (a) to find out whether the homograph effect in the different cells of the experimental design is due to a shift of the entire RT distribution, restricted to the right tail of this distribution (i.e., the longest RTs), or reflects a combination of these two effects (section 3.3.2), and (b) to find out how the magnitude of the homograph effect evolves across the range of response times (sections 3.3.3 and 3.3.4).

3.3.2 Analysis 2: ex-Gaussian analyses

As discussed in section 3.1.2, several papers in the past decade have emphasized that statistical analyses of psycholinguistic data depend too much on comparisons between condition means (e.g., Balota et al., 2008). Focusing on the mean implies that one loses individual variation among participants out of sight. This does not mean that one collapses data over participants (as in a chi-square test, for instance), such that all differences between participants are wiped out. ANOVAs, for instance, have been developed to avoid throwing away information that is tied to the different observational units (participants, items, etc.). The basic purpose of an ANOVA is to find out whether the difference between the means can be 'projected' onto each observational unit (here: each participant) without leaving a large amount of unexplained variance. For instance, when two conditions are compared in a psycholinguistic experiment, an ANOVA assumes that the two means for each participant reflect two components: (a) the difference between the grand means, i.e., the means that result from averaging across the participant means in each condition, and (b) the residual error, i.e., the deviation between a participant's observed means and the grand (condition) means. By dividing the variance in the participant means that is due to the manipulation (i.e., the difference between the grand means multiplied by the number of participants) by the residual error variance, i.e., the difference between the condition means for each participant that is not captured by the grand means (because it is smaller or larger), the test assesses whether the variance due to the experimental manipulation is sufficiently larger than the variance due to the residual error. The latter, the denominator in an ANOVA, represents the interaction between the independent variable and the random effect of the participants. Hence, in a sense, participants are even crucial units in parametric tests. However, that is not the reason why some researchers criticize the overrated importance of the mean. Participant means are not included in many parametric tests for the sake of focusing on the entire RT distribution. Rather they only serve the purpose of separating the 'signal' caused by the independent variable from the random error variance (the 'noise') that is introduced by the participants. This specific role of the participant means emphasizes that the real focus is on the condition means and not on the entire RT distribution.

As mentioned above, Balota and co-workers have written several contributions in which he argues for a different approach (e.g., Balota et al., 2008; Balota & Yap, 2011). They argue that psycholinguists ignore too often the well-known fact that RTs generally do not form a good example of a normal distribution. Recall that such right-skewed distributions can be mathematically analyzed as the superposition of two distributions: a normal (Gaussian) distribution and an exponential distribution, the latter being required to account for the larger number of long RTs than in a normal distribution. The goal of the ex-Gaussian analyses reported below is to estimate the mean (***mu***) and standard deviation (***sigma***) of the Gaussian distribution and a single parameter of the exponential distribution (***tau***), which is at the same time its mean and standard deviation. This makes it possible to identify the source of an effect. Indeed, it is important to know whether a significant effect is due to the central tendency of the normal distribution (such that it represents an average effect, i.e., a shift in the entire RT distribution, reflected in mu) or is due to an overrepresentation of long RTs in the right tail of the distribution (reflected in tau). In this context, Roelofs, Piai, and Rodriguez (2011) refer to work by Ridderinkhof and colleagues in the non-linguistic domain (Ridderinkhof, Scheres, Oosterlaan, & Sergeant, 2005; Ridderinkhof, Van den Wildenberg, Wijnen, & Burle, 2004). This research suggests that attentional mechanisms are often the cause of inhibition. As such mechanisms are relatively slow and, hence, the inhibition they cause takes time to build up, this inhibition especially surfaces in the longer RTs.

The link with the experiments in the current study is straightforward. As the homograph inhibition effect originates in the decision stage of an experimental task and may also be affected by attentional factors (entirely or partially), it is important to know whether this effect marks a shift of the entire RT distribution for the control condition, is restricted to the longer RTs (when attentional mechanisms become more prominent), or reflects the combination of these two effects. Recall that the latter scenario obtains for many psycholinguistic variables (see Balota et al., 2008). Hence, we performed ex-Gaussian analyses to our data.

As in the previous section, we performed an omnibus analysis to increase the power of our analysis and to avoid the risk for Type I errors. Recall that a single set of IH-control pairs

Table 3.5.
Mean mu values for each combination of Word Type, Stimulus Composition, and Experimental Task

LDT	pure L2	L1-L2	L2-L3
Control	540	513	531
Homograph	553	611	559
Effect	13	98	28
GNGT	pure L2	L1-L2	L2-L3
Control	546	514	478
Homograph	583	629	513
Effect	37	115	35

was used across all six experiments. It would have been unwise not to make our statistical analyses benefit from this property. We calculated the three parameters mentioned above for all participants, both for their RTs on the IHs and their L2 controls. Hence, in each of the six experiments, each participant had two mu values, two sigma values, and two tau values. These parameters were calculated with the R function mexgauss from the retimes package (Massidda, 2013). We analyzed these values with linear mixed effects models, to find out which parameter(s) were affected by a homograph effect, and whether this effect interacted with the other two design factors (List Composition and Experimental Task).

Analyses of mu

Table 3.5 shows the mean values of mu as a function of the three design factors that defined the experiments. Model selection showed that no model outperformed the model that (only) contained the fixed factors Word Type (homograph vs. control), Stimulus Composition (pure L2 list, L1-L2 list, L2-L3 list), and their interaction. Participants and items were included as crossed random factors. The factor Experimental Task had no effect.

Table 3.6 shows the output of the LMM analysis, using the control items in the L1-L2 list as the reference level. A significant homograph inhibition effect was found ($t = 8.60$, $p < .0001$). This effect was significantly larger than the inhibitory effects in the pure L2 list ($t = -4.54$, $p < .0001$) and the L2-L3 list ($t = -4.28$, $p < .0001$). When the LMM was re-run with the pure L2 list as its reference level, the inhibition effect barely missed significance ($t = 1.96$, $p = .052$). Using the L2-L3 list as the reference level, the inhibition effect was highly significant ($t = 2.66$, $p = .009$). The effects for the pure L2 list and the L2-L3 list did not differ from each other ($t < 1$).

Table 3.6.
Estimates for the fixed-effect predictors of mu in the ex-Gaussian analysis

	Estimate β	SE β	t	p
Intercept	513.91	13.77	37.33	< .0001
Homograph	106.01	12.33	8.60	< .0001
English	28.99	19.89	1.46	0.15
English-French	-11.63	19.28	-0.60	0.55
Homograph:English	-80.85	17.81	-4.54	< 0.001
Homograph:English-French	-73.90	17.26	-4.28	< .0001

Note. The intercept refers to control words in the pure English condition in the LDT task.

Table 3.7.
Mean sigma values for each combination of Word Type and Experimental Task

	LDT	GNGT
Control	70	70
Homograph	104	124
Effect	34	54

Analyses of sigma

Model selection showed that no model outperformed the model that (only) contained the factors Word Type, Experimental Task, and their interaction. The factor List Composition did not improve any model in the model selection process. Table 3.7 shows the mean values of sigma as a function of Word Type and Task.

Table 3.8 shows the output of the LMM analysis on the sigma values, using the control items in the LDT as the reference level. The homograph effect in the LDT was highly significant ($t = 5.33$, $p < .0001$): IHs gave rise to a larger standard deviation in the Gaussian component of the RT distribution than their controls. The interaction term revealed that this effect was even larger in the go/no-go task ($t = 2.36$, $p = .02$). It was not surprising, then, to find a larger homograph effect when the go/no-go task was used as the reference level for the factor Experimental Task ($t = 8.96$, $p < .0001$).

Table 3.8.
Estimates for the fixed-effect predictors of sigma in the ex-Gaussian analysis

	Estimate β	SE β	t	p
Intercept	70.33	4.98	14.12	< 0.001
Homograph	33.48	6.28	5.33	< 0.001
go/no-go	-0.64	6.91	-0.09	0.93
Homograph:go/no-go	20.49	8.70	2.36	0.02

Note. The intercept refers to control words in the LDT.

Table 3.9.
Mean tau values for each combination of Word Type, Stimulus Composition, and Experimental Task

LDT	pure L2	L1-L2	L2-L3
Control	108	116	80
Homograph	131	123	96
Effect	23	7	16
GNGT	pure L2	L1-L2	L2-L3
Control	125	118	94
Homograph	158	162	166
Effect	33	44	72

Analyses of tau

Model selection yielded a model that contained the factors Word Type, Experimental Task, List Composition, and the interaction between Word Type and Task. Table 3.9 shows the mean tau values for each cell in the design behind the six experiments.

Table 3.10 shows the output of the LMM analysis on the tau values, using the control items from the pure L2 list in the LDT as the reference level. The homograph effect was not

Table 3.10.
Estimates for the fixed-effect predictors of tau in the ex-Gaussian analysis

	Estimate β	SE β	t	p
Intercept	108.91	8.33	13.07	< 0.001
Homograph	15.01	9.11	1.65	0.10
go/no-go	10.73	9.28	1.16	0.25
English-French	-1.06	8.37	-0.13	0.90
English-Dutch	-21.34	8.45	-2.53	0.01
Homograph:go/no-go	35.28	12.62	2.80	0.006

Note. The intercept refers to control words from the English condition in the LDT.

significant in the pure L2 list in LDT (t = 1.65, p = .10). As there was no significant interaction between Word Type and List Composition (recall that this factor was not included in the model), the homograph effects in the three lists did not differ statistically from each other. The interaction between Word Type and Experimental Task showed that the homograph effect in the go/no-go task was significantly larger than in the LDT (t = 2.80, p = .006). The latter effect itself was significant. When the go/no-go task was used as the reference level on the factor Experimental Task, this effect was highly significant in the pure L2 list (t = 5.75, p < .0001). Again, the absence of a significant interaction between Word Type and List Composition indicates that this homograph effect was equally strong in all three lists.

In short, the homograph effect had a significant effect on tau in all three lists in the go/no-go task but did not have a significant effect in any list in the LDT. The three lists did not differ significantly from each other in either task.

3.3.3 Analysis 3: Quantile analyses

Recall that a quantile analysis makes it possible to find out whether the magnitude of an effect remains constant (or not) across a number of RT bins, which are rank-ordered as a function of response speed. As has been described above, this is accomplished by rank-ordering the RTs (from small to large) for each participant in each experimental condition separately, and dividing each ordered set into the same number of bins or quantiles. Suppose there are 10 such quantiles. In that case, the means of the 10% fastest RTs for each participant represent the first quantile. The means of the 10% next fastest RTs for each participant represent the second quantile, etcetera. Finally, the means of the 10% slowest RTs are grouped in the last quantile.

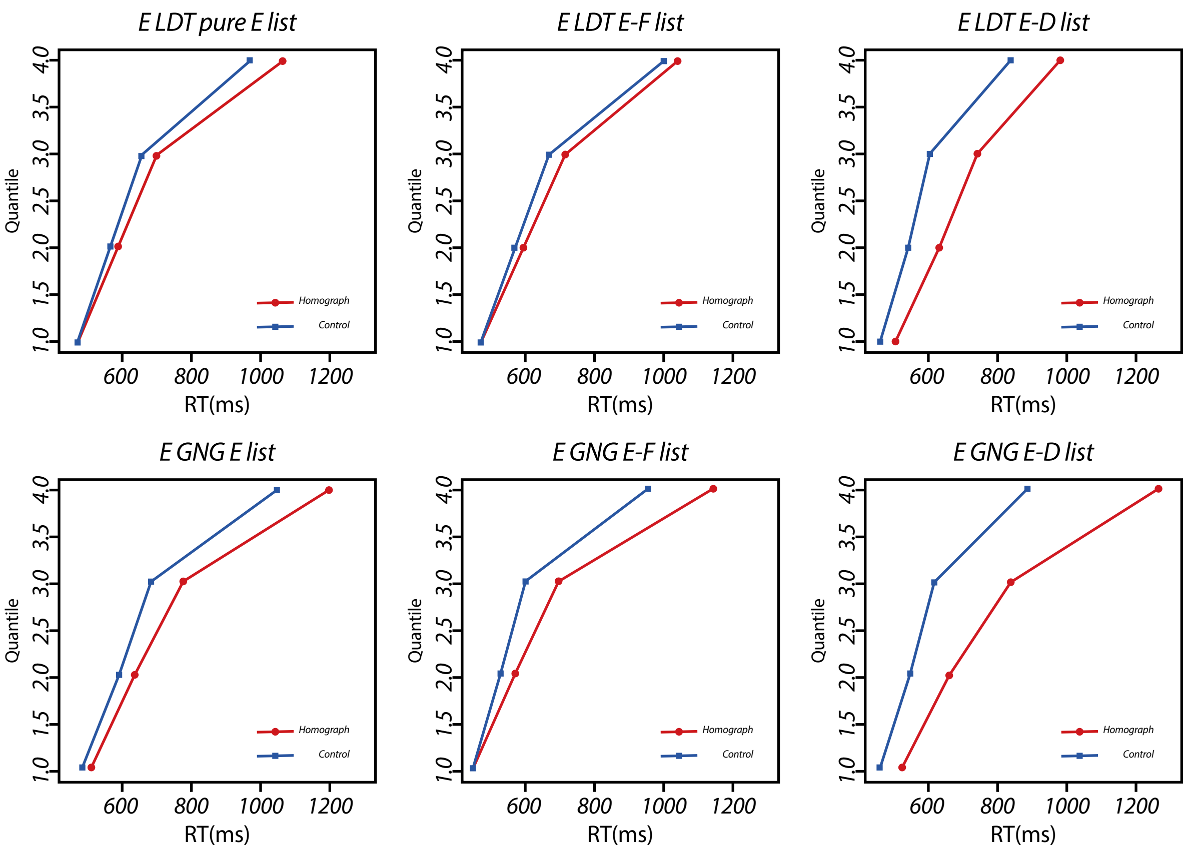
An ANOVA analysis on these participant means is then applied to test whether there is a significant interaction between the factor Quantile and the independent variable that is manipulated in the study (here: Word Type). As each participant contributed the same number of observations to each quantile, in all conditions, a significant interaction indicates that the size of the effect changes (i.e., increases or decreases) across quantiles and, hence, depends on response speed. Planned comparisons on the successive quantiles can be used to find out when the effect increases. The hypothesis behind the right-skewed RT distribution is that it will increase as RTs become longer (see above).

Note that quantile analyses are intimately related to ex-Gaussian analyses. Suppose that a manipulation affects only one parameter: mu, i.e., the mean of the Gaussian component of the RT distribution. This would mean that a change on the factor under study shifts the entire distribution, such that both short and long RTs are affected equally. This would be reflected in a non-significant interaction between the Quantile factor and the manipulation. However, when the manipulation has a larger effect on the longer RTs, the effect will only gradually emerge, and become stronger in the quantiles representing longer RTs. In the ex-Gaussian analysis, the effect will likely become significant in the analysis of tau, i.e., the mean (and standard deviation) of the exponential component of the RT distribution. In a quantile analysis, this will manifest itself in a significant interaction between the Quantile factor and the factor under manipulation (in this case, Word Type, i.e., IHs vs. controls).

A quantile analysis was performed for each of the six experiments. The RTs for each participant's IHs were rank-ordered from small to large and were then divided into four equally sized quantiles, i.e., each quantile represented 25% of the RTs. The same was done for the L2 control words. Thus a 2 (Word Type) x 4 (Quantile) design was created. The quantile plots for the six experiments are shown in Figure 3.3.

For our purpose, the most important factor was the Word Type by Quantile interaction. Note that the Quantile factor is a repeated-measures factor. Because repeated-measures designs run the danger of violating the so-called sphericity assumption[27], we used the

Figure 3.3.
Quantile plots for the experiments of the current study.



[27] The sphericity requirement is met when the variance in the difference scores between two conditions in a repeated measures design is equal for all possible pairs of repeated measures. Because this is often not the case, several correction techniques have been introduced.

Table 3.11.
p-values for the homograph effect at each quantile as a function of Task and List Composition

		Quantile 1	Quantile 2	Quantile 3	Quantile 4
LDT	pure L2	1.00	0.23	0.07	0.07
	L1L2	0.03	.000008	.000002	.00001
go/no-go	pure L2	0.18	0.07	0.001	0.03
	L1L2	.001	.000000009	0.0000004	.000003
	L2L3	1.00	0.0006	.000007	0.02

Note. p-values are only reported when the Quantile by Word Type interaction was significant. The actual p values are reported. They are put in italics when the homograph effect at the quantile is significant.

Greenhouse-Geisser correction to remedy against this problem. This made it less likely to make Type I errors. In all cases but one, this interaction reached significance (LDT – pure L2 list: $p = .04$; LDT – L1-L2 list: $p < .002$; LDT – L2-L3 list: $p = .68$; go/no-go – pure L2 list: $p = .037$; go/no-go – L1-L2 list: $p < .0001$; go/no-go – L2-L3 list: $p = .01$).

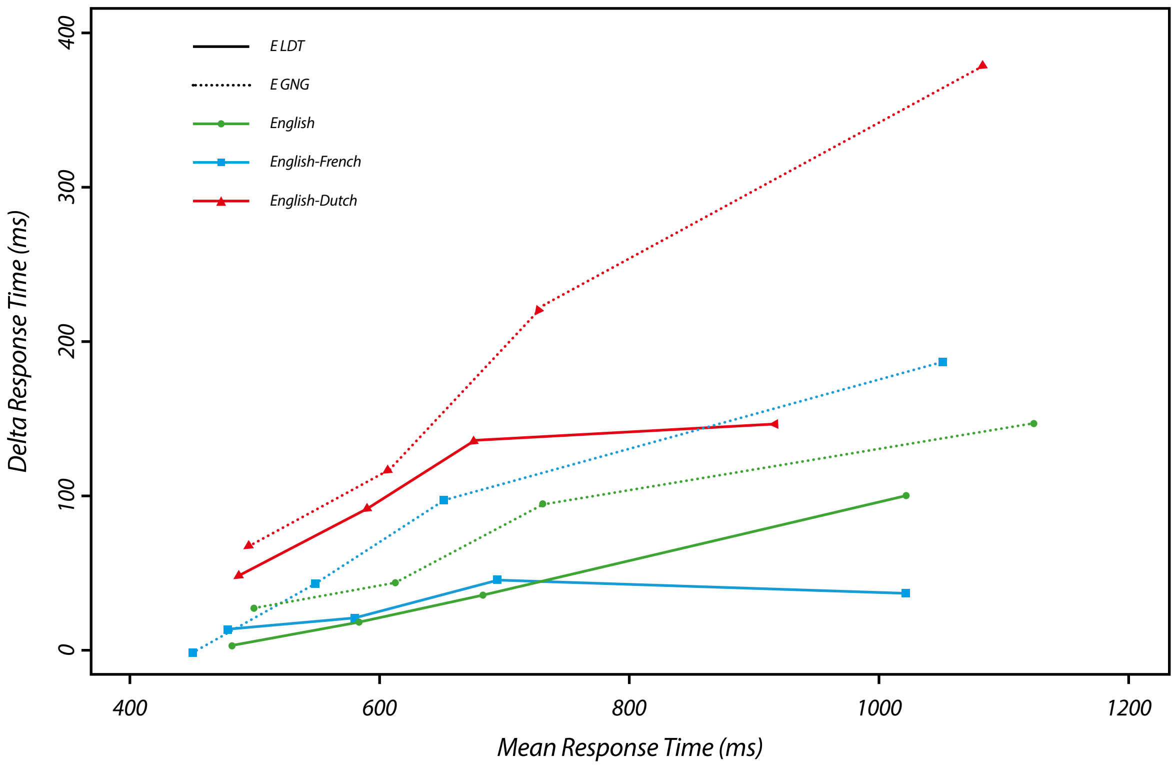
We wanted to find out at which quantile the homograph effect emerged. Hence, when the Quantile by Word Type interaction was significant we ran t-tests at each of the four quantiles, comparing the RTs for the IHs and their controls. To correct for multiple analyses on the same data set, we adjusted our significance level by applying the Bonferonni correction. As we performed four t-tests on a single data set, the significance level was set at $\alpha = 0.0125$. Table 3.11 shows the Bonferroni corrected p values for each quantile in each experiment where the interaction is significant.

In the LDT, the homograph effect at the individual quantiles only reaches significance in the L1-L2 list. The Quantile by Word Type interaction in this list indicates that the effect is relatively small in the quantile containing the fastest RTs ($p = .03$) but is highly significant in the other three quantiles (all ps < .001). In contrast to the LDT, there are many more significant homograph effects in the go/no-go task, in each of the three lists. The smallest effects are found in the pure L2 list, where significant effects only appear in the last two quantiles, i.e., in the region where the 50% slowest RTs can be found. In the L2-L3 list, the effect already becomes significant at the second quantile. Finally, in the L1-L2 list, the effect is significant at each quantile. As can be observed in Table 3.11 the L1-L2 list is the only condition where a significant effect is found at each quantile, in both tasks. In the LDT, the effect is relatively small (but significant) in the first quantile of this list ($p = .03$), but in the go/no-go task it is already highly significant in this quantile ($p < .0001$).

3.3.4 Analysis 4: Delta plot analyses

Whereas a quantile analysis tests whether the size of an effect changes across quantiles, a delta plot analysis addresses the question whether there is a 'change in the change of the effect' (when moving from one quantile to the next). This change in the effect size between two successive quantiles may become larger when moving towards the longer RTs. For instance, a 20 ms difference between the effects at quantiles 1 and 2 may develop into a 40 ms difference between the effects at quantiles 2 and 3, which may develop into an 80 ms difference between the effects at quantiles 3 and 4 (a non-linear growth of the effect). The change in the effect may also remain constant, i.e., the change in the magnitude of the effect between

Figure 3.4.
Delta plots for the experiments of the current study.



successive quantiles may increase (or decrease) with a constant. This would be the case if the change between quantiles 3 and 4 in the example were 60 ms, i.e., the effect size would increase with a constant of 20 ms when moving from one quantile to the next. As mentioned earlier, a growth of the effect size (either linear or non-linear) might result from an increase in the use of attentional mechanisms to suppress the irrelevant reading of the IH.

A constant change in the size of an effect across the RT axis will manifest itself as a straight line in a delta plot. Whenever the slope of the line becomes steeper or less steep between successive quantiles, a region has been reached where something interesting might be going on. Obviously, a delta plot only visualizes the effect. An increase or decrease in the slope's steepness is only theoretically relevant when it is statistically significant.

Calculating the slope between two quantiles is not difficult and goes back to the mathematics underlying the equation of a straight line in a two-dimensional space. The formula for the equation of a straight line is expressed as $y = \alpha + \beta x$, where α is a constant (the value at which the line crosses the y-axis when $x = 0$) and β is the slope, i.e., the number of units that y increases per unit increase of x. For instance, the equation $y = 3 + 2x$, yields the coordinates (0,3), (1,5), (2,7), etc. This is directly relevant to the calculation of the slope in a delta plot. Whenever one has two x values and two y values one can derive the slope by calculating the ratio $(y2-y1)/(x2-x1)$. For instance, given the coordinates (1,5) and (2,7) this ratio is $(7-5)/(2-1) = 2/1 = 2$, which is indeed the slope, i.e. the value of β in the linear equation above. In a delta plot, the x-axis represents the mean RT in each quantile (averaged across the critical and control items), whereas the y-axis represents the so-called delta scores, i.e., the differences between the means for the critical and control items in each quantile. Hence, the size of the effect in a quantile is plotted against the average response speed in that quantile. This makes it possible to visualize the evolution of the magnitude of the effect as a function of response speed. The delta plots for the six experiments of the current study can be found in Figure 3.4.

To calculate the slope of the line connecting to points in the delta plot, one simply applies the above logic. The slope of the straight line is the outcome of the formula (delta2 - delta1)/(mean2 - mean1). All slope values for each participant can then be analyzed by means of an ANOVA, to find out whether the slopes between all pairs of successive quantiles differ from each other. If not, there is a constant increase in the effect as RTs become longer (note that, theoretically, this 'increase' may be zero). However, if a statistical difference is found, this means that one or several slopes are steeper (or less steep) than the others. By inspecting where such changes occur on the RT axis, one can achieve insight into cognitive mechanisms whose operation is time-dependent.

For each experiment, we performed a statistical analysis on the participants' values for the three slopes connecting the four coordinates that we derived from the quantile analyses, using the above formula. No significant effect was found in five of the six cases (all $ps > .15$). The only exception was the analysis of the L1-L2 condition in the LDT: $F(2,69) = 3.18$, $p = .048$. However, closer inspection of this effect showed that the last slope is less steep than the others, whereas the opposite would be expected. As this seems hard to explain (it occurs in no other delta plot) and just hits the significance level, we do not want to speculate about what might be going on.

The general picture that emerges from the delta slope analyses is that in each of the experiments the size of the effect becomes larger as RTs become larger. Moreover, in all experiments (with the single exception just discussed) the effect grows linearly, i.e., an increase of one unit on the x-axis (i.e., the RT axis) corresponds with a constant increase in the size of the inhibition effect on the y-axis.

3.4 General Discussion

The current study was designed with three goals in mind. (1) Can one find a homograph inhibition effect for L1L2 interlingual homographs if these are embedded in a mixed list with only monolingual words in L2 and L3, when L2 is the target language? If so, does this list behave as a pure L2 list (there being no L1 words), as a mixed L1-L2 list (L3 and L1 being the non-L2 language in both types of mixed lists), or as third list type, whose effect is situated intermediate the effects for the pure L2 and mixed L1-L2 lists? This question has never been addressed before. (2) Is the homograph effect larger in a go/no-go task than in a variant of the lexical decision task. The former task arguably involves a more complex decision and response process, and, hence, leaves more room for the influx of attentional processes (which would increase the magnitude of the inhibition effect)? If so, is the effect of list composition the same in the go/no-go task as in the LDT? The effect of experimental task in its interaction with list composition has not been reported before. (3) Does the homograph effect represent a shift with respect to the entire RT distribution of the monolingual control words, is it restricted to the region of the longer RTs (because a response conflict is especially expected when the participant cannot quickly solve it), or is it a combination of both these effects? To the best of our knowledge, no one has ever looked beyond means comparisons when studying the visual recognition process of IHs.

To investigate the first question, our participants were Dutch-English-French trilinguals. Trilinguals have occasionally been studied (see Lemhöfer, Dijkstra, & Michel, 2004; Van Hell & Dijkstra, 2002) but, in general, they have received little attention in the literature. Note that the question is an entirely different one than the one that we raised with respect to cognate processing in Chapter 2, where we also studied trilinguals. There, we wondered whether the existence of a cognate in L3 affects its recognition speed in L2, and whether this effect was influenced by the experimental task. In the six experiments reported here we did not use L2-L3 IHs (English-French IHs barely exist). This makes it impossible to find out whether the existence of an IH in L3 affects its recognition as an L2 word. Rather, the question was whether the presence of many monolingual L3 words in an experiment where L2 is the target language makes it more difficult to recognize the L2 reading of L1-L2 IHs, using a pure L2 list as the point of comparison. According to the BIA+ model (Dijkstra & van Heuven, 2002), the mixed L2-L3 list will cause response bindings between these two languages and the responses only. The target-language L2 will be linked to the 'yes'/'go'-response and the non-target language L3 to the 'no'/'no-go'-response in the LDT and the go/no-go task, respectively. As participants do not encounter monolingual L1 words, this model predicts that they will not associate this language with the 'no'/'no-go' response. Hence, BIA+ predicts a null effect for L1L2 IHs in a L2-L3 list. However, as there have been reports of significant inhibition effects in a pure L2 list (see above), it may be more reasonable to expect equal homograph effects in a pure L2 list and a mixed L2-L3 list. An alternative possibility is that participants will not only associate L3 to the negative response but each language that is not the target language (e.g., also L1 when they encounter an L1 word). In that case, participants would generalize the 'no'-response to all non-target languages. Such a view predicts that the higher-frequency L1 reading of the IHs will trigger a negative response and, hence, give rise to the same inhibition effect as in the L1-L2 list. An intermediate position is also possible. This could be the case if some participants act according to the first scenario and others according to the second one (if such strategies can be developed).

The second goal was accomplished by using the experimental design behind the six experiments: an orthogonal combination of the factors Experimental Task and List Composition. Note that this was the same design as in the experiments reported in Chapter 2. Whereas both design factors have been studied in previous research, we are not aware that anyone has made a direct comparison between the effects of list composition in several tasks. If the go/no-go task requires a more complex response process than our variant of the lexical decision task, which is highly likely considering the necessity to withhold a response for negative decisions (see above), larger homograph inhibition effects are expected for this task. The rationale here is that a more complex response process will be more strongly affected by attentional mechanisms (which are slow as well). This will result in a conscious attempt to suppress the irrelevant reading of the IH, which will yield a large inhibition effect.

