INVESTIGATION OF BILINGUAL LANGUAGE PROCESSING VIA PUPILLOMETRY

KONSTANTINE KOSTIA H. PHO

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN LINGUISTICS AND APPLIED LINGUISTICS (LINGUISTICS)

YORK UNIVERSITY

TORONTO, ONTARIO

MAY 2026

ABSTRACT

The present study investigates bilingual lexical processing through the use of pupillometry, with the aim of examining how language relatedness and cognate status influence cognitive effort during word recognition. Specifically, this research addresses three core questions: whether bilingual language processing differs depending on shared phonological and morphological features between language pairs; what the relationship between pupil dilation and mental effort reveals about cognate processing; and what pupillometric measures can tell us about lexical organization in the bilingual mind.

To this end, two bilingual populations were examined: English–French bilinguals (Group EF) and English–Mandarin bilinguals (Group EM). Participants completed a Language Discrimination Task across three conditions: a No Control condition requiring language identification of real words, a Control condition incorporating non-words in a standard lexical decision paradigm, and a Cognate condition involving language identification of cognate and non-cognate items. Throughout all tasks, pupil dilation was continuously recorded using eye-tracking technology and analyzed alongside behavioural measures of reaction time and accuracy.

The study revealed systematic effects of task demands and language similarity on bilingual lexical processing. Reaction time (RT) and accuracy data indicated that the No Control condition, involving only L1–L2 identification, yielded the fastest responses and highest accuracy across both language groups, reflecting its low cognitive demands. The Control condition, which required additional evaluation of word versus non-word status, produced slower RTs and reduced accuracy, demonstrating the inhibitory cost of non-word processing. The Cognate condition, while sharing

core L1–L2 identification demands with the No Control condition, elicited intermediate RTs and accuracy, reflecting additional processing demands from cross-linguistic lexical overlap. Crucially, differences between English–French and English–Mandarin bilinguals highlighted the role of language relatedness: EF bilinguals showed stronger cross-linguistic effects and greater cognitive effort in cognate trials, whereas EM bilinguals exhibited minimal influence of cognates, suggesting that linguistic similarity governs the engagement of cross-linguistic activation.

Pupillometry data converged with these behavioural findings, providing a sensitive physiological marker of cognitive load. Peak pupil dilation occurred in higher-demand conditions (Control and Cognate), reflecting increased mental effort, while the No Control condition produced minimal dilation. Cognate processing elicited larger pupil dilations in EF bilinguals compared with non-cognates, demonstrating that even familiar cognates can increase cognitive load due to cross-linguistic lexical competition, contrary to classical facilitation accounts. In EM bilinguals, cognate status had no significant effect on pupil dilation, consistent with the reduced overlap between English and Mandarin. Together, these results indicate that bilingual lexical processing is shaped by both task demands and the degree of similarity between language pairs. The findings support revisions to the BIA+ model (see Dijkstra & Van Heuven, 2002 [1]), incorporating hierarchical interactions between L1 and L2 and accounting for language dominance, and demonstrate that pupillometry provides a powerful complement to behavioural measures for revealing subtle cognitive mechanisms in bilingual word recognition.

DEDICATION

Veni, vidi, vici

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance, support, and generosity of many individuals, to whom I owe a great deal of gratitude. First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Chandan Narayan, whose expertise, patience, and thoughtful feedback shaped this research from its earliest stages to its completion. Your guidance has been invaluable, and I am deeply grateful for the time and care you invested in this work. Equal thanks are owed to my committee members, Professor Emily Elfner and Professor Liz Smeets, whose insightful comments and constructive critiques pushed this research to a higher standard. Your perspectives were instrumental in refining both the theoretical and empirical dimensions of this dissertation. And to my external reviewers, Dr. Lara Pierce and Dr. Ashley Chung-Fat-Yim, whose input and critique have allowed me to finalize this dissertation into what it is today.

I am grateful to the three undergraduate research assistants—Alexandra Borg, Muntaha Ahmed, and Veronica Ward-Bone—who, at the time, were undergraduate students generously dedicating a full year to supporting the practical demands of this project. Your diligence, reliability, and enthusiasm were a constant source of motivation, and this research is stronger for your contributions. Each of you has since gone on to pursue your own professional and academic paths, and I wish you nothing but the very best in all that lies ahead.

A special thank you goes to Jonah Jung, a friend I have had the pleasure of knowing since we met during an exchange program in Bordeaux, France, in 2017. Jonah's trilingual fluency in English, French, and Mandarin was indispensable in the development and validation of the word lists used in this study, and I am grateful that a friendship formed so far from home proved so valuable so many years later. Your willingness to lend your expertise and your time made a meaningful difference to the quality of the stimulus materials.

My thanks extend to two colleagues in the Department of Languages, Literature and Linguistics at York University, Marianne Laplante and Yifan Wu, whose technical insight and linguistic expertise were tremendously helpful. Your willingness to test the experimental code and offer further guidance on the word lists saved countless hours and contributed meaningfully to the integrity of the experimental design.

I am grateful to Professor Gang Pan and Professor Jia Ma, whose expert advice on the selection of Mandarin cognates ensured that the stimulus materials were both linguistically valid and contextually appropriate. Your contributions were essential to the rigour of this research.

Sincere thanks are owed to Dr. Edwin Dalmaijer of the University of Bristol, whose generosity in taking the time to discuss the PyOpenGaze wrapper proved to be a turning point in the technical implementation of this study. That chance conversation was more consequential than you may have realised, and I am grateful for it.

A final thanks goes to my dad, who, from beginning to end, graciously listened to my ramblings, no matter how nonsensical and crazy I might have sounded.

To everyone who played a role, however large or small, in supporting this Ph.D. journey from start to finish, I am truly grateful. The encouragement, conversations, and kindness I received along the way meant more than I can adequately express here. Thank you for your patience, your belief in me, and for reminding me of the world beyond the dissertation. This work is dedicated, in no small part, to you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

OVERVIEW

The question of how humans process language has been at the forefront of studies in cognitive science for quite some time. Within language processing, the last 30 years alone has seen interest in questions surrounding lexical organization, bilingual memory, and competing activation, particularly in the fields of psychology and psycholinguistics (see Grosjean, 2013 [15]). The study of bilingualism has been particularly important leading to insights into the so-called **bilingual advantage** (see Bialystok et al. 2003 [16]; Bialystok & Martin 2004 [17]; Mortan & Harper 2007 [18]), or lack thereof (see van den Noort et al. 2019 [19]; Paap et al. 2018 [20]; Paap et al. 2019 [21]; Ware et al. 2020) [22]. From a purely psychological standpoint, there is contradictory behavioural evidence to support and disprove the presence of a bilingual advantage—which refers to the cognitive benefits that bilingual individuals experience, such as enhanced executive function, better multitasking ability, and improved problem-solving skills, due to the mental flexibility required to manage multiple languages.

Within the domains of Lexical Processing (LP), bilingualism, and cognition, researchers have investigated the purported bilingual advantage in terms of processing capacity by examining factors such as speed of processing and accuracy (see Bialystok et al., 2004 [23]; Costa et al., 2008 [24]; Kroll Bialystok, 2013 [25]). Using variations of lexical decision tasks and reading tasks, researchers have aimed to determine how bilingual speakers are able to process two languages with a high degree of efficiency and accuracy (see Dijkstra & Van Heuven, 2002 [26]; van Heuven et al., 1998 [27]; Marian & Spivey, 2003 [28]).

More recently, psycholinguists have begun using eye-tracking technology—either alongside or independently of tasks such as lexical decision and self-paced reading—to gauge processing (Murujosa et al. 2020 [29]). This shift toward gaze-based methodologies in bilingual language processing research is crucial because these non-invasive measures provide a fine-grained, time-sensitive index of online processing, beyond what reaction times alone can reveal. In contrast to behavioral measures that primarily capture final decision outcomes, eye-tracking allows researchers to examine moment-to-moment processing dynamics during comprehension and production. Data from eye-tracking are particularly informative because they can be leveraged to investigate cognitive strategies and linguistic priorities as they unfold in real time.

Researchers commonly analyze fixation duration (i.e., total time spent looking at a target) and pupil dilation as indices of processing load. Shorter fixation times are generally interpreted as reflecting faster lexical access and therefore more efficient processing. In contrast, larger pupil size has been associated with greater mental effort and, consequently, longer processing times (see Hess Polt [30], 1964; Janisse, 1973 [31]; Winter et al. 2021 [32]).

The present study investigates pupil dilation as an index of language processing in a bilingual context, via the **cognate effect** in on-line speech processing. The **cognate effect**, also known as the **cognate facilitation effect**, refers to the advantage that cognates have over non-cognates in speed of language recognition and production. This effect has primarily been observed during oral and written language tasks. There exists a rich literature and compelling evidence to suggest that bilinguals produce and recognize cognates faster than non-cognates. By using tasks such as lexical decision to test for cognate facilitation effects (or lack thereof), and gating paradigms to test for on-line speech processing, the aim of the present study is to examine what pupil dilation can tell us about mental effort during language processing in the context of bilinguals. Further, these studies will be designed to engage with the predictions of models of bilingual language processing such the BIA+ model (Dijkstra & van Heuven 2002 [26]), thereby allowing for a diverse (behavioural and low-level pupillometric) set of data to either support or challenge theoretical understandings of the bilingual lexical access process.

In *Chapter II: Literature Review*, I will outline what bilingualism is, highlighting key theories in bilingualism, natural language processing (NLP), and bilingual language processing. I will then provide an overview of pupillometry, beginning with its usefulness in studies of attention and cognitive load, and subsequently argue that pupillometry offers particular methodological advantages for investigating bilingual language processing, especially in comparison to traditional behavioural measures. This discussion will be followed by a review of its application in psycholinguistic research. In *Chapter III: Current Study*, I will describe the experimental framework of this study in greater detail. Specifically, I compare English–French bilinguals (henceforth, Group EF) and English–Mandarin bilinguals (henceforth, Group EM) using a series of **language discrimination tasks**, in which participants decide the language of a presented stimulus word. These tasks are administered across three conditions: No-Control (NC), Control (CONT), and Cognate (COG). I will then present the study's hypotheses and predicted outcomes. In *Chapter IV: Results and Analysis*, I will go over the core findings and results of this research, before then concluding in *Chapters V: Discussion* and *Chapter VI: Conclusion* with the broader implications of these findings for psycholinguistics and our understanding of natural language processing in bilingual contexts.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

This chapter situates the present study within the broader theoretical and empirical landscape of bilingual language processing and psychophysiological measurement. It begins by defining bilingualism and outlining the key dimensions—such as proficiency, dominance, and age of acquisition—that shape bilingual experience and influence lexical processing. The chapter then reviews major theories of language processing, with particular attention to models of lexical representation in both monolingual and bilingual speakers. Competing accounts of bilingual lexical organization, ranging from language-selective access to non-selective, integrated models, are examined in order to clarify how cross-linguistic activation, lexical competition, and control mechanisms have been conceptualized in the literature.

Building on this theoretical foundation, the chapter turns to conceptual and empirical issues central to bilingual research. It defines cognates and reviews their significance for understanding cross-language interaction, before surveying key experimental paradigms used to investigate bilingual lexical access, including **Picture–Word interference tasks**, **Auditory Word Recognition paradigms**, and the **Lexical Decision Task (LDT)**. These frameworks provide the methodological context for studying how bilinguals process and differentiate between their languages.

Finally, the chapter introduces **pupillometry** as a measure of attentional engagement and cognitive effort. After outlining the physiological basis of pupil dilation, it reviews its growing application within psycholinguistics as an online index of processing load. By integrating insights from bilingual lexical theory and pupillometric methodology, this chapter establishes the conceptual groundwork for the current study and highlights the empirical gap it seeks to address.

2.2 Defining Bilingualism

In the simplest of terms, **bilingualism** is the phenomenon of individuals being fluent in two languages. This simple definition almost entirely mirrors the definition offered by Uriel Weinreich, one of the founding fathers of studies in bilingualism, in his 1968 book *Languages in Contact: Findings and Problems* [33]. Weinreich defines bilingualism as "*The practice of alternately using two languages*" (Weinreich, 1968 [33]). However, this definition alone leaves much to be desired and does not sufficiently capture the nuances of what it means to be bilingual. In fact, this general definition poses issues for those studying the cognitive and linguistic phenomenon that is bilingualism. To this end, I highlight two crucial questions which need to be answered in order to better understand bilingualism, namely: (i) *how can we quantify fluency?* and (ii) *does being bilingual necessarily mean an individual has a perfect mastery of both languages?*

In *Language*, Bloomfield (1933) attempts to define bilingualism by taking into account those two questions:

> *In the cases where this* ***perfect foreign-language learning is not accompanied by loss of the native language, it results in 'bilingualism', native-like control of two languages****. After early childhood few people have enough muscular and nervous freedom or enough opportunity and leisure to reach perfection in a foreign language; yet bilingualism of this kind is commoner than one might suppose, both in cases like those of our immigrants and as a result of travel, foreign study, or similar association. Of course, one can-not define a degree of perfection at which a good foreign speaker becomes a bilingual: the distinction is relative.* (Bloomfield, 1933 [34])

While Bloomfield's definition certainly does answer the question of language mastery, Hoffman (1991) [35] observes that this definition is not without some level of contradiction, namely: *if one cannot define 'a degree of perfection' in bilingualism, how can we talk of 'perfect foreign-language learning'?* (Hoffman, 1992 [35]). Hoffman highlights the definition put forward by Mackey [36] [37], which incorporates both Weinreich's alternate use of language and Bloomfield's point about proficiency:

> *It seems obvious that if we are to study the phenomenon of bilingualism we are forced to consider it as something entirely relative. We must more-over include the use not only of two languages, but of any number of languages.* ***We shall therefore consider bilingualism as the alternate use of two or more languages by the same individual****.* (Mackey, 1970 [37])

Regardless of which definition of bilingualism we decide to subscribe to, it is clear that the issue of quantifying fluency remains. That said, until linguists can agree on a de facto way to quantify fluency, it would be prudent to subscribe to the view that it is not necessary that all bilinguals have full mastery of both languages since bilinguals, even balanced bilinguals[1], seldom have complete command and control over both of their languages. This is corroborated by the fact that bilingualism is hardly a symmetric phenomenon, with external factors such as **Age of Acquisition (AoA)**, **order of acquisition**, and **frequency of use** all contributing to language dominance—i.e., the degree to which a person is more proficient in, or prefers to use, one language over others in different contexts or aspects of communication (see Brysbaert & Ghyselinck, 2006 [39]). This, in turn, is likely to influence bilingual language production and processing. In other

[1] Balanced bilingualism refers to an individual who has balanced usage of and balanced proficiency in two languages (Yow & Li, 2015 [38]). Many researchers have suggested that the concept of a balanced bilingual is an abstraction only, and that most, if not all bilinguals experience some form of language dominance, wherein one of their languages is more developed than the other.

words, an individual's experience with their languages (L1 and L2) will condition the way they function vis-à-vis language production (i.e., speech) and language processing (i.e., perception).

For the purposes of this research, I distinguish between two main groups of bilingual speakers: simultaneous bilinguals (i.e., those exposed to L1 and L2 from birth or very early in development) and sequential bilinguals (i.e., those who acquire L2 after the onset of L1). Importantly, sequential bilingualism refers to acquisition order and does not necessarily imply adult L2 acquisition; Age of Acquisition (AoA) more precisely captures the timing of L2 onset along a developmental continuum. This distinction is relevant because AoA has been argued to play a role in ultimate attainment and processing efficiency, potentially due to maturational factors (Brysbaert & Ghyselinck, 2006 [39]).

In the context of the present study, the sequential bilinguals under investigation acquired their L2 during childhood rather than post-puberty, though they differ in relative proficiency and patterns of language use. Previous research suggests that simultaneous bilinguals may, under certain conditions, exhibit processing advantages—particularly in tasks requiring rapid lexical access—when both languages have been acquired early and maintained at relatively high proficiency (Yow & Li, 2015 [38]). However, such advantages would not necessarily apply equally to both languages, nor can they be assumed solely on the basis of acquisition type; they depend on proficiency, continued exposure, and context of use.

It is therefore crucial to emphasize that (i) AoA alone does not determine fluency or processing outcomes, and (ii) language dominance is a multifaceted construct encompassing relative proficiency, frequency of use, functional distribution across contexts, and patterns of exposure (see Montrul, 2015 [40]). Dominance may differ across modalities and domains, and it can shift over time. For this reason, acquisition type must be interpreted alongside more fine-grained measures of proficiency and dominance when examining bilingual language processing.

In the context of this dissertation, the preceding discussion is important because it frames the participant populations and the factors that may influence bilingual language processing. Specifically, by distinguishing between simultaneous and childhood sequential bilinguals, and by considering Age of Acquisition (AoA) alongside relative proficiency and language dominance, the study is able to investigate how these variables interact to affect processing efficiency and cognitive load. While AoA is not the sole focus of this research, understanding its potential influence helps contextualize observed differences in performance between groups. This section therefore provides the necessary background for interpreting experimental results and highlights why controlling for acquisition timing, dominance, and proficiency is essential in evaluating the cognitive mechanisms underlying bilingual language processing.

Bilingualism is therefore not necessarily defined by perfect mastery of two separate languages, but rather by an individual's ability to communicate and function effectively in two languages across a range of contexts. A definition of bilingualism must also consider the relative balance of the two languages within a speaker's mind, as bilinguals often differ in terms of language dominance and proficiency. In some cases, the two languages may be relatively balanced, whereas in others one language may be more dominant due to differences in exposure, use, or Age of Acquisition (AoA). This balance is closely related to whether the languages were acquired simultaneously or sequentially: simultaneous bilinguals are typically exposed to both languages from early childhood, often resulting in more comparable levels of proficiency, while sequential bilinguals may develop stronger dominance in the first-acquired language due to earlier and more sustained exposure. These differences in acquisition history and relative dominance can shape how the two linguistic systems are organized and accessed during processing. In sum, to define bilingualism is to understand the complexities and intricacies of the human mind when managing two separate linguistic systems.

With this understanding of bilingualism and its key dimensions in place, the following section turns to the theoretical frameworks that have been proposed to account for how language is processed and represented in both monolingual and bilingual speakers.

2.3 Theories in Language Processing

Before engaging in an in-depth discussion of bilingual language processing, it is necessary to first outline the fundamental mechanisms of natural language processing as they operate in both monolingual and bilingual speakers. Crucially, we must note two key theories of language processing: first, a majority of theories in speech production agree on two fundamental points: (a) that semantic, syntactic, and lexical forms constitute independent levels of representation, and (b) that these levels of representation are likely accessed sequentially over the course of language production (Caramazza, 1997 [41]). Secondly, the dominant view on lexical access is that the process itself involves at least two distinct steps: the first step involves selecting a semantically and syntactically specified lexical representation (i.e., lemma), followed by the selection of its corresponding lexical-phonological representation (e.g., lexeme) (see Bock, 1982 [42]; Butterworth, 1989 [43]; Burke, MacKay, Worthley, & Wade 1991 [44]; Bock & Levelt, 1994 [45]).

In the case of monolinguals, this process follows a relatively direct route; in a model described by Kay et al. (1992) [6] (see Figure 2.1), we can see a distinction of three modalities for input: auditory/speech, visual in the form of pictures, and visual in the form of orthograph. In this model, stimuli are received and analyzed at what can be taken to be the sublexical level; in other words, data has not yet been lexicalized, but is instead being analyzed and identified in what would constitute the lemma selection stage. In the next stage, the stimulus is processed by a recognition system, or by an input to the lexicon, which sends information directly back to the semantic system. This stage can be taken to be the lexicalization stage, or lexeme selection stage. Finally, there is an output stage which simply shows that the word recognition process—or lexical decision—is complete.

While Kay et. al.'s (1992) [6] model for monolingual language processing captures all the key points in major theories of speech processing, we note that their model is insufficient for capturing the complexity of bilingual language processing. In the case of bilinguals, we note that language processing is not as direct as tapping into a single repertoire and mapping across all levels of selection. In fact, bilinguals must deal with two languages, which, as evidence suggests, are consistently in parallel activation. This parallel or simultaneous activation in turn leads to what is known as competing activation.

In the study by Marian and Spivey (2003) [28], Russian–English bilingual participants completed a visual-world eye-tracking task designed to investigate parallel activation of both languages. Participants were presented with a target object (e.g., a picture of a cat) along with several distractor objects. The distractors were chosen to create three types of phonological overlap: (i) within-language competitors, whose names sounded similar to the target in the same language as the target (*cap* in English), (ii) between-language competitors, whose names sounded similar in the other language (*kot* in Russian), and (iii) simultaneous competitors, whose names overlapped with the target in both languages. The non-target language refers to the language that was not required for the current trial—i.e., the language participants were instructed not to use in identifying the target.

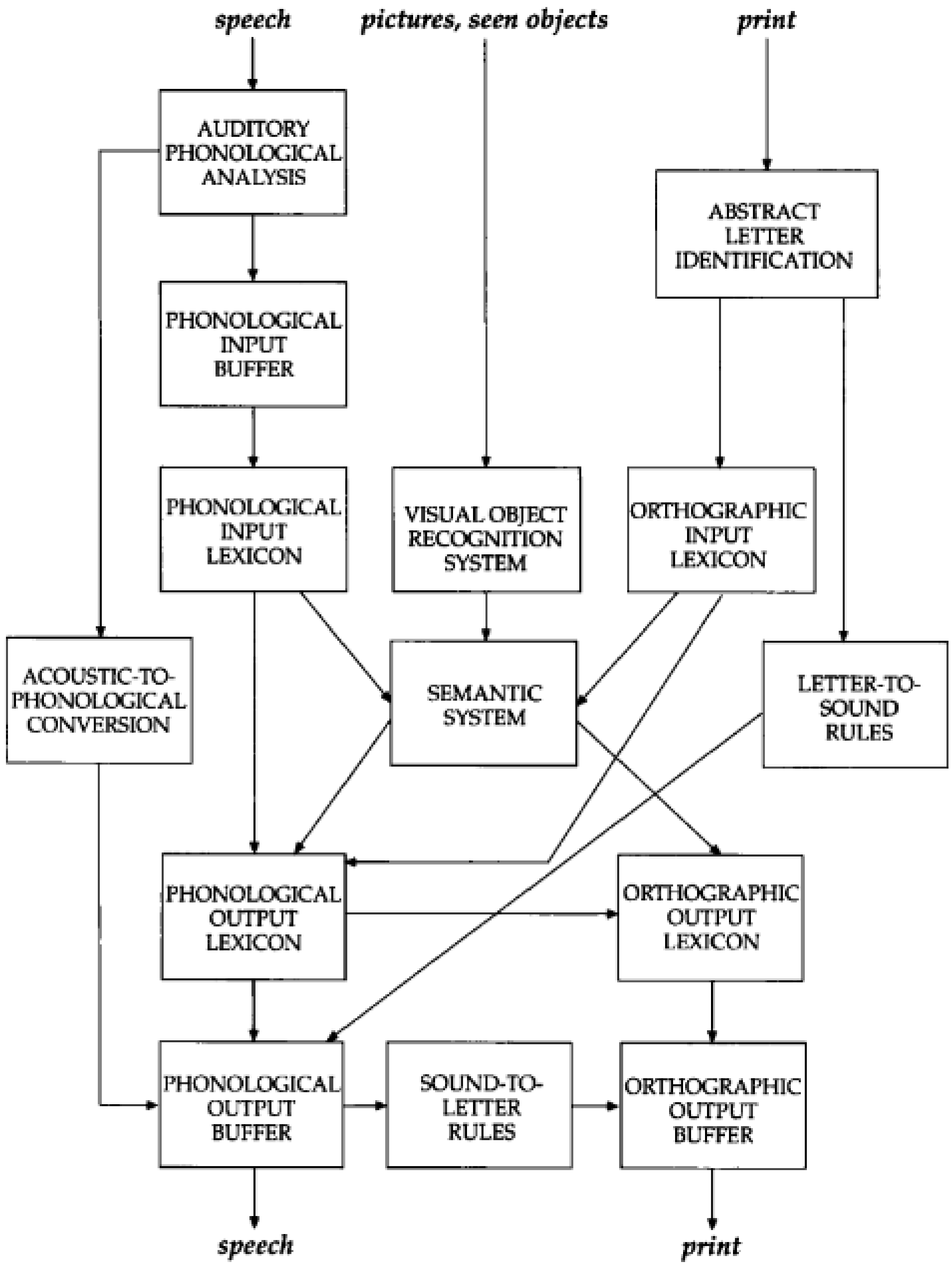


Figure 2.1: A model for monolingual language processing put forward by Kay et. al. (1992) [6]. Stimuli is divided into 3 categories based on perception modality: Speech (aural), Pictures (visual), and Print (visual/read). Based on the modality of perception, stimuli are treated differently, and are processed at different levels. However, regardless of modality, all stimuli must necessarily call back to the semantic system to parse meaning.

By comparing eye movements across these conditions, the researchers observed that when only one language needed to be active (e.g., the target was in English and participants were instructed to respond in English), competition was largely confined to within-language distractors, and there was minimal interference from the non-target language. In contrast, when the task required participants to consider both languages or allowed for potential cross-language activation, between-language competition emerged consistently alongside within-language competition. For example, participants looking for a target named *cat* in English might initially fixate on the Russian competitor *kot*, demonstrating parallel activation across languages. These results illustrate that lexical activation in bilinguals depends heavily on task demands: the target language and its context determine whether competition is confined to a single language or extends across both languages.

In short, we observe within-language competition effects in a single language processing only. When the task requires only one language to activate, the within-language competition has little to no effect on the non-target language. In contrast, if the task requires both languages to be active, thus creating a parallel activation, we observe that between-language competition occurs consistently between and within both languages (see Spivey & Marian, 1999 [46]; Marian & Spivey, 2003 [28]). However, whether competition occurs depends heavily on external factors that can influence the degree of activation of a target or non-target language, such as relative proficiency, frequency of use, or the context in which a language is presented. In other words, lexical activation in bilinguals is not fixed but varies depending on these situational and individual factors.