The third goal was to study the relationship between the magnitude of the homograph inhibition effect and response speed. To that end, we performed more fine-grained statistical analyses than linear mixed effects analyses. Balota and colleagues (e.g., Balota et al., 2008) and others have emphasized (and demonstrated) the importance of looking beyond the mean and study the entire RT distribution. They convincingly showed that this makes it possible to achieve more insight into the underlying cognitive processes of an effect than only making comparisons between the condition means. For instance, Balota et al. (2008) arrived at novel insights into the familiar semantic priming effect. Even though means comparisons invariably show faster responses in the primed condition, a study of the RT distribution indicated stronger priming effects in the region of the longest RTs when auditory targets were degraded and visual primes were clearly perceptible or even masked. This finding showed that, under such conditions, participants rely more heavily on the prime for items that are difficult to identify and, hence, take much processing time. Following Balota and co-workers we used ex-Gaussian and quantile analyses to find out where the effect of IHs is situated in the RT distribution and how it evolves as response speed becomes slower. We also used delta plot analyses, which can be derived from quantile analyses, to find out whether changes in the magnitude of the homograph effect are linear, i.e., constant from quantile to quantile, or whether the increase between two quantiles is discontinuous, i.e., non-linear. This, too, can shed light on the cognitive processes behind the homograph effect.

3.4.1 The effect for L1-L2 homographs in a L2-L3 list

Besides a single exception, a significant inhibition effect was obtained for the L1L2 IHs in all six experiments, both in RTs and in the error rates. The only exception was the effect on error rates in the pure L2 list in the LDT. Clearly, our IHs caused considerable inhibition, both across tasks and across list compositions. This is probably due to the frequency relationship between their L1 and L2 readings, the former being more frequent based on frequency counts and, hence, even more based on subjective frequency (L2 being the weaker language). The finding that the RT effect is even significant in the pure L2 list validates these items as an appropriate set for measuring homograph inhibition.

Note that these homograph effects were not obtained in separate analyses of each experiment but in an omnibus analysis, in which all experiments were analyzed simultaneously. By changing the set of control items in the intercept, such that each set of controls in the orthogonal combination of Experimental Task and List Composition was once part of the intercept, the homograph effect for each experiment could be assessed. Note that this technique does not amount to different analyses (leading to an increase in Type I errors), as changing the intercept involves no change in the model formula. By changing the set of controls in the reference level, one only changes the perspective from which one looks at the data. However, the data are modelled in the same way.

In the RT analysis, the homograph effect was significantly larger in the L1-L2 list than in the pure L2 list and the L2-L3 list, whereas the effect did not differ between the latter two lists. As the interaction between Word Type and List Composition did not engage in a third-order interaction with Experimental task, this pattern obtained for the two tasks. In the error analysis, the homograph effect was significantly larger in the L1-L2 and the L2-L3 lists than in the pure L2 list (recall that no homograph effect on the error rates was found in the latter list). The effect was also significantly larger in the L1-L2 list than in the L2-L3 list.

These findings replicate the typical effect for L1L2 IHs in a L1-L2 list when L2 is the target language (e.g., Dijkstra et al., 1998, 2000): (a) a very large inhibition effect, both in the RT pattern (LDT: 108 ms, go/no-go: 174 ms) and in the error pattern (LDT: 29.60%, go/no-go: 14.90%), and (b) an effect that is significantly larger than in a pure L2 list. However, the crucial question concerns the homograph effect in the L2-L3 list. In the RT analysis, this effect is significantly smaller than the one in the L1-L2 list and does not differ from the inhibition effect in the pure L2 list. In the error analysis, the effect is also smaller than in the L1-L2 condition, but it is larger than in the L2 list. At first sight, this might suggest an accuracy-speed trade-off, i.e., participants in the L2-L3 list make fast responses on the IHs and, hence, make more errors by making 'yes'/'go'-responses to the L1 reading of an IH. However, this seems rather unlikely, considering the relatively small error percentages in the LDT (8.10%) and in the go/no-go task (7.30%). It is more likely that some participants respond too quickly when they occasionally recognize the L1 reading of an IH. If the L1 reading of the IHs had caused a larger response conflict than in the pure L2 list, it seems reasonable that the more sensitive RT measure would have picked this up, with a larger inhibition effect in the L2-L3 list than in the pure L2 list as a result. This was not the case.

Hence, it seems fair to conclude that the inhibition effect for L1L2 IHs is not affected by the presence of monolingual L3 words when L2 is the target language. This finding is compatible with the prediction made by the BIA+ model, even though this seems not to be the case at first sight. During the experiment, the participants learn to associate the positive response ('yes'/'go') to L2 and the negative response ('no'/'no go') to L3. As L1 is not linked to an experimental response, it should not activate any response, and, hence, yield no response conflict and no inhibition effect. However, this prediction is not borne out by the data: a significant homograph inhibition effect was obtained in the L2-L3 list, in both tasks. Note that the same reasoning applies to the pure L2 list. If the bindings between languages and responses were restricted to the languages in the experiment, no homograph inhibition effect could ever be found in a pure L2 list. However, this inhibition effect is significant in the RT analysis for both tasks. Note that significant inhibition in a pure L2 list has already been reported before, even though the effect varies from experiment to experiment and is possibly be due to uncontrolled stimulus properties (see above).

However, the problem in accounting for the pattern of inhibition effects along the lines of BIA+ can easily be solved. In the BIA+ model, a distinction is made between the activation process, which causes the activation of the two IH readings, and the decision process, where the response bindings with different languages become important. In the case of our IHs in the pure L2 and the L2-L3 lists, the dominant L1 reading is likely to become activated before the less frequent and less familiar L2 reading. Hence, participants will classify the item as a non-member of L2, which calls for a negative response. However, when the L2 reading is activated, they will realize that the item is a L2 word after all. This temporary response conflict must be resolved, such that a correct 'yes'/'go'-response can be made. Compare this to the situation where half of the monolingual words are L1 words, i.e., the L1-L2 list. Here, a different account is called for. In this case, the realization that the fast activation of the L1 reading has two effects: (a) the participant initially classifies the word as a non-member of L2 and (b) the strong association between L1 and the negative response, due to the many monolingual L1 words in the list, further strengthens the decision to make a negative response. Hence, when the L2 reading becomes available, it will be much more difficult to resolve the response conflict in favor of a correct 'yes'/'go'-response than in the pure L2 and L2L3 lists. Accordingly, a larger homograph inhibition effect is obtained in the L1-L2 list than in the pure L2 list and the L2-L3 list.

The first conclusion we can draw is that the presence of many monolingual L3 words, which require a negative response (just like the L1 reading of IHs of the L1L2 type), does not make it more difficult to resolve a response conflict when the L1 reading of a IH is activated before its L2 reading. Hence, language-to-response bindings are language-specific, i.e., they do not generalize to other non-target languages. More specifically, when participants continually make 'no' or 'no go' responses to L3 words, they do not infer (either consciously or unconsciously) that any non-L2 language (e.g., L1) also requires a negative response. If that had been the case, the response conflict in the L2-L3 list would have been as large as the conflict in the L1-L2 list (or at least be an effect intermediate the inhibition effects in the pure L2 list and in the L1-L2 list).

3.4.2 The effect of experimental task

As mentioned earlier, our study is probably the first to orthogonally manipulate the factors List Composition and Experimental Task. This makes it possible to investigate whether the effect of list composition is the same across tasks (at least, the tasks used in this study). Our choice of tasks was not accidental. Nieuwenhuis et al. (2003) and Donckers and van Boxtel (2004) reported ERP data indicating that a go/no-go task causes a considerable amount of conflict monitoring. In addition, Dijkstra et al. (2000) found a much larger homograph effect in a go/no-go task than in a language decision task. As there are several suggestions that the response process is quite complex in the go/no-go task, we might find different inhibition effects in this task than in the LDT.

Even though a more complex response process might lead one to expect longer RTs in the go/no-go task, we found little support for this hypothesis (see Table 3.2). Response speed did not differ between the LDT and the go/no-go task in the pure L2 list and in the L1-L2 list ($ts < 1$), and participants made faster (rather than slower) go/no-go responses in the L2-L3 list ($t = -2.06$, $p = .04$). However, even the latter effect is only just below the significance threshold. Stronger effects were obtained in the error analyses (see Table 3.4): whereas the effect in the pure L2 list was not significant ($t = 1.53$, $p = .12$), but significantly fewer errors were made in the go/no-go task in the L1-L2 list ($t = -4.75$, $p < .0001$) and the L2-L3 list ($t = -4.32$, $p < .0001$). The latter effects are highly significant. Even though these are between-participant effects, the data suggest that participants in a go/no-go task are very cautious and attempt to avoid errors. This cautious process of decision-making and responding is compatible with the idea that a go/no-go task involves careful conflict monitoring. The more cautious participants are in responding, the more likely that they will lose time in the resolution of a response conflict, which predicts stronger inhibition effects in the go/no-go task than in the LDT.

This is indeed what we found. The homograph effect was significantly larger in the go/no-go task, but only in the RT analysis (see the interaction term Homograph-Go/NoGo in Table 3.2). As there was no third-order interaction between Word Type, Experimental Task, and List Composition, the larger homograph effects in the go/no-go task did not disturb the pattern in the homograph effects among the three lists in the LDT. Following additive-factors logic (Sternberg, 1998) one might be tempted to conclude that the effect of the go/no-go task affects a different processing level than the level where the response conflict manifests itself. For instance, one might suggest that participants are always more cautious when performing a go/no-go task, and that this increased caution gives rise to a constant increase in the homograph effect, whether the list gives rise to a mild response conflict (pure L2 list and L2-L3 list) or a severe response conflict (L1-L2 list). However, we will not elaborate on such suggestions, as there may have been insufficient power to detect the third-order interaction, and as the rationale behind additive-factors logic has also been called into doubt (Stafford & Gurney, 2011).

The second conclusion that we can draw is that a go/no-go task apparently induces a very cautious decision and response process. The task also yields significantly larger effects of homograph inhibition in the response times, at least compared to our LDT. These larger effects are likely due to the more cautious decision and response process, but it is difficult to identify the locus of this effect. Finally, the increase in the magnitude of the homograph effect in the go/no-go task is constant across our three list compositions: a pure L2 list, a L1-L2 list, and a L2-L3 list.

3.4.3 The localization of the homograph inhibition effect in the RT distribution

Balota and colleagues (e.g., Balota et al., 2008) demonstrated that it is possible to achieve more insight into the cognitive processes underlying an effect by moving beyond tests that involve mean comparisons. More particularly, they showed the added value of analyzing the RT distributions for the several conditions in an experiment. Even though we found strong evidence in favor of inhibition effects in all six experiments (with the exception of the error analysis for the L2 pure list), we followed their suggestions to zoom in on the data. Hence, we addressed a question that has not been raised thus far in this area of research: is the homograph effect that one finds by comparing the means for the IHs and their controls the result of a shift of the entire RT distribution, such that the effect is the same for both fast and slow responses, or is it the result of a subset of RTs (most likely the longest) because a response conflict (and its fingerprint in the data, i.e., inhibition) needs time to build up? As mentioned earlier, the homograph effect in the means comparison may reflect both these effects as well. To address this question, we delved deeper into the RTs and subjected them to three types of analyses: (a) ex-Gaussian analyses, (b) quantile analyses, and (c) delta plot analyses.

Recall that an ex-Gaussian analysis mathematically decomposes the actual RT distribution into two component distributions (which together yield the observed RTs): a Gaussian, i.e., normal, distribution and an exponential distribution. If there is a homograph effect in the analysis of the mu values, i.e., the mean of the Gaussian component, this indicates that the entire RT distribution for the homographs is shifted with respect to the RT distribution for the controls. If there is a homograph effect in the analysis of the tau values, i.e., the mean (and at the same time standard deviation) of the exponential component, this indicates that the homograph effect is situated in the region of the longest RTs. If a homograph effect is obtained for both mu and tau, this means that IHs both shift the entire RT distribution and have a stronger effect on the longest RTs than on the others. If a homograph effect is obtained on sigma, i.e., the standard deviation of the Gaussian component, this indicates that responses to homographs are more variable than responses to their controls.

The homograph effect on mu mimicked the effects in the linear mixed effects analysis of the RTs. In an omnibus analysis on the participants' mu values for the IHs and their controls, we found significant homograph effects for all three lists. The effect was virtually significant in the pure L2 list ($p = .052$), and significant in the L2-L3 and L1-L2 lists. Whereas the former two did not differ from each other, they were both significantly smaller than the effect in the L1-L2 list. As the factor Experimental Task did not interact with the homograph effects, the same pattern obtained for the two tasks. In other words, the homograph effect reflects (at least in part) a shift with respect to the entire RT distribution of the controls. Moreover, this shift is largest for the L1-L2 list.

The effect of tau was highly significant in both the LDT and the go/no-go task, indicating that significantly more long responses were made to IHs than to their controls. This effect was significantly larger in the go/no-go task. As this effect did not interact with the factor List Composition, its effect was the same across the three lists (in both tasks). Hence, the homograph effect on tau was equally large in the L1-L2 list (where the largest inhibition effect was found) as in the other two lists.

The homograph effect on sigma was highly significant in both tasks, but was more significant in the go/no-go task than in the LDT. As was the case for tau, this effect did not interact with the factor List Composition: it did not differ significantly across lists, in neither task. Hence, the variability caused in the RTs for the IHs is not larger when there are many L1 words in the list (L1-L2 list) than when the non-target items are either L3 words (L2-L3 list) or nonwords (pure L2 list).

Taken together, the ex-Gaussian analyses indicate that the homograph effect in the means is (a) in part due to a shift with respect to the entire RT distribution of the controls (the effect on mu) and (b) in part due to the presence of more RTs in the right end of the right-skewed RT distribution (the effect on tau). Whereas mu is affected by the factor List Composition (but not by Experimental Task), tau is affected by the factor Experimental Task (but not by List Composition). Hence, the presence of many L1 words in the L1-L2 list, causes a relatively constant increase in the response conflict and the resulting inhibition effect (with respect to the pure L2 and L2-L3 lists) across the entire RT range, i.e., a distributional shift. In contrast, a more difficult task (go/no-go) does not cause a distributional shift. Rather, its effect manifests itself at the higher end of the RT scale, such that a homograph effect is obtained in the analysis of the tau values. Arguably, a task that requires more attentional demands (because it is performed with considerable caution, i.e., the go/no-go task, see above), causes a larger increase in long RTs on IHs than a less complex task (i.e., the LDT). However, note that a homograph effect on tau was obtained for both tasks. In short, the homograph effects in a pure L2, L1-L2, and L2-L3 list are (a) partially caused by a distributional shift (indexed by mu, a shift that is largest in the L1-L2 list), in both tasks, and (b) partially caused by more long RTs on the IHs than on their controls. The latter effect is even more pronounced in the go/no-go task. Hence, as many psycholinguistic effects, the homograph inhibition effect is a composite effect (Balota et al., 2008).

Note that the effect on sigma may not directly be related to our third question but is nonetheless interesting. It shows that RTs to IHs are much more variable than RTs to their matched L2 controls. This should not come as a surprise. The magnitude of homograph inhibition will both be item-dependent and participant-dependent, such that more variability is expected in a set of IHs than in a set of monolingual words. Interestingly, the homograph effect at this parameter, too, is larger in the go/no-go task. Together with the task effect on the homograph effect in the tau values, this suggests that the decision and response process is more complex (and for that reason probably be performed with more caution, an extra source of variability) than in the LDT.

The results of the quantile analyses can be briefly summarized. In all but one case (the L2-L3 list in the LDT) the interaction between Quantile and Word Type was significant. This indicates that in all five experiments the size of the homograph effect changed as response speed becomes slower. In the t-tests on the data for the individual quantiles, two findings are of theoretical interest: (a) the homograph effect at the quantiles is more often significant in the go/no-go experiments and (b) there is only one list in which the homograph effect is highly significant across all four quantiles: the L1-L2 list, in both tasks (it is only moderately significant in the first quantile for the LDT). The former finding suggests that the inhibition effect appears sooner in the RT range when task-execution is more complex. This is probably because the homograph effect is larger in the go/no-go task than in the LDT task (see above), such that it can be detected earlier. The latter finding suggests that the presence of many L1 words affects the entire RT range. This is compatible with the finding of a stronger effect of the L1-L2 list on the mu parameter.

Finally, the delta plot analyses demonstrate that the homograph effect linearly grows as response speed increases. In other words, there is a constant increase in the size of the homograph effect as the response time increases. At first sight, this seems to contradict the finding that the homograph effect reflects (in part) a distributional shift, which suggests that the effect is constant across the entire RT range. However, the phrase 'in part' is crucial. Recall, that we found a homograph effect on tau in all experiments, i.e., in all three lists in both experimental tasks. As the observed RT distribution is the integration of the Gaussian component (where the effect of mu is situated) and the exponential component (where the effect of tau is situated), the pattern in the delta plots does make sense.

3.4.4 Conclusion

We set out to answer three questions. We are now in a position to formulate an answer to these questions. (1) Does the presence of monolingual L2 and L3 words lead to a larger effect of homograph inhibition on L1L2 IHs, even though none of the IHs occurs in L3? Answer: no, it does not. Linear mixed effects analyses indicate that the largest inhibition effect is found in a L1-L2 list, and that significant but equally-sized inhibition effects are found in a pure L2 list and a L2-L3 list. This means that language-to-response bindings emerge as the result of the specific languages in the experimental list and are not generalized to other languages, more particularly, the negative response is not generalized to all non-target languages. This is compatible with the distinction between activation and decision levels in the BIA+ model and with the importance of the decision level for explaining the magnitude of homograph inhibition. The higher-frequency L1 reading is activated before the L2 reading, which causes a temporary response conflict. If the list contains many L1 words, such that L1 becomes associated with the 'yes'/'go'-response, this further increases the response conflict, which emanates in the form of a large inhibition effect. Smaller effects of homograph inhibition are obtained when no L1 words appear in the list (pure L2 list, L2-L3 list), because the response conflict is not increased by such a language-to-response binding.

(2) Is the effect of list composition on the size of the homograph effect comparable in the go/no-go task, a task with a difficult decision and response process, and in a variant of the lexical decision task? And how do the homograph effects in the two tasks relate to each other? Answer: the pattern of inhibition effects is similar in both tasks (i.e., across the three lists) but the effect in each list is larger in the go/no-go task than in the LDT. This reflects the difficulty in the former task to monitor the response conflict, which is also indexed by the great caution in making go/no-go responses (as reflected in significantly smaller homograph effects in the error rates for the mixed conditions).

(3) How does the homograph effect relate to the RT distribution of its controls? Answer: the evidence shows that the effect is due to both a distributional shift (whose size depends on list composition) and on IHs having a larger number of long RTs in the right end of the RT distribution (whose size depends on the nature of the task). Furthermore, the effect increases

as response time increases (in the L1-L2 list for the LDT and in all lists for the go/no-go task), as indicated by the quantile analyses and the delta plots derived from them. Finally, the latter indicate that the increase in the homograph effect is constant across the RT range.

Even though a lot is known about the visual recognition of IHs, the above insights add further information to this knowledge base. We believe that none of the above three questions have been addressed in previous research. Hence, these findings shed light on the ongoing investigation of bilingual lexical processing.

APPENDIX Interlingual homographs used in this study

Interlingual homographs, control words; their mean Zipf frequencies, and translations of the Dutch meaning of the homograph.

HOMOGRAPHS					ENGLISH CONTROLS		
word	Dutch meaning	length	English Zipf	Dutch Zipf	word	length	English Zipf
boot	boat	4	4.43	4.98	toad	4	3.69
brand	fire	5	4.68	4.65	spoon	5	4.29
breed	wide	5	4.26	3.79	towel	5	3.87
brief	letter	5	4.26	4.87	broad	5	4.22
hoop	hope	4	3.59	5.57	fowl	4	3.43
kin	chin	3	3.54	3.89	row	3	4.75
kind	child	4	5.6	5.52	mute	4	3.16
leek	layman	4	3.56	4.98	nail	4	4.18
lid	member	3	4.16	4.52	jar	3	3.96
loop	course	4	3.82	5.07	girl	4	5.29
rug	back	3	3.57	4.91	ray	3	4.56
rust	peace	4	3.33	4.88	duck	4	4.53
slang	snake	5	3.56	4.33	birch	5	3.43
slap	limp	4	4	3.85	rash	4	3.61
slim	smart	4	3.8	5.05	rude	4	4.27
slot	lock	4	3.78	4.72	snag	4	3.02
toe	closed	3	4.08	5.76	fur	3	4.03
trap	stairway	4	4.22	4.72	dear	4	5.08
wet	law	3	4.76	4.91	hat	3	4.76
wit	white	3	3.67	4.52	jaw	3	3.38
mean:		3.90	4.03	4.77		3.90	4.10
sd:		0.72	0.54	0.53		0.72	0.62
range:		3-5	3.33-5.60	3.79-5.76		3-5	3.02-5.29

CHAPTER 4
How study phase and non-target language proportion affect cross-linguistic interference in homograph processing

4.1 Introduction

It has been proposed that bilinguals differ from monolinguals in at least two respects. First, bilinguals possess considerable knowledge about the lexicon, grammar, pragmatics, and other linguistic properties of a second language next to their mother tongue. Second, as a consequence of dealing with two languages time and again, they have a well-developed ability to control the use of their languages ('cognitive control'). This control is apparent in the sensitivity of their processing strategies and decision processes when handling their languages under different task demands.

Both aspects can be clarified by considering how bilinguals deal with 'false friends' or interlingual homographs (IHs) during reading. False friends are words that share their orthography across languages, but have different meanings. As an example, consider the word form ***smart***. This letter string is a word in both English and Dutch. Although the English and Dutch word forms are etymologically related, the English word nowadays means 'sadness' in Dutch. As a consequence, when bilinguals encounter the word form smart, they must select one of its two meanings, depending on whether in the task context the word must be considered as English or as Dutch. Proper lexical selection entails the application of a form of cognitive control when participants are making decisions tuned to the task at hand (Green, 1998).

In this study, we used IHs to investigate the relationship between lexical processing on the one hand and cognitive control on the other. In particular, we (i) introduced a study phase and (ii) vary the proportion of words from the non-target language in language-specific lexical decision to investigate their effect on the degree of cross-linguistic interference observed in homograph processing.

Before we zoom in on our series of four experiments, we will first consider how the occurrence of IHs in pure or mixed stimulus list contexts affects their processing. Next, we will consider how cognitive control may regulate the selection of the homograph reading that is appropriate in the experimental situation at hand.

4.1.1 Activation of interlingual homographs depends on task, stimulus list context, and lexical characteristics

Studies on false friends (e.g., Lemhoefer, Dijkstra, & Michel, 2004; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008; Van Wijnendaele & Brysbaert, 2002; Von Studnitz & Green, 2002) support the view that in early stages of bilingual word recognition, both orthographic forms of an IH become active. This co-activation entails ***language non-selective lexical access***, i.e., word forms from two languages are activated in parallel. The degree of co-activation has been found to depend on the homographs' relative subjective frequency in the two languages they belong to (e.g., Dijkstra, Timmermans, & Schriefers, 2000).

When the two counterparts of the IH are co-activated, one might expect that recognizing the target reading of a homograph is slowed down due to a conflict between their two meanings across languages. However, it turns out that this inhibition effect depends on the specific demands of the task at hand, the stimulus list context in which the homograph appears, and the homograph's lexical characteristics in the two languages.

First, when tasks require the use of both languages (e.g., language switching, Hernandez, 2009; language decision, Dijkstra et al., 2000), both readings of the IH may be relevant. This results in interference effects relative to one-language control words. In contrast, when a task is language-specific, interference from the other language is usually much smaller or absent (e.g., Dijkstra, Van Jaarsveld, & Brinke, 1998).

Second, stimulus list composition also affects the degree of IH interference observed. Presenting an IH in mixed language context (e.g., in an L2 lexical decision task) increases inhibition effects (e.g., De Groot, Delmaar, & Lupker, 2000; Dijkstra et al., 1998). Relative to this experimental context, an L2 language-specific task involving a monolingual context may result in a reduction of the interference effects and even in null-effects (e.g., De Bot, Cox, Ralston, Schaufeli, & Weltens, 1995; Dijkstra et al., 1998; Gerard & Scarborough, 1989; but see also Von Studnitz & Green, 2002, who obtained an inhibition effect; and Font, 2001, who observed a facilitation effect).

Third, the lexical characteristics of the interlingual homograph in the two languages are important. Interference effects are larger for IHs that have a higher frequency reading in the non-target language (e.g., De Groot, Borgwaldt, Bos, & van den Eijnden, 2002; Dijkstra et al., 1998; also Chapter 3 of the current dissertation).

The effects of these various factors have been interpreted in terms of differences in language activation. According to a ***relative language activation account***, words from different languages may be more or less active (e.g., Dijkstra, 2007; Grosjean, 2000). Task relevance, mixed stimulus list composition, and higher frequency of non-target language readings might contribute to activation of the competing language and to interference effects for IHs. However, testing this proposal led several authors to conclude that some of these factors affect word recognition not in terms of changes in relative language activation, but in terms of decision criteria (e.g., De Groot et al., 2000; Dijkstra et al., 2000; Linck, Hoshino, & Kroll, 2008; Marian & Spivey, 2003; Smits, Martensen, Dijkstra, & Sandra, 2006; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009).

In line with this idea, we will argue in the next section that some of the homograph effects may reflect decision level effects, rather than differences in lexical activation across experimental situations.

4.1.2 Interlingual homograph effects are affected by decision and control, which depend on tasks, context, and lexical characteristics

It has been proposed that to optimize performance in a particular word recognition task, a set of decision criteria governs the procedure for word selection. This so-called ***task schema*** (Green,

1998) relies to different degrees on various sources of available information, depending on their relevance in the experiment. However, the specific characteristics of decision criteria in the task schemas are not completely clear.

One component in the task schema concerns the instructions to participants. In an L2 lexical decision task, participants are standardly asked explicitly to decide whether a presented item is an existing word in L2 or not. However, the task assignment often is not so clear about how to handle upcoming L1 words. For instance, participants may be instructed to respond with a 'yes'-response to L2 words, while no mention is made of the presence of IHs in the experiment. In this case, the non-target reading of the IH is not explicitly connected to a certain response ('yes' or 'no'). This lack of specification may result in uncertainty on the part of the participant, and perhaps even in the absence of competition between the IH readings. As a consequence, it may result in null-effects for IHs (e.g., Kerkhofs, Dijkstra, Chwilla, & De Bruijn, 2006; Smits et al., 2006).

In line with this notion, Smits et al. (2008) proposed that participants manage the response conflict between the two homograph readings at a decision level. In a monolingual context, it is safe to ignore activated representations in the non-target language, because responses are limited to words from one language only. In contrast, in a mixed language context, words from both languages have to be considered before responding.

Dijkstra and colleagues (1998, 2000) had Dutch-English bilinguals perform an English (L2) LDT including IH and Dutch control words, but no purely Dutch words. In this task context, response times to IH and controls did not differ. In contrast, when purely Dutch words were included in the experiment, strong inhibition effects arose for IH relative to controls. In Dijkstra et al. (1998), a generalized variant of the LDT even resulted in a facilitation effect for IHs. Note that in this situation, bilinguals can rely on both readings of the IH. Thus, the IH readings do not cause any response conflict, but in fact support the same 'yes'-decision. Van Heuven et al. (2008) considered the contribution of identification and decision to participants' performance in a study involving L2 and generalized lexical decision tasks. Slower responses to IHs were observed in an L2-specific lexical decision task, reflecting response-based language conflicts, but not in generalized lexical decision, which involved stimulus-based conflicts only. In other words, the conflict was present when participants had to choose between languages, not between the readings. These observations were supported by fMRI data. Thus, behavioral and fMRI data indicate that both task demands and language intermixing can modulate decision criteria.

A related theoretical issue is whether the decision criteria for IH recognition are adapted or shift in the course of an experiment. Dijkstra, De Bruijn, Schriefers, & Ten Brinke (2000); Elston-Güttler, Gunter, & Kotz (2005); Von Studnitz & Green (2002) investigated this issue for a number of experimental situations. Dijkstra, De Bruijn, et al., (2000) presented IHs in an English lexical decision task consisting of two parts. Participants were explicitly instructed to give a 'yes'-response to English words and homographs, and a 'no'-response to Dutch words and non-words. In the first part of the study, no Dutch words occurred in the stimulus list, whereas in the second part, Dutch words were introduced. No significant RT difference with control words was found for IHs in the first part, but a significant inhibition effect arose in the second part, immediately after the first Dutch word was introduced. The authors suggested that although the difference in stimulus list composition affects word recognition, performance changes did not necessarily arise as the result of changes in relative language activation.

Probably, the participants (also) adapted their decision criteria or task schema when Dutch words were introduced. A study of Elston-Güttler et al., (2005) included IHs in a lexical decision task. The IHs served as primes (e.g., 'gift' that means poison in German) at the end of a sentence and were followed by target words that corresponded to the English translation of the German meaning. Before the task began, the researchers showed a silent movie to the participants with an accompanying narrative in either L2 or L1. On the basis of behavioral and ERP data, they noticed that participants who watched the movie in L1 responded faster to semantically related target words (e.g., 'poison'), but only in the first part of the task, the effect in the second part was not significant. This observation implies that gradually raising decision criteria were present, suggesting that participants tried to diminish the non-target reading effect by zooming in on their L2, i.e. engaging in a more language selective mode of processing. Von Studnitz and Green (2002) compared bilingual performance in two LDTs, either in a pure L2 context or in a mixed context including words from both languages. They replicated the finding that including non-target words (Dutch in their case) in the stimulus list increased the degree of homograph interference. At the same time, they showed that the interference effect decreased in the course of the experiment. In a bilingual context, the interference effect was larger in the first part of the experiment. Next, they replicated the experiments, but now informed each participant at the start or after the first part of the experiment that ambiguous words were present that exist in both languages. A reduced interference effect was found only in the first situation, where participants could adapt to the stimulus list composition from the very beginning of the experiment.

Based on these above mentioned studies, we conclude that both relative activation and decision criteria are required in an account of the available data. Similar assumptions are made by recent models of bilingual word recognition that equip a word identification account with a decision mechanism. For instance, the Bilingual Interactive Activation Plus model (BIA+, Dijkstra & van Heuven, 2002) presents a two-leveled account, assuming a word identification level tapping into a shared multilingual lexicon, and a decision level regulating the cognitive operations required for recognition and response in a particular task context. The BIA+ model proposes that decision criteria can modulate word selection. These notions have also been inspired by the Inhibitory Control model by Green (1998) for word production. In this model, task schemas may influence the processing of IHs and result in different responses depending on the schema specification.

4.1.3 The current study and predictions

In this study, we zoomed in on lexical retrieval mechanisms by introducing a study phase before the main task and by manipulating the proportion of the non-target language words in an experiment. Focus of the research was whether a preceding study phase and varying proportions of the non-target language result in different magnitudes of the IH inhibition effect. Knowing this should lead to new insights in the mechanisms involved in lexical activation of IHs.

To investigate the effect of ***study phase***, we tested how bilinguals recognize IHs in a purely English context. In Experiment 1, participants performed an English LDT in a standard fashion. In Experiment 2, a study phase was added, during which participants had to memorize the English meanings of homographs in sentences before they performed the English LDT. The comparison of the experimental results should reveal whether the homograph effect is reduced in a situation with a preceding study phase.

It was hypothesized that repeating the target meaning via the study phase puts more focus on the target language in the task at hand. Because participants learn to link IH to one language (here English) and pre-activate its relevant meaning, this should reduce the effect of the other language (here Dutch) in a later LDT. Alternatively, if no effect of study phase is observed later, this suggests that study cannot reduce the influence of the non-target language in the experiment when participants focus on the target language reading anyway. This would demonstrate that pre-activation of the L2 is not enough to overrule the dominant L1 reading of the IH. Such a result is unexpected by any model (also BIA+) that assumes context-independent pre-activation of lexical items can boost the resting level activation of word candidates, leading to faster recognition at a later stage (repetition priming).

To investigate how the ***proportion of non-target language*** items in an experiment affects the processing of interlingual homographs, the English LDT of Experiment 3 had a similar design as that of Experiment 2. However, 10% of Dutch items was added to the stimulus list.

In earlier LDT studies, either pure (only English) or mixed (50% Dutch and 50% English) lists were used, resulting in an all-or-none contrast. In the present study, we added the intermediary option of 10% Dutch items. We envisioned two possible outcomes, based on different underlying mechanisms. On the one hand, the proportion of 10% non-target language items might not be large enough to change the degree of relative language activation of the IHs or the decision criteria applied in the task. Thus, the observed IH inhibition effects would be equal irrespective of whether the percentage of L2 words is 0% or 10%. Alternatively, just 10% of Dutch non-target language items might be enough to activate Dutch more as a language competing for response. Whatever the outcome of the proportion manipulation may be, it will inform us about the limitations of language activation effects and the associated decision criteria that apply globally (across the experiment as a whole) or locally (only on consecutive trials)

4.2 Experiment 1: English LDT in pure English context

4.2.1 Method

Participants

Twenty-four Linguistics students of the University of Antwerp were tested (18 women, 6 men; range: 19-23 years of age; mean age: 20.3). Twenty-two were right-handed and two left-handed. All of them had normal or corrected-to-normal vision. All were Dutch native speakers with English as their dominant foreign language. The participants had studied English as a foreign language at secondary school for at least six years and used English regularly (mean experience with English was 8.8 years). Participation in the experiment was voluntary.

Materials and Design

The critical items for LDT were 20 Dutch-English (L1-L2) orthographically identical homographs. Originally, we selected words from the CELEX database (Baayen et al., 1995). Later we compared the frequencies with the SUBTLEX databases, once the latter became accessible (Keuleers et al., 2010; van Heuven et al., 2014)[28]. Based on the SUBTLEX databases, the frequency of the Dutch reading for Dutch-English homographs was still higher than for the English reading (see below).

[28] Note, we used the formula log10(frequency per million*1000) from van Heuven et al., 2014, to calculate Zipf values for Dutch data. It made it easier to compare different data sets.

We asked 10 randomly selected students from the same population as our participants to decide for each homograph (a) how different its meaning in the two languages was, (b) how similar its pronunciation in these languages was, and (c) how familiar the word was in the participants' L2. Based on these ratings (1 = "none", 5 = "very strong", we selected the 20 highest rated Dutch-English homographs from the set of identical words that had a higher frequency in Dutch.

All selected homographs had a higher frequency in Dutch (mean: 4.77, SD = 0.53, range: 3.79 – 5.76) than in English (mean: 4.03, SD = 0.54, range: 3.33 – 5.60). For each homograph, an English monolingual control word was selected with the same number of letters (mean: 3.90 for homographs and controls) and a matched frequency (mean: 4.10, SD = 0.62, range: 3.02 – 5.29). A t-test indicated that homographs and controls were accurately matched on frequency ($t = 0.19$; $p > 0.1$). In addition, we added 10 English filler words from the same frequency range as the other English words in the experiment. Additionally, we generated 50 English-like non-words (derived from different English words) using the Wuggy program (Keuleers & Brysbaert, 2010). All non-words were matched on number of letters with English words and homographs.

Procedure
The experiment was run on DELL computers "Optiplex 380" connected to a 15" DELL monitor in a dimly illuminated room. The stimuli were presented in the center of the screen and reaction times were registered with the DMDX program (Forster & Forster, 2003). Logitech 'Wingman precision' game controllers were used for responding and to initiate the experimental sequences (using upper buttons and START button). The task was introduced by a written instruction in English. Participants pressed one button as quickly and as accurately as possible when the presented letter string was a word in English, and another button if it was not. The response buttons were reversed for left-handed participants. A set of 12 practice trials, different from the test trials, preceded the main task. The experiment proper consisted of four blocks and took about 15 minutes. Each trial was presented after a fixation (+) sign that appeared in the middle of the screen for 500 ms. The item stayed in view until a response had been made or until a time out of 2.500 ms had passed. The next trial was initiated 500 ms after the response or time out.

4.2.2 Results and discussion

In all experiments, non-words were not analyzed. During data cleaning, two control words (i.e. ***birch, fowl***) and their paired homographs (i.e. ***slang, hoop***) were removed, because they resulted in more than 50% error responses (n=80, i.e. 10%). In addition, all error responses were excluded from the RT analyses (n=46, i.e. 5.6%). Finally, all responses outside the RT range between 300 ms and 1500 ms were removed from both the RT and error analyses (n=13, i.e. 1.6%). This left us with 769 RTs for the homographs and their controls (i.e. 92.2% of the 828 data points remaining after removing the two homograph-control pairs).

Next, we applied (generalized) linear mixed effect models to analyze the effects of homographs vs controls. We first normalized the response latencies by means of an inverse transformation (-1000/RT). We then ran analyses using the lmerTest package (Kuznetsova et al., 2016) in the R statistical software package (R Core Team, 2014) with lmer() function for RTs and glmer() function for Errors (see Baayen, Davidson, & Bates, 2008; Baayen & Milin, 2010, for a review on linear mixed effects analyses). Initially, we analyzed the various experiments separately, and then looked for relevant interactions between them. We started from the simplest regression model including random predictors for participants and items. By means of likelihood ratio tests (α=.05) we then tested whether including each new fixed parameter improved the model; if this was not the case, the parameter was not included in the model[29]. Because the first two experiments were identical in terms of task and homographs, and differed only in study phase (absent vs present), their results will be analyzed and reported together.

Furthermore, we ran post-hoc analyses involving cumulative distributions, because we were interested in the dynamics of the homograph inhibition (e.g., Ratcliff, 1979; Roelofs, 2008). In quantile analyses, we ordered all RTs for each participant from fastest to slowest within each condition and divided them in quartiles (25% quantiles). We then calculated the mean latencies for each quartile separately for critical and control items. Plotting the Quantiles across participants reveals how the RT distribution varies across conditions (for a detailed explanation see Balota et al., 2008; Bub, Masson, & Lalonde, 2006). More information came from Analyses of Variance considering Quantile and Type (control vs homograph) as within-participants and within-items factors. We used a multivariate approach with lm() and Anova() from the car package (Fox et al., 2014) in R, applying the Greenhouse-Geisser correction for potential violations of sphericity. We also applied a paired t-test with Bonferroni correction to consider in detail every Quantile pair across participants. In this way, we established which distributional pairs were significantly different across the cumulative quartile distributions.

4.3 Experiment 2: English LDT in pure English context (+study)

4.3.1 Method

Participants
Twenty-four students from the same population as in Experiment 1 were tested (12 women, 12 men; range: 19-25 years of age, mean age: 21.9 years). Twenty of them were right-handed and four left-handed. They had the same experience with English (mean experience: 10.1 years).

Materials and Design
The same stimulus items (Dutch-English homographs and their pure English controls) were administered as in Experiment 1 for English LDT. To be included in the study phase before the LDT, 40 sentences using homograph and controls were created. The sentences were checked (i) to be as short as possible, (ii) to explain the critical word, and (iii) to use the critical word as the final item. We added also short one/two word(s) clarifications in parentheses to explain the critical word, e.g., 'Keep your answer BRIEF (short)'. For use in the test phase, a distractor was chosen for every critical word. It had to be of the same word type and form as the explaining word(s) in parentheses, e.g., BRIEF (short/slow).

Procedure
The same procedure for English LDT was used as in Experiment 1. However, in this experiment, the LDT was preceded by the study phase involving the critical words (homographs and controls). The study phase consisted of two parts. In the first part, actual study, sentences describing the critical word appeared for 2.500 ms after a fixation (+) sign in the middle of the screen. The last word in every sentence (homograph or control) was printed in capitals.

[29] Per experiment: lmer (RTinverted ~ TYPE + previousRT + (1|PP) + (1|ITEM)
Interaction: lmer (RTinverted ~ TYPE*EXPERIMENT + previousRT + (1|PP) + (1|ITEM)

Participants received an instruction in English to read the sentences on the screen and memorize the meaning of the word presented in capitals by using the meaning of the sentence and the clarification of the word in the parentheses. All 40 sentences were pseudo-randomized and presented three times. In the second part, participants were tested, whether they memorized the English meaning of the homograph. Here on each trial a critical word (homograph or control) was presented for 2.500 ms in lowercase letters in the middle of the screen. On the line below the word, the clarification from the parenthesis was presented either on the right or on the left half of the screen. Simultaneously, on the other half of the screen a distractor was presented. Participants were instructed in English to choose the correct meaning of the presented word as quickly as possible. Only participants who made less than 5 errors (10% or less) in the test part continued the experiments and carried out the English LDT.

4.3.2 Results and discussion

We cleaned the data for Experiment 2 by removing two control words (i.e. ***birch, fowl***) and their matched homographs (i.e. ***slang, hoop***) that resulted in more than 50% error responses (n=80, i.e. 10%). In addition, all error responses were excluded from the RT analyses (n=25, i.e., 2.9%). Finally, all responses outside the RT range between 300 ms and 1500 ms were removed from both the RT and error analyses (n=27, i.e., 3.1%). This left us with 812 RTs for the homographs and their controls (amounting to 94.0% of the 864 remaining data points after removing two homograph-control pairs).

The final model for the RT analysis included Word Type and Experiment as fixed-effect predictors. As Previous RT did not improve the model, it was removed. Mean response times (RT) and errors (ERR) for IHs are presented in Table 4.1. Experiment 1 replicated the sometimes observed homograph inhibition effect for RTs ($\beta = 0.05$, $t=2.34$, $p<0.05$). Thus, in a purely English context, responses to Dutch-English homographs were significantly slower than to matched English controls (678 vs. 647ms). We used the ***relevel*** function to zoom in on the main homograph effect (Word Type) in the Experiment 2. The effect was also significant ($\beta = 0.04$, $t=2.32$, $p<0.05$) indicating the presence of homograph inhibition: Participants responded slower to the homographs than to English controls (688 ms vs. 662 ms).

In the error analysis for Experiment 1, Word Type was marginally significant ($\beta = 0.66$, $z=1.94$, $p=0.05$), indicating that participants made significantly more errors to Dutch-English homographs than to their English controls (7.0% vs 4.1%). The effect of errors was not significant for Experiment 2 ($\beta = 0.26$, $z=0.63$, $p>0.1$). Participants did not make significantly more errors for homographs than for English controls (3.2% vs 2.5%). However, there was no interaction between Word Type and Experiment (see Table 4.2).

Quantile analyses for Experiments 1 and 2 (see Figure 4.1) revealed no global differences ($p>0.05$) in distributions and no difference per quartile (all $p> 0.1$).

Although more errors were made on homographs preceded by a study phase (2.9%) relative to no study phase (0.7%), the effect was not significant. The RTs and Quantile analyses show also no effect of study phase.

Table 4.1.
Mean Reaction Times (RT) and Error Percentages (ERR) for interlingual homographs and their controls in the four experiments of this study

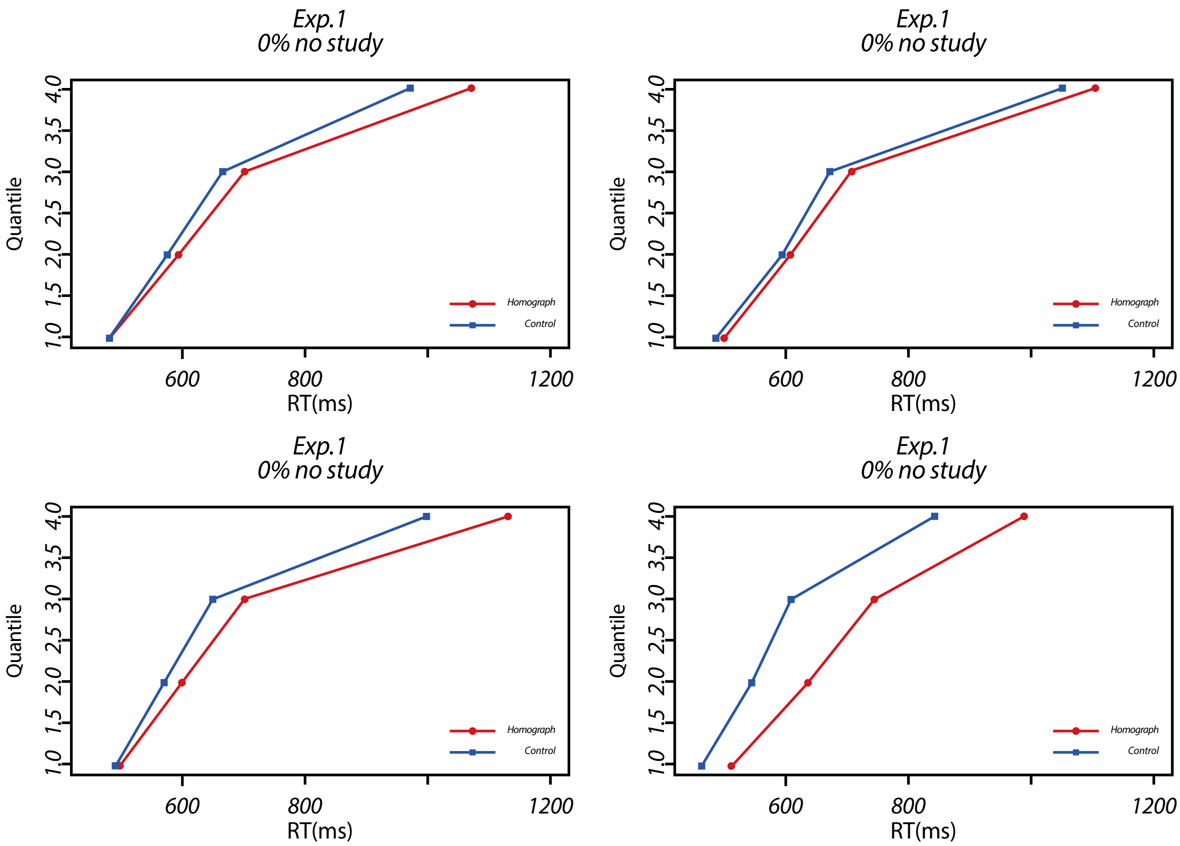
	Exp. 1 0% no study		Exp. 2 0% + study		Exp. 3 10% + study		Exp. 4 50% no study	
Word Type	RT	ERR	RT	ERR	RT	ERR	RT	ERR
English-Dutch	678	7.0	688	3.2	685	2.8	702	33.1
English-French	647	4.1	662	2.5	636	0.5	594	3.5
go/no-go	31*	2.9*	26*	0.7	49***	2.3*	108***	29.6***

Note: * $p \leq .05$ ***$p \leq 001$

Table 4.2.
Estimates for the fixed-effect predictors Word Type and Experiment in the RT analysis of the Dutch-English homographs (Experiment 1: no study phase; Experiment 2: study phase).

	Estimate β	SE β	df	t	p
Intercept	-1.6110	0.04515	67.20	-35.67	< 0.001
Word Type (Exp 1)	0.0497	0.02125	1518.0	2.34	<0.05
Word Type x Experiment	-0.0023	0.02961	1518.0	-0.08	>0.1

Figure 4.1
Quantile plot for experiments of the current study



4.4 Experiment 3: English LDT with 10% Dutch words and study phase

4.4.1 Method

Participants

Twenty-four students from the same population as the participants in the previous experiments were tested (16 women, 8 men; age range: 19-26, mean age: 21.8 years). Twenty-one of them were right-handed and 3 left-handed. They all had normal or corrected-to-normal vision. They had similar experience with English as the participants in the previous experiments. Their mean experience with English was 9.5 years.

Materials and Design

The same items and sentences were used as in Experiment 2. However, 10 English fillers and 10 English-like non-words in the LDT were replaced with Dutch words and Dutch-like non-words. In this way, at least 10% of the items were associated with the Dutch lexicon and in theory could activate Dutch in this experiment.

Procedure

The same procedure from Experiment 2 was used. Participants started with a training phase. Only participants who scored less than 5 errors (10%) in the test phase took part in the LDT.

4.4.2 Results and discussion

During data cleaning, two control words (i.e. birch, fowl) and their matched homographs (i.e. slang, hoop) were removed because they had more than 50% error responses ($n=80$, i.e. 10%). In addition, all error responses were excluded from the RT analyses ($n=14$, i.e. 1.6%). Finally, all responses outside the RT range between 300 ms and 1500 ms were removed from both RT and error analyses ($n=13$, i.e. 1.5%). This left us with 837 RTs for the homographs and their controls (i.e. 96.9% of the 864 remaining data points after removing two homograph-control pairs).

Mean RTs and error rates are presented in Table 4.1. The RT difference between homographs and matched controls was statistically significant ($\beta = 0.09$, $t=4.53$, $p<0.001$). The same held for error rate differences ($\beta = 1.89$, $t= 2.42$, $p<0.05$). However, there was no difference in the size of the homograph effect between Experiments 2 and 3 (see Table 4.3) in terms of RTs ($\beta = -0.04$, $t=-1.52$, $p>0.1$). There was only a weak trend towards a difference in errors ($\beta = 1.62$, $z=1.83$, $p=0.07$). This shows, that overall, participants displayed similar RT effects for the homographs in the two experiments, but there was a weak tendency to make more errors in the experiment with 10% Dutch items (0.7% vs 2.3%).

Quantile analyses showed no global difference ($p>0.1$), but there were significant differences comparing pairs of quantiles separately (see Figure 4.1). The analyses pointed to differences in the second quartile: $p<0.05$; q3: $p<0.01$. This finding indicates that participants were slightly more distracted by the presence of purely Dutch words in Experiment 3, resulting in somewhat slower responses.

Table 4.3.
Estimates for the fixed-effect predictors in the RT analysis of Dutch-English homographs and controls in Experiment 2 (purely English context with study) and Experiment 3 (with study phase and 10% Dutch items).

	Estimate β	SE β	df	t	p
Intercept	-1.6372	0.04304	67.30	-38.04	< 0.001
Word Type (Exp 3)	0.0884	0.01985	1583.0	4.45	< 0.001
Word Type x Experiment	-0.0431	0.02829	1584.0	-1.52	>0.1

4.5 Experiment 4: English LDT in English-Dutch context (50%)

4.5.1 Method

Participants
Twenty-four students from the same population as before were tested (22 women, 2 men; age range: 19-24 years of age, mean age: 22.3 years). Twenty-three of them were right-handed and one left-handed. They had similar experience with English as participants in the previous experiments (mean: 9.7 years).

Materials and Design
In this experiment, the same Dutch-English homographs and their controls as in the previous experiments were included. To create the 50% Dutch-English condition, we replaced English-like non-words with 50 Dutch words (requiring a 'no'-response) that were approximately matched on an item-by-item basis with all English words in number of letters and word frequency.

Procedure
The same procedure was used as in Experiment 1.

4.5.2 Results and discussion

The same type of data cleaning was used as before. Two control words (i.e. ***birch, fowl***) and their matched homographs (i.e. ***slang, hoop***) were removed, because they led to more than 50% error responses (n=80, i.e. 10%). In addition, all error responses were excluded from the RT analyses (n=156, i.e. 18%). Finally, all responses outside the RT range between 300 ms and 1500 ms were removed from both RT and error analyses (n=10, i.e. 1.2%). This left us with 698 RTs for the homographs and their controls (i.e., 80.8% of the 864 remaining data points after removing two homograph-control pairs).

Mean RTs and ERRs are presented in the Table 4.1. In the 50% condition of the Dutch-English context, a highly significant effect of Word Type was found for RTs ($\beta = 0.24$, $t=10.93$, $p<0.001$) and for error rates ($\beta = 3.07$, $z=7.01$, $p<0.001$). Thus, because words from the non-target reading of the homographs (Dutch) were linked to the 'no'-response, a highly significant homograph effect arose. The comparison with Experiment 3, including only 10% Dutch items (Table 4.4) showed that the homograph effect was significantly stronger in the experiment where the proportion of the Dutch items was high ($\beta = 0.20$, $t=6.26$, $p<0.001$). In addition, the homograph effect was significantly stronger in 50% condition than in the pure English condition (Experiment 1) on the RTs (β = -0.15, t=5.25, p<0.001). For the error rates, the interaction between Word Type and Experiment was significant in Experiment 1 (β = 2.39, z=5.54, p<0.001), but not in Experiment 3 (β = -1.08, z=-1.31, p>0.1). In the case of the glmer analysis, we did not find any difference between the experiments. However, participants made only 2.3% errors in Experiment 3 and 29.6% in Experiment 4 (see Table 4.1).

We therefore decided to employ a non-parametric test to consider this issue in detail, to avoid issues with the normality and homogeneity of the data. Our within-subject design allowed to calculate the individual causal effects per experiment (***delta y***). Delta Y (ΔY) is the difference of the number of mistakes made for homographs and for controls. For every participant, the ΔYi was calculated, where i = is the participant's order number. In this way, we can check whether the ΔY values of one group (Experiment 3) are different from the ΔY in another group (Experiment 4). The mean ΔY for Experiment 3 is 0.42 and for Experiment 4 is 5.3. To check whether the difference between the delta values is significant, we used a randomisation test. For the calculation of this test, we used the outcomes of ΔY for and Experiment as covariate for . We used 50,000 runs where stayed constant and was randomly distributed. The resulting comparison showed a strong influence of the factor Experiment on the error rates for homographs (p<0.00001). In other words, participants made significantly more errors on homographs in the condition where proportion of the dominant non-target words was high (Experiment 4).

The quantile analysis was significant with respect to the whole distribution (p<0.01) and for each quantile (Q1: p<0.05; Q2-Q4: p<0.001), indicating that homograph inhibition was evenly present across the distributional line (see Figure 4.1).

Table 4.4.
Estimates for the fixed-effect predictors in the RT analysis of the Dutch-English homographs in (Experiment 4 (with English-Dutch context) and Experiment 3 (with 10% Dutch items).

	Estimate β	SE β	df	t	p
Intercept	-1.7755	0.05133	55.80	-34.59	< 0.001
Word Type (Exp 4)	0.2419	0.02213	1474.6	10.93	< 0.001
Word Type x Experiment	-0.1551	0.02954	1470.7	-5.25	< 0.001

4.6 General Discussion

In this study, we investigated how contextual differences in terms of study phase and proportion of non-target language items in an experiment affect the recognition of IHs.

4.6.1 Study phase

We first compared the performance of Dutch-English bilinguals in two English LDTs without and with study phase in a purely English context (Experiments 1 and 2). Participants received an instruction to memorize English (target) meanings of homographs and controls, and their recollection of the meaning was tested as well. Only participants who made less than 5 errors (10%) were allowed to take part in the subsequent LDT.

Surprisingly, no effects of including a study phase in the experiment were found. Models like the BIA+ model would have expected a facilitatory effect on the recognition of the correct reading of the IH under the assumption that repetition of a lexical item may temporarily increase the resting level activation for this item relative to other items. Our results, on first sight, also contrast with the empirical findings of Poort, Warren, & Rodd (2016). These researchers found that IH processing in a Dutch sentence slowed down later processing of the same IH in an English sentence, even after a break involving a different task (a digit span task). Their results suggest that priming a homograph with a Dutch meaning indeed makes it harder to select the English meaning in an English LDT later. Thus, the difference between their and our results indicates that priming with the English word meaning in itself does not suffice to reduce the interference of the Dutch reading. Possibly, the dissociation between the two studies could again be explained in terms of an episodic component, here involving mediated priming, if it is assumed that episodic memory traces are more accessible in a native language than in a foreign language. This interpretation would explain that the homograph effect can be increased by repetition priming (or study phase) in the native language (which is one of the non-target languages in the LDT), but cannot be decreased by L2 primes (or study) inserted before the LDT is performed.

In fact, it is indeed possible that the effects during the study phase were more of an episodic than a declarative nature. Seen from this angle, the absence of an effect of the study phase is in line with a semantic priming study by Van Abbenyen (2015). In her study, participants memorized pairs of words, after which they were tested in an episodic recognition task (ERT) and an LDT. Participants showed clear effects in the ERT, where they had to remember whether the words appeared in the training phase. However, no effect arose in the LDT, where they had to decide if items were existing words. It was concluded that episodic traces can be used in relevant tasks (ERT), but play no role in irrelevant tasks (LDT). In our experiment we tested whether participants remembered the association made between the homograph and its meaning. Even participants who made almost no errors showed no effect of the study phase in the LDT.

We can find further support for this reasoning in priming studies that report that episodic effects are very sensitive to the exact experimental manipulations. For example, Oliphant (1983) designed three experimental situations involving a ***no-repetition*** and two ***repetition*** conditions. In the first condition, lexical decisions to the same words had to be made on two separate occasions. In the second condition, several words from the instruction texts that participants had to read aloud were included in a later lexical decision task. Interestingly, a facilitatory repetition effect was observed only in the first condition (entailing two subsequent lexical decisions). Oliphant concluded that the occurrence of the repetition effect depended on the subject's awareness in the different conditions. Including a study phase in our experiments did not lead to a reduced inhibition effect (although the English reading was previously activated in the study phase). Apparently, if participants do not realize that they would benefit from information in the study phase in the LDT, they will not use episodic traces in this LDT.[30]

One might argue that prime words leave residually activated lexical representations, and therefore would always facilitate the recognition of later re-occurring items. However, Forster

[30] We never mentioned that the words from the study phase would be / were used in the LDT.