To better understand the differences between monolingual and bilingual NLP, the following sections will explore the theories of NLP presented here, comparing them to models that have been proposed to represent the complexities of language processing.

2.3.1 Representation of the Monolingual and Bilingual Lexicon

Bilinguals are faced with parallel and competing activation. What this suggests is that, depending on task requirements, the lexicon of both languages may be activated at the same time (i.e, parallel), or the lexicon of one language may dominate the other (i.e., competing) (see Marian & Spivey, 2003 [28]). The nature of representation in the lexicon is critical for understanding why competing activation is occurring; selective and non-selective views offer differing predictions and provide a lens with which to examine representation and organization in the bilingual mind.

The **language-selective** view predicts that linguistic input (auditory or visual) in one language should activate only the target language. In contrast, the **language non-selective** view predicts that linguistic input in one language can induce co-activation of both languages (Wang, Hui & Chen, 2020 [47]). Based on the evidence presented in the previous section, the language non-selective view is the more plausible account of bilingual lexical access and will serve as the baseline assumption for this dissertation. In other words, for the purposes of the current study, we assume that bilinguals automatically activate lexical representations from both languages in parallel, even when only one language is contextually relevant.

It is important to note that "single-language processing" in bilinguals is a theoretical idealization rather than a reflection of empirical reality. Behavioural and eye-tracking studies consistently demonstrate cross-language activation and competition effects across modalities, including visual word recognition and sentence processing (Dijkstra et al., 1998 [48]; Marian & Blumenfeld, 2007 [49]; Libben & Titone, 2009 [50]) and spoken-word comprehension in the visual-world paradigm (Marian & Spivey, 2003 [28]). Crucially, these effects arise even in tasks that do not explicitly require selecting between languages, indicating that non-selective activation is automatic and difficult to suppress (Blumenfeld & Marian, 2013 [51]).

Taking this language non-selective view together with evidence for parallel and competing activation, we can now turn to the question of how the bilingual mental lexicon is organized under these constraints. The presence of systematic cognate facilitation, interlingual homograph interference, and phonological competition across languages supports models in which lexical representations from both languages are integrated within a shared system and interact dynamically during processing (van Heuven et al., 1998 [27] ; Dijkstra & van Heuven, 2002 [52]). From this perspective, language membership does not function as an early filter on lexical access, but rather as a later-acting constraint that modulates competition once multiple candidates have already been activated.

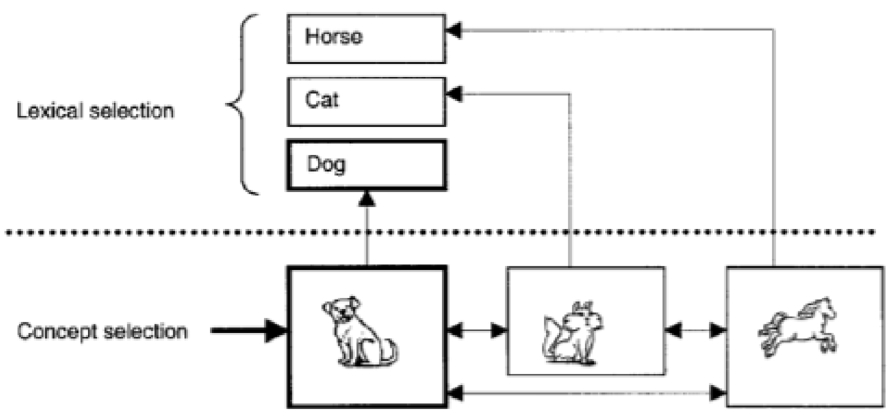


Figure 2.2: From La Heij, W. (2005) [7]. A simple model of lexical access. A picture of a dog is presented for naming. At the conceptual level, the corresponding conceptual representation is selected (concept selection). At the lexical level, several semantically related word candidates become activated. Lexical selection determines which of these words will be produced.

In the case of monolingual lexical access, the process is clear: levels of representation are sequentially accessed over the course of language production (Caramazza, 1997 [41]). Then, the process itself would involve at least two distinct steps: selection of the lemma or the conceptual representation, followed by the selection of the lexeme or the corresponding lexical-phonological representation.

While La Heij's (2005) [7] model (see Figure 2.2) is straightforward, it may allow us to model bilingual memory based on the single/blended lexicon, where concepts are mapped together between the sub-lexical and lexical levels via nodes (see Rumelhart & Norman, 1985 [53]). However, based on the evidence in favour of competing activation, it is more likely that the bilingual mind separates the two lexicons at the lexical level, and thus we can expect to see a two-way mapping between phonological representations, a two-way mapping in lexeme selection, all of which map to a single semantic representation.

Due to the complexity of bilingual memory, modelling it has become a widely debated topic. While several models have been proposed—such as the **Bilingual Interactive Activation Model** (see Figure 2.3; Dijkstra et al., 1998 [48]), **Revised Hierarchical Model** (Kroll et al., 1994 [54]; 1997 [55]), the **Sense Model** (Finkbeiner et al., 2004 [56]), and the **Distributed Feature Model** (de Groot, 1992)—some have faced criticism for either resembling the monolingual model too closely or failing to fully capture the nuances of bilingual cognition.

Of the proposed models, the Distributed Feature Model (de Groot, 1992 [57]) is perhaps the most comprehensive for capturing language processing. In this model, we can see distinct mapping between separate lexemes—representing separation of the L1 and L2 lexicon—to a shared conceptual memory, indicating that while the lexicon is separated at the lexical level, all sublexical and conceptual information is shared between them. In sum, representations of the bilingual memory are hardly as straightforward as mapping the monolingual memory.

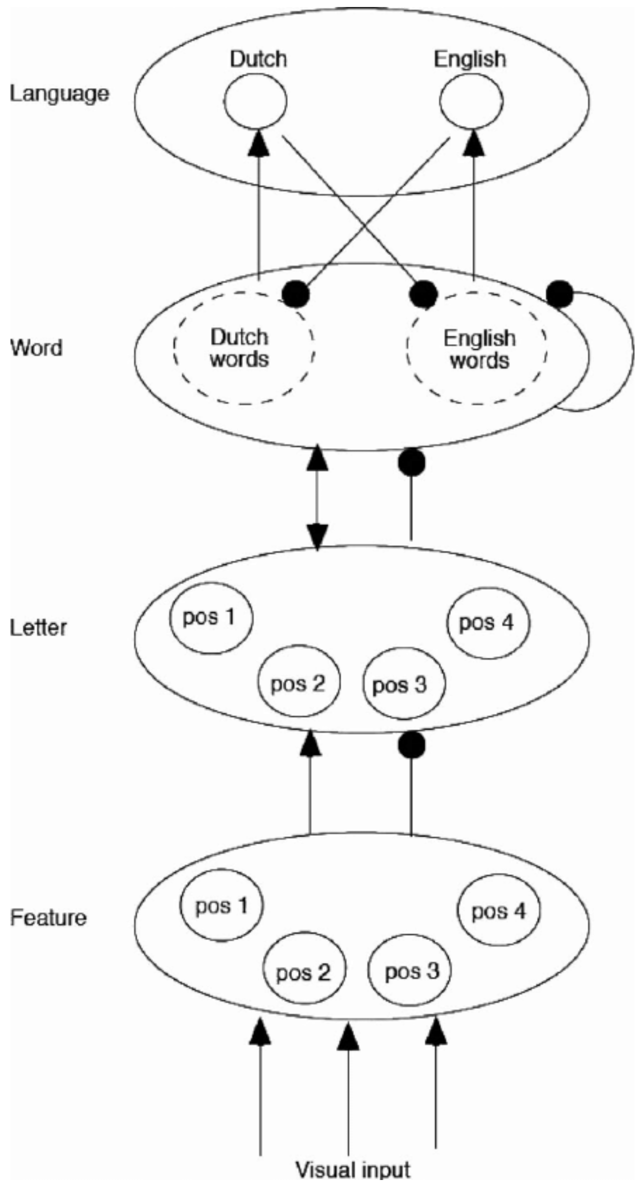


Figure 2.3: A precursor to the BIA+ model, the Bilingual Interactive Activation Model (see Dijkstra et al., 1998 [8]) shows a clear separation between the lexicons of both languages occurring only at the word, or lexeme node. This presupposes that prior to lexical activation, the brain must have a sufficient amount of data prior to making a lexical decision, thus activating the appropriate lexicon. In this model, the feature node represents phonetic input, while the letter node represents the visual input (i.e., orthographic representation).

While there is compelling evidence to show that both the bilinguals and monolinguals map between separate levels of representation, representation of the bilingual memory becomes somewhat flimsy at the level of separation between lexicons. This muddying of the separation of lexicons can be attributed to interactions between nodes at the conceptual level. For instance, in the Distributed Feature Model proposed by de Groot (1992) [57] (see Figure 2.4), the English word *father* and the Spanish word *padre* activate the same set of conceptual nodes; in other words, both the English and Spanish lexical entries trace back to the same conceptual representations. In contrast, the conceptual nodes activated by *advice* are not completely identical to those activated by *consejo*. These examples illustrate how the boundaries between lexicons may become less distinct, supporting the idea that the activation of conceptual nodes—and the subsequent retrieval of the appropriate lexical form—depends on contextual factors and task demands.

In light of the models discussed in this subsection, it is necessary to clarify the theoretical framework guiding the present study. While the Distributed Feature Model and the Revised Hierarchical Model offer important insights into conceptual mapping and cross-language links, this dissertation adopts the **BIA+ model** (see Figure 2.5; Dijkstra & van Heuven, 2002 [52]) as its primary account of bilingual lexical access.

2.3.2 Modelling Bilingual Language Processing

Contemporary theories of bilingual lexical processing recognize that language activation is sensitive to linguistic context, such as whether a task involves a single-language or mixed-language setting. Among existing models, the BIA+ framework provides a detailed account of how such contextual factors interact with lexical activation. Extending the original BIA model, BIA+ proposes an integrated lexical network in which orthographic, phonological, and semantic representations from both languages are stored within a shared system. Activation flows bottom–up from input features (e.g., letters or sounds) to word-level representations, allowing multiple lexical candidates from both languages to become activated in parallel.

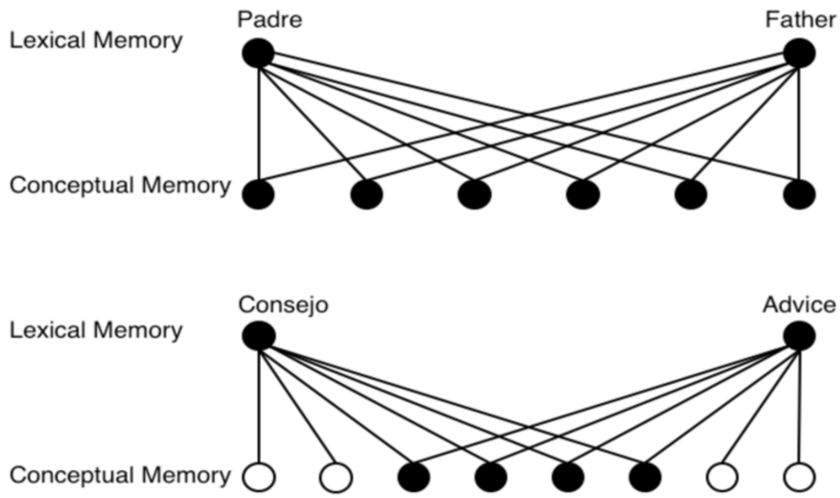


Figure 2.4: The Distributed Feature Model (de Groot (1992)) shows the mapping of lexemes from both L1 and L2, in this case, Spanish and English, relative to their conceptual or semantic representations.

Crucially, language membership information does not function as an early filter. Instead, language nodes influence processing at later stages by modulating competition among already activated lexical candidates. The “+” component of the model introduces a task/decision system, separate from the word identification system, which accounts for how contextual cues and response requirements shape observable performance without altering the architecture of lexical representation itself.

As illustrated in Figure 2.5, this distinction between the word identification system (feature, letter, word, and language nodes) and the task/decision system explains how cross-language activation can persist even under single-language instructions. This architecture aligns closely with the empirical evidence reviewed in this chapter supporting non-selective access and parallel activation.

For these reasons, the experimental design, predictions, and analyses in the following chapters are grounded in the core assumptions of the BIA+ model: integrated lexical representation, parallel activation across languages, and late-stage modulation of competition by contextual and decision-level processes.

Unlike most models of bilingual language processing, the BIA+ model accounts for the processing aspect (e.g., sublexical and lexical levels), as well as the task. The addition of the task schema was largely influenced by Green's (1998) [58] **Inhibitory Control (IC) model**, which suggests that the competing potential outputs of the lexico-semantic system are inhibited depending on the goals of the speaker. In other words, responses deemed inappropriate for the task, such as words from the non-target language, are inhibited, thus preventing their production. Furthermore, there is robust evidence to support Green's claim (1998) and to support the **Bilingual Interactive Activation Plus (BIA+) model**.

Building on the BIA+ model and the discussion of task dependency in language processing, and to further add to Marian & Spivey's (2003) [28] study, we find further evidence in studies of cognate effects; i.e., the effect by which words that share similar forms and meanings across languages are processed faster and with a higher degree of accuracy. For example, Starreveld et al. (2013 [59]) investigated bilingual lexical access in Dutch–English bilinguals using a picture–word interference (PWI) paradigm in a speech production task. Participants were asked to name pictures in a target language while ignoring simultaneously presented distractor words. Critically, some pictures depicted English–Dutch cognates, allowing the authors to examine the presence of a cognate effect, which would indicate activation of the non-target language. Distractor words varied in semantic relatedness to the target and in language membership (target vs. non-target), and naming latencies and error rates were measured to assess lexical competition. The study showed that non-target-language distractors can produce semantic interference, providing evidence that lexical candidates from both languages are activated in parallel and compete during speech planning, even when only one language is required.

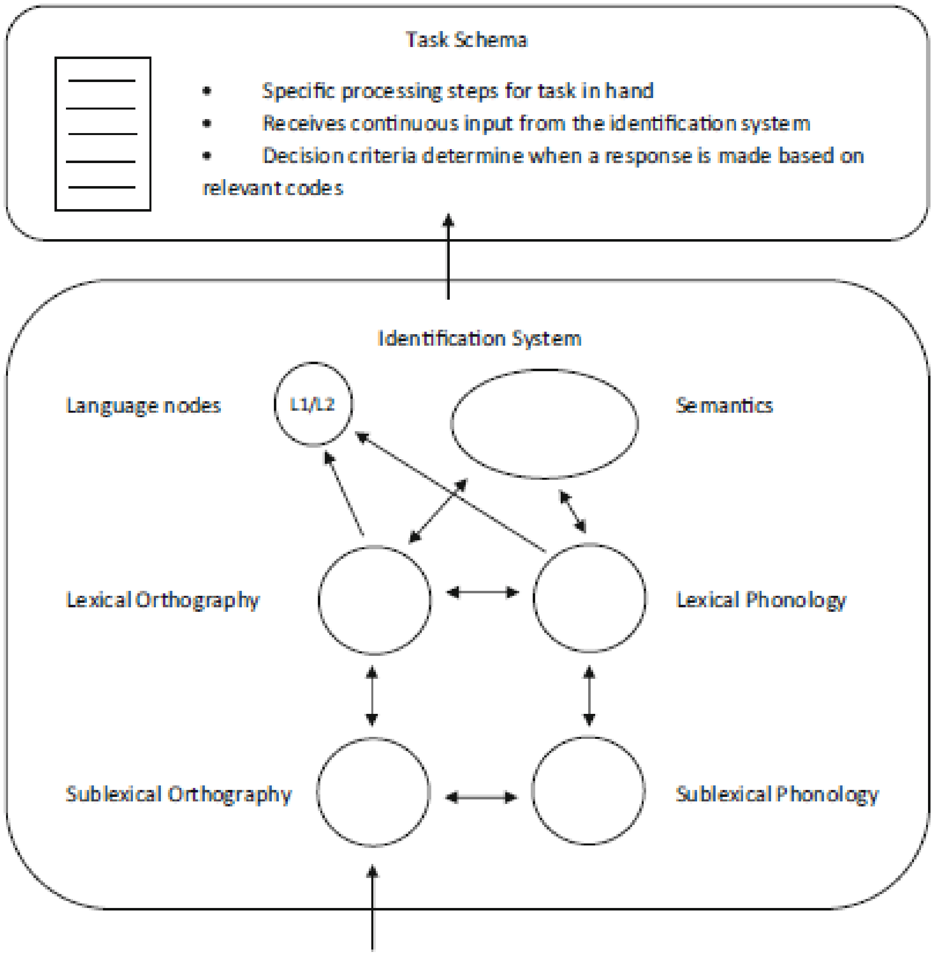


Figure 2.5: The BIA+ model for bilingual language processing (Dijkstra & van Heuven, 2002) presupposes that language selection and processing is task dependent. The introduction of the Task Schema (see Inhibitory Control (IC) Model by Green (1998)). However, much like the model for monolingual language processing, the BIA+ model indicates processing must necessarily go back to the semantic node to draw out meaning. Further, the BIA+ also accounts for modality, which other models of bilingual language processing do not necessarily do.

Starreveld et. al. (2013) [59] notes that in both experiments, a cognate effect was observed, however, it was conditioned by language and sentence constraint (i.e., whether the target word was presented in isolation or in a contextual condition). In this study, high-constraint sentences provide strong contextual clues that make the target word more predictable, reducing the influence of cross-language activation, while low-constraint sentences offer minimal context, leading to greater reliance on the language's form and meaning, which enhances the cognate effect. To this end, the cognate effect was more prominent in the following contexts: (i) the effect was greater in L2 than in L1, (ii) it was greater with low-constraint sentences than with high-constraint sentences, and (iii) there was no observable effect in the high-constraint L1 condition. From these results, we can draw the conclusion that the greater the restriction we put on context, the less competing activation there will be, therefore confirming Green's (1998) [58] and Dijkstra & van Heuven's (2002) [26] claims that bilingual language processing is largely constricted by task demands.

In summary, monolingual and bilingual lexical access are structurally quite similar, if not the same, based on representations of a lexical, sub-lexical and semantic level, with the addition of an L2, however, a second language node is necessary (see Figure 2.5). We can conclude that representations of lexical items can be subdivided into the three levels (sublexical, lexico-semantic, and language) to which the bilingual memory can map to, and that the mapping of the bilingual memory (i.e., processing) is conditioned by task demands.

Having established the theoretical architecture of bilingual lexical access—and in particular the role of parallel activation and task-dependent competition—the following section examines one of the most studied phenomena arising from this architecture: the cognate effect.

2.4 Defining Cognates

Contemporary and alternative definitions of **cognates** in bilingual processing research have shifted from strictly etymological criteria to psycholinguistic ones, often defining cognates as translation equivalents that share substantial phonological and/or orthographic overlap across languages

(Starreveld et al., 2013 [59]). Under this broader operational definition, the key factor is cross-linguistic form similarity, which has been shown to facilitate co-activation during lexical access. Importantly, even word pairs that are not strict historical cognates—such as English *beef* and French *boef*—may elicit cross-language activation due to phonological similarity. For the purposes of the present study, cognates will therefore be defined operationally as translation equivalents exhibiting significant phonological and/or orthographic overlap, regardless of strict etymological origin. This definition allows us to focus on the psycholinguistic mechanisms of cross-language activation and competition, which are central to predictions derived from the BIA+ framework.

Cognates—i.e., those words that are related in orthography—are commonly used in studies of bilingualism to evaluate lexical organization, separation, and activation across languages. Consider, for example, the study conducted by De Groot (1992) [57], which explored conceptual representations in the bilingual mind. In this study, English and Dutch cognates were used to determine how *lexemes* in each language map onto shared or separate conceptual representations in Dutch–English bilinguals. A central finding was that conceptual representations are not fixed, but are dynamically constructed depending on processing demands. Drawing on Barsalou (1982) [60], De Groot distinguishes between **context-independent properties**—core semantic features automatically activated upon encountering a word (e.g., the property "smells unpleasantly" when encountering *skunk*)—and **context-dependent properties**, which are activated only in specific contexts (e.g., the property "floats" for *basketball* in the sentence *Chris used a basketball as a life preserver when the boat sank*). Crucially, if translation equivalents that are cognates activate the same set of context-independent properties across languages, this suggests they share a stable conceptual core representation. Differences, if present, are more likely to emerge at the level of context-dependent properties. Thus, the distinction between context-independent and context-dependent activation provides a way to test whether cognates are linked to a shared semantic representation or to partially overlapping conceptual structures in the bilingual lexicon.

Cognates have also been widely used to evaluate processing efficiency in bilinguals, particularly in sentence–picture naming paradigms. In these tasks, participants read or heard a sentence that either strongly constrained the identity of an upcoming word (high-constraint context) or provided little predictive information (low-constraint context), and then named a subsequent picture in a specified language (L1 or L2). Using this design, researchers have consistently found that cognates are named more quickly and accurately than non-cognates (Dijkstra et al., 1998 [8]; Dijkstra et al., 1999 [61]; Gollan, Forster & Frost, 1997 [62]; Schwartz, Kroll & Diaz, 2007 [63]; Van Hell & Dijkstra, 2002 [64]).

Within this paradigm, several effects have been observed. First, cognates benefit from additional activation at sublexical phonological output nodes, receiving converging activation from lexical representations in both the target and non-target language. Second, stronger connections between semantic, lexical, and sublexical levels are typically observed in the dominant language (L1) compared to the weaker L2. Third, when the sentence context strongly predicted the upcoming picture (high-constraint context), overall activation within the target language system increased. Finally, in highly predictive contexts, part or all of the semantic representation of the upcoming picture could be pre-activated before the picture appeared, facilitating subsequent naming. In other words, contextual information allowed participants to anticipate the target concept, accelerating lexical access. Importantly, the magnitude of this cognate facilitation effect is modulated by factors such as the language of production (L1 vs. L2) and the predictability of the target word within the sentence context (Starreveld et al., 2013 [59]).

2.4.1 Cognates Relative to the Present Study

For the present study, the intention is to leverage the cognate effect to better investigate cognitive load and attention during language processing. Below, I will provide more detail on the task procedure and specify the criteria for what counts as a cognate in this study.

Having established a general operational definition of cognates, it is now necessary to clarify how this definition will be applied in the present study, which focuses specifically on cognate pairs in the English–French and English–Mandarin language combinations.

For English–French, cognate identification is straightforward. Given the shared alphabetic writing system and historical lexical overlap, we adopt the modern psycholinguistic definition of cognates as translation equivalents that share substantial phonological, morphological, and semantic features. Under this definition, many English–French word pairs qualify as cognates due to clear cross-linguistic form similarity.

The situation is markedly different for English–Mandarin. English and Mandarin do not share a writing system, and Mandarin's tonal phonology introduces additional surface-level divergence. Strict phonological and orthographic overlap between translation equivalents is therefore rare. If cognates were defined narrowly by visible or segmental similarity, very few English–Mandarin pairs would qualify. However, adopting a broader psycholinguistic definition that allows for phonologically adapted loanwords increases the number of potential cognates. For example, proper nouns such as "Toronto" are naturalized in Mandarin as 多伦多(*Duō lún duō*), and historically borrowed forms such as English *typhoon* and Chinese 台风(*tái fēng*) demonstrate cross-linguistic phonological an semantic relationships despite differences in script and phonetic realization.

In sum, for the purposes of the present study, cognates will therefore be defined as translation equivalents exhibiting cross-linguistic lexical overlap at the phonological and/or morphological level, even when that overlap reflects adaptation across writing systems. This modified operational definition allows for meaningful comparison across English–French and English–Mandarin while maintaining theoretical consistency within the BIA+ framework.

2.5 The Cognate Effect

Having established what cognates are and how they are operationalized for the present study, this section reviews the empirical evidence for how cognate status influences bilingual lexical processing. As noted above, cognates have traditionally been thought to facilitate processing through dual activation of overlapping representations; however, recent evidence reveals more complex dynamics, where cognates sometimes slow response times due to cross-linguistic interference and competition (Poarch & Van Hell, 2012 [65]; Brenders et al., 2011 [66]).

Bilingual language processing models such as the Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & Van Heuven, 2002 [1]) suggest that during lexical access, both language systems are co-activated. This parallel activation can produce either facilitation (i.e., if the shared representations boost activation) or interference (i.e., if the bilingual must inhibit the competing lexical candidate from the non-target language) (Green, 1998 [58]).

2.5.1 Evidence Supporting the Cognate Facilitation Effect

A considerable body of psycholinguistic evidence supports the notion that cognates are linked to a processing advantage in bilingual lexical access. The **Cognate Facilitation Effect**, whereby cognates are recognized, retrieved, or produced faster and with fewer errors than non-cognates, has been consistently observed across a range of experimental paradigms, including lexical decision, picture naming, and translation tasks (see Cristoffanini, Kirsner, & Milech, 1986 [67]; Costa, Santesteban, & Caño, 2000 [68]; Lemhöfer & Dijkstra, 2004 [69]).

The theoretical basis for this facilitation is grounded in the shared **lexical representation hypothesis**, which posits that cognates activate overlapping orthographic, phonological, and semantic representations across a bilingual's two languages (see De Groot & Nas, 1991 [70]; Dijkstra & Van Heuven, 2002 [1]). We could oversimplify this by saying that cognates are "doubly active" in both the L1 and L2: a single orthographic or phonological form simultaneously activates the corresponding semantic representations in both languages. In the **Bilingual Interactive Activation Plus (BIA+) model**, lexical access is inherently non-selective, such that words from both languages are activated in parallel. In the case of cognates, this overlap produces additive activation that speeds recognition (Dijkstra & Van Heuven, 2002 [1]). This dual-source activation enhances the likelihood that a cognate reaches the activation threshold more quickly than a non-cognate.