& Davis (1984) deny this outcome on the basis of masked priming studies. They demonstrated that the typically observed frequency effect attenuation on repetition priming disappears if participants are not aware of the prime. Also, when the word is presented again, the experimental conditions must be as close as possible to those of its first presentation to obtain repetition effects. As this was not the case in our study, we assume that episodic traces were created in the study phase that were not accessed in the LDT.

4.6.2 Proportion of non-target language items

As the second issue in our study, we investigated whether different proportions of the non-target language items in an experiment affect the size of the IH inhibition effect. First, the comparison of cumulative distributions across the experiments shows that inhibition effects appeared rather late on the distributional scale. Inhibition was observed across the whole range of the distribution only in the experiment including 50% of the non-target items.

Furthermore, although there were some differences in the mean scores of the homograph effect in the 0%, 10%, and 50% conditions, there were no interactions between the homograph effects and experiments involving 0% or 10% non-target language items. There was no increased homograph inhibition effect in the experiment with a low proportion (10%) of the dominant non-target language relative to the experiment with 0%. An increase in homograph inhibition was only found in the high proportion condition (50%). Thus, including 10% items from the non-target language in the experiment was not enough to persistently activate the non-target reading of the homograph or change the decision process.

To explain why effects were strong and found across the whole distribution in the 50% condition but not in the 10% condition, one might resort to the assumption of the BIA+ model (Dijkstra & van Heuven, 2002) that not only lexical characteristics and language activation influence homograph recognition, but also task demands and stimulus list context. Depending on the characteristics of the stimulus list context, the response criteria applied to word recognition might be adapted in the course of an experiment. Even if participants have certain expectations and are focused on a target language, they might flexibly (possibly even trial by trial) adapt their response criteria based on encountered item characteristics and stimulus list context. As a result, a low proportion of the non-target language might have only a transient and local effect on the setting of decision parameters regulating the link between lexical activation in a particular language and the associated response. In order to bring about a more persistent change in decision parameter settings, a more frequent encounter with non-target language items across the experimental session would be required.

To investigate homograph prominence (in terms of their density across trials) in more depth, we calculated the distance (number of items) between each IH and previous L1 word. Note that in Dijkstra, De Bruijn, et al. (2000), discussed in the Introduction, inhibition arose in the second part of the experiment as soon as the first Dutch word was introduced. In other words, the dominant L1 was directly activated when the first L1 monolingual word appeared.

We conducted a post-hoc analysis including distance between the L1 monolingual word and the homograph as a predictor for the inhibitory effect. Obviously, the density of the L1 words was much lower in the 10% condition (mean = 7.9; sd = 6.2; range = 1:34) than in the 50% condition (mean = 1.8; sd = 1.0; range = 1:6). In both experiments, there was no influence of L1 distance on the number of errors made on the homographs (β = 0.14, z=0.86, $p>0.1$ for 10%; and β = -0.08, z=-0.25, $p>0.1$ for 50%). However, in the RTs there was a significant effect only in the 50% condition (β = -1.08, t=-1.31, $p>0.1$), whereas the 10% condition showed no influence of the L1 (β = -0.04, t=-2.12, $p<0.05$). In other words, if the L1 monolingual word was presented in a trial closer to the homograph item, the homograph effect was stronger, but in the 10% condition there were not enough L1 monolingual words in the context to activate the Dutch reading to a larger extent (than in the 0% condition). These tests provide even more evidence that the ratio of target language / non-target language items was apparently too low in the 10% condition to change decision parameters to a sufficient extent or for a longer period of time (no effect over the whole experiment). As a direct consequence of the low density of pure L1 words in the list, the distance between the last L1 word and IH was too large to make participants doubt more about the L1 reading of the IH.

4.6.3 Conclusion

To wrap up, our set of four experiments indicate that decision criteria applied during the trilingual word recognition process do not only depend on the linguistic characteristics of presented IHs (e.g., if the non-target L1 reading is more dominant), but also on the non-target items present in the stimulus list context. The proportion of non-target language items associated with the 'no'-response, must apparently be high enough to create a lasting activation or decision effect in comparison to a monolingual list situation. Furthermore, temporarily and perhaps episodically generated participant expectations and attentional focus directed during a study phase do not per se alter more global decision criteria set in a later LDT.

APPENDIX A Interlingual homographs used in this study

Interlingual homographs, control words; their mean Zipf frequencies, and translations of the Dutch meaning of the homograph.

HOMOGRAPHS					ENGLISH CONTROLS		
word	Dutch meaning	length	English Zipf	Dutch Zipf	word	length	English Zipf
boot	boat	4	4.43	4.98	toad	4	3.69
brand	fire	5	4.68	4.65	spoon	5	4.29
breed	wide	5	4.26	3.79	towel	5	3.87
brief	letter	5	4.26	4.87	broad	5	4.22
hoop	hope	4	3.59	5.57	fowl	4	3.43
kin	chin	3	3.54	3.89	row	3	4.75
kind	child	4	5.6	5.52	mute	4	3.16
leek	layman	4	3.56	4.98	nail	4	4.18
lid	member	3	4.16	4.52	jar	3	3.96
loop	course	4	3.82	5.07	girl	4	5.29
rug	back	3	3.57	4.91	ray	3	4.56
rust	peace	4	3.33	4.88	duck	4	4.53
slang	snake	5	3.56	4.33	birch	5	3.43
slap	limp	4	4	3.85	rash	4	3.61
slim	smart	4	3.8	5.05	rude	4	4.27
slot	lock	4	3.78	4.72	snag	4	3.02
toe	closed	3	4.08	5.76	fur	3	4.03
trap	stairway	4	4.22	4.72	dear	4	5.08
wet	law	3	4.76	4.91	hat	3	4.76
wit	white	3	3.67	4.52	jaw	3	3.38
mean:		3.90	4.03	4.77		3.90	4.10
sd:		0.72	0.54	0.53		0.72	0.62
range:		3-5	3.33-5.60	3.79-5.76		3-5	3.02-5.29

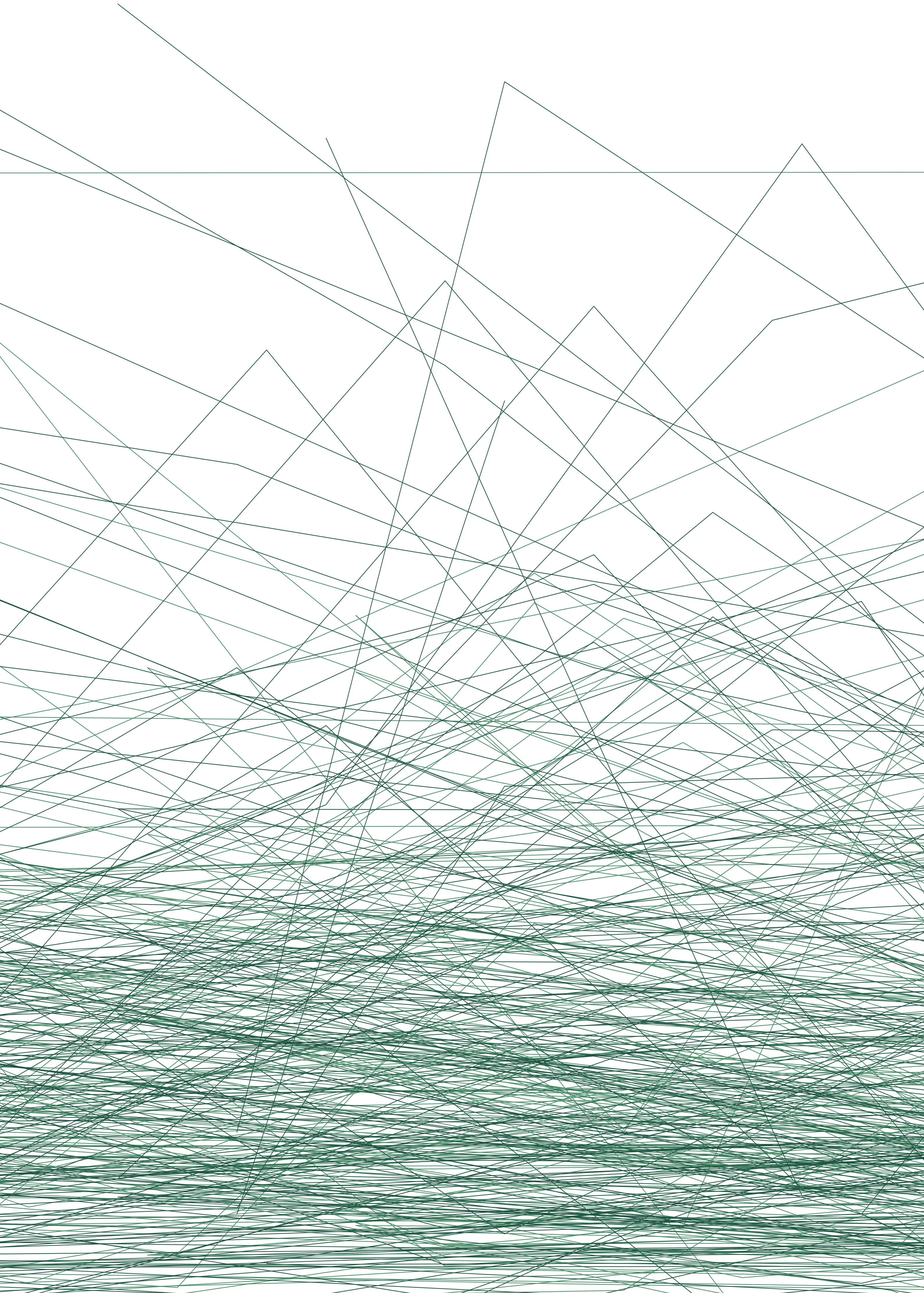
APPENDIX B Sentences used in this study
Sentences for the study phase and distractor for the test phase after / sign.

Homographs

1. I cannot walk with one BOOT (shoe) / glove
2. Bourbon is a good whisky BRAND (variety) / difference
3. A poodle is a well-known dog BREED (race) / crew
4. Keep your answer BRIEF (short) / slow
5. Her earrings are shaped like a HOOP (circle) / square
6. An uncle is next of KIN (family) / person
7. The lady is very nice and KIND (not cruel) / not clever
8. Green soup is made of LEEK (vegetable) / berry
9. He takes of the can's LID (cap) / seal
10. A bending round curve is a LOOP (round) / straight
11. The dog lays on the RUG (carpet) / couch
12. Iron eventually becomes RUST (oxidation) / silver
13. Informal language is SLANG (non-standard) / non-educated
14. In the fight he gave him a SLAP (punch) / buddy
15. She eats much but stays SLIM (not thick) / not huge
16. He put the coin in the SLOT (hole) / channel
17. I hit the stone with my TOE (foot) / knee
18. The mouse is caught in a TRAP (ambush) / bullet
19. After the rain it is WET (not dry) / not soft
20. That joker has a lot of WIT (funny) / serious

Controls

1. That's not a frog but a TOAD (amphibian) / turtle
2. You eat soup with a SPOON (cutlery) / glass
3. After showering she took a TOWEL (cloth) / food
4. The swimmer's shoulders are BROAD (not small) / not long
5. At Christmas dinner we eat FOWL (poultry) / plane
6. Everybody is waiting in a ROW (line) / crowd
7. It was quiet as the tv was on MUTE (silent) / loud
8. The manicurist fixed her broken NAIL (finger) / elbow
9. There are some cookies in the JAR (vase) / bottle
10. The yellow dress fitted the GIRL (not boy) / not parent
11. The walls reflected every RAY (sunshine) / darkness
12. In the lake we saw a DUCK (bird) / mammal
13. The leafs are falling from the BIRCH (tree) / flower
14. Stress gives him a RASH (irritation) / burst
15. He apologized for being so RUDE (not polite) / not ignorant
16. They had to solve the SNAG (difficulty) / innocence
17. The rabbit has a soft FUR (coat) / paper
18. I would do anything for you my DEAR (sweet) / angry
19. He was wearing a HAT (helmet) / jacket
20. He lost a tooth and broke his JAW (mouth) / belly



CHAPTER 5
Do bilinguals use shared cognitive control mechanisms in linguistic and non-linguistic tasks?

5.1 Introduction

In the everyday life of our quickly changing society, people must process different types of information all the time, selecting what they need and ignoring what is irrelevant. This is not unlike the problems that bilinguals have in their daily conversations when they must select the one language appropriate to the situation at hand while inhibiting the other. Different research papers (e.g., Bialystok, Craik, & Luk, 2008) even argue that bilinguals are so constantly trained in inhibiting an irrelevant language that this has led to more general advantages in their cognitive abilities with respect to attending to and selecting information. It seems likely that such a ***bilingual cognitive advantage*** depends on the specific experiences of the bilinguals with their languages. These considerations lead to an important issue that spans different cognitive modalities: Will bilinguals score better on non-linguistic tasks when they have more experience in suppressing irrelevant language information?

To clarify this issue, we must first consider the advantages or disadvantages associated with bilingualism relative to monolingualism. From a historical perspective, the societal view on bilingualism has not always been positive. Bilingual studies early in the 20th century found that, compared to monolinguals, bilinguals had smaller vocabularies, poorer grammatical skills, and also scored worse on verbal and non-verbal intelligence tests (e.g., Saer, 1923; Grabo, 1931). However, other researchers, like Peal and Lambert (1962) claimed that constantly switching between languages leads to optimisation of the 'mental flexibility' of bilinguals, allowing them to perform better on a number of cognitive tasks. Their study presented results for several tasks that measured nonverbal intelligence. In these tasks, bilingual children outperformed monolingual children. Later research on bilinguals confirmed that bilingualism might result in certain advantages in cognitive or linguistic performance. For example, studies found that bilingual children were more aware than monolingual children of the conventional nature of the language and its symbols, words, and structures (e.g., Bialystok, 1988; Cummins, 1978). More recent studies (e.g., Bialystok, 1992) focused not only on the differences between monolinguals and bilinguals, but also on cognitive mechanisms that linguistic and non-linguistic tasks may have in common. For instance, Bialystok (1992) proposed that metalinguistic awareness and cognitive control are linked to a domain-general mechanism governing both linguistic and non-linguistic processing.

In this paper, we consider the possibility of modality over-arching mechanisms in more detail. We will first discuss and apply two non-linguistic tasks (Simon task and AX-Continuous Performance Task) that engage inhibitory mechanisms in different ways. On the basis of the collected data, we will then compare the mechanisms in these tasks to the language suppression mechanism in the bilingual L2 lexical decision task. To set the stage for this theoretical comparison, we first review the main findings for all tasks in the following sections.

5.1.1 Simon task

A frequently used task to test cognitive control is the Simon task (Simon & Rudell, 1967). In this task, participants press a button in response to items on a computer screen while ignoring the spatial position they appear in. The classical task includes two colored dots, one red and one green, that are presented on the right or on the left side of the screen. The right response button is labeled red and the left button is labeled green. A congruent situation then arises when the dot appears on the same side as the response button to be pressed, an incongruent situation when it is not. To solve this task optimally, participants must ignore irrelevant or misleading information, i.e., they must apply ***interference suppression.*** Although different variants of this task are mentioned in the literature, all variants focus on this stimulus-response compatibility.

The Simon task is mainly used to compare the cognitive processing of bilinguals and monolinguals and assess the 'bilingual advantage'. Unfortunately, the results have not been consistent across studies. Some researchers have indeed obtained bilingual advantages (e.g., Bialystok, 2006; Linck, Hoshino, & Kroll, 2008; Martin-Rhee & Bialystok, 2008), whereas other found mixed results (e.g., Bak, Nissan, Allerhand, & Deary, 2014; Keijzer & Schmid, 2016; Prior & Gollan, 2011; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016), or no advantages at all (e.g., Kirk, Fiala, Scott-Brown, & Kempe, 2014; Namazi & Thordardottir, 2010; Paap & Greenberg, 2013).

In order to explain these mixed results for the Simon task, researchers have introduced conditions that were more cognitively demanding (e.g., Costa, Hernández, & Sebastián-Gallés, 2008) or required a combination of different executive processes for optimal performance (e.g., Bialystok, 2006, 2011). More reliable bilingual advantages were found in more difficult tasks (Bialystok & Martin, 2004; Craik & Bialystok, 2006). This suggests that bilinguals use more compound mechanisms in suppressing irrelevant information, resulting in cognitive advantages only if the task becomes more difficult and thus combines different simpler mechanisms. For example, Bialystok et al. (2006) observed a bilingual advantage in a Simon task with four colors that involved a more demanding high switch condition. They also showed that the bilingual advantage can be unstable, in the sense that the effect may disappear in the last block of an experiment. This suggests that the difference between bilinguals and monolinguals can be reduced depending on practice, however the overall RTs were still significantly shorter for bilingual participants. Shorter RTs suggest that, while using the same type of cognitive control, bilinguals still need less time to manage their responses.

On the basis of these studies, one may conclude that situations with higher conflict (i.e., more difficult tasks) demand more monitoring efforts, which may therefore demonstrate more prominent bilingual advantages. The studies by Elston-Güttler, Gunter, & Kotz (2005); and Von Studnitz & Green (2002) support the view that bilinguals flexibly use the decision criteria to optimize their decisions depending on context. More recent research (e.g., Colzato et al., 2008; Costa et al., 2008; Morales, Gómez-Ariza, & Bajo, 2013; Morales, Yudes, Gómez-Ariza, & Bajo, 2015) suggests that especially better processes of monitoring and goal maintenance contribute to the bilingual advantage. This implies that proactive control processes (e.g., Braver, 2012), or even the dynamic adjustment of the proactive and reactive mechanisms, are involved in the cognitive control tasks (e.g., Green & Abutalebi, 2013; Kroll & Bialystok, 2013; Morales et al., 2013, 2015). In the view of these researchers, proactive control is linked to monitoring, while reactive control is linked to inhibition and response suppression. In other words, proactive control adapts to the context for achieving the goal, but because it is continuously operating, it overloads working memory. Reactive control is focused on conflict detection in case inhibitory control is required. In this way, reactive control keeps working memory free for other task components. Because some researchers assume that bilinguals derive their bilingual advantage from the exact dynamic adjustment of proactive and reactive control, the AX-Continuous Performance Task was proposed.

5.1.2 AX-CPT

The AX-Continuous Performance Task (CPT) was used by Morales et al. (2013, 2015) to study executive control capacities in bilinguals, in particular the relative dynamics of proactive and reactive control. The researchers applied a modified version of the original task (Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956) in which participants (bilinguals and monolinguals) had to press YES on every X that follows an A (probe and cue respectively), and NO in every other case (labeled as BX, AY, and BY trials). Although all participants responded at similar speeds to BX and BY trials, there were slower responses and more errors on AY trials. Because of the high rate with which AX trials occur, participants are apparently prepared to respond YES after every A cue (a proactive control mechanism); to suppress the response in other cases, an inhibitory mechanism is needed (reactive control). In the case of an AY probe, proactive and reactive control must be adjusted in order to reduce the chance of making errors. The AY condition showed an advantage for bilinguals relative to monolinguals with respect to RTs (Morales et al., 2013) and in ERP recordings (Morales et al., 2015). In fact, both studies by Morales and colleagues revealed a better coordination of proactive and reactive control in bilinguals than in monolinguals. This implies that both groups of participants used reactive control in similar way, but that bilinguals employed a more efficient proactive strategy to perform the task. In the ERP recordings, bilinguals exhibited more negative N2 amplitudes to AY trials; a more negative N2 is thought to reflect a stronger conflict detection (Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003; Yeung & Cohen, 2006). Morales et al. (2013, 2015) concluded that monolinguals rely on either proactive or reactive control, whereas bilinguals selectively adjust and engage both mechanisms in response to task demands.

5.1.3 Lexical decision task

Different researchers state (e.g., Abutalebi & Green, 2007; Kan & Thompson-Schill, 2004) that, like in non-linguistic tasks, bilinguals use selection mechanisms in linguistic tasks in order to reduce non-target language influences. The mandatory use of these mechanisms may then have the consequence that bilinguals become more competent in cognitive tasks than monolinguals. This idea is also in line with different bilingual processing models (Inhibitory Control (IC), Abutalebi & Green, 2007; Green, 1998; Green & Abutalebi, 2013; Bilingual Interactive Activation (BIA) and BIA+, Dijkstra & van Heuven, 1998, 2002; van Heuven, Schriefers, Dijkstra, & Hagoort, 2008) that argue reactive cognitive control is essential for selecting words and languages.

With respect to lexical processing in general, many bilingual studies have shown that words from all languages of the bilingual are simultaneously activated and accessed in a language non-selective way. This conclusion has received empirical support in the domains of word reading (both in isolation and in sentences), listening, and word production (e.g., Brysbaert, 1998; Colomé & Miozzo, 2010; Marian & Spivey, 2003; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009; Van Hell & Dijkstra, 2002; see also Chapters 2 and 3 of the current dissertation). To test the language non-selective lexical access hypothesis, studies often used words from different languages that overlap either in spelling and meaning (i.e., "cognates", like ***water-water; baker-bakker*** in English and Dutch), only in spelling (i.e., "interlingual homographs", like ***brand*** meaning 'fire' in Dutch), or only in phonology (i.e., "interlingual homophones" like ***marker-marka*** meaning 'stamp' in Russian). In all these cases, the formal overlap may cause selection problems during word retrieval.

In order to solve such selection problems, mechanisms of cognitive control must be applied. Two mechanisms that have been proposed are (a) the inhibition of non-target readings; (b) the use of language membership to select the right word candidate. The lexical decision task (LDT) is widely used to test if conflict is present on the activation level. In this task, participants decide whether a presented stimulus is a word in a certain language or not. When interlingual homographs are presented in a mixed-language lists, the homograph inhibition effect observed in the LDT increases. Given that the above mentioned empirical results vary depending on stimulus list composition and task demands, it is likely that bilinguals use a flexible decision system to adapt their recognition process to global and local context requirements. In previous studies (Chapter 2 with cognates, and Chapter 3 with interlingual homographs), we have observed that relative language activation plays an important role in the increasing or reducing the degree of lexical conflict.

In line with this theoretical position, Green and Abutalebi (2013) propose an Adaptive Control Hypothesis that depends on the specific characteristics of a language context. A ***single-language*** context is a situation in which, for example, one language is used at home and another at work. A ***dual language*** context is present when two languages are used in the same context, but by different speakers. In this situation, a high demand is set on language control. Finally, a ***dense code-switching*** context arises when it is safe to use two languages in the same conversation with the same person. This situation sets lower demands on language control. Dual-language use should make higher demands on executive control than dense code-switching, and should have more impact on developing control skills. Daily practice in verbal inhibitory control should also lead to bilingual advantage for solving non-verbal cognitive control tasks.

According to influential inhibitory bilingual processing models (like the IC and BIA models), the activation of word forms in a non-target language can be reduced, thus facilitating the retrieval of the target word. The IC model proposed by Green (1998), focusing on word production, holds that representations of the relevant language are activated while those of an irrelevant language are inhibited. When a word is sought to express a lexical concept, lemmas linked to the concept are activated. In the case of bilingual speakers, lemmas in two languages become activated. To select the target word to be uttered, suppression of the other lemma occurs. The model specifies that the language inhibition observed in this case is reactive rather than proactive, which means that inhibition occurs only after both lemmas have already become active. The degree of inhibition is dependent on L2 proficiency, which is reflected in the relative degree of activation of lexical candidates from both languages.

Thus, according to the IC model, selecting and retrieving a lexical item does not only require the activation of that item, but also the inhibition of the corresponding non-target language item. When this inhibitory selection mechanism is used time and again, it can lead to generally improved interference suppression in non-linguistic tasks. The competition between word candidates from the two languages can then be differentially inhibited in a top-down fashion on the basis of language membership. In other words, a stronger language (like the native language) may then be suppressed to a larger extent than a weaker language (like a second language), in order to speed up the selection of the correct target word. In this view, inhibition mechanisms in language production can overlap with domain general mechanisms used for non-verbal inhibition, because the two types of processes are similar and not language-specific but domain-general. Various studies support this hypothesis (e.g., Bialystok, Martin, & Viswanathan, 2005; Costa et al., 2008; De Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014).

For example, in the study by De Bruin et al. (2014), Dutch-English-German trilinguals were examined in an fMRI scanner. They showed more activation when they switched to their second (English) and third (German) language in comparison with the non-switch trials. However, there was no difference in switching to the first language. The authors concluded that (i) language switching is related to domain general inhibition, and (ii) multilinguals use inhibition while switching between the languages.

In the recognition domain, the BIA model by Dijkstra and van Heuven (1998) proposes that a visually presented word activates its sublexical, lexical, and semantic representations. In addition, the language membership representation of the activated word becomes activated. This representation provides top-down inhibition to all non-target language words and this is different in the later version of the model, the extended BIA+ model (Dijkstra & van Heuven, 2002). This model not only includes more detailed phonological and orthographic lexical and sublexical representations, but it also leaves out explicit top-down inhibition from the language nodes. In this respect, the BIA model is more compatible with the IC model. Both models hold that competition between word candidates in the two languages can be differentially regulated via top-down inhibition on the basis of language information. Notwithstanding their differences, all three models agree that both inhibition and activation take place in the word selection process.

5.1.4 Current study

In the current study, we investigated whether bilinguals use the same or similar mechanisms in linguistic and non-linguistic tasks. Furthermore, we assessed whether effects observed in non-linguistic tasks can predict effects observed in linguistic tasks. Based on the literature discussed above, we assumed that different bilingual language experiences will result in different proactive-reactive dynamics. We tested whether bilinguals differ in reactive control (Simon task) depending on their relative language proficiency or rather in the use of proactive-reactive dynamics (as measured in the AX-CPT task).

In two experiments, unbalanced Dutch-English (L1-L2) bilinguals performed a non-linguistic task (Simon or AX-CPT) and a linguistic task (L2 LDT) involving interlingual homographs in L1-L2 mixed-language stimulus lists.

Our hypothesis was that observing a correlation of non-linguistic and linguistic task effects constitutes evidence that the same or at least similar cognitive control mechanisms underlie linguistic and non-linguistic tasks. Observing such a correlation also indicates that the cognitive mechanisms are not domain-specific, but domain-general instead. In contrast, any differences in the predictors from the Simon task and AX-CPT are a sign of differences in cognitive suppression used in the tasks. With respect to this point, we assume that Simon task is simpler and benefits mainly from reactive control, while the critical AY condition in the AX-CPT demands the use of both proactive and reactive dynamics to be performed correctly.

5.2 Experiment 1: Simon Task and English LDT

5.2.1 Method

Participants

Twenty-four linguistics students of the University of Antwerp were tested (20 women, 4 men; range: 19-28 years of age, mean: 21.4 years). Of these, 21 were right-handed and 3 left-handed. They all had normal or corrected-to-normal vision. They were Dutch native speakers and had English as their dominant foreign language. The participants had studied English as a foreign language at secondary school for at least six years (mean experience 10.5 years) and used L2 regularly for studies and their spare time (watching movies, reading on the internet).

Materials and Design

The critical items for LDT were 20 Dutch-English (L1-L2) orthographically identical homographs. Originally, we selected words from the CELEX database (Baayen et al., 1995). Later we compared their frequencies with the SUBTLEX databases, once these became accessible (Keuleers et al., 2010; van Heuven et al., 2014)[31]. Based on the SUBTLEX databases, the word frequency of the Dutch reading for Dutch-English homographs was still higher than for the English reading (see further).

We asked 10 randomly selected students from the same population as our participants to assess for each homograph (a) the difference of their meanings in the two languages, (b) the similarity of their pronunciations in these languages, and (c) the familiarity of the word in the participants' L2. Based on these ratings (1 = "none", 5 = "very strong", we selected the 20 highest rated Dutch-English homographs from the set of identical words that had a higher frequency in Dutch.

All selected homographs had a higher frequency in Dutch (mean: 4.77, SD = 0.53, range: 3.79 – 5.76) than in English (mean: 4.03, SD = 0.54, range: 3.33 – 5.60). For each homograph, an English monolingual control word was selected with the same number of letters (mean: 3.90 for homographs and controls) and a matched frequency (mean: 4.07, SD = 0.60, range: 3.02 – 5.29). A t-test indicated that homographs and controls were accurately matched on frequency ($t = 0.19$; $p > 0.1$). In addition, we added 10 English filler words from the same frequency range as the other English words in the experiment.

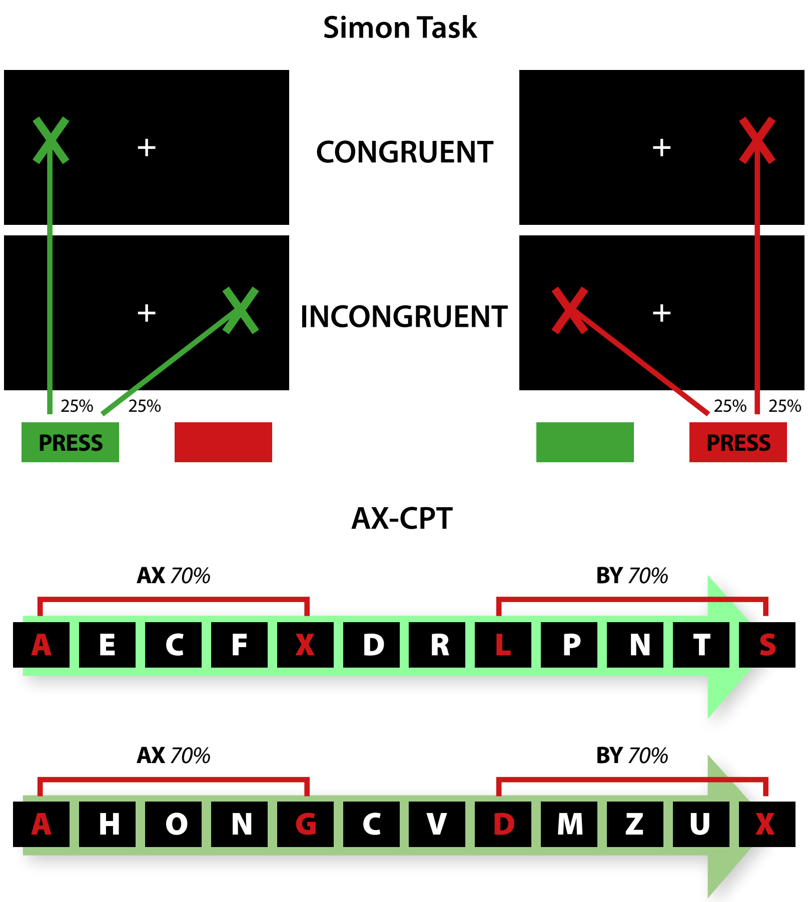
To create an English-Dutch context in this experiment, we added a set of NO-items consisting of 50 Dutch words, approximately matched on an item-by-item basis to all English words in number of letters and word frequency.

[31] Note, we used the formula log10(frequency per million*1000) from van Heuven et al., 2014 to calculate Zipf values for Dutch data. It made it easier to compare different data sets.

Procedure
The experiment was conducted in a dimly lit room on DELL computers (type Optiplex 380) connected to 15" DELL monitors. Reaction times were registered by means of the DMDX program (Forster & Forster, 2003). A game controller (type Logitech Wingman precision) was used for responding (using the left and the right front buttons) and for initiating a new item block (using the START button). The experiment was split in two parts. The first was concerned with the non-linguistic task (Simon task), the second with the linguistic task (English LDT). Before each part, participants read instructions in English about the required responses in the task.

The first part of the experiment involved a version of Simon task from Simon & Rudell (1967). Instead of a colored dots, we presented a colored "X" (either red or blue) as a target, to make this task as comparable as possible with the AX-CPT task (see Experiment 2), in which the response was also made on an "X". The participants had to press the red (right) button for a red "X" and the blue (left) button for a blue "X" as quickly and as accurately as possible, irrespective of its position on the screen (see Figure 5.1). The task consisted of 12 randomized practice trials, followed by two blocks of 40 experimental trials each. Half of all the trials were congruent (i.e., the correct response button was on the same side as the stimulus) and half were incongruent (i.e., the correct response button was reversed). Each trial began with a fixation (+) sign that appeared in the middle of the screen for 500 ms. Next, the colored "X" was presented on the left or on the right part of the screen where it stayed in view until a response had been made or until a time out of 2,500 ms had passed. The next trial was initiated 500 ms after the response or time out.

Figure 5.1.
Example of the experimental conditions in the non-linguistic tasks



After finishing the Simon task, participants received a new set of written instructions in English. Participants had to decide as quickly and as accurately as possible whether a letter string presented on the screen was a word in English. In that case, they had to "press the right upper button". If the letter string was a non-word, they had to "press the left upper button". It was not specified which other languages other than English might occur in the experiment to avoid the buildup of any expectations about the non-target language. The two buttons were reversed for left-handed participants. A set of 12 practice trials, different from the test trials, preceded the main task. The experiment itself was divided into four blocks of 20 trials each preceded by 2 dummy items. Each trial was presented in the center of the screen after a fixation (+) sign with a duration of 500 ms. The item stayed in view until a response had been made or until a time out of 2.500 ms had passed. The next trial was initiated 500 ms after response or time out. The whole experiment took about 25 minutes.

5.2.2 Results and discussion

The data from 3 participants were excluded, because they made more than 15% errors in the LDT. The data for both tasks were analyzed with (generalized) linear mixed effects models (see Baayen, Davidson, & Bates, 2008; Baayen & Milin, 2010, and references there for more details). We used the inverse transformation (RTi = -1000/RT) to normalize the data. For analyses, we used R statistical software (R Core Team, 2014) and the linear mixed model in the lmerTest package (Kuznetsova et al., 2016). Response times (RT) were analyzed with lmer function and error rates (ERR) with glmer function. We started with the simplest model including a random intercept for participants and items (if relevant), and then added one fixed-effect parameter at a time. The subsequent models were compared using likelihood ratio tests with α=.05 as a level of significance. If the model including a new fixed parameter was significantly better than the simpler one, we proceeded with the more complex model and followed the same procedure for the next fixed effect. We first analyzed both tasks separately for the main effects, and then calculated the mean congruency effect per participant[32] and included it as a predictor (fixed factor) for the LDT analysis.

In the Simon task, error responses (n= 34, i.e., 2%) and responses outside the RT range between 300 and 1500 ms (n=155, i.e., 9.2%) were removed. This left us with 1491 RT for the congruent and incongruent responses for both colors (i.e. 88.8 % of the 1680 remaining data points after removing 3 participants).

The effect of congruency was highly significant for RTs (ß = 0.18, t = 6.24, $p < 0.001$), and for Errors (ß = 0.91, z = 1.98, $p < 0.05$) indicating that participants react much slower and make more errors for incongruent stimuli (see Table 5.1).

In the LDT data we removed error responses (n=125, i.e. 14.8%) and responses outside the RT range between 300 and 1500 ms (n=14, i.e. 1.7%). This left us with 701 RT for the homographs and their controls (i.e. 83.4% of the 840 remaining data points after removing 3 participants).

The effect of Word Type (homograph vs. control) was significant for RTs (ß = 0.27, t = 12.21, $p < 0.001$) and errors (ß = 3.57, z = 8.67, $p < 0.001$), replicating the often found result of slower RTs (768 ms vs 637 ms) and more mistakes (28.1 vs 1.7) for homographs when both

[32] Congruency effect = Mean RT on congruent items – Mean RT on incongruent items

Table 5.1.
Mean Latencies (RT) and errors (ERR) from Simon Task, AX-CPT and English Lexical decision tasks (LDT).

	Experiment 1				Experiment 2				
	Simon Task		LDT		AX-CPT			LDT	
	Congruent	Incongruent	Homograph	Control	AY	BY	BX	Homograph	Control
RT (ms)	414	437	768	637	407	355	411	703	594
ERR (%)	1.2	2.9	28.1	1.7	2.5	0.8	5.0	36.7	5.2

languages are included in the stimulus list. For the next step, we calculated the congruency effect from Simon task for each participant separately subtracting the congruent mean value from the incongruent one. Because both RTs and errors were significant in the Simon task, we added both values (CongruencyRT and CongruencyERR, respectively) as a factor to test interaction with the homograph effect. However, there was no interaction with any of the added factors, neither for RTs (WordType x CongruencyRT, $ß = -0.0006$, $t = -0.76$, $p > 0.1$; WordType x CongruencyERR, $ß = 0.83$, $t = 1.43$, $p > 0.1$), nor for the errors (WordType x CongruencyRT, $ß = 0.02$, $z = 1.33$, $p > 0.1$; WordType x CongruencyERR, $ß = 9.89$, $z = 0.79$, $p > 0.1$).

In other words, absence of the interaction between LDT and Simon task indicates some differences in the mechanisms used to solve the selection problem in two different situations. We will discuss this point in more detail in the General Discussion.

5.3 EXPERIMENT 2: AX-CPT and ENGLISH LDT

5.3.1 Method

Participants
Twenty-five students from the same population as in Experiment 1 were tested (19 women, 6 men; range: 19-23 years of age, mean: 20.3 years). Twenty-two were right-handed and 3 left-handed. All participants had normal or corrected-to-normal vision. They had a comparable experience with English (mean: 8.8 years).

Materials and Design
The same items were used as in Experiment 1.

Procedure
The experiment was conducted in a dimly lit room on the same equipment as Experiment 1. The experiment was split in two parts, involving a non-linguistic task (AX-CPT) and a linguistic task (English LDT). Before each part, participants read an instruction in English about the response requirements.

The first part of the experiment was a version of AX-CPT, adapted from Morales et al. (2013). The letters were displayed on a cue-probe basis including two neutral letters presented between the trials (see Figure 5.1). Participants were instructed to keep in mind the cue (that could be either the letter "A" or any other letter except "X", "K", or "Y", because of their visual similarity with "X") until they saw the probe (that could be either the letter "X" or some other letter, but not "A", "K", or "Y"). Every time they saw the cue "A", followed by the probe "X", they were instructed to press "YES" button. In any other possible combination, they had to press the "NO" button. In the interval between each cue and probe, presented in red, participants saw 3 distractor letters (that could be any letter except "A", "X", "K", or "Y") presented in white. Participants were instructed to press "NO" to each distractor. Because we adapted this experiment to individual differences, all letters were presented until the response was made or until the time-out of 1,000 ms had passed. The task was preceded by a practice block composed of 14 trials including all 4 possible experimental conditions AX (8), BX(2), AY(2), and BY(2). Participants were provided with feedback on accuracy after each practice trial. Only participants who made less than 5 errors on the whole practice set, took part in the complete experiment and were included in the analyses. Each experimental block consisted of 100 trials. AX trials occurred in 70% of the cases and each of the other conditions in 10%. (see Figure 5.1).

After finishing the AX-CPT task, participants received a new set of written instructions for the LDT in English. For the English LDT, a procedure identical to that in Experiment 1 was applied.

5.3.2 Results and discussion

We first excluded participants from further analysis by considering their response accuracy on AX trials. Only if participants scored higher than 80%, did we assume that the connection between A and X was properly made. As a result, the data of one participant were excluded because he scored less than 80%.

For AX-CPT error responses (n=20, 2.8%) and RTs falling outside the range between 100 and 900 ms (n=22, 3%) were excluded. This left us with 678 observations for the 3 critical conditions (94.2% from the 720 remaining data points). For the lmer analyses, we took the BY condition as the base line, because in this condition no conflict should be observed.

The RT analysis showed that participants were significantly slower in the BX condition than in the BY condition (ß = 59.61, t=4.74, $p<0.001$) and also slower in the AY condition (ß = 45.50, t=3.64, $p<0.01$). The difference between BX and AY conditions was not significant (ß = 14.11, t=1.12, $p>0.1$). In sum, it appears that both critical conditions cause similar conflict during the decision process and can be taken as predictors for the homograph effect in the second part of the experiment.

The error comparison revealed a difference between BY and BX, but not between BY and AY, nor between BX and AY. In Table 5.1 we see clear that the number of errors for AY condition is somewhat in between the BY and BX, however the difference between the latter is much bigger. It indicates that participants make more errors in the BX condition, but the number of errors is still low like in the two other conditions.

In the LDT data, we removed errors (n=201, 20%) and RTs falling outside the range between 300 and 1500 ms (n= 11, 1.1%). This left us with 748 RTs for homographs and their controls (78% from the 960 remaining data points). A significant homograph effect was observed with participants reacting significantly slower (ß = 0.24, t=9.76, $p<0.001$) and making more errors (ß = 0.24, z= 10.88, $p<0.001$) on homographs than on control words.

Next, we calculated the RT differences for BY and BX conditions (BYminBX); as well for BY and AY conditions (BYminAY) for each participant separately. We added both values to our analyses as a fixed factor to look for the interaction with the homograph effect. BYminBX showed no interaction for the homograph effect ($ß = -0.0003$, $t = -1.34$, $p > 0.1$). BYminAY, however, showed significant interaction with the homograph effect ($ß = 0.0008$, $t = 1.96$, $p = 0.05$). This is taken as evidence that the decision processes involved in the two conditions use different mechanisms. Apparently, the mechanism applied in the AY condition has more in common with the mechanism involved in the homograph recognition in LDT than the mechanism used in the BX condition.

We also compared the two LDTs with each other. Neither RTs ($ß = 0.03$, $t=0.94$, $p>0.1$), nor Errors ($ß = 0.81$, $z=1.72$, $p>0.05$) showed any significant differences. This comparison indicates that there is no difference in the observed homograph effects in the two English LDTs, allowing us to compare the mechanisms used in different tasks.

5.4 General Discussion

We compared the performance of bilinguals on non-linguistic (AX-CPT and Simon task) and linguistic (LDT) tasks to find out if they use similar processing and control mechanisms in both. Participants who were better in suppressing the irrelevant reading of interlingual homographs were found to also be better in the differential dynamics required for the AX-CPT task (specifically, the AY condition). The implication is that general mechanisms exist that are used in both tasks.

Although there were clear main effects in the Simon task and in the BX condition of the AX-CPT, these did not correlate with the degree of homograph interference in the LDT. Apparently, the mechanisms for suppressing the incongruent items in the Simon task and an irrelevant response in the BX condition are somewhat different from those used in the LDT. In the following section, we consider the mechanisms involved in these tasks in more detail.

5.4.1 Simon task

In the Simon task, participants must link certain response buttons to certain colors (right with red and blue with left). Note that these buttons are also marked with the respective colors. The visual stimulus causes an unambiguous decision about the color, i.e. on the dimension 'color' there is no competition between two possible responses and thus no problem at the cognitive level. The simple decision activates a fixed motor 'program', required to push the response button that belongs to the right color. The visual stimulus, however, simultaneously and automatically activates the motor program to push the response button that on the left-right dimension is in the same position as the stimulus. This is a form of (unconscious) response priming, a process without cognitive intervention. The stimulus position varies throughout the experiment between the two extreme positions on the left-right dimension (i.e., in 50% of the cases the stimulus is presented on the left and in the other 50% of the cases on the right), and the positions of the response buttons also differ in the left-right dimension. Thus, an automatic perception of alignment or misalignment occurs between those two sources of information (stimulus and buttons). In the incongruent condition, there is a conflict between the motor program primed by the stimulus position and the motor program selected by the cognitive system that causes a delay in response.

5.4.2 AX-CPT

The AX-CPT is a more demanding task than the Simon task, because it creates a context that is cognitively more challenging. Within the experiment, both conditions (AY and BX) result in an equally strong inhibition with respect to the control condition BY, but AY correlates with the observed homograph inhibition effect, while BX does not. Although we might assume that similar cognitive mechanisms are used in both situations, there are some crucial differences. The cue (A) in condition AY activates the 'yes'-response, because in 87.5% of all cases A is effectively followed by X. However, in the condition AY, the Y (any letter but X, or similar to X) triggers the 'no'-response. In the condition BX, the participant learns throughout the experiment that if there is no A, no X can be found. This provides a still more reliable cue than the A, because the absence of an A in 100% of the cases means that a 'no'-response should be made. However, the probe (X) activates the 'yes'-response in 87.50% of the responses if it is presented.

We propose that situations involving cognitive response competition, like AY and BX, should be differentiated as follows. In the AY condition, A strongly preactivates the 'yes'-response and causes competition with the 'no'-response because the target item is not X. As a consequence, participants have to suppress the activated 'yes'-response and press "no", because there is no X. In this situation, two responses conflict with each other. This is similar to the case deciding between the two readings of an interlingual homograph. As a consequence, the AY condition is a significant predictor of the magnitude of homograph inhibition.

In the BX condition, there is also a significant competition effect. However, the mechanism involved is supposed to be different. The B cannot pre-activate any other letter representation, because participants are not trained on any other letter combination. In contrast, associative links for AX arise, because one specific letter (A) often occurs together with another specific letter (X). Therefore, the first letter pre-activates the target letter automatically and quickly. In the case of cue B, however, there is not a specific following letter but a category of letters (basically all letters except A and X, which are the two critical items in the experiment). It is therefore very likely that B preactivates the 'no'-response, because there is no associative path at all. In contrast, if an X appears, the participant should be able to answer «no» immediately, not based on pre-activation but because there has been no A (that information is missing in their cognitive system). Instead, if there is an X on the screen, they usually have to give the answer «yes» (in 7 of 8 cases). Here there is no possibility for response conflict because participants did not pre-activate the 'yes'-response, but the presence of X makes them check in their short-term memory if they have not really seen an A. This search process causes delay. However, the delay does not arise as a result of response competition and thus is not a form of inhibition due to suppression of a response.

5.4.3 LDT

Our results shed more light on control mechanisms in the LDT. The observed correlation of homograph inhibition with the AY condition in the AX-CPT indicates that there are similar underlying mechanisms in the two tasks. Our theoretical analysis suggests that in the AX-CPT there is a pre-activation of the 'yes'-response in the AY condition that has to be suppressed later. The homographs in LDT appear to make use of a similar suppression mechanism: when both readings of the homograph are automatically activated, the irrelevant one has to be suppressed. This reasoning is in line with the BIA+ model (Dijkstra & van Heuven, 2002) and pertains to the relationship between activation and decision activities (see Chapter 6).

However, more research is needed to investigate the exact mechanisms involved in different non-linguistic tasks and their correlation with the linguistic ones. It is also important to consider the issue with respect to language production. There are several studies that investigated similarities in the bilingual language control and executive control mechanisms in production (e.g., Bialystok et al., 2008; Branzi, Calabria, Lucrezia, & Costa, 2016; Calabria, Hernández, Branzi, & Costa, 2012; Cattaneo et al., 2015) and these also do not always observe a strong between-task correlation. For example, Calabria et al. (2012) tested Spanish-Catalan bilinguals and compared switch cost in a linguistic naming study with the switching costs in non-linguistic tasks and observed different result patterns throughout the experiment. This indicates that the two tasks were performed based on different underlying processes.

Even the neuroimaging literature does not give unequivocal results (e.g., Abutalebi et al., 2013). In the study of Abutalebi et al. (2013), participants performed both a linguistic switching task and a non-linguistic conflict resolution task. Activation in different brain areas during the performance of the two tasks was compared. It was concluded that some brain areas used in solving these tasks did overlap, but the results were not completely convincing. In another study, Prior and Gollan (2011) found similarities in the size of the switch cost in two tasks, but only for bilinguals who used switching often in everyday life. Furthermore, Cattaneo et al. (2015), comparing Catalan-Spanish bilinguals with Parkinson's disease to healthy individuals in linguistic and non-linguistic tasks, suggest a link between the control mechanisms used for these tasks. However, a later study of Branzi et al. (2016) focusing on similar switch experiments arrives at the conclusion that although there is a clear overlap between control systems in linguistic and non-linguistic tasks, the implied mechanisms are used differently.

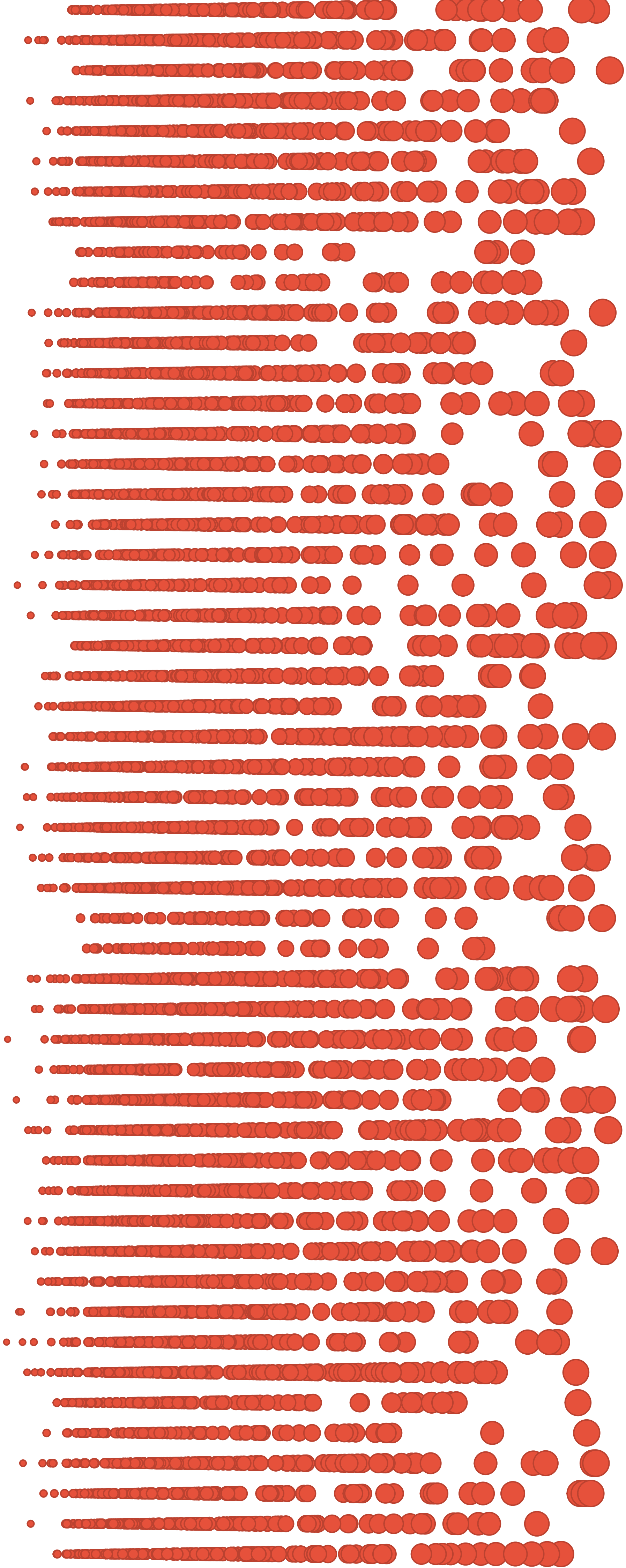
5.4.4 Conclusion

On the basis of our study, we conclude that inhibitory processes involved in linguistic task (LDT) and the non-linguistic tasks (AX-CPT) may to some extent be subserved by common mechanisms. However, this only appears to be the case when proactive-reactive dynamics are involved (i.e., the AY condition in AX-CPT). In contrast, reactive inhibition in the AX-CPT task (observed in the BX condition) and the inhibitory mechanisms involved in the Simon task appear to involve different mechanisms for suppression. In sum, more research is needed to investigate the domain general mechanisms involved in both types of tasks.

APPENDIX A Homographs used for this study

Interlingual homographs, control words; their mean Zipf frequencies, and translations of the Dutch meaning of the homograph.

HOMOGRAPHS					ENGLISH CONTROLS		
word	Dutch meaning	length	English Zipf	Dutch Zipf	word	length	English Zipf
boot	boat	4	4.43	4.98	toad	4	3.69
brand	fire	5	4.68	4.65	spoon	5	4.29
breed	wide	5	4.26	3.79	towel	5	3.87
brief	letter	5	4.26	4.87	broad	5	4.22
hoop	hope	4	3.59	5.57	fowl	4	3.43
kin	chin	3	3.54	3.89	row	3	4.75
kind	child	4	5.6	5.52	mute	4	3.16
leek	layman	4	3.56	4.98	nail	4	4.18
lid	member	3	4.16	4.52	jar	3	3.96
loop	course	4	3.82	5.07	girl	4	5.29
rug	back	3	3.57	4.91	ray	3	4.56
rust	peace	4	3.33	4.88	duck	4	4.53
slang	snake	5	3.56	4.33	birch	5	3.43
slap	limp	4	4	3.85	rash	4	3.61
slim	smart	4	3.8	5.05	rude	4	4.27
slot	lock	4	3.78	4.72	snag	4	3.02
toe	closed	3	4.08	5.76	fur	3	4.03
trap	stairway	4	4.22	4.72	dear	4	5.08
wet	law	3	4.76	4.91	hat	3	4.76
wit	white	3	3.67	4.52	jaw	3	3.38
mean:		3.90	4.03	4.77		3.90	4.10
sd:		0.72	0.54	0.53		0.72	0.62
range:		3-5	3.33-5.60	3.79-5.76		3-5	3.02-5.29



CHAPTER 6
General Conclusions

6.1 The main results

This chapter summarizes the main findings of this dissertation and discusses them in light of their importance for bilingual research and their implications for multilingual word recognition models. The aim of the current dissertation was to study in more detail than before how multilinguals process interlingual words that share their orthography and semantics across the languages (cognates) or only their orthography (interlingual homographs). In particular, we focused on the role of a number of linguistic and non-linguistic factors that affect the recognition of such interlingual words, including stimulus list context, task demands, and language dominance. Additionally, we studied the cognitive control mechanisms involved in the suppression of the irrelevant meaning of interlingual homographs and we compared these mechanisms to those used in non-linguistic tasks.

Several questions were posed in the studies of the current dissertation. Do identical cognates have multiple representations in the shared multilingual lexicon? What is the influence of different linguistic and non-linguistic factors on the processing of cognates and interlingual homographs? How are inhibitory processes distributed across an experimental session and amongst participants? Do participants use similar cognitive control mechanisms in linguistic and non-linguistic tasks? All these questions were considered by zooming in on the contribution and combination of lexical and post-lexical processes in bilingual word recognition.

The experimental chapters that form the core of this thesis, approach the issues from diverse angles in terms of stimulus materials, experimental design, research techniques, statistical analyses, and participants. First, an important asset of this research lies in the use of a shared set of stimulus materials across experiments that differ only slightly in terms of design aspects. The careful, detailed manipulation of independent variables increases the power of cross-experimental analyses and generated an unprecedented possibility for theoretical comparison. Furthermore, by studying performance in two related but slightly different experimental tasks (e.g., lexical decision task (LDT) and go/no-go task (GNGT)), shared core characteristics of multilingual word recognition could be differentiated from task-specific components. Finally, by comparing three languages rather than two (as was done in the past), we were able to disentangle the effects of competitor words in languages that were either stronger (L1) or weaker (L3) than the target language (L2). Additionally, we focused on the cognitive control mechanisms in the processing of interlingual homographs, tested them in different settings and compared to non-linguistic tasks to investigate domain-general account.

In the following part of the thesis, we will go through the different chapters and consider the different theoretical issues. After summarizing the main empirical results per research chapter, we can then interpret them in the light of multilingual word recognition models.

Chapter 1 is a literature review on word recognition in multilinguals. The central issue is whether their process of lexical activation is language-specific or non-language-specific. Two word types are ideal for addressing this question: cognates and interlingual homographs.

Many experiments have shown that interlingual words activate the lexical information in all languages in which they occur. In other words, multilinguals appear to be unable to suppress their process of lexical activation by means of a top-down mechanism. The decision which activated information is required to make a correct response in a specific task is only taken at the postlexical stage. That conclusion can be drawn from many behavioral and neurocognitive experiments. BIA+, a model that has been developed by Dijkstra and Van Heuven (2002), can account for most experimental effects because it makes a distinction between an activation level and a decision level. Whereas the former operates in a non-language-specific way, the latter is sensitive to experimental-specific factors, such as target language, list composition, and the nature of the experimental task.

In Chapter 2, we reported on six experiments with trilinguals in which three factors were orthogonally manipulated: cognate type (L1-L2 and L2-L3), stimulus list composition (English, English-French and English-Dutch), and task (LDT and GNGT). The aim was to clarify how identical cognates are represented and processed by trilinguals. Across the set of experiments, relative to matched one-language control words, cognates yielded facilitation effects, null effects, and even inhibition effects depending on the type of cognate, stimulus list composition, and task. Although participants reacted faster to Dutch-English (L1-L2) cognates than to pure English (L2) or even pure Dutch (L1) cognates, they reacted equally fast to the English-French (L2-L3) cognates and to their English (L2) controls, but much slower to the pure French (L3) words. On the whole, these findings are compatible with other studies on cognate recognition (e.g., Lemhöfer & Dijkstra, 2004; Peeters, Dijkstra, & Grainger, 2013; Van Hell & Dijkstra, 2002). We interpreted our findings in terms of an account (a) in which cognates have different (morphological) representations in the three languages that were investigated and the speed at which these representations are activated is determined by language dominance, and (b) in which the selection of multiple activated representations in different languages is affected by decision level factors. A decision level account is required, because different outcomes were observed in the two tasks, the generalized lexical decision task (GLDT) and the go/no-go task (GNGT). In fact, not only the size but even the direction of cognate effects was found to be task-dependent. When we consider the empirical literature as a whole, facilitation effects were found in tasks where both languages were present and could be used for the response, while inhibition effects arose in tasks where responses for items from different languages competed for selection (e.g., Brenders, van Hell, & Dijkstra, 2011; Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Font & Lavaur, 2004; Peeters et al., 2013). Our findings extend the results from earlier studies by showing a difference between tasks in terms of the decision processes they require. Specifically, in a generalized LDT the recognition of the word can immediately give rise to a 'yes'-response. However, in the GNGT the recognized word must be classified further to make the correct response. The observation that the process of cognate recognition can best be explained by assuming different representations in different languages, also strengthens recent conclusions concerning a research question that has thus far received little attention.