Empirical support for this comes from reaction time studies showing that bilinguals perform faster lexical decisions for cognates compared to non-cognates in both their dominant and non-dominant languages (see De Groot & Nas, 1991 [70]; Van Hell & Dijkstra, 2002 [71]). These findings have been replicated across numerous language pairs, such as English–Dutch, Spanish–English, and French–German, suggesting that facilitation is a generalizable property of bilingual lexical systems, not specific to particular linguistic combinations (Lemhöfer & Dijkstra, 2004 [69]).

Electrophysiological studies provide converging evidence. For instance, Midgley, Holcomb & Grainger (2008) [72] demonstrated reduced N400 amplitudes, a neural marker of semantic integration, for cognates relative to non-cognates, suggesting more efficient semantic access. These

ERP results reveal that facilitation arises at early stages of lexical processing, even before overt response generation, supporting a bottom-up activation account.

Facilitative effects also emerge in language production tasks. In picture-naming and translation paradigms, cognates are retrieved and articulated more rapidly, particularly in the bilingual's weaker language, due to parallel activation from both lexicons (Costa et al., 2000 [68]; Poarch & Van Hell, 2012 [65]). This supports the view that shared representations create a "boost" effect, whereby activation in one language's lexicon automatically reinforces activation in the other.

Adding to this, bilingual proficiency and language dominance modulate the strength of this effect. Balanced bilinguals, who maintain high proficiency in both languages, tend to exhibit stronger facilitation effects than unbalanced bilinguals, as their integrated lexical systems allow for more robust cross-language activation (Green & Abutalebi, 2013 [73]). These patterns indicate that cognate facilitation is not a mere artifact of word form similarity but reflects deeper mechanisms of cross-linguistic interaction that enhance efficiency in bilingual word recognition.

Collectively, this body of work demonstrates that under most conditions emphasizing semantic access or language co-activation, cognates provide measurable benefits to bilingual processing. However, as the following section discusses, these benefits can invert under specific task demands that prioritize language selectivity, lexical inhibition, or form discrimination, thereby yielding interference rather than facilitation.

2.5.2 Evidence Supporting the Cognate Interference Effect

The **Cognate Interference Effect** refers to the observation that cognates can slow processing or increase error rates, particularly in comprehension-based or recognition tasks requiring precise language selection (Brenders, Van Hell & Dijkstra, 2011 [66]; Van Hell & Dijkstra, 2002 [71]).

From a theoretical standpoint, this interference arises due to cross-linguistic competition during lexical access. In the BIA+ framework, activation of both language representations occurs automatically. When two highly similar lexical candidates are simultaneously activated, the bilingual's cognitive control system must inhibit the non-target language entry (Green, 1998 [58]). This inhibitory process consumes time and cognitive resources, producing longer response latencies. Green's **Inhibitory Control Model** explicitly predicts that greater lexical overlap (especially in cases with cognates) intensifies competition, resulting in slower selection of the target representation.

Empirical data supporting interference come from lexical decision tasks in which bilinguals are slower and less accurate when responding to cognates than to non-cognates. In a **visual lexical decision task** wherein participants had to decide as quickly and accurately as possible whether letter strings were real Dutch (L1) words or nonwords, Van Hell & Dijkstra (2002) [52] found that while cognates facilitated recognition in non-selective contexts—contexts in which lexical representations from multiple languages are activated simultaneously (e.g., a Dutch–English bilingual encountering the word film, which exists in both languages)—they slowed down responses when participants were explicitly instructed to respond in a single target language, suggesting that language selectivity modulates the effect's direction.[2] Similarly, Brenders et al. (2011) [66] observed that bilinguals took longer to recognize morphologically complex cognates, consistent with additional decomposition and suppression demands.

Eye-tracking studies have also documented increased fixation times and regressive saccades on cognates, indicating sustained lexical competition and difficulty resolving cross-linguistic over-

[2]Critically, the items/tokens used for this experiment were Dutch words that were cognates with English (L2), cognates with French (L3), or non-cognate controls. The motivation for choosing such a range of tokens was to test cross-language activation during exclusively L1 processing.

lap (Duyck et al., 2007 [74]). These findings suggest that interference may not stem purely from lexical activation but from executive control processes needed to manage competing linguistic representations.

Importantly, task modality shapes whether cognates facilitate or interfere. In language production (e.g., picture naming, translation), cognate facilitation is dominant because co-activation boosts retrieval efficiency. In comprehension or recognition tasks (e.g., lexical decision, reading), where the bilingual must inhibit one language while responding in another, interference becomes more likely (Green & Abutalebi, 2013 [73]).

Another critical factor is language distance, or the degree of structural, orthographic, and phonological similarity between the bilingual's two languages. Cognate interference tends to intensify in language pairs with greater typological distance, such as English and Mandarin, where writing systems—and to a lesser degree, phonology—diverge. This distance likely forces bilinguals to rely more on semantic similarity than on form overlap, producing conceptual-level interference rather than orthographic facilitation (Degani & Tokowicz, 2010 [75]). For instance, while the English *coffee* and the Mandarin 咖啡(pronounced [kʰɑfe̥ɪ˥] or *kāfēi* in *Pinyin*) do not share morphological or orthographic features (even in *Pinyin*), they are nonetheless phonologically similar, as *kāfēi* is a borrowing from English, and both forms refer to the same underlying concept. Thus, by presenting an English–Mandarin bilingual with any of the forms "coffee", 咖啡, or "kāfēi", activation of a shared conceptual representation and phonological features is expected, potentially engaging lexical representations in both languages despite minimal form overlap. In contrast, closely related languages like English and French may show milder interference due to partial orthographic convergence.

Finally, physiological evidence from pupillometry and neuroimaging further supports inhibitory accounts. Studies using pupil dilation as an index of cognitive load show greater dilation during cognate processing in tasks requiring selective attention (Kuchinsky et al., 2013 [76]). Taken together, evidence disproving the cognate facilitation effect demonstrates that bilingual lexical processing is highly context-dependent. Rather than universally aiding recognition, cognate status can hinder performance when the bilingual must resolve between competing lexical entries across languages. This duality underscores the need to consider both task demands and language relationships in interpreting cognate effects.

The complexity of the cognate effect, and the conditions under which it facilitates or hinders processing, raises broader questions about how bilingual language processing can be reliably studied in a controlled setting. The following section examines these questions by reviewing the experimental paradigms most commonly used to investigate bilingual lexical access.

2.6 Conceptual and Theoretical Discussions in Bilingual Processing

The following section addresses conceptual issues surrounding the study of bilingual language processing in a controlled setting. Having already established the theoretical framework and the nature of the cognate effect, the focus here turns to the experimental paradigms most commonly used to investigate bilingual lexical access, with particular attention to their respective strengths and limitations. To this end, I will review three of the most common experimental frameworks—the **Picture-Word Interference task**, the **Auditory Word Recognition task**, and the **Lexical Decision Task**—used for studies of this nature.

2.6.1 Experimental Frameworks for Studying Bilingual Language Processing

Studying bilingual language processing experimentally presents challenges that do not arise in monolingual research, precisely because any task involving a bilingual participant must contend with two linguistic systems that are simultaneously active. The three paradigms reviewed here—the **Picture-Word Interference task**, the **Auditory Word Recognition task**, and the **Lexical Decision Task**—each offer distinct advantages for probing lexical access, but each also carries limitations that become particularly consequential in a bilingual context, as will be discussed in turn.

2.6.1.1 Picture-Word Interference Task

The **picture-word interference (PWI) task** was adapted from the classic Stroop task to investigate semantic effects in word processing. Like the Stroop task, PWI is designed to reveal interference effects when conflicting information must be ignored during cognitive processing. However, instead of naming colors, participants in the PWI task name pictures while ignoring superimposed distractor words (see Figure 2.6). For example, participants might see a picture of a *banana* with the distractor word *apple* (same semantic category) or *dog* (unrelated category) written across it. Rosinski (1977) found that distractor words from the same semantic category as the target picture (e.g., *banana* and *apple*) produced more interference than unrelated words or nonsense trigrams (e.g., *banana* and *dog*), which produced equivalent levels of interference. Furthermore, this effect was consistent across age groups, from second graders to college students, suggesting that semantic interference occurs similarly across development. Rosinski concluded that pictures and words access a shared semantic space, such that competition arises when target and distractor items belong to the same semantic category.

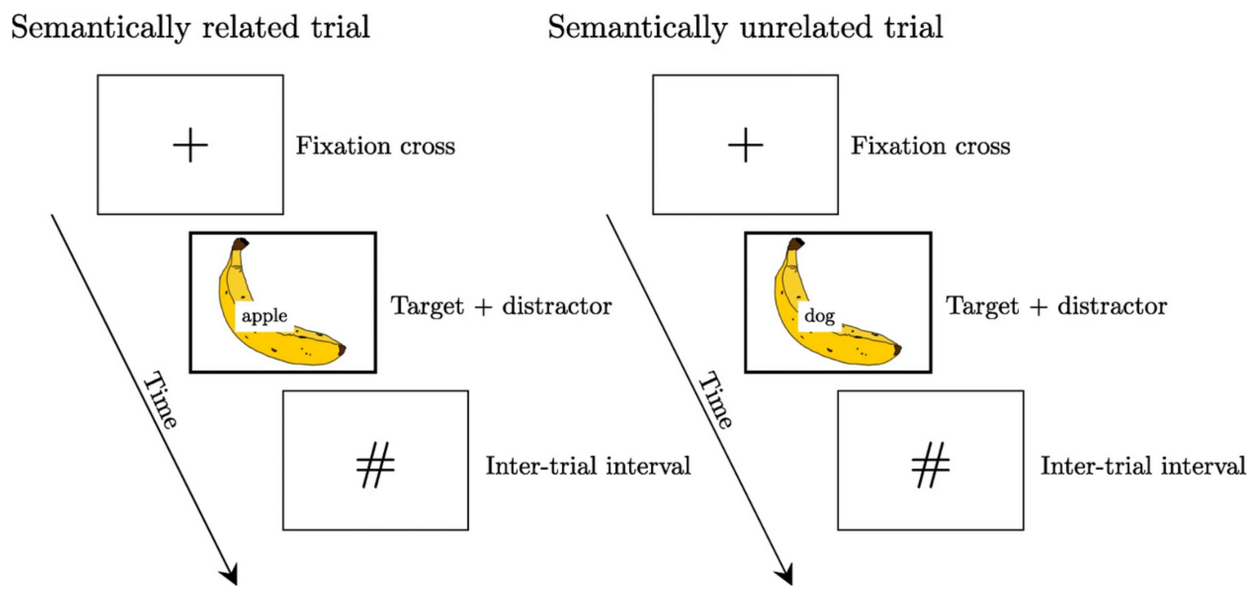


Figure 2.6: Fuhrmeister & Bürki (2022) [9], this figure provides an example of trials in a picture-word interference task. On the left, we see an example for a semantically related trial, wherein the distractor word is semantically related to the image (e.g., *banana* - image, and *apple* - word). On the right, we see an example for a semantically unrelated trial, wherein the distractor word is not semantically related to the image (e.g., *banana* - image, and *dog* - word). Typically, participants would require more time to name pictures in the semantically related task condition.

Often cited as the first to report using this method, Rosinski's (1977) [77] work paved the way for modern-day research in language processing. In fact, time and again, PWI has been shown to be a robust and reliable method for studying language production, particularly in examining the processes involved in naming, and the influence of semantic and phonological relationships between pictures and words (see Starreveld & La Heij, 1995 [78]; Schnur et al., 2006 [79]; Piai, Roelofs & Schriefers, 2015 [80]; Starreveld & La Heij, 2016 [81]).

However, PWI has its limitations. Firstly, there are ongoing discussions about whether semantic interference arises from competition at the lexical level or from post-lexical response processes. Mulatti & Coltheart (2011) [82] offered the **Response Exclusion Hypothesis** (REH), suggesting that interference emerges not from competition among lexical entries, but rather that distractor words intrude into an articulatory buffer; this then creates a hurdle or "roadblock" in processing that must be cleared before naming can proceed. In other words, the REH suggests that when presented with a picture and a distractor word, the brain first processes both the picture and the word, generating potential responses for each. But, because the speech production system is assumed to have a single-channel output buffer, only one response can be produced at a time. Further, the REH proposes that the neural system actively excludes the response associated with the distractor word, even if it's a well-formed word, to allow the target picture's name to be spoken. This hypothesis and the data supporting it (see Mahon et al., 2012 [83]; Hall, 2011 [84]; Aristei, Zwitserlood & Rahman, 2012 [85]) challenge the assumption that PWI reflects pure lexical selection mechanisms.

Building on this, the picture-word interference (PWI) paradigm also raises questions about construct validity, specifically, whether the bottom-up processing of distractors in the task truly reflects natural lexical access. In PWI, distractors are exogenous stimuli that participants are instructed to ignore while preparing a verbal response. This setup imposes specific attentional and inhibitory demands that are not typically encountered in everyday language production. In other words, the way language is processed in this task depends on the task requirements themselves—that is, the processing is task-dependent. Importantly, this distinction between bottom-up feature-driven processing and top-down contextual modulation can be conceptualized in terms of the framework illustrated in Figure 2.7. Consequently, observed effects may reflect these task-specific control mechanisms rather than the underlying mechanisms of lexical access (Lowry, Dubé & Schotter, 2021 [86]).

A second limitation to using PWI to study bilingual language processing is how it handles the complex cross-language dynamics in the bilingual mind. To this end, I would like to highlight three (3) key points to exemplify this limitation: (i) the **Translation-Equivalent Facilitation effect**, (ii) the ambiguous nature of the **cognate effect** (Wu & Thierry, 2011 [87]), and (iii) the difficulties surrounding the separation of **Semantic vs. Lexical Activation levels** (Wu & Thierry, 2011 [87]).

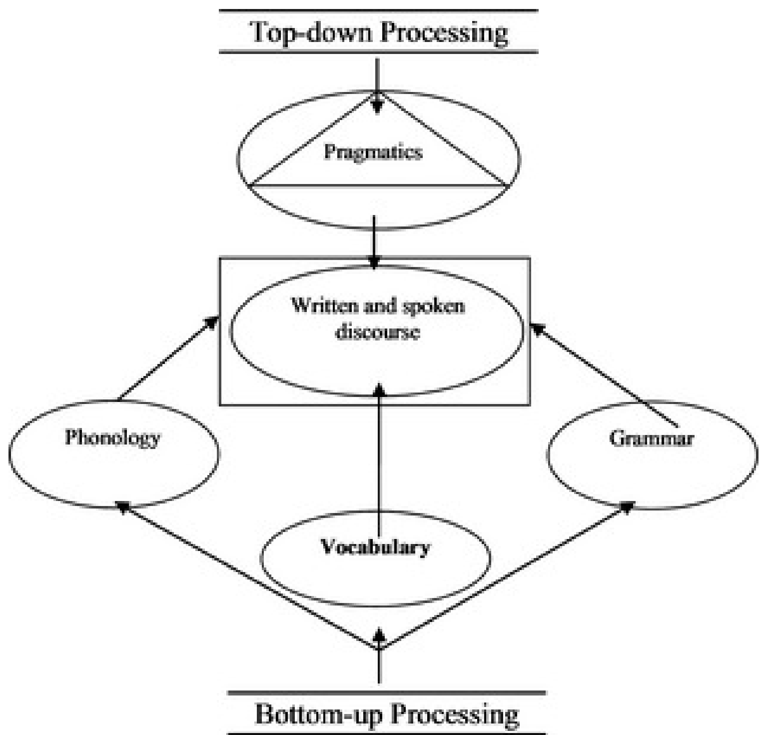


Figure 2.7: From *Vocabulary and language knowledge* (Celce-Murcia & Olshtain, 2001 [10]), this model breaks down and illustrates the fundamentals of top-down processing and bottom-up processing. Regarding PWI, the separation of lexical and semantic activation would remain ambiguous as we can not definitively be certain on which level activates first based on task requirements.

Melinger (2018) [88] remarks that the presentation of a distractor that is a direct translation of the target picture in the non-target language often facilitates naming rather than producing interference. with the Translation-Equivalent Facilitation effect. This finding is especially compelling as it runs counter to what a purely competitive model—a model which a simple PWI on monolingual speakers would generally follow—would predict. In other words, where we may be looking to induce an inhibitory effect, we may instead elicit a facilitatory effect on processing, thereby providing a task-dependent and generally unrealistic illustration of how bilingual language processing typically occurs within the PWI paradigm. Tying this into the second and third points I had put forward, this Translation-Equivalent Facilitation effect overlaps directly with any facilitatory influences of cognates on processing. As previously noted, **cognates** are those words which are similar across languages both semantically and phonologically, and it is because of these similarities that we can observe accelerated naming in this task. However, within the PWI context, it is difficult to determine whether this facilitation is due to overlapping semantic or phonological representations—or both—making it challenging to isolate levels of activation.

In summary, using the PWI does not constitute a robust enough experimental framework for accurately observing NLP in a bilingual context. The simple fact that participant performance is very much dependent on the complex task demands (i.e., discerning between a distractor word vs. naming the target image) makes for unrealistic ecological validity which would inevitably elicit inaccurate data.

2.6.1.2 Auditory Word Recognition Task

The **Auditory Word Recognition (AWR) Task** typically involves presenting spoken words to participants and measuring the recognition processes, typically through response latencies—or reaction time (RT)—and accuracy scores. A popular experimental framework for studying language processing, AWR has been used in a variety of studies, including **Event-Related Potential (ERP)** studies and eye-tracking studies. In addition, this experimental framework has been used in research on word recognition with auditory interference or noise (Morini & Newman, 2020 [89]). More relevant to bilingual language processing, this experimental framework was used by researchers such as Spivey & Marian (1999) [46], Marian, Blumenfeld & Boukrina (2008) [90]; Blumenfeld

& Marian (2013) [51]; Lagrou et al. (2011) [91]) to show cross-language competition even in native-language listening (see Figure 2.8).

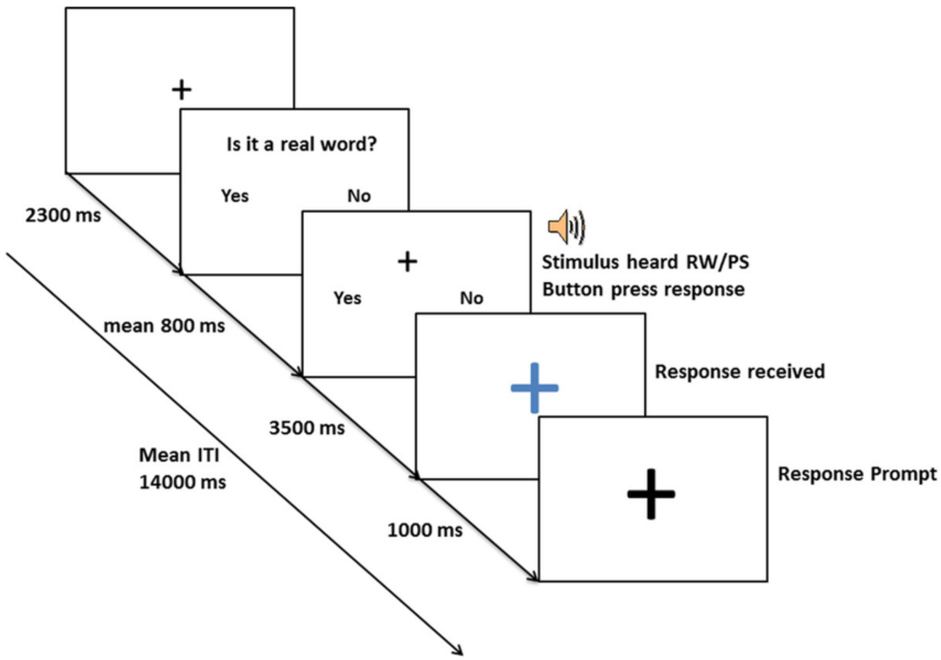


Figure 2.8: Illustration of the Auditory Word Recognition Task process as it appears in Roxbury, McMahon & Copland (2014) [11]. At the onset, the listener is presented with a visual cue (or fixation point). This is followed by the screen asking if the word is a real word or not. The fixation point then returns in lieu of the question while the YES/NO options remain on the screen. During this time, the stimulus word is heard and the listener must use a button press to respond. Once the response has been recorded, the fixation point returns on-screen and the trial times out (see Figure 2.8).

Crucially, AWR has given rise to the **Marslen-Wilson model**, otherwise known as the **cohort model** (see Marslen-Wilson & Welsh, 1978 [92]). This model of lexical access suggests primarily that people begin word activation before hearing the entire word such that when an individual hears speech segments in real-time, each speech segment (e.g., phoneme) "activates" every word in the lexicon that begins with that segment, ruling out more words as subsequent segments are heard until only the word matching the input stimuli is left (see Figure 2.9). Although the findings borne from AWR have contributed significantly to our fundamental understanding of language processing, it remains that this method is not the most robust, especially when dealing with bilingual language processing.

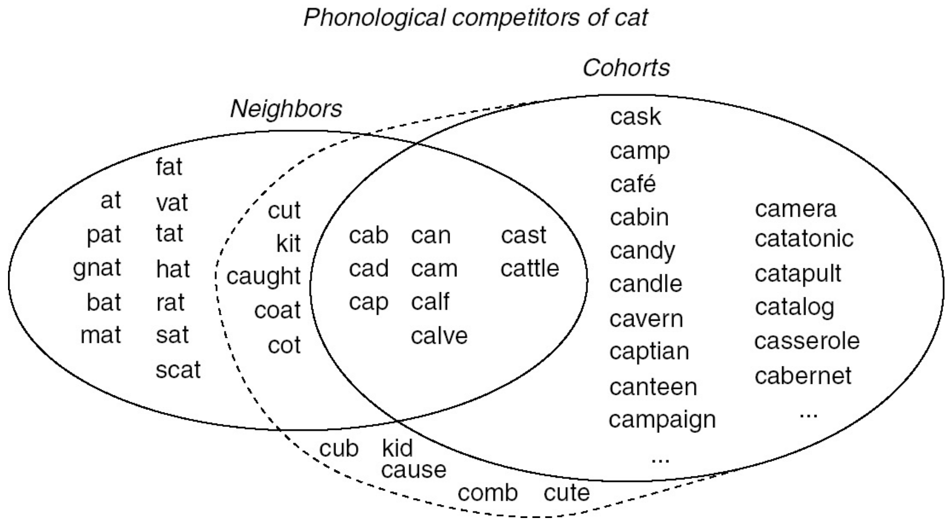


Figure 2.9: From Magnuson, Mirman Harris (2012) [12], the above illustrates and exemplifies the competing activation of lexical items put forward by the Cohort Model. In this case, we observe the cohort and neighbourhood competitors for the word *cat* (see Figure 2.9).

On the practical front, we must consider modality and stimuli limitations, and cross-language competition. Perhaps the most obvious issue with AWR in a bilingual context is the limitations of modality and stimuli; in a monolingual context, a listener only has one linguistic system from which to make their decisions—consequently, any phoneme which may not be included within that system likely gets treated as "foreign" and therefore does not trigger activation. In contrast, bilinguals must necessarily deal with two linguistic systems in tandem, thus having access to two phonemic inventories. In cases of phonemes which belong only to one inventory, processing results will likely be skewed since faster reaction times would be expected. Take, for instance, a comparison of Mandarin and English: because Mandarin, being a tonal language, is vastly different in phonology from English, a listener is likely to recognize and react faster in identifying words in in each of their languages than if the language pairs were similar in phonology, as is the case with English in

French. While this fact alone demonstrates the flimsiness of AWR in a bilingual context, we must also take into account the inflated lexical competitor set that bilingual speakers will have to face.

In line with the phonological argument, because of the two linguistic systems occupying the bilingual mind, bilingual speakers have access to two lexicons. As such, they face additional competition, as cohort competitors from both the L1 and the L2 are active during recognition (see Figure 2.9). This larger pool of similar-sounding lexical candidates would, therefore, slow processing and recall. This competing activation across the L1 and L2 necessarily entails further moderating factors, namely proficiency, immersion effects, and cultural and lexical variations. By immersion effects, we refer to the extent to which a bilingual is currently or habitually immersed in a particular language environment; such immersion can influence activation levels and competition patterns during processing. For instance, L2-dominant bilinguals are more likely to show onset competition in both the L1 and the L2, but rhyme competition typically emerges only in the L2 (Lagrou, Hartsuiker & Duyck, 2011 [91]). Cultural variations, on the other hand, encompass differences across language communities—such as norms for word usage, frequency patterns, or pragmatic conventions—which may qualitatively alter processing outcomes and challenge the universal applicability of experimental findings (Goldinger, 1996 [93]).

On the theoretical front, we must consider task and decision system influences, as well as cognitive control and inhibition differences. Recall that according to the BIA+ model, word identification occurs independently of task decisions and largely remains language-nonselective (see Figure 2.8). Because the task/decision subsystem is not predicted to influence word activation, and given the modality by which stimuli are presented in AWR, the experimental task demands do not reflect natural instances of processing. This point is further supported by the previous argument on differences or similarities in linguistic systems, namely phonemic inventories, which would result in different responses. Again, because of the particularities of this experimental framework, results obtained from this method do not necessarily reflect true natural processing. Furthermore, the task-dependent nature of processing outlined in the BIA+ model suggests that, even though word identification is largely automatic and non-selective, participants are likely to adjust their cognitive strategies to meet the specific demands of the AWR task (see Figure 2.8).

Closing the discussion on why AWR is not the most robust method for observing bilingual language processing and further building on the previous points highlighted here, we need to also discuss the topic of cognitive control. Research studies have suggested that bilinguals reconfigure their linguistic and non-linguistic cognitive processes. In fact, AWR was employed for the investigation of the relationship between the resolution of linguistic competition and non-linguistic cognitive control in bilinguals and monolinguals, Blumenfeld et al. (2016) [94] found that only the bilingual group recruited similar inhibition mechanisms during both linguistic tasks (i.e., AWR) and non-linguistic tasks (e.g., Stroop task). Although the results of Blumenfeld et al. (2016) [94] revealed compelling information on how bilinguals are more consistent with competition resolution and the recruitment of cognitive control compared to monolinguals, their findings did not reveal much in terms of language processing. To this end, we note that while AWR is a useful method for studying cognitive control and competition resolution in linguistic and non-linguistic tasks, its robustness is compromised in bilingual contexts due to cross-language activation, heavy reliance on executive control and task modality issues. Together, these factors raise questions about how well lab-based auditory word recognition tasks capture real-world bilingual lexical processing.