Chapters 3 and 4 presented detailed investigations of the inhibitory control processes that bilinguals use in the recognition of the Dutch-English (L1-L2) interlingual homographs. In Chapter 3, the presence of inhibition was systematically compared across two tasks (English LDT and GNGT) and three stimulus list contexts (English, English-French, and English-Dutch). As expected, the homograph effect was context- and task-dependent. Although no difference was observed in the RT-results for pure English and English-French context conditions, effects in the English-Dutch context were much stronger. A direct comparison of the results for the English LDT and the GNGT makes clear that latter type of task is more demanding, thus yielding much stronger inhibition effects. Apparently, the conflict underlying inhibition does not only affect the relative competition of multiple meanings for selection, but also differentially affects the decision making mechanisms in the two tasks.

In Chapter 3 we also performed analyses that went further than comparisons between condition means (like e.g., Balota, Yap, Cortese, & Watson, 2008; Roelofs & Piai, 2017). In ex-Gaussian analyses, in which the RT distribution is mathematically decomposed into a Gaussian distribution (mean: mu; standard deviation: sigma) and an exponential distribution (mean: tau), we found that list composition affected mu, while the experimental task affected tau. The homograph effect on mu was significant in all experiments but was significantly larger in the English-Dutch list. The homograph effect on tau was significantly larger in the GNGT. The effect on mu indicates that the inhibition effect affected all responses (both the fast and the slow ones), which led to a shift of the entire RT distribution. That shift was larger in the English-Dutch list than in the pure English and the English-French lists. The effect on tau is an effect on the right tail of the RT distribution, i.e., the region of the longest RTs. The effect of the experimental task on tau indicates that the cognitively more demanding GNGT causes longer RTs on the interlingual homographs, which implies a larger inhibition effect at the decision level. A significant homograph effect was also found on sigma. This effect was significant in both tasks but significantly larger in the GNGT than in the LDT. It did not interact with list composition. The interlingual homographs caused significantly more variability in the RTs than their L2 controls. This suggests that, at the individual response level, there is much variation in the size of the homograph effect, probably both at the item level (e.g., as a function of participants' familiarity with both readings) and at the participant level (as familiarity with the L2 reading can vary considerably across participants). The stronger effect in the GNGT also suggests that task properties affect the decision level, where variability in the size of the interlingual competition across items and participants will cause a larger inhibition effect. In addition to the ex-Gaussian analyses, quantile analyses indicated that the homograph effect became significantly larger as the RTs became longer. Delta plot analyses showed that the effect grew linearly, i.e., that the increase in the inhibition effect was constant per RT unit.

In Chapter 3, there was a purely English context condition in which English target words appeared. In this condition, there were no pure Dutch words in the stimulus list context and only the interlingual homographs were not exclusively English words. In contrast, another context condition in the same study was mixed with respect to language. This English-Dutch context condition contained 50% Dutch words that obviously could activate the Dutch language. The consequence of this incorporation of Dutch words in the stimulus list was found to indirectly boost the Dutch reading of the homographs and resulted in a stronger homograph inhibition effect. In Chapter 4, we investigated whether a low proportion of Dutch items in the experiment (i.e. 10%) might activate the Dutch homograph reading to an extent laying somewhere in between the 0% and 50% conditions. The results showed that including 10% Dutch control words in the stimulus list was not enough to activate the Dutch homograph reading more than in a 0% condition, in contrast to the 50% condition that did produce a larger homograph inhibition effect. An additional issue that was addressed was whether the study phase might increase or decrease the interlingual homograph effect, assuming that studying the stimulus items before the experiment proper could be seen as a form of repetition priming. However, when the participants studied English sentences involving the homographs before the main experiment, this did not lead to any change in inhibition effects in a purely English context condition. This lack of effect demonstrates that short-memory connections do not necessarily effectively change long-term connections in the shared multilingual lexicon.

In sum, considering the findings for interlingual homographs as a whole, we learned that not only linguistic characteristics (i.e. language dominance, linguistic context etc.) influence the processing of interlingual words, but also task demands and stimulus list characteristics. This is in line with previous research (e.g., Dijkstra et al., 1998). The more detailed analyses of the RT distributions for homographs and their L2 controls also allowed us to find out that the effect of list composition influences all RTs in the RT distribution, i.e., causes a shift in the Gaussian component in the distribution, whereas the effect of task difficulty is restricted to the longest RTs in that distribution. We have also learnt that the homograph effect increases with increasing response times and that this increase is linear. Finally, it has become clear that the size of the homograph effect significantly increases when 50% of pure Dutch words are added to the list ('no'-responses) whereas the effect does not change when that percentage is only 10%.

Furthermore, in Chapter 5 we obtained findings with respect to the bilinguals' cognitive control that suggest that interlingual conflicts may call for suppression by domain general processes that play a role not only in linguistic, but also in non-linguistic tasks (e.g., Bialystok, Craik, & Luk, 2008). In Chapter 5 we focussed on the cognitive control issues in interlingual homograph processing to consider the similarity of cognitive control mechanisms in linguistic and non-linguistic cognitive tasks. In addition to an English LDT in an English-Dutch context, we applied a Simon task and AX-Continuous Performance Task (AX-CPT). In all three task situations, there was a clear main effect (homograph effect in LDT and congruency effect in AX-CPT and Simon task). However, similar control mechanisms were only observed for AX-CPT (AY condition) and LDT. The type of suppression mechanism in AX-CPT seems to be related to that used for responding to the correct (task relevant) reading of the homograph in lexical decision. The absence of a correlation between cognitive effects in the LDT and the Simon task is not surprising. Given the task demands, the Simon task is bound to use suppression mechanisms that differ from those used in the case of interlingual homographs in LDT. Probably the conflict in the former task is situated in the later (non-cognitive) phase of the response process, whereas the effect in the latter task is situated at the earlier level of response selection. However, further research is required here because different studies obtained inconsistent results for the Simon task (e.g., Paap & Greenberg, 2013). In other words, there is a possibility that bilinguals use different mechanisms for solving linguistic and non-linguistic tasks. In addition, different theoretical positions were taken concerning domain general and domain specific mechanisms in linguistic and non-linguistic tasks (see, e.g., Coderre & van Heuven, 2014; Duñabeitia et al., 2014).

In the next section, I will interpret the reviewed thesis results in terms of a widely-used model of bilingual word recognition, the BIA+ model (Dijkstra & van Heuven, 2002). In general, our findings support this model (and also the IC model by Green, 1998), which proposes two stages of bilingual language processing: a preconscious lexical activation level (subserved by the lexical identification system), where representations are activated; and a decision level (subserved by the task/decision system), involving more conscious, task-induced processes. According to the BIA+ model, not only linguistic characteristics influence word recognition, but also context information and task demands. Furthermore, the findings of our series of studies on cognates support the BIA+ model in showing that, similarly to interlingual homographs, cognates have multiple morphological representations for each language. This assumption is necessary to formulate a coherent account of our empirical results, namely, that participants respond differently to interlingual words (cognates or interlingual homographs) depending on stimulus list context and task demands. Finally, our cognitive control study clarifies the nature of the inhibitory homograph processes in some detail. When we combine all results obtained in our homograph, cognate, and cognitive control studies, we end up with an extended version of the BIA+ model, as depicted in Figure 6.1. In this Figure, we have added multiple languages and indicate the effects of the linguistic context and instructions (task demands). For simplicity, we refer to different languages with one label; however, it must be kept in mind that lexical representations in each language have orthographic, phonological, and semantic properties.

6.2 Implications for the BIA+ model

6.2.1 Activation of lexical representations

When a word that exists in several languages at the same time (i.e., it is an interlingual homograph or cognate) is presented to a reader, each of its readings will be activated to an extent that depends on the word reading's linguistic characteristics and subjective frequency. In addition, the speed and degree of activation of the reading in question co-depends on the amount of cross-linguistic overlap it has with its other reading at different sublexical and lexical levels (orthography, phonology, and semantics). Following the assumptions of BIA+, the lexical activation process begins at a word's (sublexical and lexical) orthographic representations and then quickly proceeds to phonological and semantic representations.

A number of studies have already focused on the consequences of orthographic, phonological, and semantic overlap for the processing of cognates and interlingual homographs (e.g., Dijkstra, Grainger, & van Heuven, 1999). These studies have generally reported that more overlap results in stronger cross-linguistic effects, especially from L1 to L2. In our experiments, we intentionally focused on orthographically identical words with close phonology to maximize the observed effects. Depending on their subjective frequency, phonological, and semantical representations are activated earlier or later in time. The implication of this is that L2 lexical representations are generally activated a bit more slowly than L1 lexical representations ("temporal delay assumption" in BIA+, also see Ferrand & Grainger, 1993). This view is also supported in ERP and N400 studies. For example, Kutas & Federmeier (2011) found that bilinguals show N400 priming and semantic congruity effects in both their languages, with the timing (and to a lesser extent the amplitude) of these effects being a function of language proficiency and age of acquisition: effects appearing later and smaller for less well-learned languages.

Context characteristics, like the order of items in a stimulus list or its composition, play an important role during target word selection, because to result in correct responses, the temporal deadlines of the word selection process must be fine-tuned to the local and global context at hand. In the experiments of Chapters 2 and 3, we observed smaller effects in a monolingual English context and not task-relevant bilingual language context (e.g., in an English-French context for Dutch-English interlingual words). The temporal delay assumption was in line with the language dominance effects found in Chapter 2: Here the effects from L1 on L2 were stronger than the effects from L3 on L2. In Chapter 3, we found significant effects in all test conditions (even in the monolingual English context), reflecting that the subjective frequencies of the readings in L1 are higher than in L2. Further support found in Chapter 2 shows that in the generalized LDT participants reacted much faster to cognates than to English or Dutch control words. This indicates that participants respond on the basis the fastest available codes, which would make them less sensitive to cross-linguistic phonological and semantic effects that take a considerable amount of time to be established. These observations were also made in several earlier studies (e.g., Dijkstra et al., 1999; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; Lemhoefer, Dijkstra, & Michel, 2004).

***Figure 6.1:** Extended BIA+ model for*

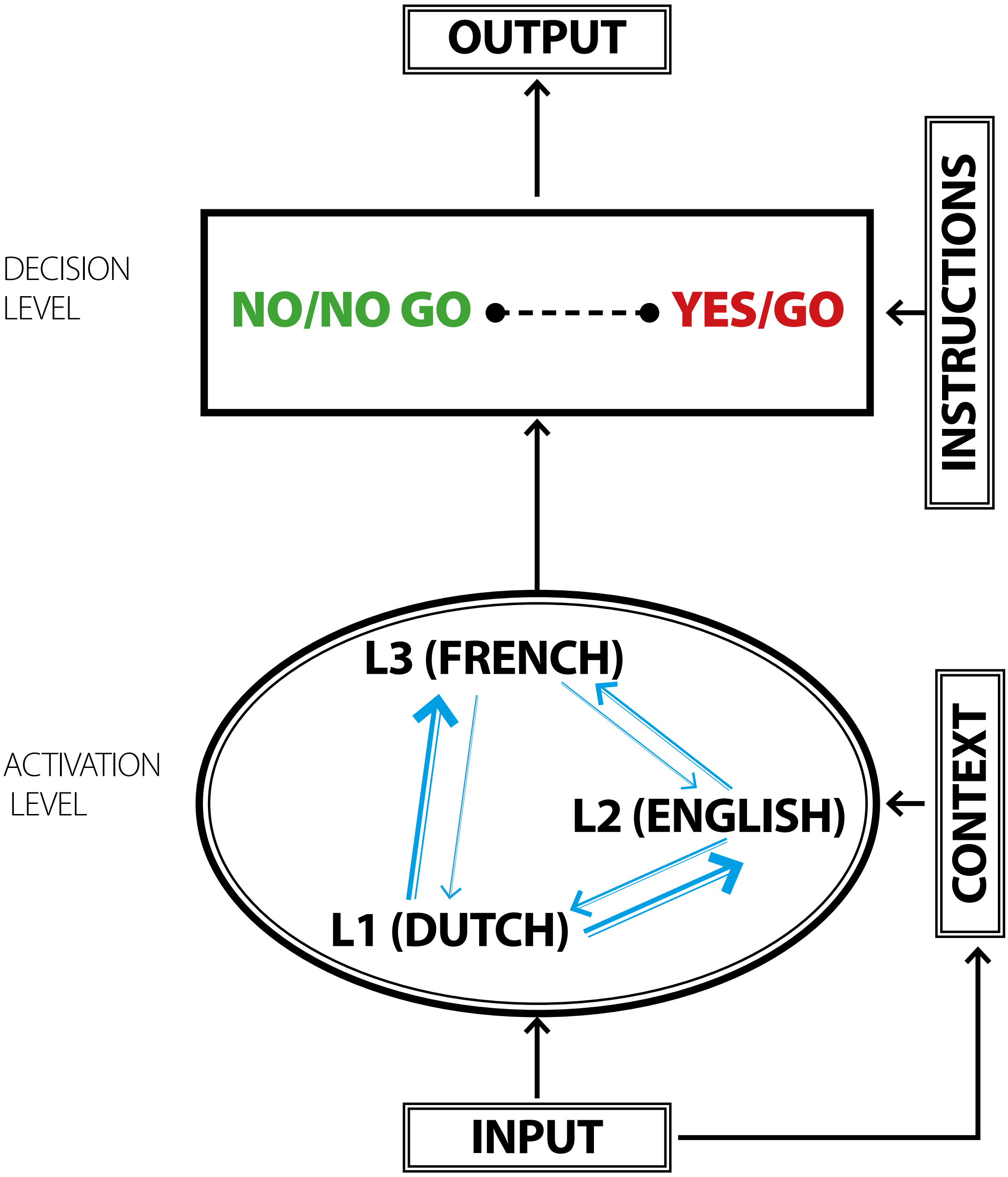


***Figure 6.1:** Extended BIA+ model for multilingual word recognition. The model includes different lexical representations on the ACTIVATION LEVEL and potential response conflict on the DECISION LEVEL. Potential effects of (linguistic) CONTEXT and INSTRUCTION influences are added at the appropriate positions of the model.*

Taking the data from Chapters 2 and 3 together, the contours arise of a multilingual processing model that is dynamic and time-sensitive to interactions between different types of codes/representations. In early stages of processing, orthographic codes become available, quickly followed by phonological and semantic codes. For cognates and false friends, resonance between activated codes is quickly established. The speed of activation of the different codes depends on the specific subjective frequency of usage of items, which is directly linked to the relative L1-L2-L3 proficiency of the participants. A more proficient language like L1 (Dutch) is activated more quickly, and a less proficient language like L3 (French) is activated more slowly, than an intermediately proficient language like L2 (English). When target items must be selected from sets of activated items, their relative activation levels at different moments in time are checked. We found that items of different languages can be co-activated, even if they differ in the degree to which they are activated.

Another very interesting finding in Chapters 2 and 3 was that words in a language appear to be activated to an extent that is unrelated to the presence of absence of other languages. For instance, item effects in the experiments with English-French context, involving Dutch-English interlingual words, were not significantly stronger than those observed in a purely English context.

The studies reported in Chapter 2 yielded important information on the representation of identical cognates. The data patterns can only be understood by assuming that identical cognates, with their cross-linguistic overlap in orthography, phonology and semantics, still have language-specific (but linked) representations that compete for selection in word recognition.

These studies further indicate that for trilingual participants the language membership of words becomes available quickly, but often only after the orthographic codes themselves have become active. The data shown here appear to be more in line with the BIA+ model than the BIA model (Dijkstra & van Heuven, 1998), because they do not show clear top-down effects of the language membership information of words on their activation in different contextual circumstances.

In our studies, we purposely included trilingual participants to be able to contrast the effects of stronger and weaker languages; we also contrasted the effects of different types of experimental tasks. As such, our results extend findings for people speaking two languages (bilinguals) and clarify the interactions between languages by incorporating a third one. It remains to be tested how the effects of syntactic and semantic sentence level characteristics interact with the dynamic system underlying word activation (e.g., Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Duyck, Assche, Drieghe, & Hartsuiker, 2007).

6.2.2 Decision level (non-linguistic influences)

The previous section considered how special and control words from different languages become activated during multilingual word recognition. However, it would be a mistake to consider the processes that lead to a response in our experimental tasks only in terms of word identification. As pointed out by models like BIA+, there is an important task/decision system that determines how the activated representations in the word identification system should be used (read-out) in order to arrive at a response in the task or even the trial at hand.

In the present research, we have considered how the activation flows of words in different languages are used in different tasks, such as the LDT and the GNGT. We observed that in somewhat different experiments the recognition latencies for interlingual homographs and cognates can differ, and we were able to unravel the systematic contribution of task demands, stimulus list composition, and item characteristics to the obtained differences. Furthermore, we investigated how the decision stage can be affected by instruction. On the whole, there was no evidence that non-linguistic expectations affected the participants' activation or decision processes when a study phase was included before the main experiment. This was not even the case when the participants learned in the study phase about the "correct meaning" of an interlingual homograph that was the later target reading of the homographs in the main experiment (Chapter 4). Instead of being affected by study phase, participants adopted decision criteria that depended on the actual experimental context with its potential conflicts during item selection. In sum, there were no differences in results between the experiments with or without study phase.

During the decision stage of the word recognition task in question, the correct response must be selected based on the set of activated lexical representations. This can only be done by resolving the competition and conflicts between the alternatives. When non-target (language) items are highly active, the resolution of competition takes time and is more prone to mistakes. One experimental task in which this was very visible was the GNGT. In the GNGT experiments in Dutch-English context with Dutch-English cognates (Chapter 2) or homographs (Chapter 3), non-target language activation led to strong inhibitory effects, even for cognates.

6.2.3 No interaction between the word identification and the decision system

As was already alluded to, the results of the present thesis do not support the existence of what may be called 'top-down' influences from the decision level to the lexical activation level (also De Groot, Delmaar, & Lupker, 2000; Dijkstra, Timmermans, & Schriefers, 2000). We obtained strong evidence that the recognition processes are primarily based on bottom-up activation, in other words, they are signal-driven. Thus, although we propose that the word identication system is interactive in the sense that different types of representations (orthographic, phonological, and semantic) can affect and strengthen each other in a process of resonance, the relationship between the word identification system and the decision system is only one-directional: Lexical activation is used, but not affected, by the decision system. The underlying argument here is that a theoretical distinction must be made between language as a cognitive domain and the cognitive control system supervising activity in various cognitive domains (the distinction also appears to be present in the brain, e.g., in the distinction of dACC / pre-SMA and LIPC activity[33]).

Based on the general observations in Chapters 2 and 3, we can say more about specific activations and interactions in the word identification system. For cognates and homographs, characteristics of the stimulus list in which they appear were found to influence the activation of their corresponding reading. For example, effects for Dutch-English interlingual words were

[33] dACC – dorsal anterior cingulate cortex; pre-SMA – anterior portion of supplementary motor area; LIPC – left inferior prefrontal cortex

weaker or absent in a purely English (L2) context and much stronger when the L1 (Dutch) was also included in the experiment. This suggests that the specific activation of lexical items in a language is sensitive to stimulus list, task, and item characteristics and may wax and wane in the course of an experimental session. At the same time, the experimental instructions (and practice sets) given to participants affect the way in which they use the dynamic activations in the word recognition system, adapting their response criteria to the specific trial conditions, previous trials, and, of course, the task demands more generally.

6.3 Future research directions

This study presents extensive research on interlingual homograph and cognate recognition mechanisms. However, some research issues related to each chapter should be considered in the future research.

In Chapter 2, we found support for the existence of multiple representations for identical cognates, and for differently sized cognate effects from L1 and L3 on L2. However, it is still important to investigate if effects in a task with L1 as target language may completely disappear when we present L1-L2 and L1-L3 cognates in pure L1 context, and if the effects also depend on language dominance (implying a stronger influence of L2 than of L3).

In Chapter 4, we did not obtain stronger inhibition effects when 10% rather than 0% non-target language items occurred in the experiment. In relation to this finding, at least two theoretical questions can be posed. First, what proportion of non-target language items is minimally necessary for yielding an inhibition effect for homographs? Second, is the interlingual homograph effect (and the associated non-target activation) observed in the presence of non-target language items gradual or all-or-none? Said differently, will the effect linearly grow with the proportion of non-target language items? Or does the effect suddenly appear when a certain proportion of non-target language items is reached?

Chapter 5 compared linguistic competition / control mechanisms to non-linguistic cognitive control mechanisms. It showed that there may be some shared domain general mechanisms in both tasks (cf. De Bruijn, Dijkstra, Chwilla, & Schriefers, 2001). However, more research is needed to strengthen this conclusion and to compare the linguistic control mechanisms to those in other cognitive tasks. This research would allow us to specify when exactly we can speak about shared or similar cognitive control mechanisms. It is highly likely that different types of tasks that require cognitive control are linked to different kinds of inhibitory control, as proposed by Miyake and Friedman (2012). Thus, a correlation would be observed between certain non-linguistic and certain linguistic tasks, but not for other task combinations.

6.4 Concluding remarks

This dissertation consists of a theoretical review and a core of several empirical chapters with detailed research on word recognition processes in multilinguals. It clarifies hitherto not completely understood aspects of the recognition process of frequently used item types like cognates and interlingual homographs. More specifically, it adds to our knowledge of currently 'hot' issues with respect to identical cognate processing and provides an extensive experimental overview and clarification of inhibitory processes in interlingual homograph recognition. The thesis findings, based on intricate manipulations of cognate and interlingual homograph characteristics show clear influences of task demands and stimulus list characteristics on lexical processing. Although these effects had already been shown, the current set of experiments also investigated the interaction between the two factors, which has never been done in earlier research. Moreover, in Chapter 3, besides the comparison between the condition means, several fine-grained analyses were performed to relate the homograph effect to the reaction time distribution. These analyses, which have never been performed in the context of the process of visual word recognition by multilinguals, yielded insights into the evolution of the effect across the entire reaction time distribution. The current set of results teaches us in some detail how the system of multilingual lexical identification operates under various experimental circumstances. Across the thesis, the presented research was theoretically framed in terms of the BIA+ model for multilingual word processing. The empirical data collected in the thesis were in line with many predictions of this model. Nevertheless, the thesis proposes some important extensions to the BIA+ model and provides more detailed information about the functioning of some of its components.

REFERENCES

Abutalebi, J., Della Rosa, P. A., Ding, G., Weekes, B., Costa, A., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. ***Cortex***, 49(3), 905–911. http://doi.org/10.1016/j.cortex.2012.08.018

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. ***Journal of Neurolinguistics***. http://doi.org/10.1016/j.jneuroling.2006.10.003

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 18(2), 234.

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. ***The Mental Lexicon***, 5(3), 436–461. http://doi.org/10.1075/ml.5.3.10baa

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

Baayen, R. H., & Milin, P. (2010). Analyzing Reaction Times. ***International Journal of Psychological Research***, 3(2), 12–28. http://doi.org/10.1287/mksc.12.4.395

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (Release 2). Linguistic Data Consortium. Philadelphia (PA): University of Pennsylvania.

Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence cognitive aging? ***Annals of Neurology***. http://doi.org/10.1002/ana.24158

Balota, D. a., & Yap, M. J. (2011). Moving Beyond the Mean in Studies of Mental Chronometry: The Power of Response Time Distributional Analyses. ***Current Directions in Psychological Science***, 20(3), 160–166. http://doi.org/10.1177/0963721411408885

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. ***Journal of Memory and Language***, 59(4), 495–523. http://doi.org/10.1016/j.jml.2007.10.004

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. ***Journal of Memory and Language***, 26(6), 658-672.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2007). Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. ***Journal of Experimental Psychology. Learning, Memory, and Cognition***, 33(5), 931–49. http://doi.org/10.1037/0278-7393.33.5.931

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. ***Developmental Psychology***. http://doi.org/10.1037/0012-1649.24.4.560

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. ***Developmental Psychology***. http://doi.org/10.1037/0012-1649.28.4.654

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. Canadian Journal of Experimental Psychology. http://doi.org/10.1037/cjep2006008

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. ***Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale***. http://doi.org/10.1037/a0025406

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. ***Journal of Neurolinguistics***, 21(6), 522–538. http://doi.org/10.1016/j.jneuroling.2007.07.001

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. ***Developmental Science***. http://doi.

org/10.1111/j.1467-7687.2004.00351.x

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. ***International Journal of Bilingualism***. http://doi.org/10.1177/13670069050090010701

Branzi, F. M., Calabria, M., Lucrezia, M., & Costa, A. (2016). Acta Psychologica On the overlap between bilingual language control and domain-general executive control. ***ACTPSY***, 166, 21–30. http://doi.org/10.1016/j.actpsy.2016.03.001

Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. ***Trends in Cognitive Sciences***. http://doi.org/10.1016/j.tics.2011.12.010.The

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. ***Journal of Experimental Child Psychology***, 109(4), 383–396. http://doi.org/10.1016/j.jecp.2011.03.012

Brysbaert, M. (1998). Word recognition in bilinguals: Evidence against the existence of two separate lexicons. ***Psychologica Belgica***.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? ***Bilingualism: Language and Cognition***, 13(3), 359–371. http://doi.org/10.1017/S1366728909990344

Bub, D. N., Masson, M. E. J., & Lalonde, C. E. (2006). Cognitive control in children: stroop interference and suppression of word reading. ***Psychological Science : A Journal of the American Psychological Society*** / APS, 17(4), 351–357. http://doi.org/10.1111/j.1467-9280.2006.01710.x

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*** (2006), 67(6), 1214–1241. http://doi.org/10.1080/17470218.2013.853090

Bybee, J. L. (1985). Morphology: A study of the relation between meaning and form. Philadelphia: John Benjamins.

Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2012). Qualitative differences between bilingual language control and executive control : evidence from task-switching MONOLINGUALS, 2(January), 1–10. http://doi.org/10.3389/fpsyg.2011.00399

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. Bulletin of the Psychonomic Society, 13(4), 212-214.

Cattaneo, G., Calabria, M., Marne, P., Gironell, A., Abutalebi, J., & Costa, A. (2015). The role of executive control in bilingual language production: A study with Parkinson's disease individuals. ***Neuropsychologia***. http://doi.org/10.1016/j.neuropsychologia.2014.11.006

Chomsky, N. (1965), Aspects of the Theory of Syntax, Cambridge, Massachusetts: MIT Press.

Clark, H. H. The language-as-fixed-effect-fallacy: A critique of language statistics in psychological research. ***Journal of Verbal Learning and Verbal Behavior***, 1973, 12, 335-359.

Coderre, E. L., & van Heuven, W. J. B. (2014). The effect of script similarity on executive control in bilinguals. ***Frontiers in Psychology***, 5, 1–16. http://doi.org/10.3389/fpsyg.2014.01070

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon.

Colomé, A., & Miozzo, M. (2010). Which words are activated during bilingual word production? ***Journal of Experimental Psychology. Learning, Memory, and Cognition***. http://doi.org/10.1037/a0017677

Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 34(2), 302–312. http://doi.org/10.1037/0278-7393.34.2.302

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***, (1998), 1–23. http://doi.org/10.1017/S1366728916000213

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. ***Cognition***, 106(1), 59–86. http://doi.org/10.1016/j.cognition.2006.12.013

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. ***Trends in Cognitive Sciences***. http://doi.org/10.1016/j.tics.2006.01.007

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. ***Journal of Cross-Cultural Psychology***, (9), 131–149.