2.6.1.3 Lexical Decision Task

In contrast to the auditory word recognition task—which requires participants to recognize or identify spoken words as they unfold over time, making it sensitive to incremental lexical activation and phonological competition rather than just lexical access and decision processes—the **Lexical Decision Task (LDT)** is a widely used paradigm in psycholinguistics and cognitive psychology for investigating word recognition and lexical access. In a typical LDT, participants are presented with a string of letters (either visually or auditorily). Upon presentation of the stimulus, they must decide as quickly and accurately as possible whether the string forms a real word (e.g., “apple”) or a non-word (e.g., “blert”) (see Figure 2.10). Reaction time (RT) and accuracy—key dependent variables for this experimental framework—are typically measured in order to quantify processing and other behavioural data. Crucially, this method is used to infer mental representations of words and processing efficiency, such that faster and more accurate decisions are believed to indicate more efficient access to lexical memory.

LDT appears to be the most common method used when investigating language processing. It has been used to investigate monolingual language processing—as in Meyer & Schvaneveldt’s (1971) [95] work on semantic priming—and in studies on bilingual language processing, being used to investigate **cross-language activation** (see Kroll, Bobb & Hoshino, 2014 [96]; Linck, Hoshino & Kroll, 2008 [97]; Nakayama et al., 2016 [98]), **lexical access** (see Zunini et al., 2020 [99]; Rojas et el., 2023 [100]), **cognitive control** (see Friesen, Chung-Fat-Yim & Bialystok, 2016 [101]; Bialystok, Craik & Luk, 2008 [102]; Nikolaev, Higby & Hyun, 2019 [103]), **word processing** (see Yap et al., 2016 [104]; Berberyan, van Rijn & Borst, 2021 [105]; Ferrand et al., 2011 [106]) and

language switching effects (see von Studnitz & Green, 1997 [107]; Broersma, Carter, Donnelly & Konopka 2020 [108]; Soares & Grosjean, 1984 [109]).

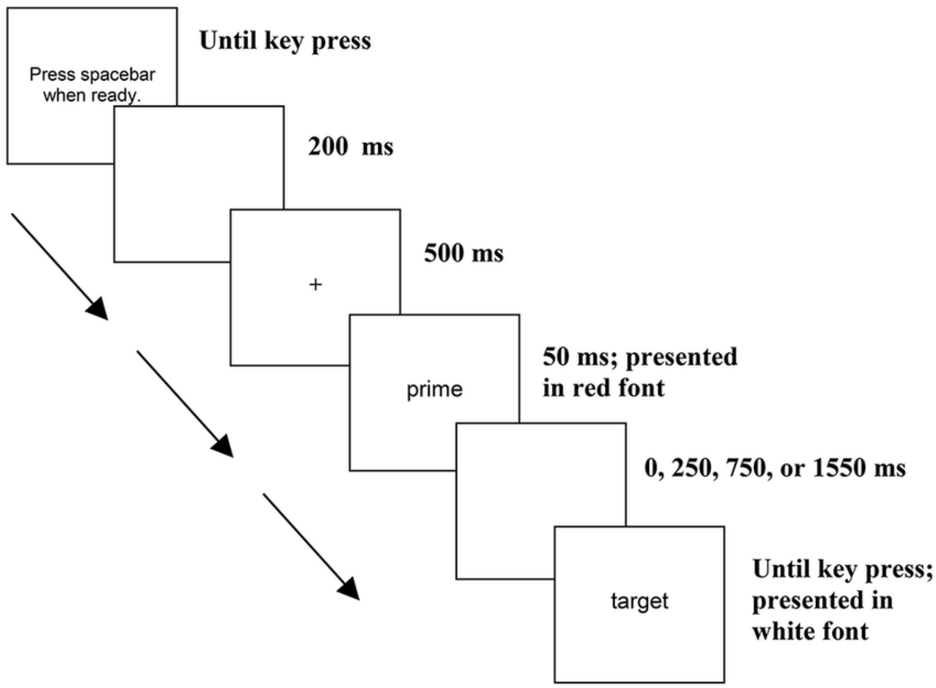


Figure 2.10: Illustration of a standard Lexical Decision Task process as it appears in Jones et al,. 2017 [13]. As the name suggests, participants are presented with stimuli and must make a decision on the string—i.e., they must decide if it is a real word or not, or if it is L1 or L2, etc. LDT has been used in a variety of studies and using the auditory or, more typically, the visual modality.

Further lending to the credibility of LDT as a robust experimental method for studying language processing, it was through LDT that McClelland & Elman (1986) [110] developed the TRACE model of speech perception. This model (see Figure 2.11) is a connectionist framework designed to account for the dynamic and interactive processes involved in recognizing spoken language. It consists of three hierarchical levels (e.g., features, phonemes, and words) each composed of interconnected processing units. Activation flows both bottom-up, from acoustic features to lexical representations, and top-down, allowing higher-level word representations to influence the perception of lower-level phonemes and features. Within each level, units compete through lateral inhibition, which enables the system to resolve ambiguity by suppressing competing alternatives. The model operates in real time and is sensitive to the temporal nature of speech, making it capable of simulating a variety of empirical findings in spoken word recognition. Crucially, the TRACE model highlights the interactive nature of speech processing, further emphasizing the role of feedback from the lexical to the phonemic levels. This mechanism allows the model to explain context-driven phenomena such as the **phoneme restoration effect** (see Warren, 1970 [111]) and **lexical competition**.

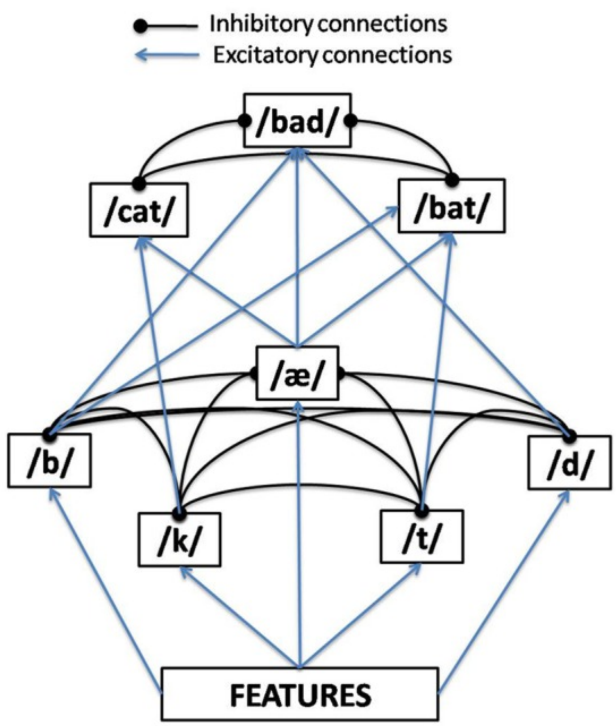


Figure 2.11: Illustration of speech processing via the TRACE model as per Hannagan, Magnuson & Grainger, 2013 [14]. This illustration demonstrates the sequential and overlapping activation of phonemes, including inhibitatory and excitatory connections, which occurs during speech recognition.

Although the lexical decision task (LDT) has proven effective for investigating monolingual lexical access, its application to bilingual language processing presents several methodological and theoretical limitations. One major concern is the **shallow processing bias** inherent in the task. Because LDT emphasizes speed and accuracy in word recognition, participants—particularly bilinguals—may rely on task-specific strategies that prioritize surface-level cues, such as orthographic patterns, rather than engaging in deeper, semantically rich processing. Consequently, the cognitive mechanisms measured by LDT may not fully reflect those involved in natural language comprehension and production.

Another significant issue is that LDT may underestimate or obscure cross-language activation. In real-world contexts, bilinguals often activate both their L1 and L2 simultaneously. However, the isolated and decontextualized stimuli typical of LDT may not sufficiently elicit such bilingual processing dynamics. While some studies have reported cross-language priming effects within LDT, these effects tend to be modest and highly sensitive to experimental design. The limited ecological validity of LDT further constrains their relevance: real-world bilingual language use involves pragmatic interpretation, code-switching, and discourse-level integration; processes largely absent from typical LDT paradigms.

Individual differences—such as language proficiency, dominance, and age of acquisition (AoA)—also play a crucial role in bilingual performance on LDTs. These factors can introduce confounds that complicate data interpretation, as they interact with task demands in complex ways. For example, the need to inhibit the non-target language can artificially increase executive control demands, producing results that reflect task-specific constraints rather than natural bilingual processing.

Lexical frequency effects present another challenge. LDTs often rely on monolingual frequency norms, which may not accurately capture bilingual language experience (Colomé, 2002 [112]; Dijkstra & Van Heuven, 2002 [52]; Van Heuven & Dijkstra, 2010 [113]). This mismatch can obscure meaningful comparisons unless bilingual-specific frequency norms are used, though such resources remain limited. Accounting for these individual differences and lexical factors is therefore essential for interpreting LDT outcomes in bilingual populations.

Despite these limitations, LDT remains a relatively reliable and controlled method for studying bilingual lexical processing compared to alternatives such as auditory word recognition (AWR) or picture-word interference (PWI) tasks. LDTs provide greater experimental precision, enabling researchers to manipulate and isolate specific lexical variables (e.g., frequency, orthographic similarity, language membership) while minimizing extraneous perceptual or motor influences. In contrast, auditory tasks can introduce variability in speech rate, accent, and clarity, while PWI paradigms may be confounded by visual processing demands or conceptual ambiguity. Moreover, LDTs are easily standardized across participants and languages, facilitating large-scale comparisons and statistical modeling. Their simplicity and control make them particularly valuable for investigating lexical access mechanisms, even if they sacrifice some ecological validity relative to more naturalistic paradigms. That said, even the most controlled LDT cannot capture the moment-to-moment fluctuations in cognitive effort that accompany lexical processing—a gap that physiological measures, and pupillometry in particular, are well positioned to address.

While RT and accuracy data obtained from paradigms such as LDT provide valuable behavioural evidence, they capture only the output of lexical processing rather than its cognitive dynamics in real time. The following section introduces pupillometry as a complementary physiological measure capable of indexing the moment-to-moment cognitive effort underlying bilingual word recognition.

2.7 Pupillometry

Pupillometry is a method in which pupil diameter is measured. Pupillometry is a marker of both sympathetic (SNS) and parasympathetic (PSNS) networks (Alshanskaia et al., 2019 [114]). For the pupil light reflex, we can expect to see antagonistic actions (or inhibition) of the iris sphincter and dilator muscles which are conditioned by the parasympathetic and sympathetic networks. This is noteworthy because increases in pupil size have been associated with the effort to acquire more visual information (Hess & Polt, 1964 [30]).

Pupillometry is not a novel method in science. In fact, the method of measuring pupil dilation has been used by clinicians and scientists for over 100 years (Loewenfeld, 1993 [115]) to evaluate cases of brain injury and impairment, as well as monitor neurological statuses and trends. Given the link between neurology, cognitive functions and pupil dilation is so robust, it is not surprising that this method is also used in psychological studies (see Simpson & Hale, 1969 [116]; Kahneman & Beatty, 1966 [117]; Kafkas & Montaldi, 2011 [118]).

In traditional studies, pupil dilation was typically measured subjectively—that is, clinicians would use a penlight or flashlight to manually evaluate pupil reactivity. This method proved to be both inaccurate and inconsistent, as measurements were often based on visual estimates rather than precise data. Today, pupil dilation is measured using automated methods with pupillometers. Even so, there remains debate over which technology is more effective: infrared pupillometers or mobile visual-spectrum automated pupillometers.

In the following subsections, I will discuss pupillometry as an index of attentional engagement, highlighting its application in psychological research. I will then examine how pupillometry has been used in research in psycholinguistics.

2.7.1 Pupillometry as an Index for Attentional Engagement

In medical sciences, assessing the pupillary light reflex has become a reliable way of measuring pupillary reactivity across many neurological diseases, including strokes, brain injuries, and tumoral herniation. To oversimplify, measuring pupillary reactivity is done by shining a light into the patient's eye to make the pupil constrict. Upon removing that light, the pupil should dilate again and return to its baseline condition. The pupil's reaction is then numerically graded on a three-point scale which corresponds to how quick the pupillary reflex is.

To obtain more consistent and accurate measurements, NeurOptics Inc. developed an algorithm called the **Neurological Pupil Index (NPi)**, which removes subjectivity from pupillary evaluation. Using a pupillometer, a patient's pupil measurements—such as size, latency, constriction velocity, and dilation velocity—are recorded. These measurements are then compared against a normative model of pupillary reactivity and graded by the NPi on a 0-to-4.9 scale, where 4.9 corresponds to the normative model and lower scores indicate abnormal or weaker reactivity.

In studies of cognitive load, long- and short-term memory, and attentional engagement, pupil dilation has been shown to correspond with higher cognitive engagement. Psychologists have linked pupil dilation to memory processes such as encoding, retrieval, and recognition. For example, in a study by Hess & Polt (1964) [30], adult participants (four men and two women) were shown images of adults and babies. Pupils of both sexes dilated when viewing images of people of the opposite sex, and female participants exhibited greater pupil dilation than males when shown images of babies or mothers with babies. The researchers concluded that while pupil dilation is a normal response to changes in light, it is also linked to arousal and emotional responses. Similarly, in an image recognition study, infants showed greater pupil dilation when viewing faces and greater constriction when shown geometric shapes (Fitzgerald, 1968 [119]).

In that same vein, results from studies on cognitive load demonstrate robust evidence supporting the theory that pupillary responses can reflect activation of the brain during cognitive tasks (i.e., greater pupil dilation is associated with increased activation or the brain, and vice versa). A study on short-term serial memory—wherein participants heard a string of words and were asked to repeat that string—dilation was observed after the items were heard, and constriction followed once the words were repeated (Kahneman & Beatty, 1966 [117]). Other studies, such as Kahneman & Treisman (1969) [120], found that pupil dilation is associated with cognitive load during memory tasks, where dilation increases as the difficulty of the task rises. Similarly, Hess & Polt (1960) [121] demonstrated that pupil size is influenced by cognitive engagement, with dilation occurring when individuals are mentally absorbed in tasks. Van der Wel & Van Steenbergen (2018) [122] reviewed numerous studies showing that pupil dilation correlates with mental effort in cognitive control tasks, supporting the idea that cognitive load affects pupil size. Bradley & Lang (2000) [123] also noted that pupil dilation reflects both emotional and cognitive demands, further extending the understanding of how attention impacts pupil responses. Samuels & Hamilton (1990) [124] confirmed that cognitive load during memory retrieval tasks leads to greater pupil dilation, while Beatty (1982) [125] expanded this by showing that mental workload increases pupil size across various tasks. Lastly, O'Boyle & O'Hara (1992) [126] found that higher cognitive workload in visual search tasks was marked by greater pupil dilation, further linking pupil size with cognitive effort. In sum, the more difficult the task, greater pupil diameter would be observed from the time between the solution and the completion of the task.

While pupillometry has long been established as a reliable diagnostic tool in medical contexts—particularly in assessing neurological integrity through the pupillary light reflex—it has also proven to be an insightful, non-invasive method for measuring attentional engagement in psychological research. The transition from clinical applications to cognitive science reflects the versatility and precision of this method. With innovations like the NPi improving objectivity in medical assessment, similar advances in psychophysiological measurement are enabling researchers to better quantify cognitive effort, emotional arousal, and memory processes. Across studies, the consistent relationship between pupil dilation and mental activity—whether during memory encoding, emotional arousal, or task engagement—demonstrates that pupillary responses serve as a robust physiological correlate of brain function. As the methodological rigor of pupillometry continues to improve, its interdisciplinary value across neuroscience, medicine, and psychology is likely to expand, offering increasingly nuanced insights into both clinical states and fundamental cognitive processes.

2.7.2 Pupillometry in Psycholinguistics

Pupillometry—the measurement of pupil size and reactivity—has emerged as a valuable tool in psycholinguistics for investigating real-time cognitive processes involved in language comprehension and production. The technique is grounded in the well-established finding that pupil dilation reflects mental effort, attentional engagement, and processing load (Beatty & Lucero-Wagoner, 2000 [127]). Unlike overt behavioral measures such as reaction times or accuracy, pupillometry offers a continuous and sensitive index of cognitive effort, allowing researchers to track fluctuations in processing load throughout the time course of linguistic tasks. While it is well known that pupil dilation correlates with increased cognitive effort during mental tasks, pupillometry has not yet become a standard method in research in Language Processing. However, with recent technological advancements in eye-tracking and physiological measurement, this method is gaining momentum in the field (Schmidtke, 2017 [128]).

Despite being relatively underutilized, existing pupillometry data in psycholinguistics reinforce previously established findings: increased pupil dilation consistently corresponds with increased processing effort. A more targeted application within this domain is the use of **Task-Evoked Pupillary Responses (TEPRs)**—a robust neurophysiological marker of cognitive load (Beatty, 1982 [129]; Beatty & Lucero-Wagoner, 2000 [127]). TEPRs have been particularly effective in studies examining sentence comprehension and syntactic complexity, offering a means to differentiate the processing costs associated with various types of phrasal structures (Just & Carpenter, 1993 [130]). These responses have been utilized in a range of experimental paradigms, including on-line verbal processing (Scheepers et al., 2013 [131]; Sauppe, 2017 [132]), sentence processing (Murujosa et al., 2020 [29]), and speech shadowing tasks (Tommola & Hyönä, 1990 [133]). However, as noted by Murujosa et al. (2020) [29], TEPRs have been rarely used in conjunction with picture naming tasks, and even less frequently—if at all—employed to assess cognitive load during lexical decision tasks, leaving a gap in the literature regarding how pupillometry might inform our understanding of lexical access in bilinguals and monolinguals alike.

One of the key advantages of pupillometry in psycholinguistic research is its fine-grained temporal resolution. Because the pupil responds to changes in cognitive demand within several hundred milliseconds, pupillometry is particularly well-suited to tracking incremental language processing. This makes it ideal for investigating phenomena such as syntactic ambiguity, where increased dilation has been observed during the reanalysis of garden-path sentences (Engelhardt et al., 2010 [134]). These effects demonstrate how pupillometry can capture the additional mental effort involved when linguistic input violates expectations or imposes greater integrative demands.

Pupillometry has also been effectively used in research on lexical access and semantic integration. Words that are low in frequency or semantically anomalous typically elicit larger pupil dilations than high-frequency or contextually congruent words, reflecting the increased effort required for retrieval and integration (Kuchinsky et al., 2013 [76]). This makes pupillometry particularly useful for studying subtle variations in lexical processing, especially in populations or tasks where reaction time or error rate measures may not be sufficiently sensitive.

In the context of bilingual language processing, pupillometry offers compelling insights into the cognitive demands of managing two languages. Increased pupil dilation has been documented during language switching and when bilinguals process input in their less dominant language, reflecting greater executive control involvement (Schneider et al., 2016 [135]). These findings align with the inhibitory control model of bilingualism and provide a real-time physiological measure of language control mechanisms, contributing evidence that complements more traditional behavioural data.

Another growing area of interest is predictive language processing, where pupillometry has been used to detect violations of expectation. When listeners or readers encounter unexpected words in a sentence, they exhibit larger pupil dilations, suggesting that the cognitive system is allocating additional resources to resolve the mismatch between expectation and actual input (Brothers et al., 2017 [136]). This application reinforces pupillometry's value for examining how predictions are formed and updated during comprehension.

Pupillometry has also been employed to study language processing across the lifespan, providing non-invasive insights into developmental and age-related differences. For instance, older adults often show increased dilation during syntactic processing, which has been interpreted as evidence of compensatory effort in the face of declining efficiency (Payne & Stine-Morrow, 2012 [137]). In contrast, in developmental studies, pupillometry is increasingly used to explore the maturation of linguistic and attentional systems in children.

Moreover, pupillometry is uniquely suited to studying language processing in naturalistic and low-demand contexts. Because it does not rely on task responses, it can be used in story listening, dialogue, or passive viewing paradigms, making it especially useful for populations who may be less responsive to explicit experimental instructions, such as children, older adults, or individuals with language impairments.

In sum, although pupillometry has not yet reached widespread adoption in language processing research, it holds considerable promise as a non-invasive, temporally sensitive, and cognitively informative measure. Its growing application, particularly through TEPRs, underscores its utility in studies of sentence comprehension, language switching, and prediction. While gaps remain in its integration with paradigms such as picture naming or lexical decision tasks, the method's ability to capture implicit processing effort in real time makes it an invaluable addition to the psycholinguistic toolkit as the field continues to move toward more nuanced and ecologically valid approaches.

Taken together, the theoretical, empirical, and methodological perspectives reviewed in this chapter converge on a common point: bilingual language processing is a dynamic, cognitively demanding process that calls for sensitive, multi-faceted measurement approaches. The following summary draws these threads together before the next chapter outlines how the present study is designed to address the gaps identified here.

2.8 Summary

This chapter has provided a comprehensive overview of bilingualism and the theoretical, conceptual, and experimental frameworks used to investigate bilingual language processing. Bilingualism is a multifaceted construct encompassing variations in proficiency, dominance, age of acquisition (AoA), and immersion, all of which influence cognitive and linguistic performance (Lagrou, Hartsuiker & Duyck, 2011 [91]; Colomé, 2002 [112]; Dijkstra & Van Heuven, 2002 [52]). These individual differences highlight the necessity of carefully accounting for participant variables when interpreting bilingual research outcomes.

We explored key theories in language processing, including models of lexical access, cross-language activation, and cognitive control mechanisms. Notable frameworks such as the BIA+ model (Dijkstra & van Heuven, 2002 [52]), the TRACE model (McClelland & Elman, 1986 [110]; Hannagan, Magnuson & Grainger, 2013 [14]), and the Cohort model (Marslen-Wilson, 1978 [92]; Magnuson, Mirman & Harris, 2012 [12]) illustrate both monolingual and bilingual representations of the lexicon, emphasizing that lexical and semantic activation is interactive, incremental, and sensitive to task demands. These models underscore that bilingual lexical access is not merely parallel processing of two independent language systems, but a dynamic and competitive system influenced by both bottom-up and top-down processes.

The chapter also critically reviewed experimental paradigms in bilingual research, including the Picture-Word Interference (PWI) task (Rosinski, 1977 [77]; Starreveld & La Heij, 1995 [78], 2016 [81]; Mahon et al., 2012 [83]), the Auditory Word Recognition (AWR) task (Spivey & Marian, 1999 [46]; Marian, Blumenfeld & Boukrina, 2008 [90]; Blumenfeld Marian, 2013 [51]; Lagrou, Hartsuiker & Duyck, 2011 [91]), and the Lexical Decision Task (LDT) (Meyer & Schvaneveldt, 1971 [95]; Kroll, Bobb & Hoshino, 2014 [96]; Zunini et al., 2020 [99]). Each paradigm provides unique insights into mechanisms such as semantic interference, phonological competition, and cross-language activation, yet each has limitations regarding ecological validity, task-dependency, and shallow processing biases. For instance, LDT may underestimate cross-language activation due to its decontextualized stimuli, while AWR and PWI are susceptible to modality-specific or task-specific control effects.

In addition, the chapter highlighted the role of pupillometry as a sensitive, objective measure of cognitive load and attentional engagement (Hess & Polt, 1964 [30]). This technique provides an alternative or complementary approach to traditional behavioural measures, offering a window into real-time processing dynamics that may be particularly useful in bilingual research. The adoption of pupillometry illustrates a broader trend toward multimodal and precise measurement techniques in psycholinguistics.

Overall, bilingual language processing is shaped by a complex interplay of linguistic, cognitive, and task-related factors. Theoretical models, experimental paradigms, and methodological tools each contribute distinct perspectives, but must be interpreted in the context of individual differences, task demands, and the ecological validity of the experimental setup. Crucially, while the LDT remains the most controlled and widely used paradigm for studying bilingual lexical access, its reliance on overt behavioural measure, namely reaction time and accuracy, means that the cognitive effort underlying lexical decisions remains largely invisible. Pupillometry offers a means of addressing this gap directly, providing a continuous, non-invasive index of processing load that can be recorded in parallel with behavioural responses. Yet as the review of pupillometry in psycholinguistics has shown, this method has rarely been applied alongside LDT in bilingual populations, leaving open the question of whether pupillometric responses are sensitive to the kind of cross-linguistic activation that LDT studies have consistently documented. It is precisely this gap that the present study seeks to address, by combining a modified LDT with pupillometric measurement across two bilingual groups differing in the degree of overlap between their languages. The following chapter outlines the methodology employed to do so.

CHAPTER III

CURRENT STUDY

3.1 Introduction

The preceding chapter established that bilingual lexical access is non-selective and parallel, that cognate status modulates processing efficiency in ways that depend on task demands and language relatedness, and that pupillometry remains an underutilized but sensitive index of cognitive effort in lexical decision paradigms, particularly in bilingual populations. The present chapter moves from that theoretical foundation to empirical testing, outlining the design and rationale of a study that directly addresses this gap.

At its core, the study investigates whether bilingual language processing varies as a function of typological proximity between a bilingual's two languages. To explore this issue, two bilingual populations are compared: English–French bilinguals (Group EF), whose languages share substantial orthographic, phonological, and morphological overlap, and English–Mandarin bilinguals (Group EM), whose languages differ markedly across these domains. This comparison enables a systematic examination of whether shared linguistic features modulate cross-linguistic activation, cognate facilitation, and the degree of cognitive control required during lexical access. The present study extends prior behavioural work by incorporating pupillometry as a real-time physiological index of processing effort, the rationale for which is developed in the following section.

Accordingly, this chapter presents the theoretical and methodological background underlying the study, outlines its purpose and predictions, and describes the experimental procedures, thereby establishing the empirical foundation for the analyses and discussion that follow.

3.2 Theoretical and Methodological Background

3.2.1 Bilingual Populations and the Language Discrimination Task

The experimental framework for this research was designed to investigate bilingual language processing in two distinct participant groups: English-French bilinguals (Group EF) and English-Mandarin bilinguals (Group EM). The core paradigm employed is the **Language Discrimination Task** (henceforth LDT), which is derived from the Lexical Decision Task. The motivation for using a Language Discrimination Task is to adapt the procedures of the lexical decision paradigm to probe bilingual language processing. Unlike a standard lexical decision task, which requires participants to decide whether a string of letters is a word or a non-word, a language discrimination task asks participants to identify the language to which a word belongs. While both tasks involve word-level judgements, the language discrimination task specifically targets language membership and cross-language activation in bilinguals, rather than general lexical access.

The inclusion of both Group EF and Group EM bilinguals is particularly interesting because it allows for comparison between two typologically distinct bilingual pairings. English and French share substantial orthographic, phonological, and morphological similarities, including cognate overlap and alphabetic script, while English and Mandarin differ markedly in orthographic system (alphabetic vs logographic), phonological structure, and morphological typology. Comparing these two bilingual populations thus enables examination of whether and how typological proximity influences lexical access and control processes. Prior research has demonstrated that lexical access may be facilitated in bilinguals whose languages share phonological and morphological structures, leading to greater cross-linguistic activation and faster word recognition (Costa, Caramazza, &

Sebastián-Gallés, 2000 [138]; Lemhöfer & Dijkstra, 2004 [69]). In contrast, bilinguals managing typologically distant languages often exhibit reduced lexical competition but greater cognitive control demands when switching or recognizing L2 items (Li, Jiang & Gor, 2017 [139]).

3.2.2 Pupillometry as a Measure of Cognitive Effort

The rationale for employing pupillometry as a core methodological tool is grounded in its proven sensitivity to cognitive effort and attentional dynamics during language processing. Pupil dilation has been repeatedly shown to track processing load during linguistic tasks involving syntactic ambiguity, semantic integration, and lexical retrieval. For instance, Engelhardt, Ferreira & Patsenko (2010) [134] demonstrated that sentences containing temporary syntactic ambiguities elicited larger pupil diameters than unambiguous sentences, indicating increased processing effort. Similarly, Kuchinke et al. (2023) [76] found that pupil dilation decreased when contextual predictability reduced semantic ambiguity, linking smaller pupil size to lower cognitive demands. These findings underscore the suitability of pupillometry for capturing subtle variations in linguistic processing effort that may not be evident in reaction time or accuracy data alone.

In bilingual contexts, pupillometry has been effectively employed to reveal differences in word retrieval effort, language dominance, and attentional control. Schmidtke (2014) [140] found that bilinguals displayed larger and later pupil dilations than monolinguals during a word–picture matching task, with less proficient bilinguals showing particularly pronounced effects. Likewise, Ershaid, McLaughlin & Lallier (2025) [141] observed that highly proficient bilinguals exhibited larger pupil responses when producing words in their less dominant language, even when their behavioural responses were faster, suggesting compensatory allocation of attentional resources. Relatedly, Yao, Connell & Politzer-Ahles (2023) [142] demonstrated that bilinguals listening to affective words showed greater pupil responses in their dominant language, indicating that emotional salience and language proficiency interact to modulate pupillary reactions. Together, these findings affirm the methodological and theoretical value of pupillometry as a real-time indicator of cognitive load during bilingual lexical processing.

3.3 Research Questions and Hypotheses

3.3.1 Research Questions

A central theoretical issue motivating the present study concerns whether bilingual lexical systems are functionally separated or integrated. If lexical representations are strictly language-selective, such that each language is accessed independently, then typological similarity between languages should exert minimal influence on processing; under this account, bilinguals of related and unrelated language pairs would be expected to show comparable behavioural and pupillometric patterns once proficiency is controlled. Conversely, if lexical access is non-selective and involves parallel cross-language activation, then greater structural overlap between languages should modulate processing dynamics, yielding stronger cognate effects and distinct patterns of cognitive effort across groups. The present study seeks to clarify these conditions by addressing the following three research questions:

1. *Can pupillometry serve as a reliable index of cognitive effort in bilingual language discrimination tasks, and does pupillary response to cognate stimuli differ across bilingual groups?*

2. *To what extent does typological similarity between a bilingual's two languages modulate behavioural and physiological indices of lexical processing, as measured by reaction time, accuracy, and pupil dilation?*

3. *To what extent do task demands — operationalized here as the three experimental conditions (NC, CONT, COG) — modulate behavioural and physiological indices of lexical processing, and are these modulations consistent with the task/decision system predictions of the BIA+ model?*

The current study addresses these three research questions by combining behavioural measures (e.g., reaction time and response accuracy) with pupillometric data collected during a language discrimination task. Together, these measures allow us to examine both whether pupillometry is sensitive to the cognitive demands of cognate processing in bilinguals, and whether the degree of typological overlap between a speaker's two languages shapes those demands. The findings from both measures are subsequently interpreted in relation to what they reveal about lexical organisation in the bilingual mind, specifically, whether the two languages are represented as functionally integrated or separated systems.

3.3.2 Hypotheses

The present study advances the following hypotheses concerning bilingual lexical processing, cognate effects, and pupillometric indices of cognitive effort. These hypotheses are grounded in prior work on bilingual lexical access, cross-linguistic activation, and the sensitivity of pupil dilation to processing load.

3.3.2.1 Language Relatedness, Lexical Organization, and Processing Effort

Bilinguals of typologically related languages (Group EF) and typologically unrelated languages (Group EM) will differ systematically in behavioural performance and pupillometric responses, reflecting differences in cross-linguistic activation, lexical organization, and language control mechanisms.

Due to shared orthographic, phonological, and morphological features between English and French, Group EF is expected to exhibit greater cross-linguistic activation and facilitation during lexical access, particularly in tasks requiring language membership judgements. This facilitation is predicted to manifest behaviourally as faster reaction times and higher accuracy, and physiologically as moderate pupil dilation, reflecting efficient processing despite increased cross-language competition arising from overlapping lexical representations.

In contrast, Group EM is predicted to show reduced cross-language activation due to typological distance between English and Mandarin, resulting in comparatively slower reaction times and lower accuracy. At the physiological level, more clearly separated lexical systems are expected to reduce representational competition but increase reliance on active language control, yielding larger and more sustained pupil dilation despite comparable or slower behavioural performance. Together, these patterns are hypothesized to reflect fundamental differences in lexical organization and control demands as a function of language relatedness.

3.3.2.2 Cognate Processing Effects

Cognate stimuli will elicit facilitation effects in Group EF but not in Group EM, reflecting differences in cross-linguistic similarity and lexical representation.

For Group EF, cognates are expected to produce faster reaction times and higher accuracy relative to non-cognate items across language discrimination conditions. This facilitation is hypothesized to arise from shared or highly overlapping lexical representations, whereby orthographic, phonological, and semantic similarity across English and French jointly activate corresponding lexical entries in both languages. Such co-activation is predicted to lower activation thresholds for cognate items, thereby accelerating lexical access and decision-making despite the presence of cross-language competition.

Contrastively, Group EM is not expected to exhibit robust cognate facilitation effects. Given the substantial typological distance between English and Mandarin, including differences in script, phonological structure, and morphological organization, cognate-like items in this language pairing lack sufficient formal overlap to trigger parallel activation across lexicons. As a result, lexical access in Group EM is expected to rely more heavily on language-specific representations and control mechanisms, yielding little to no behavioural advantage for cognate stimuli. Any cognate-related effects observed in this group are predicted to be attenuated, inconsistent, or driven primarily by semantic similarity rather than shared form.

Together, these predicted patterns would support models of bilingual lexical processing that posit graded, similarity-based cross-language activation, in which cognate facilitation is strongest when languages share form-based representations and weakest when lexical overlap is minimal.

3.3.2.3 Pupillometric Indices of Cognitive Effort and Behavioural Correlates

Pupil dilation will vary systematically as a function of task condition and language pair and will correlate with behavioural performance, serving as an index of cognitive effort during bilingual lexical processing.

Across both bilingual groups, the Cognate (COG) condition is expected to elicit pupillary responses that are distinct from those observed in the Control (CONT) and No Control (NC) conditions, though the direction of this effect is predicted to vary as a function of language relatedness. For Group EF, where orthographic and phonological overlap between languages is substantial, cognate processing may sustain heightened cross-linguistic activation across the decision window, potentially manifesting as reduced post-stimulus pupillary recovery rather than an overall increase in peak dilation. For Group EM, where lexical overlap is minimal, the COG condition is expected to impose a comparably high processing load to the other conditions, with less condition-specific modulation of pupillary response. The Control condition, which pairs real L1 and L2 words with phonotactically plausible non-words, is expected to introduce moderate lexical competition through word-status verification demands. The No Control condition, containing only unambiguous L1 and L2 words, is expected to impose the least lexical competition and serve as a processing baseline. The elevated pupillary response predicted for the Cognate condition reflects the additional processing demands associated with language selection when stimuli share form and meaning across languages, generating heightened cross-linguistic lexical competition relative to the other two conditions.

Furthermore, pupillometric measures are expected to correlate with behavioural outcomes such that larger and longer-lasting pupil dilation will be associated with slower reaction times and reduced accuracy, particularly in the Cognate condition and in Group EM. These behavioural/physiological relationships are expected to be strongest in task conditions requiring explicit language discrimination, supporting the interpretation of pupil dilation as a sensitive, real-time physiological marker of cognitive effort that complements traditional behavioural indices of bilingual lexical processing.

Together, these hypotheses generate a coherent set of empirical predictions that the experimental design is specifically structured to test. The following section describes the methodology employed to do so, detailing the participant populations, stimulus materials, experimental conditions, and data collection procedures that together operationalize the theoretical constructs outlined above.

3.4 Methodology

3.4.1 Participants

Participants were recruited from the general student body of York University in accordance with university ethical procedures. Participants were informed, first through recruitment posters (see items E.3, E.4, E.5, and E.6) which prompted prospective participants to answer an initial questionnaire. Then, assuming the individual had met the basic requirements to participate in this study, they were contacted by email, asked to confirm their interest and willingness to participate and a meeting would be booked. Prior to attending the lab, each participant was informed by email that the experiment would require 60-75 minutes of their time, which included breaks. During this time, they would be performing three lexical discrimination tasks for their language group. Finally, they were informed that they would receive a $15CAD Amazon e-giftcard for their participation.

Participants were selected based on the following criteria: (i) must demonstrate full bilingualism within their language group (e.g., Group EM or Group EF), (ii) have normal or corrected vision (glasses or contact lenses were accepted), and (iii) be available and comfortable attending the Speech Acoustics and Perception Lab (SAPLAB) at York University's Keele Campus in person to complete the experiment.

To operationalize “full bilingualism”, participants were required to have been exposed to and actively using their L2 from an early age. Specifically, only individuals who had learned their L2 before the age of six were eligible, ensuring early language acquisition and minimizing the effects of late Age of Acquisition (AOA), which, as discussed in *Chapter II: Literature Review*, can influence bilingual lexical processing and dominance. While formal proficiency tests were not administered, participants completed a Language and Social Background Questionnaire (LSBQ) (from Anderson, Mak, Keyvani Chahi & Bialystok (2018) [143]; see item E.7) that assessed self-reported language fluency, daily language use, and other relevant demographic information such as place of birth and residence. This questionnaire provided a preliminary measure of L1/L2 dominance and confirmed that participants were highly proficient in both languages. By combining early AOA criteria with self-reported fluency and use, we ensured that the sample reflected balanced bilinguals in whom effects of language dominance or delayed acquisition were minimized.

Once scheduled, participants visited the lab to review general research information and sign a consent form, ensuring they were fully aware of the study’s purpose, procedures, and the data being collected. This preparatory step allowed us to gather crucial background information that could influence responses in each experimental task while maintaining ethical and informed participation.

In total, 50 participants were selected, 25 for Group EM, and 25 for Group EF. A complete breakdown of participant background, including AoA, Country of birth, and Native/Non-Native status can be found in A.1 and A.2. Generally speaking, participant ages ranged between 18 years and 60 years of age and included a mix of undergraduate, graduate and post-doctoral individuals.[3] For Mandarin speakers, the average age of participants was 25.6 years, while for French speakers,

[3] Given the primary interest and scope of this study, age was not a factor which was controlled for. Crucially, we were looking for healthy participants with no history of neurological issues (diagnosed or not), who had normal or corrected vision and who were bilingual in Group EF or Group EM.

the average age was of 21.5 years. Of the Mandarin speakers, 22 participants were L1 native speakers of Mandarin while 3 considered themselves as heritage speakers. Further, of the Mandarin speakers, 23 participants were born outside of Canada, having moved to Canada as children or during their early teen years[4], while 2 reported they were born and raised in Canada. In contrast, of the French speakers, 18 participants considered themselves L1 native French speakers while 7 participants self-reported as being L2 heritage speakers. Further, 17 participants were born in Canada while 8 participants reported being born outside of Canada, again immigrating during their childhood or early teenage years (see preceding note regarding English proficiency verification criteria for participants born outside of Canada).

Table 3.1: A summary breakdown of the distribution of participants based on their L1 and L2, as well as the average age of acquisition for non-native French and Mandarin speakers

	French L1 (native)	**French L2 (heritage)**	**Mandarin L1 (native)**	**Mandarin L2 (heritage)**
Total Number of speakers	18	6	22	3
Average Age of Acquisition	0	7.4	0	4.75

With respect to Group EM participants specifically, although familiarity with *Pinyin* was not assessed, no participant reported being uncomfortable with the *Pinyin* presentation; in fact, all participants reported familiarity with the system through both formal instruction or day-to-day usage. This is noteworthy, as it provides an important point of validity for the use of *Pinyin* as the writing system for Mandarin stimuli in the present study. Notably, *Pinyin* is a standardized

[4]For participants born outside of Canada, English proficiency was verified through the Language and Social Background Questionnaire (LSBQ). To meet the study's full bilingualism criterion, each of these participants satisfied at least one of the following conditions: (i) at least one parent was a native or highly proficient English speaker and used English in the home; (ii) the participant attended an English-medium or internationally accredited school prior to immigrating to Canada; or (iii) English held official or widely used second-language status in their country of origin, providing sustained institutional exposure prior to immigration. These criteria ensured that English acquisition began early and was not limited to post-immigration classroom instruction, thereby satisfying the Age of Acquisition (AoA) threshold of onset before age six outlined in the participant eligibility criteria above.

romanization system and is widely used across Mainland China, Taiwan, and among diaspora communities worldwide. As such, young Mandarin speakers today are broadly accustomed to encountering and using *Pinyin* in a variety of everyday contexts—from texting, where *Pinyin* input is commonly used as shorthand in lieu of stroke-based character entry, to passive exposure through social media, advertisements, billboards, and video-sharing platforms. This widespread familiarity suggests that the use of *Pinyin* in the present study did not constitute an unfamiliar or artificial reading condition for Group EM participants, and that stimuli presented in this format were likely processed with a reasonable degree of fluency across the group.

Finally, and following a review of the participants, one Group EF speaker had to be excluded as they did not fully meet the requirements to participate in this study. This was due to the fact that upon meeting the participant and going through the LSBQ with them, it was quickly noticed that this individual was not, in fact, bilingual and had instead stated they were on the basis that one of their parents was Franco-Canadian. As a result, only the data from 24 Group EF participants was considered for the final analysis. In total, 49 participants—25 Group EM and 24 Group EF participants—were considered for this experiment.

3.4.2 Materials

3.4.2.1 Word List

Before developing the experimental code—which consisted of individual Python scripts for each task condition paired with a Gazepoint eye-tracking wrapper (see Dalmaijer, Mathôt & Van der Stigchel [144])—care was taken to establish the word lists that would serve as the stimulus materials for this research. The intention was to create one 50-word word list for each test condition for each language group, with tokens distributed as evenly as possible within each condition to minimize any biasing effects during the Language Discrimination tasks.

The experiment comprised three task conditions, each designed to probe a different aspect of bilingual lexical processing. In the **No Control (NC)** condition, word lists contained only real words in English and the participant's L2 (French or Mandarin), with no non-words or cognates present; participants were required simply to identify the language of each stimulus. In the **Control (CONT)** condition, real L1 and L2 words were presented alongside phonotactically plausible non-words, requiring participants to first assess word status before making a language judgement. In the **Cognate (COG)** condition, real L1 and L2 words were presented alongside cross-linguistically overlapping cognate items, introducing the additional challenge of resolving lexical competition arising from shared form and meaning across languages. Table 3.2 provides a preliminary breakdown of token distribution across these three conditions for each language group.

Table 3.2: Preliminary distribution of tokens across languages and task conditions

	No Control	**Control**	**Cognate**
English	25	17	17
French/Mandarin	25	17	17
Non-word	N/A	16	N/A
Cognate	N/A	N/A	16

To ensure we did not choose abstract concepts or words which could potentially be misinterpreted, only concrete nouns and verbs were chosen for the word lists. Words exceeding three syllables were excluded to minimize potential confounding effects of lexical length on reaction time (RT) measurements. Beyond these points, words in English, French and Mandarin were generally chosen arbitrarily based on commonly known and used words. Non-words were arbitrarily generated with the only requirements being that they followed general English or French phonotactic patterns—for Mandarin, this would correspond to general *Pinyin* conventions—and that the word did not exceed three syllables. Tables 3.3 and 3.4 provide examples of words that were included in each experimental task conditions. A full word list is included at the end in the *Appendix C: Word List*.

Table 3.3: Examples of words that were included in each task condition (NC, CONT, COG) for Group EF

Token	Token Type
CALL	English Word
AVOIR	French Word
AUTILLER	Non Word
CHAIR	Cognate

Table 3.4: Examples of words that were included in each task condition (NC, CONT, COG) for Group EM

Token	Token Type
DAY	English Word
DE	Mandarin Word
BAIO	Non Word
LAKE	Cognate

A trilingual (English, French, and Mandarin) speaker was consulted during the creation of the word list. As for cognates, these were selected from among high-frequency words in each language, operationalized as items falling within the upper frequency bands of widely used subtitle-based lexical frequency norms for English, French, and Mandarin (SUBTLEX-US: Brysbaert & New, 2009 [145]; Lexique 3 for French: New et al., 2004 [146]; SUBTLEX-CH: Cai & Brysbaert, 2010 [5]). These corpora provide standardized frequency estimates based on large subtitle datasets and are commonly used in psycholinguistic research to approximate everyday lexical exposure. Given the typological differences between the language pairs, cognate status was operationalized differently for the English–French (EF) and English–Mandarin (EM) groups. For Group EF, cognates were defined primarily as orthographically identical or near-identical lexical items that share semantic overlap across English and French (e.g., PAIN). Given the shared Latin script and substantial lexical overlap between the two languages, orthographic transparency served as the primary criterion, with semantic relatedness verified through native speaker consultation. Only items with clear semantic correspondence across languages were retained in the final list.

For the EM group, however, the distinct writing systems of English and Mandarin preclude orthographic overlap as a defining criterion. Cognates were therefore defined on the basis of phonological similarity and semantic correspondence rather than orthographic form. Specifically, items were selected where the Mandarin lexical item (presented in *Pinyin*) bore a clear phonological resemblance to an English lexical item and shared a related meaning (e.g., kāfēi – “café”). These items were reviewed with a native Mandarin speaker to ensure that the perceived phonological similarity was plausible for bilingual speakers and that semantic overlap was sufficiently transparent.

This difference in operationalization reflects structural differences between the language pairs rather than inconsistency in selection criteria. In the EF group, orthographic overlap provides a salient cross-linguistic cue due to the shared alphabetic system, whereas in the EM group, phonological overlap serves as the primary cross-linguistic link because Mandarin employs a logographic writing system. The study therefore defines cognates functionally—as lexical items exhibiting cross-linguistic form similarity (orthographic or phonological) alongside semantic relatedness—while adapting the specific criteria to the typological characteristics of each language pair.

Each word list was then carefully reviewed with a native speaker of Mandarin and with another native speaker of French. Following these reviews, three words were excluded from the English-Mandarin lists, and one was added to the French. Tables 3.5 and 3.6 provide a distribution of the tokens based on task for the final word lists for Group EF and Group EM

Table 3.5: Final distribution of tokens for Group EF across all three test conditions

	No Control	Control	Cognate
English	25	17	18
French	25	17	17
Non-word	N/A	16	N/A
Cognate	N/A	N/A	16

Table 3.6: Final distribution of tokens for Group EM across all three test conditions

	No Control	Control	Cognate
English	23	17	18
Mandarin	25	17	19
Non-word	N/A	14	N/A
Cognate	N/A	N/A	14

3.4.2.2 Coding

Once the stimulus materials were finalized, the focus shifted to the technical implementation of the experimental paradigm. The intention was to create a straightforward Language Discrimination task (LDT) that would require minimal input from the participant while ensuring that all relevant data was captured. As mentioned previously, this experiment consisted of two major components: first, there was the task component, which would track accuracy and response times to each stimuli. The second was the eye-tracking component, which would account for all eye-related data during the trial periods.

With regard to LDT, the structure of the code was as follows: Tokens (words) were presented in random order based on the pre-determined word list. Before each token presentation, there was a fixation mark lasting 8 seconds. This duration was chosen to ensure that the participant's gaze and pupils would be able to readjust and return to a neutral or resting position.[5] After the fixation

[5]It is widely known that pupil dilation is involuntary and will occur at any given point (see Mathôt, 2018 [147]) . During these 8 seconds, it is not expected that the participant will remain at baseline, but rather that they reach baseline dilation.

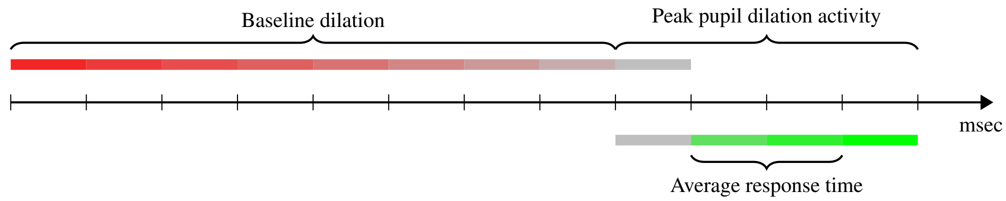


Figure 3.1: Timeline for token presentation during LDT based on trends in pupil dilation. Assume each tick represents $1000msec$. The grey-red bar from $0msec$ and $8,000msec$ corersponds with the fixation point presentation and represents the baseline or resting period. At $8,000msec$, the fixation point would time out and the token would be displayed. The grey overal represents the first $1,000msec$ of the token presentation. At this point, we expect to observe first dilation activity. By $9,000msec$, the participant would have recognized the token and would have likely already responded. At this point, we expect to observe peak or maximum pupil dilation activity. If no response was captured, the remaining trial time—between $9,000msec$ and $10,000msec$—would correspond to reflection or processing time.

point, a token would be displayed for a total of 5 seconds or until user input was recorded. Previous research has shown that people take an average of 400 and $500msec$ to process words via a visual modality during rapid recognition tasks (see Hoedemaker, Gordon & Peter 2017 [148], Balota et al., 2008 [149]; Kinoshita et al., 2015a,b [150]). Thus, the reason for displaying each token for 5 seconds was to ensure the participant would have enough time to read, recognize the word and provide a response, especially in conditions wherein more thought and reflection would be predicted, therefore leading to longer reaction time (RT). Figure 3.1 illustrates the model timeline for pupil dilation activity as predicted for this research.

Once the model for each trial was created, the experimental LDT code was set up to capture behavioural data (e.g., RT and response) and timestamps at three crucial points, outlined here:

1. The experiment code was designed to log response or "key inputs", which corresponded with the arrow key selection made by the participant.

2. The experiment code was designed to log Reaction Time (RT), which corresponded with the total amount of time (in milliseconds) between the stimuli presentation and the participant's recorded response

3. The experiment code was designed to track eye data at three points: (i) at the fixation point presentation or *CrossTimeStamp*, (ii) at the stimulus presentation or *WordTimeStamp*, and (iii) at the response time or *KeyInputTimeStamp*

A total of six LDT scripts were written, one for each condition (e.g., NC, Control, Cognate) and one for each language group. Finally, the data captured by the LDT code would be compiled into a single spreadsheet, labelled "Lexical Decision Task Response", each time it was run. With the behavioural data pipeline established, it was then necessary to implement a parallel system for capturing ocular data in real time and synchronizing it with the LDT output. This was achieved through the integration of an open-source eye-tracking wrapper, described in the following subsection.

3.4.2.3 Opengaze.py

A Gazepoint GP3 eye tracker was used to record all eye-related data, including fixation and gaze location (i.e., tracking the participant's gaze relative to an X–Y axis division of the monitor). To ensure that the Gazepoint GP3 functioned seamlessly alongside the experimental program, I implemented the wrapper developed by Dalmaijer, Mathôt & Van der Stigchel (2014) [144]. This wrapper was part of their open-source, cross-platform toolbox for eye-tracking experiments. Its primary purpose was to allow the eye tracker to operate in parallel with the Language Discrimination Task (LDT), ensuring synchronized recording of ocular and behavioural data.

While the experimental script logged stimulus presentation and response timestamps, the wrapper simultaneously generated a separate output file dedicated exclusively to ocular measurements. These measurements were sampled at $17 msec$ intervals from the onset of the task, matching the native sampling rate of the Gazepoint GP3. At the conclusion of each trial, a CSV file containing the complete eye-movement record was produced. This parallel recording system provided an additional verification layer, allowing cross-referencing of ocular and behavioural data in cases of potential timing discrepancies. Further details regarding data verification procedures are discussed in the Data Preparation section. Importantly, running both systems concurrently across all experimental conditions maximized data capture and ensured precise alignment between pupil dilation metrics and behavioural responses (reaction time and accuracy).