Davis, C., Sánchez-Casas, R., García-Albea, J. E., Guasch, M., Molero, M., & Ferré, P. (2010). Masked translation priming: Varying language experience and word type with Spanish–English bilinguals. ***Bilingualism: Language and Cognition***, 13(2), 137. http://doi.org/10.1017/S1366728909990393

De Bot, K., Cox, A., Ralston, S., Schaufeli, A., & Weltens, B. (1995). Lexical processing in bilinguals. ***Second Language Research***. http://doi.org/10.1177/026765839501100101

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

De Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. ***NeuroImage***, 90, 348–359. http://doi.org/10.1016/j.neuroimage.2013.12.049

De Groot, A., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical Decision and Word Naming in Bilinguals: Language Effects and Task Effects. ***Journal of Memory and Language***, 47(1), 91–124. http://doi.org/10.1006/jmla.2001.2840

De Groot, A. M., Delmaar, P., & Lupker, S. (2000). The processing of interlexical homographs in translation recognition and lexical decision: support for non-selective access to bilingual memory. ***The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology***, 53(2), 397–428. http://doi.org/10.1080/713755891

De Jong, N. H., Schreuder, R., Harald Baayen, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. ***Language and Cognitive Processes***, 15(4–5), 329–365. http://doi.org/10.1080/01690960050119625

De Moor, W. (1998). Visuele woordherkenning bij tweetalige personen [Visual word recognition in bilinguals]. Unpublished master's thesis, University of Ghent.

Dijkstra, T. (2007). The multilingual lexicon. In M. G. Gaskell (Ed.), The Oxford handbook of psycholinguistics (pp. 250–265). Norfolk: Oxford University Press.

Dijkstra, T., Jaarsveld, V., Ten, B., Gollan, Forster, & Frost. (1999). Foreign Language Knowledge and Native Language Performance. ***Jared & Kroll Nas***, 9(4), 780–789.

Dijkstra, T., De Bruijn, E., Schriefers, H., & Ten Brinke, S. (2000). More on interlingual homograph recognition: language intermixing versus explicitness of instruction. ***Bilingualism: Language and Cognition***, 3(1), 69–78. http://doi.org/10.1017/S1366728900000146

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. ***Journal of Memory and Language***, 41(4), 496–518. http://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. ***Journal of Memory and Language***, 62(3), 284–301. http://doi.org/10.1016/j.jml.2009.12.003

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., & Harald Baayen, R. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. ***Language and Cognitive Processes***, 20(1–2), 7–41. http://doi.org/10.1080/01690960444000124

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On Being Blinded by Your Other Language: Effects of Task Demands on Interlingual Homograph Recognition. ***Journal of Memory and Language***, 42(4), 445–464. http://doi.org/10.1006/jmla.1999.2697

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

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). http://doi.org/10.1017/S1366728902003012

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

Donkers, F. C. L., & Van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. ***Brain and Cognition***, 56(2 SPEC. ISS.), 165–176. http://doi.org/10.1016/j.bandc.2004.04.005

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? ***Experimental Psychology***, 61(3), 234–251. http://doi.org/10.1027/1618-3169/a000243

Duyck, W. (2005). Translation and Associative Priming With Cross-Lingual Pseudohomophones: Evidence for Nonselective Phonological Activation in Bilinguals. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 31(6), 1340–1359. http://doi.org/10.1037/0278-7393.31.6.1340

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. ***Journal of Experimental Psychology. Learning, Memory, and Cognition***, 33(4), 663–679. http://doi.org/10.1037/0278-7393.33.4.663

Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. ***Behavior Research Methods, Instruments, & Computers***, 36(3), 488–499. http://doi.org/10.3758/BF03195595

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. a. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. ***Cognitive Brain Research***, 25(1), 57–70. http://doi.org/10.1016/j.cogbrainres.2005.04.007

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. ***Psychological Science***, 19(12), 1201–1206. http://doi.org/10.1111/j.1467-9280.2008.02224.x

Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. ***Bulletin of the Psychonomic Society***, 31(2), 119–122. http://doi.org/10.1111/j.1467-9280.2006.01821.x

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. Journal of Experimental Psychology. ***Learning, Memory, and Cognition***, 32(5), 1075–1089. http://doi.org/10.1037/0278-7393.32.5.1075

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? ***Bilingualism: Language and Cognition***, 9(2), 153. http://doi.

org/10.1017/S1366728906002501

Font, N. (2001). Rôle de la langue dans l'accès au lexique chez les bilingues : influence de la proximité orthographique et sémantique interlangue sur la reconnaissance visuelle de mots, Thèse de Doctorat non publiée, Université Paul Valéry, Montpellier.

Font, N., & Lavaur, J.-M. (2004). Effets de la fréquence du voisinage orthographique interlangue lors de la reconnaissance visuelle de mots chez les bilingues. ***L'Année Psychologique***, 104(3), 377–405. http://doi.org/10.3406/psy.2004.29673

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 10(4), 680–698. http://doi.org/10.1037/0278-7393.10.4.680

Forster, K., & Dickinson, R. (1976). More on the language-as-fixed effect: Monte-Carlo estimates of error rates for F1, F2, F', and minF'. ***Journal of Verbal Learning and Verbal Behavior***, 15, 135–142.

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. ***Behavior Research Methods, Instruments, and Computers***, 35(1), 116–124. http://doi.org/10.3758/BF03195503

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., … Zeileis, A. (2014). car: Companion to Applied Regression. In: ***R Package Version*** 2.0-21.

Gerard, L. D., & Scarborough, D. L. (1989). Language-Specific Lexical Access of Homographs by Bilinguals. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 15(2), 305-315. http://doi.org/10.1037/0278-7393.15.2.305

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-Language Intrusion Errors in Aging Bilinguals Reveal the Link Between Executive Control and Language Selection. ***Psychological Science***, 22(9), 1155–1164. http://doi.org/10.1177/0956797611417002

Grabo. R. P. (1931). A study of comparative vocabularies of junior high school pupils from English and Italian speaking homes. Bulletin No. 13. Washington, D.C.: U.S. Office of Education.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. ***Psychological review***, 103(3), 518.

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

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. ***Journal of Cognitive Psychology***, 25(July 2015), 1–16. http://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (1982). Life with Two Languages: An Introduction to Bilingualism. Cambridge, Mass: Harvard University Press, 370pp.

Grosjean, F. (2000). The bilingual's language modes. In J. L. Nicol (Ed.), One mind, two languages: Bilingual language processing (pp. 1–22). Oxford, England: Blackwell Publishing.

Hernandez, A. E. (2009). Language switching in the bilingual brain: What's next? ***Brain and Language***, 109(2–3), 133–140. http://doi.org/10.1016/j.bandl.2008.12.005

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? ***Acta Psychologica***, 127(2), 277–288. http://doi.org/10.1016/j.actpsy.2007.06.003

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

Kan, I. P., & Thompson-Schill, S. L. (2004). Selection from perceptual and conceptual representations. ***Cognitive, Affective & Behavioral Neuroscience***. http://doi.org/10.3758/CABN.4.4.466

Keijzer, M. C. J., & Schmid, M. S. (2016). Individual differences in cognitive control advantages of elderly late Dutch-English bilinguals. ***Linguistic Approaches To Bilingualism***.

http://doi.org/10.1075/lab.14032.kei

Kennette, L. N., & Van Havermaet, L. R. (2012). Interlingual homograph recognition by bilinguals: A new paradigm. ***The New School Psychology Bulletin***, 9(2), 7–16. http://doi.org/10.1037/e534932013-002

Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & De Bruijn, E. R. a. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. ***Brain Research***, 1068(1), 170–183. http://doi.org/10.1016/j.brainres.2005.10.087

Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. ***Behavior Research Methods***, 42(3), 627–633. http://doi.org/10.3758/BRM.42.3.627

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. ***Behavior Research Methods***, 42(3), 643–650. http://doi.org/10.3758/BRM.42.3.643

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. ***Journal of Cognitive Psychology*** (Hove, England). http://doi.org/10.1080/20445911.2014.929580

Kroll, J. F., & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. ***Journal of Cognitive Psychology*** (Hove, England), 25(5), 497–514. http://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. ***Journal of memory and language***, 33(2), 149-174.

Krott, A., Baayen, R. H., & Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. ***Linguistics***, 39(1; ISSU 371), 51-94.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). ***Annual Review of Psychology***, 62, 621–647. http://doi.org/10.1146/annurev.psych.093008.131123.Thirty

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. ***R Package Version*** 2.0-33.

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

Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO:Cognate effects in trilingual word recognition. ***Language and Cognitive Processes***, 19(5), 585–611. http://doi.org/10.1080/01690960444000007

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), 533–50. http://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. http://doi.org/10.1037/0278-7393.34.1.12

Li, P., & Farkas, I. (2002). 3 A self-organizing connectionist model of bilingual processing. ***Advances in psychology***, 134, 59-85.

Libben, M. R., & Titone, D. a. (2009). Bilingual lexical access in context: evidence from eye movements during reading. ***Journal of Experimental Psychology. Learning, Memory, and Cognition***, 35(2), 381–390. http://doi.org/10.1037/a0014875

Linck, J. a, Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical porcesses and inhibitory control. ***Mental Lexicon***, 3(3), 349–374. http://doi.org/10.1075/ml.3.3.06lin.

Loebell, H., & Bock, K. (2003). Structural priming across languages. ***Linguistics***, 41(5), 791–824. http://doi.org/10.1515/ling.2003.026

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. ***Bilingualism: Language and Cognition***, 6(2), 97–115. http://doi.org/10.1017/S1366728903001068

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. ***Bilingualism: Language and Cognition***, 11(1), 81–93. http://doi.org/10.1017/S1366728907003227

Massidda, D. (2013). Retimes: Reaction time analysis. ***R package Version*** 0.1-2.

McClelland, J. L., & Rumelhart, D. E. (1980). An Interactive activation model of the effect context in perception, part I., (5).

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2012). Effects of cognate status on word comprehension in second language learners: an ERP investigation. ***Journal of Cognitive Neuroscience***, 23(7), 1634–1647. http://doi.org/10.1162/jocn.2010.21463.Effects

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

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. ***Current Directions in Psychological Science***, 21(1), 8–14. http://doi.org/10.1177/0963721411429458

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. ***Journal of Cognitive Psychology***, 25(5), 531–546. http://doi.org/10.1080/20445911.2013.807812

Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. ***Neuropsychologia***, 66, 157–169. http://doi.org/10.1016/j.neuropsychologia.2014.11.014

Morton, J. (1969). Interaction of information in word recognition. ***Psychological Review***, 76(2), 165–178. http://doi.org/10.1037/h0027366

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. ***Journal of Memory and Language***, 53(4), 496–512. http://doi.org/10.1016/j.jml.2005.07.003

Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. ***Journal of Memory and Language***, 72(1), 59–84. http://doi.org/10.1016/j.jml.2013.12.004

Nachev, P., Wydell, H., O'Neill, K., Husain, M., & Kennard, C. (2007). The role of the pre-supplementary motor area in the control of action. ***NeuroImage***, 36(SUPPL. 2), T155–T163. http://doi.org/10.1016/j.neuroimage.2007.03.034

Namazi, M., & Thordardottir, E. (2010). A working memory, not bilingual advantage, in controlled attention. ***International Journal of Bilingual Education and Bilingualism***. http://doi.org/10.1080/13670050.2010.488288

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. ***Applied Psycholinguistics***, 28, 661–677. http://doi.org/10.1017/S014271640707035X

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. ***Cognitive, Affective & Behavioral Neuroscience***, 3(1), 17–26. http://doi.org/10.3758/CABN.3.1.17

Oliphant, G. W. (1983). Repetition and recency effects in word recognition. ***Australian Journal of Psychology***, 35, 393–403. http://doi.org/10.1080/00049538308258751

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. ***Cognitive Psychology***, 66(2), 232–258. http://doi.org/10.1016/j.cogpsych.2012.12.002

Paradis, M. (1978). Aspects of Bilingualism, Columbia, SC.: Hornbeam Press, 315 pp.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. ***Psychological Monographs: General and Applied.*** http://doi.org/10.1037/h0093840

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. ***Journal of Memory and Language***, 68(4), 315–332. http://doi.org/10.1016/j.jml.2012.12.003

Poort, E. D., Warren, J. E., & Rodd, J. M. (2016). Recent experience with cognates and interlingual homographs in one language affects subsequent processing in another language. ***Bilingualism: Language and Cognition***, 19(1), 206–212. http://doi.org/10.1017/S1366728915000395

Prior, A., & Gollan, T. H. (2011). Good Language-Switchers are Good Task-Switchers: Evidence from Spanish–English and Mandarin–English Bilinguals. ***Journal of the International Neuropsychological Society***, 17(4), 682–691. http://doi.org/10.1017/S1355617711000580

R Core Team. (2014). R Core Team (2014). R: A language and environment for statistical computing. R ***Foundation for Statistical Computing, Vienna, Austria***. URL http://www.R-Project.org/.

Raaijmakers, J., Schrijnemakers, J., & Gremmen, F. (1999). How to deal with "the language as fixed effect fallacy": common misconceptions and alternative solutions. ***Journal of Memory and Language***, 41, 416–426.

Raaijmakers, J. G. W. (2003). A further look at the language- as-fixed-effect fallacy. ***Canadian Journal of Experimental Psychology***, 57, 141–151.

Ratcliff, R. (1978). A theory of memory retrieval. ***Psychological Review***, 85, 59–108.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. ***Psychological Bulletin***, 86(3), 446–461. http://doi.org/10.1037/0033-2909.86.3.446

Ridderinkhof, K. R., Scheres, A., Oosterlaan, J., and Sergeant, J. (2005). Delta plots in the study of individual differences: new tools reveal response inhibition deficits in AD/HD that are eliminated by methylphenidate treatment. J. Abnorm. Psychol. 114, 197–215.

Ridderinkhof, R. K., Van den Wildenberg, W. P. M., Wijnen, J., and Burle, B. (2004). "Response inhibition in con- flict tasks is revealed in delta plots," in Cognitive Neuroscience of Attention, ed. M. I. Posner (New York: The Guilford Press), 369–377.

Roelofs, A. (2008). Dynamics of the attentional control of word retrieval: analyses of response time distributions. ***Journal of Experimental Psychology***. General, 137(2), 303–323. http://doi.org/10.1037/0096-3445.137.2.303

Roelofs, A., & Piai, V. (2017). Distributional analysis of semantic interference in picture naming. ***The Quarterly Journal of Experimental Psychology***, 70(4), 782–792. http://doi.org/10.1080/17470218.2016.1165264

Roelofs, A., Piai, V., & Rodriguez, G. G. (2011). Attentional inhibition in bilingual naming performance: Evidence from delta-plot analyses. ***Frontiers in Psychology***, 2(AUG), 1–10. http://doi.org/10.3389/fpsyg.2011.00184

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. ***Journal of Consulting Psychology***. http://doi.org/10.1037/h0043220

Rouder, J. N., Lu, J., Speckman, P., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. ***Psychonomic Bulletin & Review***, 12(2), 195–223. http://doi.org/10.3758/BF03257252

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. ***Psychological Review***, 89(1), 60–94. http://doi.org/10.1037/0033-295X.89.1.60

Saer, D. J. (1923). The Effect of Bilingualism on Intelligence. ***British Journal of Psychology***, 14(1), 25–38.

Sandra, D. (2009). Perspectives on language and cognition: from empiricism to rationalism and back again. In Cognition and pragmatics/Sandra, D.,Östman, J., Verschueren J. [edit.] - Amsterdam, Benjamins,1-15.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. ***Journal of Memory and Language***, 56(2), 153–171. http://doi.org/10.1016/j.jml.2006.10.002

Schreuder, R., & Baayen, R. H. (1997). How simplex complex words can be. ***Journal of Memory and Language***, 37, 118–139.

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

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: the effect of an irrelevant cue on information processing. ***The Journal of Applied Psychology***. http://doi.org/10.1037/h0020586

Smits, E., Martensen, H., Dijkstra, T., & Sandra, D. (2006). Naming interlingual homographs: Variable competition and the role of the decision system. ***Bilingualism***, 9(3), 281. http://doi.org/10.1017/S136672890600263X

Smits, E., Sandra, D., Martensen, H., & Dijkstra, T. (2009). Phonological inconsistency in word naming: Determinants of the interference effect between languages. ***Bilingualism: Language and Cognition***, 12(1), 23. http://doi.org/10.1017/S1366728908003465

Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: the effect on lexical access. ***Memory & Cognition***, 12(4), 380–386. http://doi.org/10.3758/BF03198298

Stafford T, Gurney K.N. Additive Factors Do Not Imply Discrete Processing Stages: A Worked Example Using Models of the Stroop Task. ***Frontiers in Psychology.*** 2011;2:287. doi:10.3389/fpsyg.2011.00287.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 30, 276-315.

Sternberg S. (1998). "Discovering mental processing stages: the method of additive factors," in An Invitation to Cognitive Science: Methods, Models, and Conceptual Issues, 2nd Edn, eds Scarborough D., Sternberg S., editors. (Cambridge, MA: MIT Press;), 702–863

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. ***Journal of experimental psychology***, 18(6), 643.

Van Abbenyen, L. (2015). Do Episodic Memory Representations Mediate Masked Priming Effects In The Lexical Decision Task ? A Confrontation Of The Prospective And Retrospective Accounts. PhD dissertation,. University of Antwerp.

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. ***Journal of Memory and Language.***

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. ***Psychological Science***, 20(8), 923–927. http://doi.org/10.1111/j.1467-9280.2009.02389.x

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.

van Heuven, W., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effect in bilingual word recognition. ***Journal of Memory and Language***, 39, 458–483.

van Heuven, W. J. B., 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. http://doi.org/10.1080/17470218.2013.850521

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

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: phonological priming from the second to the first language. ***Journal of Experimental Psychology. Human Perception and Performance***, 28(3), 616–627. http://doi.org/10.1037/0096-1523.28.3.616

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. ***Bilingualism: Language and Cognition***. http://doi.org/10.1017/S1366728914000352

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. ***Memory & Cognition***, 35(5), 938–952.

Von Studnitz, R., & Green, D. (2002). Interlingual homograph interference in German-English bilinguals: Its modulation and locus of control. ***Bilingualism: Language and Cognition***, 5(1), 1–23. http://doi.org/10.1017/S1366728902000111

Yeung, N., & Cohen, J. D. (2006). The Impact of Cognitive Deficits on Conflict Monitoring. ***Psychological Science***. http://doi.org/10.1111/j.1467-9280.2006.01680.x

Nederlandse samenvatting

1 De belangrijkste resultaten

Hier vatten we de belangrijkste bevindingen van dit proefschrift samen en leggen de link met hun belang voor tweetalig onderzoek en hun implicaties voor meertalige woordherkenningsmodellen. Het doel van dit proefschrift was om meer in detail te bestuderen hoe multilingualen interlinguale woorden verwerken die hun orthografie én semantiek over twee talen met elkaar delen (cognaten) of alleen hun orthografie (interlinguale homografen). In het bijzonder richten we ons op de rol van een aantal linguïstische en niet-linguïstische factoren die de herkenning van interlinguale woorden beïnvloeden, waaronder de stimuluslijstcontext, de taakvereisten en de dominantie van een taal. Daarnaast hebben we de cognitieve controlemechanismen bestudeerd die betrokken zijn bij de onderdrukking van de irrelevante betekenis van interlinguale homografen en hebben we deze mechanismen vergeleken met de mechanismen van cognitieve controle die gebruikt worden in niet-linguïstische taken.

In de experimenten van het huidige proefschrift staan een aantal vragen centraal. Hebben identieke cognaten meerdere representaties in het gedeelde meertalige lexicon? Wat is de invloed van verschillende linguïstische en niet-linguïstische factoren op de verwerking van cognaten en interlinguale homografen? Wat is het verband tussen de grootte van homograafinhibitie en de responssnelheid? Gebruiken proefpersonen vergelijkbare cognitieve controlemechanismen in linguïstische en niet-linguïstische taken? Al deze vragen werden onderzocht door in te zoomen op de bijdrage en combinatie van lexicale en post-lexicale processen tijdens het proces van meertalige woordherkenning.

De experimentele hoofdstukken die de kern van dit proefschrift vormen, benaderen de kwesties vanuit verschillende invalshoeken in termen van stimulusmateriaal, experimenteel design, onderzoekstechnieken, statistische analyses en proefpersonen. Een eerste belangrijke bijdrage van dit onderzoek betreft het gebruik van dezelfde set stimulusmateriaal in diverse experimenten, die slechts in geringe mate van elkaar verschilden in termen van hun opzet. De zorgvuldige, gedetailleerde manipulatie van onafhankelijke variabelen vergroot de kracht van statistische analyses over experimenten heen en zorgt voor een tot dusver niet benutte mogelijkheid tot theoretische vergelijking. Bovendien konden gedeelde kernkenmerken van het proces van meertalige woordherkenning onderscheiden worden van taakspecifieke componenten door de data in twee gerelateerde maar licht verschillende experimentele taken met elkaar te vergelijken (bv. lexicale-decisietaak (LDT) en go/no-go-taak (GNGT)). Ten slotte, door drie talen met elkaar te vergelijken in plaats van twee (zoals in het verleden meestal het geval was) konden we de effecten van woordcompetitie van de dominantere taal (L1) met die van de zwakkere taal (L3) op de doeltaal (L2) met elkaar vergelijken. Daarnaast hebben we ook de focus gericht op de cognitieve controlemechanismen bij de verwerking van interlinguale homografen en die vergeleken met de controlemechanismen in niet-linguïstische taken om na te gaan of een verklaring van homograafinhibitie op basis van domeinonafhankelijke inhibitieprocessen mogelijk is.

We zullen eerst alle hoofdstukken overlopen en de theoretische vragen op een rijtje zetten. Na de samenvatting van de belangrijkste empirische resultaten per hoofdstuk, zullen we die interpreteren in het licht van meertalige woordherkenningsmodellen.

Hoofdstuk 1 is een overzicht van de vakliteratuur over woordherkenning bij meertaligen. De centrale vraag is of het proces van lexicale activatie taalspecifiek of niet-taalspecifiek verloopt. Twee woordtypes lenen zich idealiter tot het beantwoorden van die vraag: cognaten en interlinguale homografen. Uit veel experimenten is gebleken dat interlinguale woorden de lexicale informatie activeren in alle talen waarin ze voorkomen. Meertaligen kunnen dus hun proces van lexicale activatie niet via een 'top down'-mechanisme onderdrukken. Pas op postlexicaal niveau wordt beslist welke geactiveerde informatie nodig is om correct te reageren binnen de specifieke taak. Dat blijkt zowel uit vele gedragsexperimenten als uit neurocognitieve experimenten. BIA+, het model dat door Dijkstra en Van Heuven (2002) werd ontwikkeld, kan de meeste experimentele effecten verklaren omdat het een onderscheid maakt tussen een activatiefase en een beslissingsfase. Terwijl de eerste fase op een niet-taalspecifieke wijze opereert, is de tweede fase gevoelig voor de experiment-specifieke factoren, zoals de doeltaal, de lijstsamenstelling en de aard van de experimentele taak.

In Hoofdstuk 2 rapporteerden we zes experimenten met trilingualen waarin drie factoren orthogonaal gemanipuleerd werden: cognaattype (L1-L2 en L2-L3), stimuluslijstsamenstelling (Engels, Engels-Frans en Engels-Nederlands) en taak (LDT en GNGT). Het doel was om na te gaan hoe identieke cognaten door trilingualen worden gerepresenteerd en verwerkt. Over de experimenten heen veroorzaakten de cognaten facilitatie-effecten, nuleffecten en zelfs inhibitie-effecten, ten opzichte van hun Engelse (L2) controles, afhankelijk van het soort cognaat, de stimuluslijstsamenstelling en de taak. Hoewel de proefpersonen sneller op de Nederlands-Engelse (L1-L2) cognaten reageerden dan op zuivere Engelse (L2) of zelfs zuivere Nederlandse (L1) woorden, reageerden ze even snel op de Engels-Franse (L2-L3) cognaten als op hun Engelse (L2) controles, maar waren ze veel trager in hun reacties op puur Franse (L3) woorden. In het algemeen zijn deze bevindingen compatibel met andere studies met betrekking tot cognaatherkenning (bv. Lemhöfer & Dijkstra, 2004; Peeters, Dijkstra, & Grainger, 2013; Van Hell & Dijkstra, 2002). We hebben onze bevindingen geïnterpreteerd in termen van een verklaring (a) waarin cognaten verschillende (morfologische) representaties hebben in de drie onderzochte talen en de activatiesnelheid van die representaties bepaald wordt door de dominantie van de taal, en (b) waarin de selectie van meerdere geactiveerde representaties in verschillende talen wordt beïnvloed door factoren op het beslissingsniveau. Een verklaring op basis van het beslissingsniveau is vereist omdat we verschillende uitkomsten observeerden in de twee taken: een gegeneralizeerde LDT en een GNGT. Niet alleen de grootte maar zelfs de richting van de cognaateffecten was taakafhankelijk. Als we de empirische literatuur als een geheel beschouwen, werden facilitatie-effecten gevonden in taken waar beide talen aanwezig waren en gebruikt konden worden voor het bepalen van een respons, terwijl inhibitie-effecten zich voordeden in taken waarbij responsen op items uit verschillende talen competitie veroorzaakten op selectieniveau (bv. Brenders, Van Hell, & Dijkstra, 2011; Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Font & Lavaur, 2004; Peeters et al., 2013). Onze bevindingen breiden de resultaten uit eerdere studies uit omdat ze een verschil aan het licht brengen tussen taken in termen van de beslissingsprocessen die ze vereisen. Meer specifiek, in een gegeneraliseerde LDT kan de herkenning van het woord onmiddellijk aanleiding geven tot een 'ja'-respons, terwijl het herkende woord in de GNGT verder geclassificeerd moet worden om een juiste respons te kunnen maken. Ook de vaststelling dat cognaatherkenning het best verklaard wordt door de aanname van verschillende representaties in verschillende talen, versterkt recente conclusies m.b.t. een voorlopig weinig onderzochte onderzoeksvraag.

In Hoofdstuk 3 en Hoofdstuk 4 werd gedetailleerd onderzoek gerapporteerd naar de inhibitorische controleprocessen die meertaligen gebruiken voor de herkenning van

Nederlands-Engelse (L1-L2) interlinguale homografen. In Hoofdstuk 3 werd de aanwezigheid van inhibitie systematisch vergeleken tussen twee taken (Engelse LDT en GNGT) en drie samenstellingen van de stimuluslijst (Engels, Engels-Frans en Engels-Nederlands). Zoals verwacht was het homograafeffect context- en taakafhankelijk. Hoewel er geen verschil werd waargenomen tussen de RT-resultaten in de puur Engelse en Engelse-Franse contexten, waren de effecten in de Engels-Nederlandse context veel sterker. Een vergelijking van de resultaten voor de Engelse LDT en de GNGT maakte duidelijk dat het laatste type taak veeleisender is, waardoor veel sterkere inhibitie-effecten ontstaan. Blijkbaar beïnvloedt het conflict dat tot inhibitie leidt niet alleen de relatieve concurrentie tussen meerdere betekenissen maar ook de beslissingsmechanismen in de twee taken.

In Hoofdstuk 3 voerden we ook analyses uit die verder gingen dan vergelijkingen tussen de conditiegemiddeldes (zoals bv. Balota, Yap, Cortese, & Watson, 2008; Roelofs & Piai, 2017). In ex-Gaussiaanse analyses, waarin de RT-distributie mathematisch opgesplitst wordt in een Gaussiaanse distributie (gemiddelde: mu; standaarddeviatie: sigma) en een exponentiële distributie (gemiddelde: tau), vonden we dat de lijstsamenstelling een effect had op mu, terwijl de taak een effect had op tau. Het homograafeffect op mu was significant in alle experimenten maar was significant groter in de Engels-Nederlandse lijst. Het homograafeffect op tau was significant groter in de GNGT. Het effect op mu betekent dat het inhibitie-effect alle responsen beïnvloedde (de snelle en de trage), wat tot een verschuiving van de hele RT-distributie leidde. Die verschuiving was groter in de Engels-Nederlandse lijst dan in de puur Engelse en de Engels-Franse lijst. Het effect op tau is een effect op de rechterstaart van de RT-distributie, m.a.w. de regio van de hoogste RT's. Het effect van de experimentele taak op tau betekent dat de cognitief meer belastende GNGT voor meer hogere RT's zorgt op de interlinguale homografen, wat een grotere inhibitie op beslissingsniveau impliceert. Er werd ook een significant homograafeffect op sigma gevonden. Dat effect was significant in beide taken maar significant groter in de GNGT dan in de LDT. Het interageerde niet met de lijstsamenstelling. De interlinguale homografen veroorzaakten significant meer spreiding in de RT's dan hun controles in L2. Dat suggereert dat er op individueel responsniveau veel variatie is in de grootte van het homograafeffect, wellicht op itemniveau (bv. als functie van de vertrouwdheid met beide lezingen) en op proefpersoonniveau (aangezien de vertrouwdheid met de L2-lezing wellicht sterk kan variëren tussen proefpersonen). Naast de ex-Gaussiaanse analyses bleek uit quantile-analyses dat het homograafeffect significant vergrootte naarmate de RT's groter werden. Deltaplotanalyses lieten zien dat het homograafeffect lineair groeide, m.a.w. dat de toename in inhibitie constant was per eenheid toename in de RT.