Methodologically, the study incorporated several safeguards to ensure the validity and reliability of the pupillometric data. Each participant underwent eye-tracker calibration prior to testing, and pre-trial fixation baselines were collected to account for inter-individual variability in pupil size. Data preprocessing procedures included blink detection and removal, interpolation of missing samples, and normalization of pupil size relative to baseline. The dependent variables comprised behavioural measures (reaction time and accuracy) and pupillometric indices (baseline-adjusted peak dilation, latency to peak, and sustained dilation duration). Mixed-effects regression models were employed to examine both within-participant effects (NC vs. CONT vs. COG conditions) and between-group differences (Group EF vs. Group EM), with pupillometric measures incorporated to assess their relationship to behavioural outcomes. By integrating behavioural and physiological indices, the study aims to provide a comprehensive account of how language similarity, lexical competition, and cognitive control interact during bilingual lexical access.

3.5 Experimental Procedures

The experiment was conducted in the SAPLAB at York University's Keele Campus, in a quiet, well-lit room with a south-facing windows directly left of the computer and eye-tracker. During testing, the blinds were kept closed to minimize excess lighting. Participants were seated at the computer, with the Gazepoint GP3 eye tracker positioned just below standard eye level beneath the monitor. To reduce tracking errors, a chin rest was used to stabilize the participant's head throughout the session. For each test condition, participants were provided with instructions and encouraged to ask questions before beginning. The eye tracker was recalibrated at the start of each condition to further minimize potential tracking errors.

The experiment consisted of three task conditions. In the No Control (NC) condition, participants simply decided whether a visually presented orthographic stimulus was a real word in their L1 (English) or their L2. The Control (CONT) condition required participants to judge whether the token was a real word or a non-word, following a standard lexical decision paradigm. In the Cognate (COG) condition, participants first determined the language of the token (L1 or L2), and for cognates (i.e., cross-language words sharing meaning or form) they additionally indicated whether the item was a real word in L1 or L2.

Throughout the experiment, two types of data were collected simultaneously. First, offline behavioural judgment data were recorded, including participants' responses, reaction times (RTs), and accuracy, reflecting explicit decisions about word recognition and language membership. Second, online pupillometry data were continuously recorded via the eye-tracking system, capturing real-time physiological responses to each stimulus. The eye tracker was calibrated to measure baseline pupil size, peak dilation, and dilation duration. These pupillometric indices—including average pupil size, maximal deviation from baseline, and sustained dilation—were analyzed alongside the behavioural measures to explore how cognitive effort varies across conditions and between bilingual groups, providing a more comprehensive view of lexical processing dynamics.

The following sub-sections describe the procedure for each test condition from start to finish.

3.5.1 Condition 1: No Control

The No Control (NC) condition was the first of the three tasks that participants were asked to complete, and included only words in English and French (see distribution in Table 3.5), and English and Mandarin (see distribution in Table 3.6). For this task, participants were invited to sit in front of the computer, and the Gazepoint tracker was adjusted and calibrated for the participant. Instructions for this task were clearly explained as follows:

1. Once we launch this task, you will see a black screen pop up. The same instructions I am listing will be on the screen for you to read.
2. You will see a red cross appear in the middle of the screen for a few seconds; this is called the fixation point. During this time, you may blink, just make sure to bring your gaze back to this cross.
3. Once the cross disappears, you will see a word. Please hit the left arrow key if you believe the word is English and right if you believe the word is French/Mandarin.
4. Once you key in your answer, the process will repeat again until you are done. This task will last 5-8 minutes.
5. In the event you do not answer, the word will time out and you will move on to the next fixation and word.

6. Please hit the space bar to launch the task once you feel ready to begin.

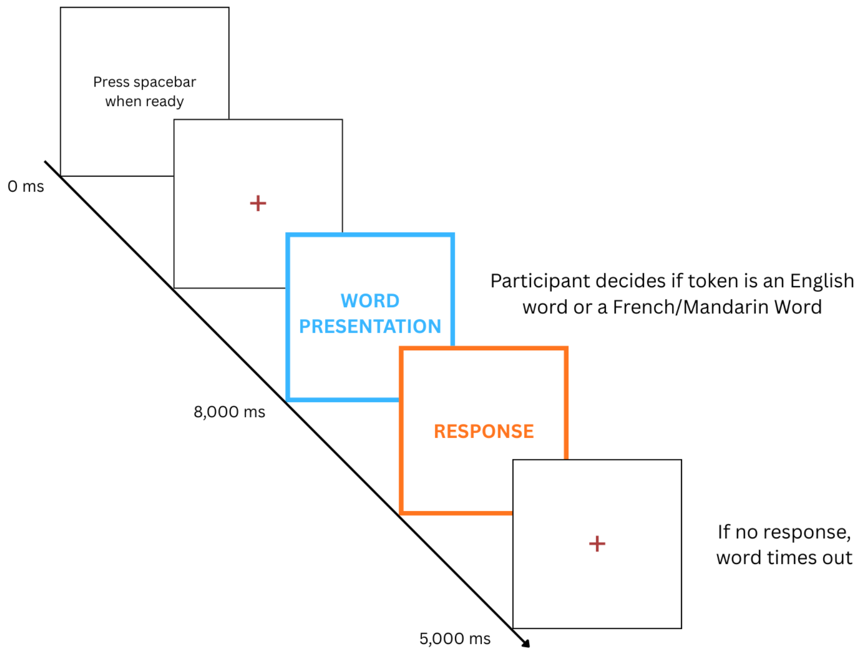


Figure 3.2: Experimental framework for the No Control Task. In this task condition, participants are presented first with a fixation point for 8 seconds. Then they are presented with the stimulus. During this time, the participant must decide if the word is an **English word** or a **French/Mandarin word**.

After explaining the instructions and answering all questions, the task code was launched and the participant was invited to begin. The No Control task consisted of 48 tokens for Group EM and 50 tokens for Group EF, and followed a fixation-point and stimulus-presentation cycle, as illustrated in Figure 3.2. The entire task took on average 7 minutes to complete. Upon completion, the participant was invited to step away and take a quick break.

3.5.2 Condition 2: Control

The Control task constituted the second condition in the experimental sequence, introducing an additional response option (i.e., the non-word judgement) that required participants to discriminate not only between languages but between real words and phonotactically plausible letter strings. This task included a mix of English, French and Non-words (see distribution in Table 3.5) for Group EF, and English, Mandarin and Non-words (see distribution in Table 3.6) for Group EM. The same setup was used for this task and the process was the same. Instructions were presented to the participants as follows:

1. Once we launch this task, you will see a black screen pop up. The same instructions I am listing will be on the screen for you to read.
2. You will see a red cross appear in the middle of the screen for a few seconds; this is called the fixation point. During this time, you may blink, just make sure to bring your gaze back to this cross.
3. Once the cross disappears, you will see a word. Please hit the left arrow key if you believe the word is English, right if you believe the word is French/Mandarin and press the up arrow key if you believe the word is not real.
4. Once you key in your answer, the process will repeat again until you are done. This task will last 5-8 minutes.
5. In the event you do not answer, the word will time out and you will move on to the next fixation and word.
6. Please hit the space bar to launch the task once you feel ready to begin.

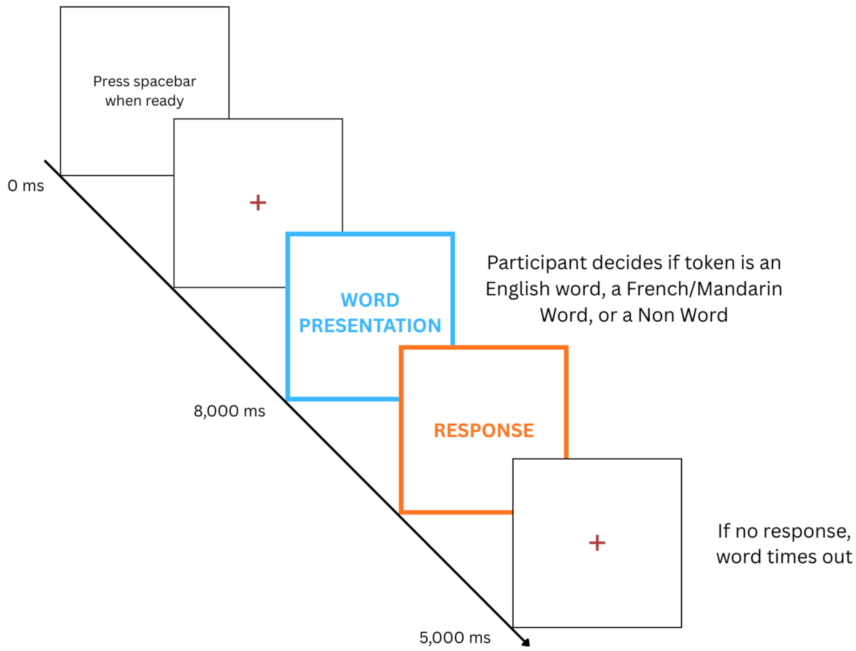


Figure 3.3: Experimental framework for the No Control Task. In this task condition, participants are presented first with a fixation point for 8 seconds. Then they are presented with the stimulus. During this time, the participant must decide if the word is an **English word**, a **French/Mandarin word**, or a **non word**.

After explaining the instructions and answering all questions, the task code was launched and the participant was invited to begin. The Control task included a total of 48 tokens for Group EM and 50 tokens for Group EF and followed a fixation-point and stimulus-presentation cycle, as illustrated in Figure 3.3. The entire task took on average under 8 minutes to complete. Upon completion, the participant was invited to step away and take a quick break.

3.5.3 Condition 3: Cognates

The Cognate task served as the final experimental condition, presenting the most complex language membership judgement of the three tasks by introducing items whose form overlapped across both of the participant's languages. The task order (i.e., NC followed by CONT, and finally COG) was deliberately structured rather than randomized across participants. This sequencing was motivated by a desire to minimize potential cross-task priming effects. Specifically, presenting the Cognate condition last helped prevent participants from becoming consciously or unconsciously attuned to cognates earlier in the experiment, which could have influenced their responses in the No Control or Control tasks. By placing the COG task at the end, participants completed the simpler lexical and language discrimination trials first, establishing baseline performance measures without prior exposure to cognate-related cues. This approach ensured that any effects observed in the Cognate condition reflected genuine lexical processing differences rather than carry-over effects from earlier tasks.

This task included a mix of English and French words, along with cognates (see distribution in Table 3.5) for Group EF, and English and Mandarin words along with cognates (see distribution in Table 3.6) for Group EM. Notably, all tokens in this task condition were presented in all-caps using the Roman alphabet, without accents. For Group EM, this meant that even Mandarin cognates were rendered according to standard romanization (*Pinyin*) conventions, resulting in words that appeared neutral or ambiguous; they did not strongly resemble either English or Mandarin orthography.

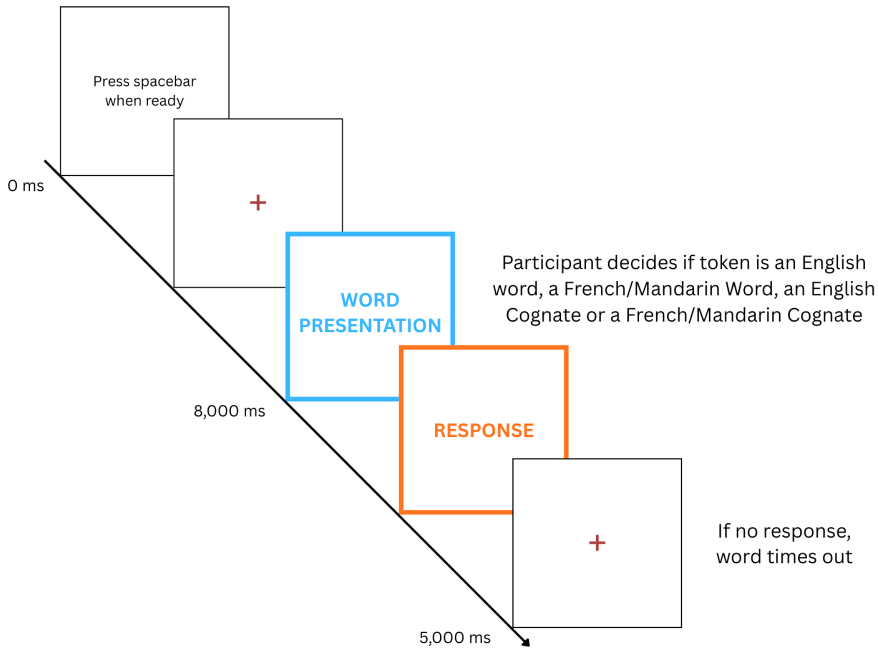


Figure 3.4: Experimental framework for the Cognate Task. In this task condition, participants are presented first with a fixation point for 8 seconds. Then they are presented with the stimulus. During this time, the participant must decide if the word is an **English word**, a **French/Mandarin word**, an **English cognate**, or a **French/Mandarin cognate**.

The same setup was used for this task and the process was the same (see Figure 3.4). Instructions were presented to the participants as follows:

1. Once we launch this task, you will see a black screen pop up. The same instructions I am listing will be on the screen for you to read.

2. You will see a red cross appear in the middle of the screen for a few seconds; this is called the fixation point. During this time, you may blink, just make sure to bring your gaze back to this cross.
3. Once the cross disappears, you will see a word. Please hit the left arrow key if you believe the word is English and right if you believe the word is French/Mandarin.
4. For words that look like either Mandarin/French or English, go with your first instinct and select left for English or right for Mandarin/French based on your assessment.
5. Once you key in your answer, the process will repeat again until you are done. This task will last 5-8 minutes.
6. In the event you do not answer, the word will time out and you will move on to the next fixation and word.
7. Please hit the space bar to launch the task once you feel ready to begin.

After explaining the instructions and answering all questions, the task code was launched and the participant was invited to begin. The Cognate task included a total of 51 tokens for Group EM and 51 tokens for Group EF. The entire task took on average 8 minutes (rounded to the closest minute) to complete.

CHAPTER IV

RESULTS AND ANALYSIS

4.1 Data Review and Cleaning

Previous literature (see Valtakari et al. (2021) [151]; Cuve et al. (2022) [152]) has touched on the reliability of the Gazepoint GP3 eyetracker. Of note, they highlight issues surrounding low sampling rate and possibility of data loss as being primary offenders when it comes to the overall quality of the data. Being cognizant of these risks, the decision was made to review each data file against its respective debug files to ensure all lines corresponded with the three key timestamps and to reduce the number of error trackings if at all. A line of data was considered an error track if it met any of the following criteria:

1. If the pupil dilation data were impossible figures (e.g., values under $2mm$ or 9.3 pixels, or over $8mm$ or 37 pixels);
2. If the pupil dilation data were blank (i.e., no data recorded);
3. If the eye-tracking data contained the same values across multiple lines.

Since the debug file captured all entries recorded at $17msec$ intervals, we were able to correct some error tracks by looking at the token (i.e. the individual word) prior to or after the one extracted by the Debug extraction script. All error tracks that could not be manually corrected were removed from the participant's results files.

Group	ID	ITEM	Correct NC	RT - NC	Correct CONT	RT - CONT	Correct COG	RT - COG
		1	1	838				
		2	1	1103				
		3	0	1408				
		4	RIGHT	1402				
		5	LEFT	1101				
		6	RIGHT	2407				
		7	LEFT	1096				

Figure 4.1: Final Response data file with response and RT populated for all participants across both language pairs

Following a manual review of each participant and their files, all files were consolidated into two Excel spreadsheets, which would become the analysis files. The first file was for Response and RT only, and took into account the task condition (e.g., NC, CONT, and COG), the tokens (i.e., word) and their word status (e.g., whether the token or word was a real word, a non-word, or a cognate) for each task condition, the language group and the participant number (see Figure 4.1). The second file was created for the analysis of the left and right pupil dilation across timestamps, the language group, and the participant number (see Figure 4.2).

Group	ID	ITEM	NC PD1 L	NC PD1 R	NC PD2 L	NC PD1 R	NC PD3 L	NC PD3 R	CONT PD1 L	CONT PD1 R	CONT PD2 L	CONT PD2 R	CONT PD3 L	CONT PD3 R	COG PD1 L	COG PD1 R	COG PD2 L	COG PD2 R	COG PD3 L	COG PD3 R
		1	0.00408	0.00431	0.00455	0.0046	0.00451	0.00466												
		2	0.00525	0.00542	0.0057	0.00598	0.00527	0.00542												
		3	0.00561	0.00564	0.00497	0.00515	0.00526	0.0054												

Figure 4.2: Final eyetracking data file with response and RT populated for all participants across both language pairs

For both files, language groups were coded as either “1” (for English-French) or “2” (for English-Mandarin). Tokens were sorted in alphabetical order and then assigned a number “1” through 50. For the Response and RT file, a correct response (i.e., correctly identified the language) was coded as “1”, while an incorrect response (i.e., incorrectly identified language) was coded as “0” in the NC condition. For the Control condition, a response code of “1” corresponded to real words while responses coded as “0” indicated non-words. For the Cognate condition, correct responses (i.e., correctly identified the language) were coded as “1”, incorrect responses (i.e., incorrectly identified language) were coded as “0”, cognates identified as English were coded as “2”, and cognates identified as French or Mandarin were coded as “3”.

For the eye-tracking data file, left and right pupil measures were separated according to the task condition and the timestamps; for instance, *NC PD1 L* would correspond to *No Control*, *Timestamp 1* (fixation), *Left pupil*.

Table 4.1: Coding pattern for responses in the Group EF and Group EM No Control task condition.

Token	Real English	Real French
BOY	1	0
ABEILLE	0	1

Token	Real English	Real Mandarin
CLOSE	1	0
DAXUE	0	1

Table 4.2: Coding pattern for responses in the Group EF and Group EM Control task condition.

Token	Real English	Real French	Non-Word
BE	1	0	0
BERGER	0	1	0
AUTILLER	0	0	1

Token	Real English	Real Mandarin	Non-Word
CRANE	1	0	0
XIAOBA	0	1	0
OPNER	0	0	1

Table 4.3: Coding pattern for responses in the Group EF and Group EM Cognate task condition. The value “0” indicates an incorrect response, “1” a correct response, “2” a cognate identified as English, and “3” a cognate identified as French or Mandarin.

Token	Real English	Real French	Cognate: EN	Cognate: FR
BARK	1	0	0	0
AVOIR	0	1	0	0
POUR	0	0	2	3

Token	Real English	Real Mandarin	Cognate: EN	Cognate: CH
DAY	1	0	0	0
CHUSHI	0	1	0	0
KAFEI	0	0	2	3

4.2 Experimental Outcomes: Behavioural Data and Pupillometry

This chapter presents the experimental results across the three task conditions (i.e., No Control (NC), Control (CONT), and Cognate (COG)) for both language groups (EF and EM). The results were organized by condition and outcome measure. For each condition, reaction time (RT) and response accuracy were examined first, followed by pupil dilation as an index of cognitive load. Where relevant, findings from the mixed-effects models are reported in terms of fixed effects (language group, trial, and their interaction) and random effects (by-participant variability). Although participant background variables—including language dominance and age of onset of bilingualism (see *Chapter II: Literature Review*)—were collected via questionnaire, they are not the primary focus of the present analyses. These variables were referenced only where they meaningfully informed patterns of between-participant variability; a more comprehensive examination of individual differences is reserved for future work.

Preliminary analyses were conducted in Excel to obtain an initial descriptive overview of RT and accuracy patterns across conditions. These were followed by inferential analyses using mixed-effects modeling in R (version 4.5.1; R Core Team), allowing both fixed and random sources of variability to be accounted for simultaneously. Type III Wald chi-square tests were used to assess whether RT and accuracy were significantly influenced by TASK_CONDITION, LANGUAGE_GROUP, and their interaction. Comparable modelling procedures were applied to the analysis of pupil dilation over time. The results presented below were described with reference to the theoretical frameworks introduced in *Chapter II: Literature Review* and *Chapter III: Current Study*, with particular attention to how task condition and language background relate to lexical processing and cognitive load.

As outlined in Section 3.4: Methodology, stimulus selection was designed to minimize potential confounds across conditions. Only concrete nouns and verbs were included, and words exceeding three syllables were excluded in order to reduce potential lexical length effects on reaction time. Non-words were constructed in accordance with the phonotactic conventions of each respective language (English, French, and Mandarin). For the Cognate condition, items were selected on the basis of frequency and cross-linguistic similarity, with attention to maintaining reasonable comparability in phonological and morphological structure across languages. Representative examples are provided in Tables 3.3 and 3.4, and the complete word lists are available in *Appendix C*.

Together, these selection criteria were implemented to ensure that observed differences in reaction time and accuracy reflected the experimental manipulations rather than uncontrolled linguistic variability.

Table 4.4: A summary of the average Reaction Time (RT) in msec for Group EF and Group EM across the No Control, Control and Cognate conditions.

Condition	**EF RT (ms)**	**EM RT (ms)**
No Control	1,075	1,036
Control	1,447	1,377
Cognate	1,107	1,224

Table 4.4 presents the average reaction time (RT) in milliseconds for each language group across all three experimental conditions. The No Control condition elicited the fastest responses in both groups (EF: $1,075ms$; EM: $1,036ms$). Reaction times increased substantially in the Control condition (EF: $1,447ms$; EM: $1,377ms$). In the Cognate condition, Group EF showed reaction times $(1,107ms)$ comparable to the NC baseline, whereas Group EM demonstrated slower responses $(1,224ms)$ relative to their NC performance.

Table 4.5: A summary of the average net accuracy scores for Group EF and Group EM across all conditions, calculated as the proportion of correct responses relative to total responses.

Condition	Group EF Accuracy	Group EM Accuracy
No Control	97.67%	97.13%
Control	91.32%	78.33%
Cognate	98.43%	91.57%

Table 4.5 presents net accuracy scores—calculated as the proportion of correct responses relative to total responses—across all three conditions. The NC condition yielded the highest overall accuracy (EF: 97.67%; EM: 97.13%), the Control condition produced the lowest scores in the initial net analysis (EF: 91.32%; EM: 78.33%), and the Cognate condition showed apparent high accuracy (EF: 98.43%; EM: 91.57%) that, as discussed in *Section 4.2.3*, changes substantially when token types are taken into account.

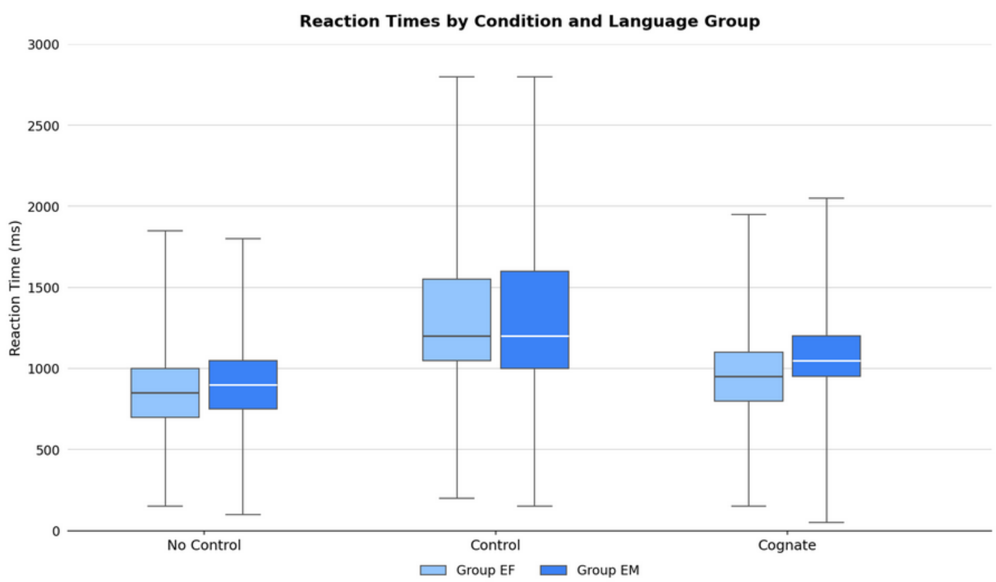


Figure 4.3: A comparison of Reaction Time by Condition and Language group.

As illustrated in Figure 4.3, descriptive comparisons of RT across conditions and language groups reveal consistent patterns. Group EF exhibited slightly longer average RTs than Group EM across the NC and COG conditions, while Group EM yielded longer RT in the CONT condition.

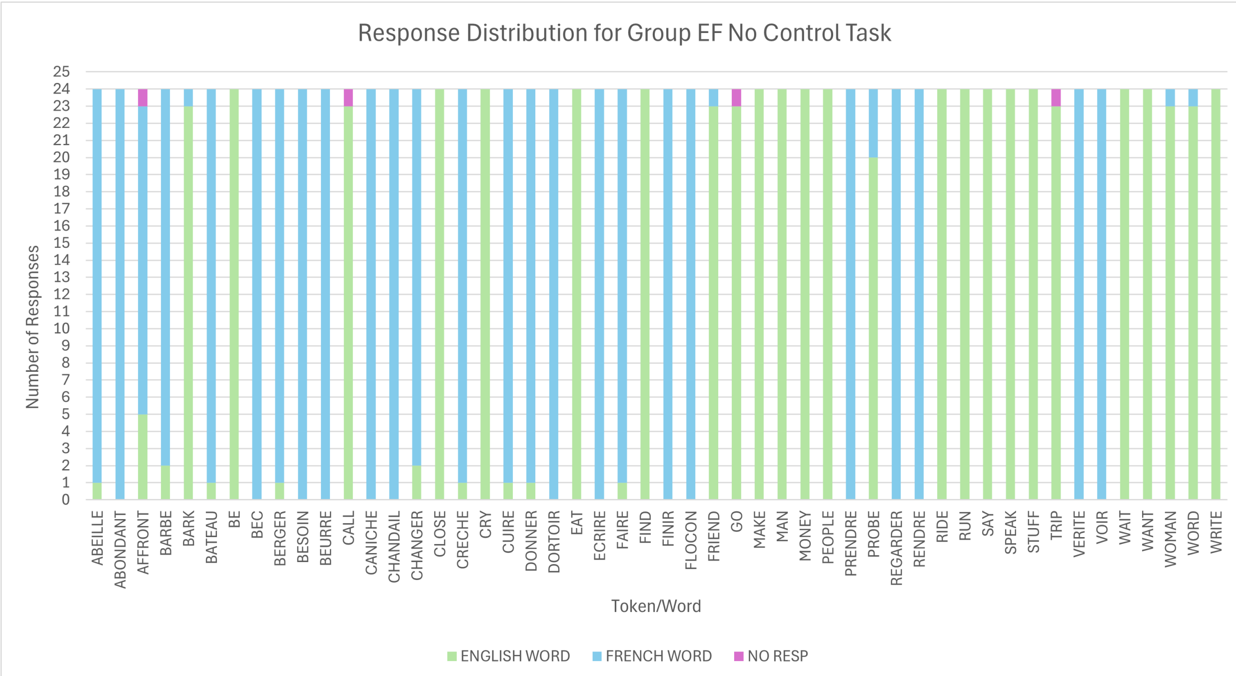
Furthermore, conditions associated with longer RTs also showed higher error rates: RT increased from NC to Control, and error rates increased in parallel, with the highest error rates observed in the Cognate condition.

Figures 4.4, 4.5, and 4.6 provide token-level overviews of response distributions for each language group across all three conditions. In these graphs, a 'LEFT' response indicates an English selection, a 'RIGHT' response indicates a French or Mandarin selection, and 'NO RESP' indicates that no response was made.

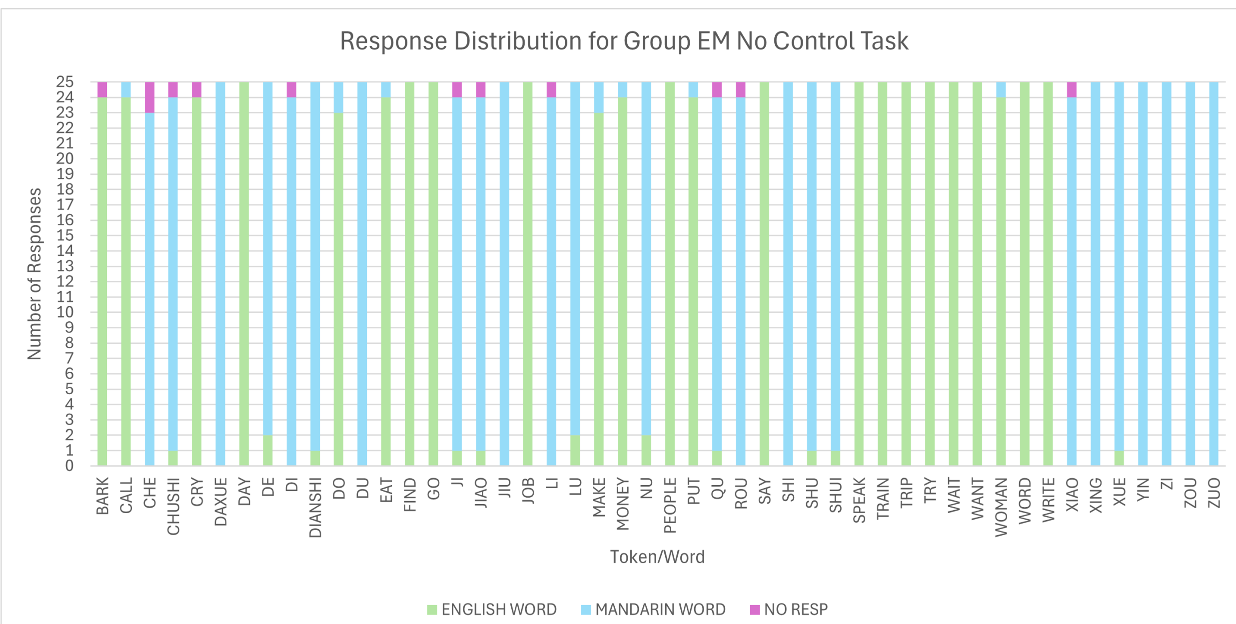
To statistically evaluate the descriptive differences across conditions and language groups shown above, a linear mixed-effects model (LMM) was first fit to the full dataset using the `lmer` function from the `lme4` package (Bates et al., 2015 [153]) in R. The model included TASK_CONDITION, LANGUAGE_GROUP, and their interaction as fixed effects, with participant included as a random intercept:

$$\text{RT} \sim \text{TASK\_CONDITION} \times \text{LANGUAGE\_GROUP} + (1\text{—PARTICIPANT})$$

The model revealed a significant main effect of CONDITION ($\chi^2 = 202.07$, $df = 2$, $p < .001$), confirming that RT differed significantly across the three conditions. A significant main effect of LANGUAGE_GROUP was also observed ($\chi^2 = 14.02$, $df = 1$, $p < .001$), as was a significant CONDITION × LANGUAGE_GROUP interaction ($\chi^2 = 21.18$, $df = 2$, $p < .001$), indicating that the effect of condition on RT differed across language groups. Pairwise comparisons confirmed that all three condition contrasts were statistically significant ($p < .01$), with the NC–Control difference being the largest ($482ms$).

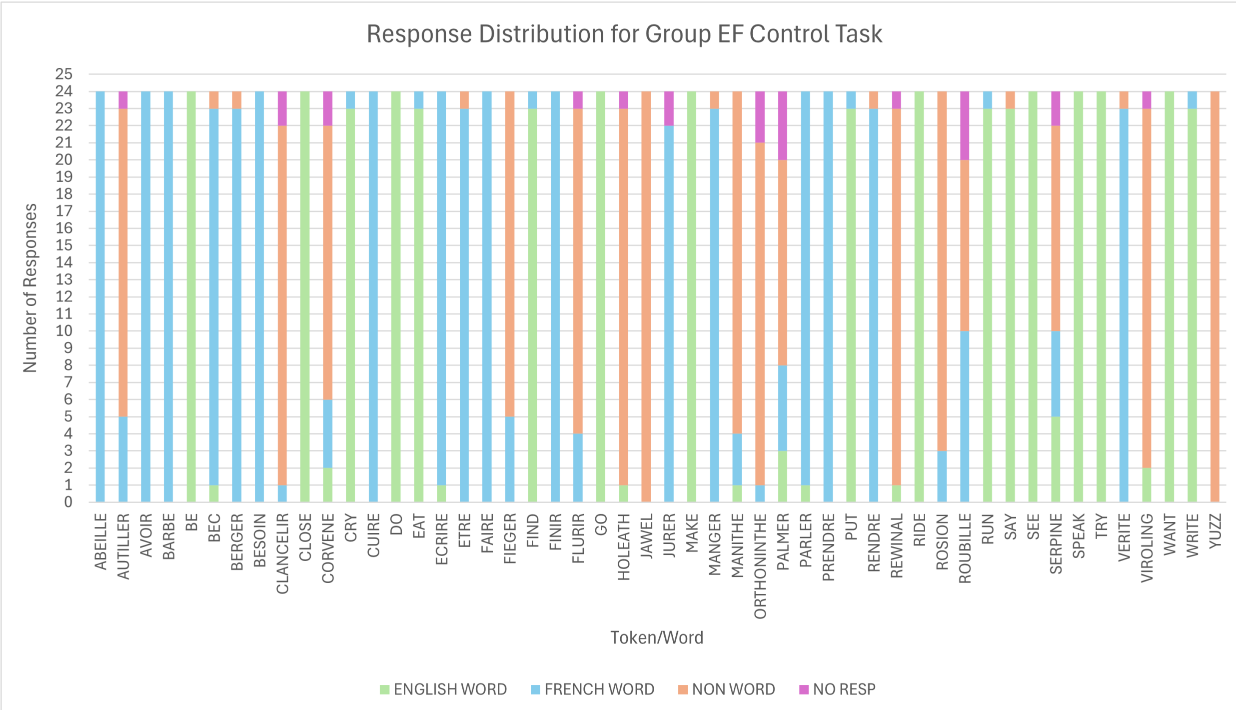


(a) Group EF — NC Condition

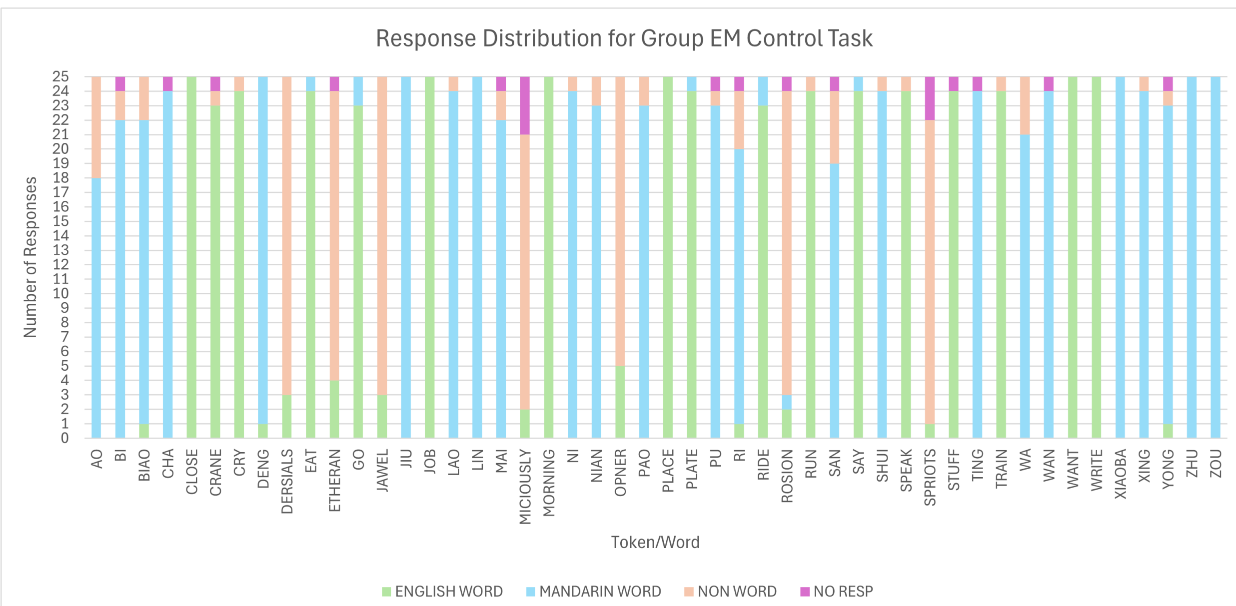


(b) Group EM — NC Condition

Figure 4.4: Distribution of responses per trial for each language group in the NC Condition.

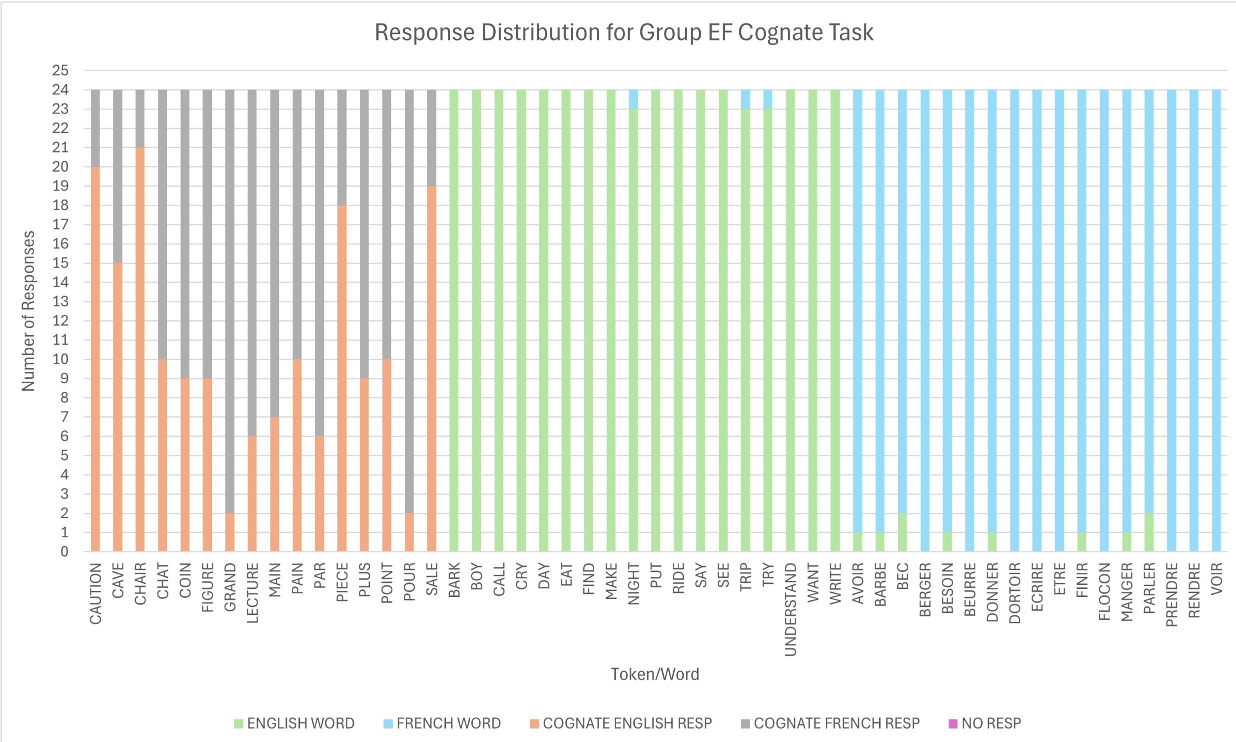


(a) Group EF — Control Condition

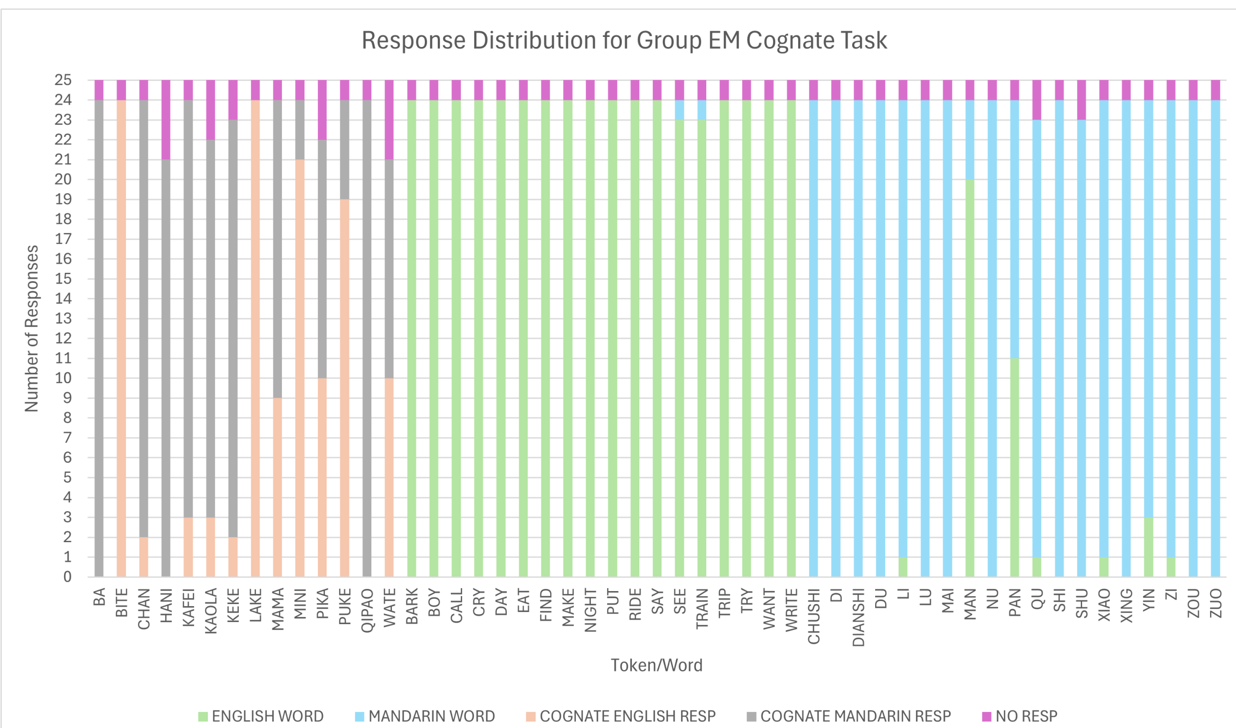


(b) Group EM — Control Condition

Figure 4.5: Distribution of responses per trial for each language group in the Control Condition.



(a) Group EF — Cognate Condition



(b) Group EM — Cognate Condition

Figure 4.6: Distribution of responses per trial for each language group in the Cognate Condition.

Having established a significant overall effect of condition on RT, the following sections examine each condition in turn, beginning with the No Control condition as a processing baseline. In this chapter, results for RT and response accuracy are presented according to the corresponding task condition, and pupillometric results are presented at the end.

4.2.1 Condition 1: No Control

4.2.1.1 Reaction Time (RT)

To examine the effect of language group and trial number on RT in the NC condition, a linear mixed-effects model (LMM) was fit using the `lmer` function from the `lme4` package (Bates et al., 2015 [153]) in R, using the formula:

$$\text{RT} \sim \text{LANGUAGE\_GROUP} \times \text{TRIAL} + (1\text{—PARTICIPANT}) \qquad \text{(NC Model)}$$

Random slopes for LANGUAGE_GROUP and TRIAL were initially included for participants; however, the maximal model failed to converge. Following standard practice, the random-effects structure was simplified to random intercepts only for participants.

In this model, RT served as the dependent variable. The fixed effects included LANGUAGE_GROUP and TRIAL, where TRIAL refers to the sequential order in which stimuli were presented within each condition (indexed 1-50), capturing potential order-of-presentation effects on response time — hereafter referred to as trial order. Participant was included as a random intercept to account for between-participant variability. Inspection of scaled residuals (Min = -2.69, Median = -0.21, Max = 6.77) indicated mild positive skew, consistent with raw RT distributions; results should be interpreted with appropriate caution. A Type III Wald chi-square test (see Table 4.6) was used to evaluate the significance of each fixed effect in the NC model.

Table 4.6: Analysis of Deviance Table (Type III Wald chi-square tests) for RT in the NC condition. Significance codes: ** $p < .01$; *** $p < .001$.

	Chisq	**Df**	**Pr(¿Chisq)**
(Intercept)	455.32	1	$< .001$ ***
language_group	0.48	1	0.49
trial	9.69	1	0.002 **
language_group:trial	2.43	1	0.12

As shown in Table 4.6, trial number significantly predicted RT ($\chi^2 = 9.69$, $p = .002$), indicating that response times varied across the course of the trial sequence. In contrast, neither LANGUAGE_GROUP ($\chi^2 = 0.48$, $p = .490$) nor the LANGUAGE_GROUP:TRIAL interaction ($\chi^2 = 2.43$, $p = .119$) reached significance, suggesting that while RT shifted across trials, the two language groups did not differ in how their response times evolved over the course of the NC task. This is consistent with the interpretation of the NC condition as a processing baseline with minimal between-group differences.

4.2.1.2 Response Accuracy

To examine response accuracy in the NC condition, a generalized linear mixed-effects model (GLMM) was fit using the `glmer` function from the `lme4` package (Bates et al., 2015 [153]) in R, with a binomial link function to account for the binary nature of the accuracy outcome (correct = 1, incorrect = 0):

$$\text{ACCURACY} \sim \text{LANGUAGE\_GROUP} \times \text{TRIAL} + (1\text{—PARTICIPANT})$$

As with the RT model, random slopes were initially specified but removed due to non-convergence of the maximal model. The random-effects structure was therefore simplified to a random intercept for participants only.

Model fit statistics are summarized in Table 4.7. The random intercept variance for participants (0.51; SD $= 0.71$) indicates some between-participant variability, reflecting individual differences in response tendencies, although overall accuracy remained near ceiling.

Table 4.7: Summary of model fit statistics and random intercept variance estimates for the generalized linear mixed-effects model (NC condition).

Metric	**Estimate**	**Notes**
Akaike Information Criterion (AIC)	272.4	Model fit index
Bayesian Information Criterion (BIC)	301.3	Model fit index
Log-likelihood	-131.2	Model likelihood
Participant random intercept variance	0.5008	Between-participant variability
Participant random intercept SD	0.7077	Standard deviation

Table 4.8: Analysis of Deviance Table (Type III Wald chi-square tests) for accuracy in the NC condition. Significance codes: * $p < .05$; ** $p < .01$; *** $p < .001$.

	Chi-sq	**Df**	$Pr(> Chisq)$
(Intercept)	40.73	1	< .001 ***
language_group	2.70	1	0.10
trial	9.01	1	0.003 **
language_group:trial	4.27	1	0.039 *

As shown in Table 4.8, trial significantly predicted response accuracy ($\chi^2 = 9.01$, $p = .003$). Additionally, there was a significant interaction between TRIAL and LANGUAGE_GROUP ($\chi^2 = 4.27$, $p = .039$): while accuracy generally improved across the trial sequence, this benefit was attenuated for Group EM. Although the main effect of LANGUAGE_GROUP alone was not significant ($\chi^2 = 2.70$, $p = .100$), the interaction suggests that the two groups differed in how accuracy evolved across the task, potentially reflecting differences in cognitive strategy or sensitivity to trial progression under minimal interference.

A secondary analysis was conducted to complement the net accuracy scores reported in Table 4.5. Whereas the net analysis calculated accuracy as the proportion of correct responses out of total responses at the participant level, this secondary analysis treated each individual trial as a binary observation (correct = 1, incorrect = 0) and computed accuracy across all trials regardless of trial type. This approach weights each trial equally and is less sensitive to imbalances in response distribution across trial categories. As shown in Table 4.9, this method produced slightly higher estimates, with Group EF reaching 99.0% and Group EM 98.9%, reinforcing the interpretation that the NC condition represents a baseline of optimal performance with minimal between-group differences. These findings provide a reference point for evaluating the effects of interference (CONT) and cognate complexity (COG).

Table 4.9: A summary of accuracy scores for Group EF and Group EM across all conditions, treating each individual trial as a binary observation (correct = 1, incorrect = 0) regardless of trial type. This differs from Table 4.5, which reports net accuracy at the participant level.

Condition	**Group EF Accuracy**	**Group EM Accuracy**
No Control	99.0%	98.9%
Control	94.3%	79.6%
Cognate	66.7%	68.7%

4.2.2 Condition 2: Control

4.2.2.1 Reaction Time (RT)

As shown in Table 4.4, the Control condition produced the longest overall reaction times of all conditions, with Group EF averaging $1,447ms$ and Group EM averaging $1,377ms$. To examine the sources of this variability, a linear mixed-effects model (LMM) was fit using the `lmer` function from the `lme4` package (Bates et al., 2015 [153]) in R:

$$\text{RT} \sim \text{LANGUAGE\_GROUP} \times \text{TRIAL} + (1\text{—}\text{PARTICIPANT}) \qquad \text{(CONT Model)}$$

Random slopes for LANGUAGE_GROUP and TRIAL were initially included but excluded following non-convergence of the maximal model, consistent with standard practice in mixed-effects modelling.In this model, RT served as the dependent variable, with LANGUAGE_GROUP and TRIAL as fixed effects, and participant as a random intercept. A Type III Wald chi-square test (Table 4.10) was used to evaluate the significance of each fixed effect. To support interpretation of the significant effects, fixed-effect estimates are reported in Table 4.11.

Table 4.10: Analysis of Deviance Table (Type III Wald chi-square tests) for RT in the Control condition. Significance codes: ** $p < .01$; *** $p < .001$.

	Chisq	**Df**	**Pr(>Chisq)**
(Intercept)	68.75	1	$< .001$ ***
language_group	8.99	1	$< .001$ **
trial	2.06	1	0.15
language_group:trial	11.67	1	$< .001$ ***

Table 4.11: Fixed-effect estimates for RT in the Control condition. Reference group: Group EF.

	Estimate	**SE**	**t value**	**Pr(> | t |)**
(Intercept: Group EF)	1242.20	149.81	8.29	$< .001$
Group EM	605.14	201.83	2.998	0.003
Trial Order	2.66	1.86	1.44	0.152
Group EM $\times$ Trial Order	-9.16	2.68	-3.42	$< .001$

These results indicate two key findings. First, there was a significant difference in RT between the two language groups, with Group EF exhibiting longer reaction times ($\beta = 605.14$ ms relative to Group EM). Second, the significant interaction between LANGUAGE_GROUP and TRIAL ($\chi^2 = 11.67$, $p < .001$) suggests that RT evolved differently across the trial sequence for the two groups. Specifically, the negative interaction estimate ($\beta = -9.16$) indicates that reaction times changed across the trial sequence in a different manner for Group EM compared to Group EF. Importantly, trial alone did not significantly predict RT overall ($p = 0.152$), indicating the absence of a general fatigue or ordering effect across participants. However, the significant interaction demonstrates that the evolution of RT across the experiment was group-specific rather than uniform across both language groups.

4.2.2.2 Response Accuracy

A GLMM was fit for response accuracy in the Control condition using the same specification as the NC condition. Table 4.12 presents the Type III Wald chi-square results. As with the preceding models, random slopes were initially specified but excluded following non-convergence of the maximal model, with the random-effects structure simplified to a random intercept for participants only.

Table 4.12: Analysis of Deviance Table (Type III Wald chi-square tests) for accuracy in the Control condition. Significance codes: * $p < .05$; ***$p < .001$.