In Hoofdstuk 3 was er een zuivere Engelse contextconditie, waarin enkel Engelse doelwoorden aangeboden werden. In die conditie waren er geen zuivere Nederlandse woorden in de stimuluslijstcontext aanwezig en waren alleen de interlinguale homografen geen puur Engelse woorden. In de gemengde Engels-Nederlandse contextconditie werden 50% Nederlandse woorden gebruikt. Die konden uiteraard het Nederlandse lexicon activeren. Het gevolg van deze opname van Nederlandse woorden in de stimuluslijst bleek indirect de Nederlandse lezing van de homografen te versterken en resulteerde in een sterker inhibitie-effect voor de homografen. In Hoofdstuk 4 gingen we na of een kleiner percentage van de Nederlandse items in het experiment (namelijk 10%) ook de Nederlandse lezing van een homograaf zou kunnen versterken, waardoor het effect van homograafinhibitie mogelijk tussen de 0% en 50% condities zou liggen. Uit de resultaten bleek dat 10% Nederlandse woorden in de stimuluslijst niet genoeg was om de Nederlandse lezing van een homograaf meer te activeren dan in een 0%-conditie, in tegenstelling tot 50% Nederlandse woorden, die een grotere inhibitie-effect veroorzaakten. Een bijkomende vraag was of een studiefase het homograafeffect zou doen toenemen of afnemen, waarbij de assumptie was dat het opfrissen van de Engelse betekenis van de homografen voor het experiment gezien kan worden als een vorm van repetitiepriming. Echter, het studeren van Engelse zinnen waarin een homograaf voorkwam had geen invloed op de inhibitorische effecten in een zuivere Engelse context. Dit gebrek aan effect laat zien dat verbindingen in het kortetermijngeheugen niet noodzakelijk de langetermijnverbindingen beïnvloeden in het gedeelde meertalige lexicon.

Samenvattend, op basis van de totaliteit van onze bevindingen met interlinguale homografen hebben we geleerd dat niet alleen taalkenmerken (dat wil zeggen taaldominantie, taalcontext, enz.) de verwerking van interlinguale woorden beïnvloeden, maar ook taakvereisten en stimuluslijstkenmerken. Dat is in overeenstemming met eerder onderzoek (bv. Dijkstra et al., 1998). Door de meer gedetailleerde analyse van de RT-distributies voor homografen en hun L2-controles hebben we ook kunnen vaststellen dat het effect van stimuluslijstsamenstelling alle RT's binnen de RT-distributie beïnvloedt, d.w.z. voor een verschuiving van de Gaussiaanse component binnen die distributie zorgt, terwijl het effect van taakmoeilijkheid zich beperkt tot de langste RT's in die distributie. We hebben ook geleerd dat het homograafeffect toeneemt naarmate de responstijd toeneemt en dat die toename lineair is. Ten slotte is gebleken dat het homograafeffect significant in grootte toeneemt als er 50% Nederlandse woorden aan de lijst worden toegevoegd ('neen'-responsen) maar niet als dat percentage slechts 10% bedraagt.

In Hoofdstuk 5 werd onderzocht of meertaligen domeinonafhankelijke onderdrukkingsmechanismen gebruiken, die ze niet alleen toepassen voor het onderdrukken van interlinguale conflicten (linguïstische taken) maar ook voor het onderdrukken van conflicten in niet-linguïstische taken (bv. Bialystok, Craik, & Luk, 2008). We hebben gefocust op het proces van cognitieve controle bij de verwerking van interlinguale homografen om de gelijkenis tussen cognitieve controlemechanismen in linguïstische en niet-linguïstische cognitieve taken te onderzoeken. Naast een Engelse LDT in een Engels-Nederlandse context, hebben we bij eenzelfde groep proefpersonen een Simontaak of een AX-Continuous Performance Taak (AX-CPT) afgenomen. In de drie taaksituaties was er een duidelijk inhibitorisch hoofdeffect van de experimentele conditie (homograafeffect in LDT en congruentie-effect in AX-CPT en Simontaak). Echter, vergelijkbare controlemechanismen werden alleen gevonden in de LDT en de AX-CPT (AY-conditie, waar een conflict ontstaat tussen de aangeboden doelstimulus, nl. Y, en de verwachte stimulus, nl. X). Het type onderdrukkingsmechanisme in de AX-CPT lijkt gerelateerd te zijn aan het mechanisme dat gebruikt wordt om te reageren op de juiste (taakrelevante) lezing van een homograaf in de LDT. Het ontbreken van een correlatie tussen cognitieve effecten in de LDT en de Simontaak is niet verwonderlijk. Gezien de taakvereisten moeten er wel andere onderdrukkingsmechanismen gebruikt worden in de Simontaak dan in het geval van homograafverwerking in de LDT. Wellicht situeert het conflict in de eerste taak zich in de latere (niet-cognitieve) fase van het responsproces, terwijl het effect in de laatste taak zich op het vroegere niveau van responsselectie situeert. Echter, verder onderzoek is nodig omdat verschillende studies contradictorische resultaten hebben gevonden voor de Simontaak (bv. Paap & Greenberg, 2013). Met andere woorden, het is mogelijk dat tweetaligen verschillende mechanismen gebruiken voor de onderdrukking van irrelevante informatie in linguïstische en niet-linguïstische taken. Daarnaast werden verschillende theoretische posities aangenomen met betrekking tot domeinonafhankelijke en domeinspecifieke mechanismen voor linguïstische en niet-linguïstische taken (zie bv. Coderre & van Heuven, 2014; Duñabeitia et al., 2014).

In de volgende paragraaf zullen we de resultaten interpreteren in termen van een veelgebruikt model voor tweetalige woordherkenning, het Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & van Heuven, 2002). In het algemeen ondersteunen onze bevindingen dit model (en ook het Inhibitory Control model van Green, 1998), dat twee fasen van tweetalige taalverwerking voorstelt: een voorbewust lexicaal activatieniveau (waarvoor het lexicale identificatiesysteem verantwoordelijk is), waar woordrepresentaties worden geactiveerd, en een beslissingsniveau (waarvoor het taak- en beslissingssysteem verantwoordelijk is), waarbij meer bewuste, taakgeïnduceerde processen betrokken zijn. Volgens het BIA+ model beïnvloeden niet alleen linguïstische kenmerken het proces van woordherkenning, maar ook contextinformatie en taakvereisten. De bevindingen van onze cognaatstudies ondersteunen het BIA+ model en tonen aan dat cognaten verschillende (morfologische) representaties in verschillende talen hebben, zoals interlinguale homografen. Die assumptie is noodzakelijk om een coherent verhaal te creëren voor onze empirische resultaten, namelijk dat proefpersonen verschillend reageren op interlinguale woorden (cognaten of homografen), afhankelijk van de stimuluslijstsamenstelling en de taakvereisten. Ten slotte verduidelijkt onze studie van het proces van cognitieve controle enigszins de aard van de inhibitorische homograafprocessen. Wanneer we alle resultaten in onze experimenten combineren – die voor homografen, cognaten en processen van cognitieve controle – kunnen we een uitbreiding van het BIA+ model voorstellen, zoals afgebeeld in Figuur 1. In deze figuur hebben we meerdere talen toegevoegd en de effecten van de linguïstische context en instructies (taakvereisten) aangegeven. Voor de eenvoud verwijzen wij naar verschillende talen met een label. Men moet echter wel in gedachten houden dat lexicale representaties in elke taal orthografische, fonologische en semantische kenmerken hebben.

2 Implicaties voor het BIA+ model

2.1 Activatie van lexicale representaties

Wanneer een woord dat in meerdere talen bestaat (een interlinguale homograaf of cognaat) aan een lezer wordt aangeboden, wordt elk van zijn lezingen geactiveerd in een mate die afhangt van de linguïstische karakteristieken en de subjectieve frequentie van elke lezing. Daarnaast is de snelheid en mate van activatie van de relevante lezing afhankelijk van de cross-linguïstische overlap met de andere lezing op verschillende sublexicale en lexicale niveaus (orthografie, fonologie en semantiek). Volgens de assumpties van BIA+ begint het lexicale activatieproces van een woord bij zijn (sublexicale en lexicale) orthografische representaties maar breidt het zich dan snel uit naar zijn fonologische en semantische representaties.

Een aantal studies hebben al gefocust op de gevolgen van orthografische, fonologische en semantische overlap voor de verwerking van cognaten en interlinguale homografen (bv. Dijkstra, Grainger, & van Heuven, 1999). Die studies hebben over het algemeen vastgesteld dat een grotere overlap resulteert in sterkere cross-linguïstische effecten, vooral van L1 op L2. In onze experimenten hebben we opzettelijk gefocust op orthografisch identieke woorden met een gelijkaardige fonologie om de effecten te maximaliseren. Afhankelijk van hun subjectieve frequentie worden fonologische en semantische representaties eerder of later in de tijd geactiveerd. De implicatie hiervan is dat lexicale representaties in L2 over het algemeen iets langzamer worden geactiveerd dan lexicale representaties in L1 (de temporal delay assumptie in BIA+, zie ook Ferrand & Grainger, 1993). Die visie wordt ook ondersteund door ERP- en N400-studies. Bijvoorbeeld, Kutas & Federmeier (2011) hebben vastgesteld dat tweetaligen in beide talen primingeffecten en semantische congruentie-effecten vertonen op de N400, waarbij de timing (en in mindere mate de amplitude) van deze effecten een functie is van taalvaardigheid en de verwervingsleeftijd: effecten zijn kleiner voor talen die minder goed gekend zijn.

Contextkenmerken, zoals de volgorde van de items in een stimuluslijst of de samenstelling van de lijst, spelen een belangrijke rol tijdens de selectie van het doelwoord.

***Figuur 1:** Uitgebreid BIA+ model voor multilinguale woordherkenning*

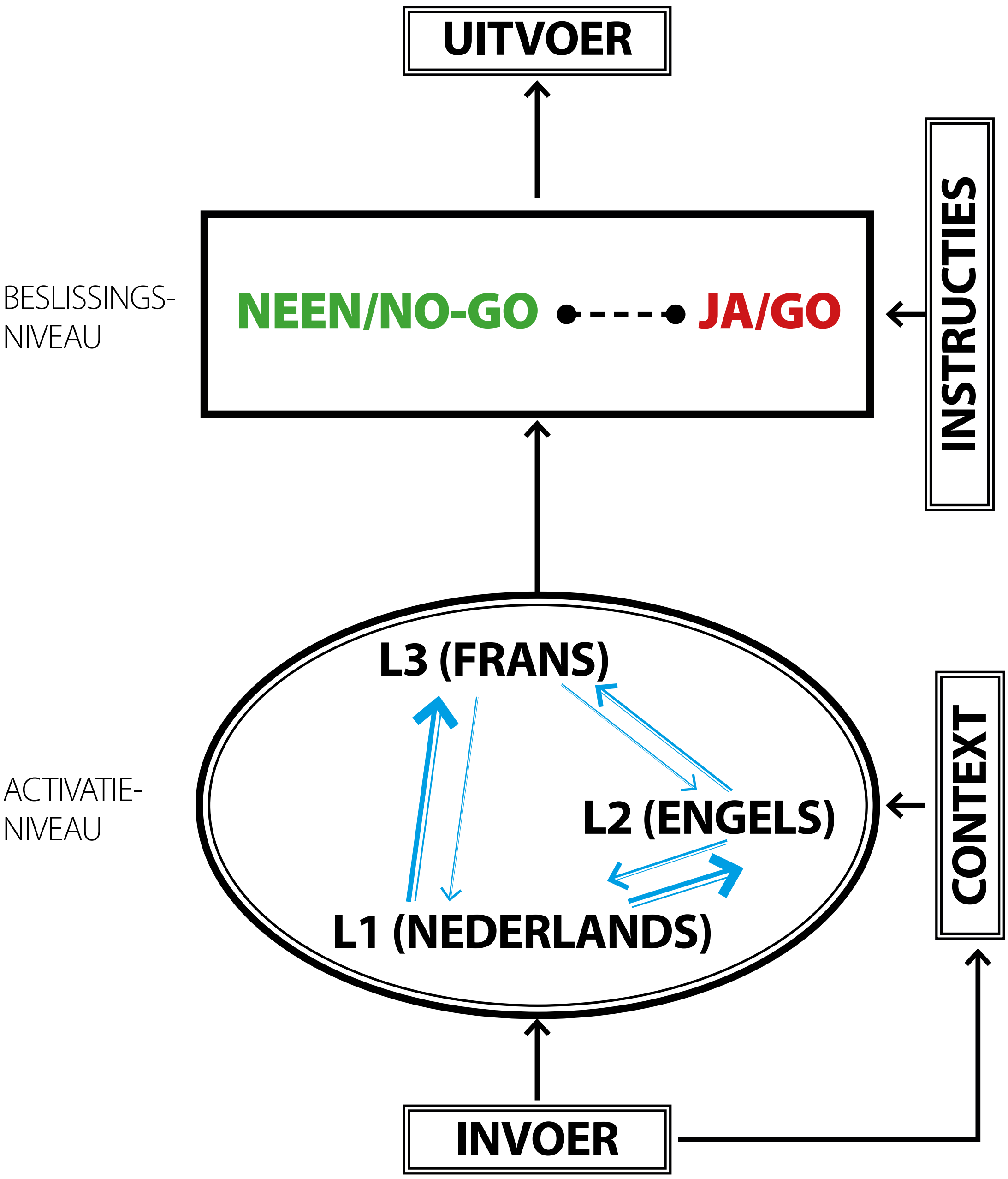


***Figure 1:** Uitgebreid BIA+ model voor multilinguale woordherkenning. Het model omvat verschillende lexicale representaties op het ACTIVATIENIVEAU en een potentieel responsconflict op het BESLISSINGSNIVEAU. Potentiële effecten van de invloed van de (linguïstische) CONTEXT en de INSTRUCTIE zijn in het model toegevoegd.*

Om de juiste reacties te maken, moeten de temporele deadlines van het woordselectieproces afgestemd worden op de lokale en globale context. In de experimenten van Hoofdstuk 2 en Hoofdstuk 3 hebben we kleinere effecten geobserveerd in een monolinguale Engelse context en in een niet-taakafhankelijke tweetalige taalcontext (bv. in een Engels-Franse context voor Nederlands-Engelse interlinguale woorden). De temporal delay assumptie is compatibel met de effecten van taaldominantie die in Hoofdstuk 2 werden gevonden. Daar waren de effecten van L1 op L2 sterker dan de effecten van L3 op L2. In Hoofdstuk 3 hebben we in alle samenstellingen van de stimuluslijst significante effecten gevonden (zelfs in de monolinguale Engelse context), wat aantoont dat de subjectieve frequentie van de interlinguale homografen hoger is in L1 dan in L2. Verdere ondersteuning is te vinden in Hoofdstuk 2, waar proefpersonen in de gegeneraliseerde LDT sneller reageerden op cognaten dan op puur Engelse of puur Nederlandse controlewoorden. Dit geeft aan dat proefpersonen op basis van de snelst beschikbare codes reageren, waardoor ze minder gevoelig zijn voor cross-linguïstische fonologische en semantische effecten, die veel tijd in beslag nemen. Die observaties werden ook in een aantal eerdere studies gedaan (bv. Dijkstra et al., 1999; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; Lemhöfer, Dijkstra & Michel, 2004).

Tezamen schetsen Hoofdstuk 2 en Hoofdstuk 3 de contouren van een meertalig verwerkingsmodel dat dynamisch en tijdsgevoelig is voor interacties tussen verschillende types representaties. In de vroege stadia van de verwerking worden orthografische codes beschikbaar, snel gevolgd door een activatie van de fonologische en semantische codes. Voor cognaten en valse vrienden komt de resonantie tussen de geactiveerde codes snel tot stand. De activatiesnelheid van de verschillende codes hangt af van de specifieke subjectieve frequentie van de items, die rechtstreeks verband houdt met de vaardigheid van de deelnemers in hun L1, L2 en L3. Ten opzichte van een intermediaire taal, zoals L2 (Engels), wordt een taal waarin iemand taalvaardiger, zoals L1 (Nederlands), sneller geactiveerd en een taal waarin iemand minder taalvaardig is, zoals L3 (Frans), langzamer geactiveerd. Wanneer doelitems geselecteerd moeten worden uit sets van geactiveerde items, worden hun relatieve activatieniveaus op verschillende momenten in de tijd gecheckt. Uit onze resultaten bleek dat items van verschillende talen gecoactiveerd kunnen worden, zelfs als ze verschillen in de mate waarin ze geactiveerd zijn.

Een andere bijzonder interessante conclusie in Hoofdstuk 2 en in Hoofdstuk 3 is dat woorden in een taal worden geactiveerd in een mate die niet afhankelijk is van de aanwezigheid van andere talen. Bijvoorbeeld, effecten op Nederlands-Engelse interlinguale woorden waren niet significant sterker in de experimenten met Engelse-Franse context dan in een zuiver Engelse context.

De in Hoofdstuk 2 gerapporteerde studies leverden belangrijke informatie op over de representatie van identieke cognaten. De datapatronen kunnen alleen worden begrepen door aan te nemen dat identieke cognaten, met hun cross-linguïstische overlap in orthografie, fonologie en semantiek, nog steeds taalspecifieke (maar met elkaar verbonden) representaties hebben die tijdens het proces van woordherkenning concurreren voor selectie.

Uit deze studies blijkt dat bij drietalige proefpersonen het taallidmaatschap van woorden snel beschikbaar komt, maar vaak pas nadat de orthografische codes zelf geactiveerd zijn. Onze gegevens zijn meer compatibel met het BIA+ model dan met het BIA-model (Dijkstra & van Heuven, 1998) omdat ze geen duidelijke topdowneffecten laten zien van het taallidmaatschap van woorden op de activatie van die woorden in verschillende contextuele omstandigheden.

In onze studies hebben we bewust drietalige proefpersonen getest om de effecten van sterkere en zwakkere talen op het proces van woordherkenning in L2 met elkaar te kunnen vergelijken. We hebben ook de effecten van verschillende experimentele taken vergeleken. Bijgevolg breiden onze resultaten eerdere bevindingen voor tweetaligen uit en verhelderen ze de interacties tussen talen door een derde taal in het onderzoek te betrekken. Het moet nog verder onderzocht worden hoe de effecten van syntactische en semantische kenmerken op zinsniveau interageren met het dynamische woordactivatiesysteem (bv. Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010b; Duyck, Assche, Drieghe, & Hartsuiker, 2007).

2.2 Beslissingsniveau (niet-linguïstische invloeden)

In het vorige deel werd besproken hoe interlinguale woorden en controlewoorden uit verschillende talen geactiveerd worden tijdens het proces van meertalige woordherkenning. Het zou echter fout zijn om de processen die in onze taken tot een respons leiden alleen in termen van woordidentificatie te beschrijven. Modellen als BIA+ poneren dat er een belangrijk taak/beslissingssysteem moet worden onderscheiden dat bepaalt hoe de representaties die het woordidentificatiesysteem activeert gebruikt moeten worden (uitgelezen) om tot een respons te komen in de experimentele taak of voor een specifieke trial in de lijst.

In het huidige onderzoek hebben we nagegaan hoe de activatiestromen die zich voordoen bij woorden in verschillende talen gebruikt worden in verschillende taken, zoals de LDT en de GNGT. We hebben kunnen vaststellen dat in sterk op elkaar lijkende experimenten de herkenningstijden voor interlinguale homografen en cognaten kunnen verschillen, en we hebben de systematische bijdrage van taakvereisten, stimuluslijstsamenstelling en itemkenmerken op de verschillen tussen de effecten kunnen ontrafelen. Daarnaast hebben we onderzocht hoe de beslissingsfase door de instructies kan worden beïnvloed. In het algemeen waren er geen aanwijzingen dat niet-linguïstische verwachtingen de activatie- of beslissingsprocessen van de proefpersonen beïnvloedden. Dit was zelfs niet het geval als de proefpersonen in een studiefase die aan het eigenlijke experiment voorafging de 'juiste betekenis' van de interlinguale homografen in Engelse zinnen kregen aangeboden, m.a.w. de betekenis die correspondeerde met de betekenis in de doeltaal, die later in het experiment aangeboden werd (Hoofdstuk 4). In plaats daarvan pasten ze de beslissingscriteria aan in functie van de feitelijke experimentele context, wat tot potentiële conflicten tijdens de selectie uit verschillende geactiveerde representaties leidde. Bijgevolg waren er geen verschillen tussen de resultaten in de experimenten met of zonder studiefase.

Tijdens de beslissingsfase van de woordherkenningstaak moet de juiste reactie worden gekozen op basis van de set geactiveerde lexicale representaties. Dit kan alleen als de competitie en de conflicten tussen de alternatieven wordt opgelost. Wanneer items in de niet-doeltaal sterk geactiveerd zijn, duurt de oplossing van het conflict langer en is dat proces meer vatbaar voor fouten. Eén experimentele taak waarin dit sterk tot uiting kwam, is de GNGT. In de GNGT-experimenten in de Nederlands-Engelse context met Nederlands-Engelse cognaten (Hoofdstuk 2) of homografen (Hoofdstuk 3) leidde niet-doeltaalactivatie tot sterke inhibitorische effecten, zelfs voor cognaten.

2.3 Geen interactie tussen het woordidentificatie- en het beslissingssysteem

Zoals hierboven al aangehaald werd, bieden de resultaten van het huidige proefschrift geen ondersteuning voor het bestaan van 'top-down'-invloeden van het beslissingsniveau op het lexicale activatieniveau (zie ook de Groot, Delmaar, & Lupker, 2000; Dijkstra, Timmermans, & Schriefers, 2000). We hebben daarentegen sterke evidentie gevonden dat de herkenningsprocessen voornamelijk gebaseerd zijn op een proces van bottom-upactivatie, m.a.w., dat ze datagedreven zijn. Hoewel we voorstellen dat het woordidentificatiesysteem zelf interactief is, in die zin dat verschillende types representaties (orthografische, fonologische en semantische) elkaar kunnen beïnvloeden en versterken in een resonantieproces, is de relatie tussen het woordidentificatiesysteem en het beslissingssysteem unidirectioneel: lexicale activatie wordt gebruikt door het beslissingssysteem maar wordt er niet door beïnvloed. Het onderliggende argument hierbij is dat er een theoretisch onderscheid moet worden gemaakt tussen taal als cognitief domein en het systeem van cognitieve controle dat de activiteit in verschillende cognitieve domeinen superviseert (dit onderscheid lijkt ook aanwezig te zijn in het brein, bv. in het onderscheid van dACC / pre-SMA en LIPC activiteit[34]).

Op basis van de algemene opmerkingen in Hoofdstuk 2 en Hoofdstuk 3 kunnen we meer zeggen over specifieke activaties en interacties in het systeem voor woordidentificatie. Voor cognaten en homografen bleken de eigenschappen van de stimuluslijst waarin ze verschijnen invloed te hebben op de activatie van hun bijbehorende lezing. Bijvoorbeeld, effecten voor Nederlands-Engelse interlinguale woorden waren zwakker of afwezig in een zuiver Engelse (L2) context en veel sterker wanneer de L1 (Nederlands) ook in het experiment werd opgenomen. Dit suggereert dat de specifieke activatie van lexicale items in een taal gevoelig is voor de samenstelling van de stimuluslijst, de experimentele taak en itemkenmerken, en dat die activatie kan toenemen en afnemen in de loop van een experimentele sessie. Tegelijkertijd beïnvloeden de experimentele instructies (en oefensets) die aan proefpersonen worden gegeven de manier waarop zij de dynamische activaties in het woordherkenningssysteem gebruiken, meer bepaald door hun antwoordcriteria aan te passen aan de specifieke experimentele omstandigheden, de vorige items en de taakvereisten.

3 Mogelijkheden voor toekomstig onderzoek

Hoewel deze studie uitgebreid onderzoek naar de herkenningsmechanismen van interlinguale homografen en cognaten rapporteert, moeten bepaalde aspecten in de diverse hoofdstukken dieper onderzocht worden in toekomstig onderzoek.

In Hoofdstuk 2 vonden we ondersteuning voor het bestaan van meerdere representaties voor identieke cognaten en voor verschillende groottes van de cognaateffecten van L1 en L3 op L2. Het is echter belangrijk om ook te onderzoeken of die effecten in een taak met L1 als doeltaal volledig verdwijnen wanneer L1-L2 en L1-L3 cognaten in een pure L1-context gepresenteerd worden en of de effecten ook dan afhankelijk zijn van taaldominantie (wat een sterkere invloed van L2 zou impliceren van L2 dan van L3).

In Hoofdstuk 4 vonden we geen sterker inhibitie-effect wanneer 10% items uit de niet-doeltaal (L1) aangeboden werden dan wanneer 0% van die items werden aangeboden. Met

[34] De Engelse afkortingen voor de dorsale cortex cingularis anterior (dACC); het anterieure deel van de supplementaire motorische cortex (pre-SMA) en de linkse inferieure prefrontale cortex (LIPC).

betrekking tot die bevinding kunnen tenminste twee theoretische vragen gesteld worden. Ten eerste, welk percentage van items uit de niet-doeltaal is nodig om een inhibitie-effect voor homografen te vinden? Ten tweede, is het homograafeffect (en de bijbehorende activatie van de niet-doeltaal) gradueel of een alles-of-nietskwestie? Met andere woorden, neemt het effect lineair toe met het aantal niet-doeltaalitems? Of verschijnt het effect plotseling wanneer een bepaalde proportie van de niet-doeltaalitems is bereikt?

In Hoofdstuk 5 vergeleken we controlemechanismen voor competitie in het talige domein met processen van cognitieve controle in het niet-linguïstische domein. De uitkomsten laten zien dat er in beide taken mogelijk domeinonafhankelijke mechanismen gebruikt worden (zie De Bruijn, Dijkstra, Chwilla, & Schriefers, 2001). Echter, meer onderzoek is nodig om deze conclusie te versterken en de linguïstische controlemechanismen te vergelijken met die in andere cognitieve taken. Dat zou ons toelaten om te specificeren wanneer we precies kunnen spreken over gedeelde of soortgelijke mechanismes voor cognitieve controle. Het is zeer waarschijnlijk dat verschillende soorten taken die cognitieve controle vereisen gekoppeld zijn aan verschillende soorten inhibitorische controle, zoals voorgesteld door Miyake en Friedman (2012). In die zin zou een correlatie kunnen worden waargenomen voor de combinatie van bepaalde niet-linguïstische en bepaalde linguïstische taken, maar niet voor andere combinaties van taken.

4 Concluderende opmerkingen

Dit proefschrift bestaat uit een theoretisch overzicht van de vakliteratuur en verschillende empirische hoofdstukken met gedetailleerd onderzoek naar woordherkenningsprocessen in meertaligen. Het verduidelijkt aspecten van het herkenningsproces van veelgebruikte itemtypes, zoals cognaten en interlinguale homografen, die tot op heden niet volledig begrepen werden. Meer specifiek draagt het bij tot onze kennis over een aantal 'hot issues' van dit ogenblik, zoals de verwerking van identieke cognaten, en biedt het een uitgebreid experimenteel overzicht en verheldering van inhibitorische processen bij interlinguale homograafherkenning. De bevindingen van dit proefschrift, die gebaseerd zijn op diverse manipulaties met cognaten en interlinguale homografen, brengen duidelijke invloeden aan het licht van taakvereisten en stimuluslijstkenmerken op lexicale verwerking. Hoewel die effecten al bekend waren, werd in de huidige experimenten ook de interactie tussen beide factoren bestudeerd, wat in eerder onderzoek nooit gebeurd was. Bovendien werden in Hoofdstuk 3, naast de vergelijking van de conditiegemiddeldes, ook diverse types analyses uitgevoerd om het homograafeffect te koppelen aan de reactietijddistributie. Die analyses, die nooit eerder werden uitgevoerd in de context van het proces van visuele woordherkenning bij meertaligen, leverden inzichten op in de evolutie van het effect over de volledige distributie van reactietijden. De huidige reeks resultaten leert ons in detail hoe het systeem van meertalige lexicale identificatie functioneert onder verschillende experimentele omstandigheden. Doorheen het proefschrift werd het gepresenteerde onderzoek theoretisch gekaderd binnen het BIA+ model voor meertalige woordverwerking. De empirische gegevens die in het proefschrift werden verzameld, waren in overeenstemming met de meeste voorspellingen van dit model. Bovendien stelt het proefschrift enkele belangrijke uitbreidingen voor het BIA+ model voor en levert het meer gedetailleerde informatie op over de werking van sommige van zijn componenten.