	Chi-sq	**Df**	$Pr(> Chisq)$
(Intercept)	28.77	1	$< .001$ ***
language_group	5.58	1	0.018 *
trial	1.74	1	0.19
language_group:trial	0.27	1	0.61

A significant main effect of LANGUAGE_GROUP was observed ($\chi^2 = 5.58$, $p = .018$), indicating that accuracy differed between the two language groups. The fixed-effect estimate for Group EM was $\beta = -1.88$ ($SE = 0.79$, $z = -2.36$, $p = .018$), confirming that Group EM exhibited significantly lower accuracy than Group EF in the Control condition. Neither TRIAL ($\chi^2 = 1.74, p = .188$) nor the LANGUAGE_GROUP:TRIAL interaction ($\chi^2 = 0.27, p = .61$) reached significance, indicating that trial progression did not differentially affect accuracy across groups in this condition. The random-effects structure revealed a participant intercept variance of 0.02 (SD = 0.14), indicating relatively low between-participant variability in accuracy.

These results demonstrate a clear effect of language group on response accuracy in the Control condition. As shown in Tables 4.5 and 4.9, Group EF achieved substantially higher accuracy than Group EM under both scoring methods (net: 91.32% vs 78.33%; binary: 94.3% vs 79.6%). This divergence suggests greater difficulty for Group EM participants in suppressing interference from non-word stimuli, pointing to language-group-specific differences in inhibitory control or lexical competition under increased task demands. This pattern is further consistent with the summary accuracy data reported in Table D.2 in the *Appendix D*.

4.2.3 Condition 3: Cognate

4.2.3.1 Reaction Time (RT)

As shown in Table 4.4, the Cognate condition produced intermediate reaction times for Group EF (1,107 ms) and the slowest reaction times of any condition for Group EM (1,224 ms). To examine these patterns, a linear mixed-effects model (LMM) was fit using the `lmer` function from the `lme4` package (Bates et al., 2015 [153]) in R:

$$\text{RT} \sim \text{LANGUAGE\_GROUP} \times \text{TRIAL} + (1\text{—PARTICIPANT}) \qquad \text{(COG Model)}$$

Random slopes were initially included for participants but excluded due to non-convergence, following the same simplification procedure applied in the NC and Control condition models. In this model, RT served as the dependent variable, with LANGUAGE_GROUP and TRIAL as fixed effects, and participant as a random intercept. A Type III Wald chi-square test was used to evaluate the significance of each fixed effect (see Table 4.13).

Table 4.13: Analysis of Deviance Table (Type III Wald chi-square tests) for RT in the Cognate condition. Significance codes: * $p < .05$; *** $p < .001$.

	Chisq	Df	Pr(¿Chisq)
(Intercept)	39.48	1	$< .001$ ***
language_group	6.52	1	0.011 *
trial	0.29	1	0.59
language_group:trial	4.19	1	0.041 *

Both LANGUAGE_GROUP ($\chi^2 = 6.52$, $p = .011$) and the LANGUAGE_GROUP:TRIAL interaction ($\chi^2 = 4.19$, $p = .041$) were significant, indicating that participants in different language groups exhibited significantly different RTs, and that the effect of trial progression on RT depended on language group membership. Trial alone did not significantly predict RT ($\chi^2 = 0.29$, $p = .588$), suggesting no general ordering effect across participants. Notably, Group EM showed the longest average RTs in the Cognate condition (see Table 4.4), suggesting some form of interference arising from cross-linguistic activation between English and Mandarin.

In summary, these RT findings suggest that lexical access is influenced by language group, and that trial-level features play a more pronounced role for Group EM speakers, possibly reflecting increased sensitivity to lexical frequency or processing complexity.

4.2.3.2 Response Accuracy

In the Cognate condition, response accuracy dropped relative to both the NC and Control conditions (see Table 4.9). Performance decreased in both language groups, with accuracy dropping to 66.7% for Group EF and 68.7% for Group EM when scored against binary correct/incorrect responses. A GLMM was fit for response accuracy using the same specification as the previous conditions. Table 4.14 presents the Type III Wald chi-square results.

Table 4.14: Analysis of Deviance Table (Type III Wald chi-square tests) for accuracy in the Cognate condition. Significance codes: ** $p < .01$; *** $p < .001$.

	Chi-sq	**Df**	$Pr(> Chisq)$
(Intercept)	10.11	1	0.001 **
language_group	40.72	1	$< .001$ ***
trial	3.53	1	0.060
language_group:trial	43.70	1	$< .001$ ***

As shown in Table 4.14, there was a strong main effect of LANGUAGE_GROUP ($\chi^2 = 40.72$, $p < .001$) and a significant LANGUAGE_GROUP:TRIAL interaction ($\chi^2 = 43.70$, $p < .001$). The fixed-effect estimate for Group EM was $\beta = -4.72$ ($SE = 0.74$, $z = -6.38$, $p < .001$), confirming that Group EM performed significantly worse than Group EF. The significant interaction further indicates that accuracy patterns shifted differently across the trial sequence for Group EM than for Group EF. Trial alone approached but did not reach significance ($\chi^2 = 3.53$, $p = .060$).

Importantly, these findings stand in contrast to the initial net-accuracy analysis (Table 4.5), in which the Cognate condition appeared to support a cognate facilitation effect, with both groups showing high accuracy scores (EF: 98.43%; EM: 91.57%). However, this interpretation changes when individual trials are explicitly incorporated. When accuracy was recalculated as binary correct/incorrect responses (Table 4.9), scores dropped sharply for both groups (EF: 66.7%; EM: 68.7%), indicating that cognates introduce additional complexity that negatively affects response accuracy once full trial information is taken into account. This pattern suggests that cognate processing may involve heightened cross-linguistic activation or competition, effects not fully captured by net accuracy scores alone. For the purposes of the present analysis, the data presented in Table 4.9 are therefore treated as the primary indicator of response accuracy, though Table 4.5 will be revisited later for comparative and interpretive purposes.

4.2.4 Pupil Dilation

Table 4.15 summarises average pupil diameter (in millimetres) recorded at three time points per trial—baseline (prior to stimulus presentation), at stimulus onset, and at participant response—for each language group and task condition. This preliminary overview was generated from raw dilation values prior to inferential modelling, and serves to establish the overall temporal trajectory of pupillary responses across conditions.

As shown in the baseline column, pupil diameter was relatively stable prior to stimulus onset across all conditions and language groups, ranging from approximately 4.41 to $5.02mm$. At stimulus presentation, dilation increased consistently across all conditions and both groups; an expected pattern reflecting cognitive engagement during lexical processing (Mathôt, 2018 [147]; Mathôt & Vilotijević, 2022 [154]). A partial recovery in dilation was observed at the response time point, though values generally remained elevated above baseline.

Of particular note is the pattern observed in the Cognate (COG) condition. Relative to NC and CONT, the reduction in pupil size from stimulus onset to response was comparatively smaller in the COG condition across both language groups. For example, in English-French COG trials, dilation increased from $4.90mm$ (baseline) to $5.31mm$ (stimulus) but decreased only to $5.23mm$ at response—a recovery of just $0.08mm$. In NC trials for the same group, the comparable recovery was approximately 0.16–$0.20mm$. This attenuated recovery in the COG condition suggests that cognate processing sustained higher cognitive engagement across the decision window, rather than facilitating rapid lexical resolution. This observation directly motivated the dependent variable and modelling approach described in the following sections.

Table 4.15: Average dilation measurements (in mm, rounded to 2 decimal places) at baseline, stimulus presentation, and response, for each language group and task condition.

		Baseline	**@Stimulus**	**@Response**
NC	English	4.87mm	5.38mm	5.22mm
	French	5.00mm	5.47mm	5.27mm
	English	4.67mm	5.105mm	4.91mm
	Mandarin	4.58mm	5.135mm	4.91mm
CONT	English	4.99mm	5.48mm	5.36mm
	French	4.97mm	5.48mm	5.32mm
	Non-Word	5.02mm	5.50mm	5.42mm
	English	4.48mm	4.99mm	4.77mm
	Mandarin	4.47mm	4.92mm	4.712mm
	Non-Word	4.56mm	5.00mm	4.72mm
COG	English	5.00mm	5.25mm	5.06mm
	French	4.92mm	5.21mm	5.01mm
	Cognate	4.90mm	5.31mm	5.23mm
	English	4.41mm	4.82mm	4.60mm
	Mandarin	4.53mm	4.88mm	4.63mm
	Cognate	4.46mm	4.87mm	4.73mm

4.2.4.1 Modelling Approach

The modelling approach for pupil dilation was designed to be both theoretically motivated and directly interpretable with respect to the study's research questions. The key preliminary observation (i.e., that the COG condition showed less post-stimulus pupillary recovery than NC and CONT) motivated the choice of dependent variable: the pupil dilation difference score, defined as the difference between dilation at response and dilation at stimulus onset (RESPONSE − STIMULUS) for each trial. This single continuous measure captures the change in cognitive engagement across the decision window and is directly comparable across conditions and language groups.

Pupil dilation was recorded separately for the left and right eye. These were averaged across eyes prior to modelling. There is no theoretical basis to expect differential processing between the left and right pupil during a visual language task (Mathôt & Vilotijević, 2022 [154]), and averaging reduces trial-level noise while eliminating the need to include eye as an additional model factor—improving both model parsimony and interpretability.

A single unified model was fitted across all three conditions, with Condition included as a fixed factor. This differs from the RT analyses, in which separate models were fitted per condition to allow condition-specific interpretation of group and sequence effects without cross-condition conflation. For pupil dilation, however, the primary research question concerns how conditions compare to one another in their cognitive load profiles. A unified model allows this comparison to be made directly and formally, through the Condition main effect and the Condition $\times$ Language Group interaction, without requiring informal cross-model comparisons.

Condition was dummy-coded with the Cognate (COG) Condition set as the reference level. This choice was theoretically motivated: with COG as the reference, the main effect of LANGUAGE_GROUP directly reflects whether Group EF and Group EM differ in the condition of primary theoretical interest, and the condition coefficients for NC and CONT express how much cognitive load changed relative to cognate processing. Trial order—the sequential position of each stimulus within its condition block—was included as a fixed-effect covariate to account for potential fatigue or adaptation effects. Participant was included as a random intercept to account for between-participant variability. The model was estimated using restricted maximum likelihood (REML) and specified as:

AVG_PUPIL_DIFF $\sim$ LANGUAGE_GROUP $\times$ CONDITION + TRIAL_ORDER + (1—PARTICIPANT)

Where LANGUAGE_GROUP (EF vs. EM; reference: EF), CONDITION (NC, CONT, COG; reference: COG), and their interaction were entered as fixed effects, with TRIAL_ORDER as a continuous covariate and participant as a random intercept.

Random slopes for CONDITION and TRIAL_ORDER were initially specified for participants but were excluded following non-convergence of the maximal model. The random-effects structure was therefore reduced to a random intercept for participants only, consistent with the simplification approach applied across all models in this chapter.

4.2.4.2 Difference Scores, Results & Interpretation

Table 4.16: Descriptive statistics for average pupil dilation difference scores (response − stimulus, mm), averaged across left and right eye, by language group and condition.

Language Group	Condition	N	Mean (mm)	SD (mm)
EF	COG	1,172	3.90	4.28
	NC	1,139	5.79	5.74
	CONT	1,138	5.84	6.06
EM	COG	1,275	5.46	4.52
	NC	1,192	6.20	5.84
	CONT	1,195	6.29	5.48

As shown in Table 4.16 and illustrated in Figure 4.7, pupil dilation difference scores were consistently lowest in the Cognate condition for Group EF ($M = 3.90mm$, $SD = 4.28$), compared to both NC ($M = 5.79mm$) and CONT ($M = 5.84$ mm). This mirrors the preliminary observation in Table 4.15: for Group EF, the smallest post-stimulus pupillary change occurred during cognate processing, indicating more sustained cognitive engagement. Group EM showed a comparatively flatter profile across conditions, with COG ($M = 5.46mm$) approaching NC ($M = 6.20mm$) and CONT ($M = 6.29mm$) values more closely than observed for Group EF, suggesting that all three conditions imposed relatively similar processing demands on Group EM participants.

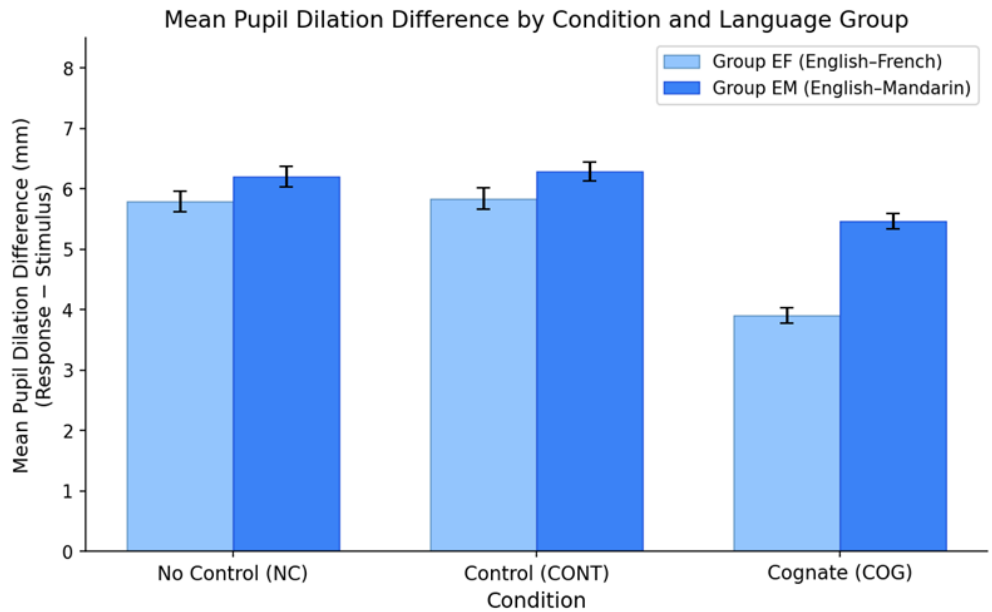


Figure 4.7: Mean pupil dilation difference scores (response − stimulus, mm) by condition and language group, with standard error bars. Lower scores in the COG condition for Group EF indicate more sustained dilation (less recovery) relative to NC and CONT.

The mixed-effects model converged successfully ($N = 7{,}111$ observations; 49 participants; *Log-likelihood* $= -20{,}058.03$). Participant-level random intercept variance was 13.09 ($SD = 3.62$), indicating moderate between-participant variability. Residual variance was 15.95 ($SD = 3.99$). Fixed-effect estimates were extracted directly from the model summary using the `lme4` package (Bates et al., 2015 [153]), reporting β coefficients, standard errors, and *z-values* rather than chi-square statistics, as the inferential focus here is on the magnitude and direction of individual fixed effects rather than omnibus tests of model terms. Estimates are reported in Table 4.17 and visualised in Figure 4.9.

Table 4.17: Fixed-effect estimates from the mixed-effects model (AVG_PUPIL_DIFF $\sim$ LANGUAGE_GROUP $\times$ CONDITION + TRIAL_ORDER + (1|PARTICIPANT)). Reference levels: Group EF; Cognate (COG) condition. Random effects: participant intercept variance $= 13.09$ ($SD = 3.62$); residual variance $= 15.95$ ($SD = 3.99$). $N = 7{,}111$ observations; 49 participants. *** $p < .001$; ns = not significant.

Term	β **(mm)**	**SE**	z	p	
Intercept (EF, COG)	4.05	0.75	5.38	$< .001$	***
Language Group: EM	1.48	1.05	1.42	.157	ns
Condition: NC	2.01	0.17	11.96	$< .001$	***
Condition: CONT	1.76	0.17	10.47	$< .001$	***
EM × NC	−1.27	0.23	−5.46	$< .001$	***
EM × CONT	−0.94	0.23	−4.02	$< .001$	***
Trial Order	−0.003	0.003	−0.80	.422	ns

The model intercept ($\beta = 4.05$, $p < .001$) represents the estimated mean pupil dilation difference score for Group EF in the Cognate condition, and serves as the reference against which all other effects are interpreted.

Condition effects within Group EF. Both NC ($\beta = +2.01$, $p < .001$) and CONT ($\beta = +1.76$, $p < .001$) produced significantly larger pupil dilation difference scores than COG for Group EF. This means that in the NC and CONT conditions, dilation decreased more from stimulus onset to response than it did in the COG condition—indicating greater cognitive recovery in those conditions. In the COG condition, dilation remained closer to its stimulus-onset peak throughout the decision window, consistent with the preliminary observation in Table 4.15 that cognate processing yields a smaller post-stimulus pupillary decline. Taken together, these results indicate that cognate processing sustained the highest level of cognitive load for Group EF across the response window, a finding that stands in contrast to classical cognate facilitation accounts.

Language group main effect. The main effect of language group was not statistically significant ($\beta = +1.48$, $p = .157$), indicating that Group EF and Group EM did not differ overall in their pupil dilation difference scores within the Cognate condition. This null main effect should be read alongside the significant interaction terms described below.

Language Group × Condition interactions. Both interactions reached high significance: EM × NC ($\beta = -1.27$, $p < .001$) and EM × CONT ($\beta = -0.94$, $p < .001$). These negative estimates indicate that the gap between COG and the other two conditions was substantially smaller for Group EM than for Group EF. While Group EF showed a pronounced increase in difference scores moving from COG to NC and CONT, Group EM's profile was comparatively flat. This divergence is captured clearly in Figure 4.7, which plots smoothed dilation difference scores across the trial sequence for each condition and language group separately. In the left panel (Group EF), the three condition lines are clearly separated throughout the trial sequence, with the COG line sitting consistently below both NC and CONT. In the right panel (Group EM), this separation largely disappears: the three lines converge toward similar values across the sequence, with no condition producing a markedly distinct pupillary profile. This is the central finding of the pupillometry analysis: the two language groups do not simply differ in their overall dilation levels, but in how their cognitive load varies across conditions. For Group EM, all three conditions appear to impose broadly comparable processing demands; for Group EF, the Cognate condition is distinctively more taxing.

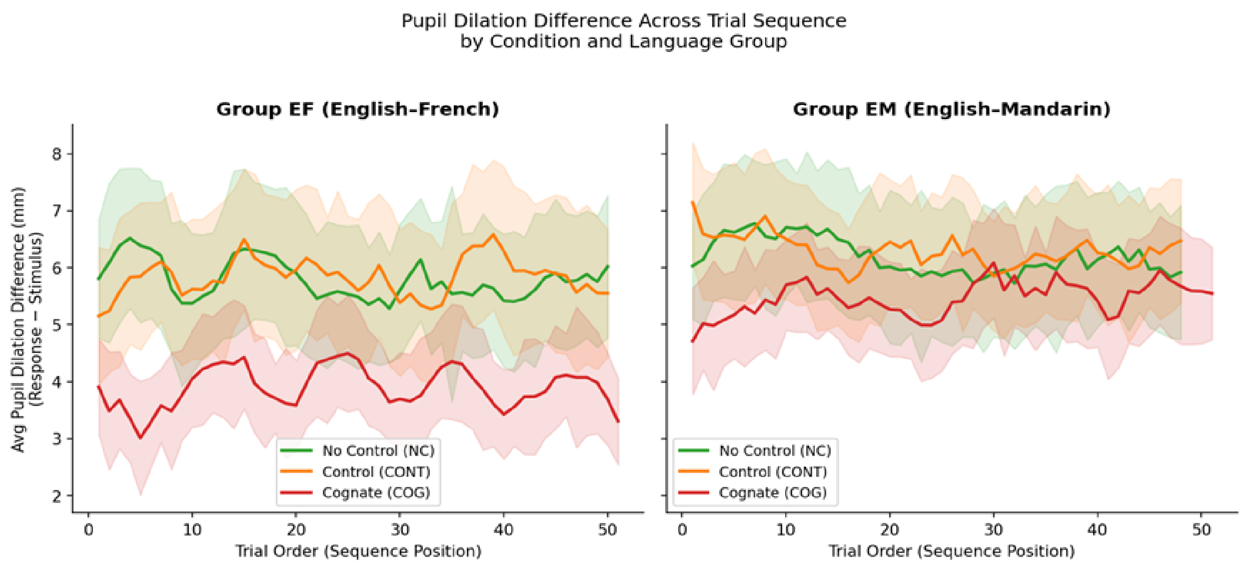


Figure 4.8: Smoothed pupil dilation difference scores across trial sequence position, by condition and language group. Shaded regions represent ±1 SE. Group EF (left panel) shows clear condition separation, with COG consistently below NC and CONT. Group EM (right panel) shows a flatter, more convergent pattern across conditions, reflecting the significant Language Group × Condition interaction.

Trial order. Trial order did not significantly predict pupil dilation difference scores ($\beta = -0.003$, $p = .422$), confirming the absence of a systematic fatigue or adaptation effect across the trial sequence. This supports the interpretation that the observed condition differences reflect genuine processing effects rather than confounds introduced by stimulus ordering.

Figure 4.9 provides a visual summary of all fixed-effect estimates and their 95% confidence intervals. The plot makes three features of the results immediately apparent. First, the two condition effects within Group EF (NC and CONT, shown in light blue) are large and precisely estimated, with confidence intervals that sit well clear of zero. Second, both interaction terms (EM × NC and EM × CONT, shown in dark blue) are negative and similarly well-estimated, quantifying how much of Group EF's condition advantage is absent in Group EM. Third, the language group main effect and trial order (shown in grey) straddle zero comfortably, confirming that neither reached significance. Taken together, the coefficient plot reinforces the interpretation that the Language Group × Condition interaction is the primary structural finding of the pupillometry analysis, and that this interaction is driven by Group EF's distinctive response to cognate stimuli rather than by any overall between-group difference in pupil dilation.

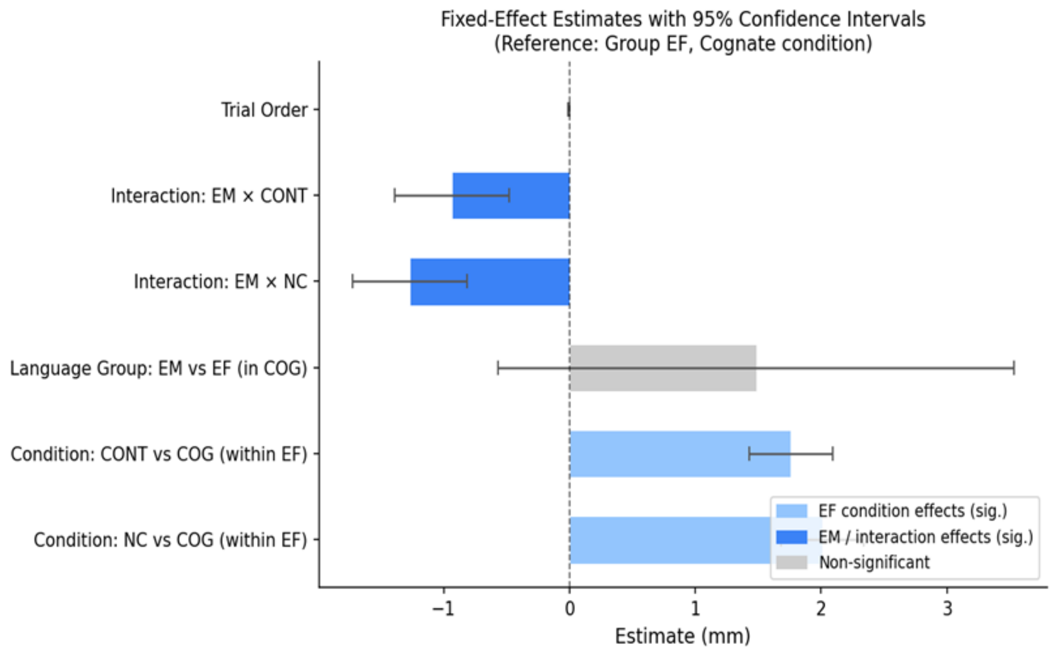


Figure 4.9: Fixed-effect estimates with 95% confidence intervals. Light blue bars represent condition effects within Group EF; dark blue bars represent EM-specific and interaction effects; grey bars are non-significant. Reference: Group EF, Cognate condition.

4.2.4.3 Summary of Pupillometry Findings

The pupillometry results provide converging psychophysiological evidence for the behavioural patterns observed in reaction time and accuracy. The preliminary descriptive analysis (see Table 4.15) established that the Cognate condition produced the smallest post-stimulus pupillary recovery across both groups; an attenuated decline from peak dilation that suggests sustained cognitive engagement rather than facilitated lexical access. The inferential model formalised and extended this observation.

The key finding is a significant Language Group × Condition interaction: Group EF showed markedly smaller difference scores in the COG condition relative to NC and CONT, while Group EM's profile was comparatively flat across all three conditions. This divergence indicates that the cognitive cost of cognate processing is not uniform across bilingual populations. For Group EF bilinguals, for whom English and French share substantial orthographic and phonological overlap, cognate processing appears to sustain heightened cross-linguistic activation and inhibitory control demands throughout the decision window. For Group EM bilinguals, the typological distance between English and Mandarin—even when stimuli are presented in phonologically transparent borrowed forms—may mean that all three task conditions impose a comparably elevated processing load, leaving less room for condition-specific modulation.

These findings align with the RT and accuracy results reported in the preceding sections. Group EM showed the slowest overall RTs in the COG condition and the largest accuracy drop relative to NC, consistent with the pupillometric evidence of uniformly high processing demands. Group EF, by contrast, showed near-NC RT performance in COG but a distinctive pupillary signature of sustained cognitive load, suggesting that their behavioural response speed belies the underlying effort involved. Together, the behavioural and pupillometric measures paint a complementary picture of bilingual lexical processing under conditions of cross-linguistic competition.

The following section considers what these patterns mean for theoretical accounts of bilingual lexical access and cognitive control.

4.3 Experimental Outcomes: The Cognate Effect

The preceding analyses established task condition-level differences in RT, accuracy, and pupil dilation across all three task conditions. This section narrows the focus to the Cognate condition specifically, examining whether cognate status—that is, whether a given stimulus was a cognate or a non-cognate real word—produced differential effects on lexical decision performance and cognitive load. This distinction is central to the study's primary research question: whether the Cognate Effect operates as a facilitatory or inhibitory mechanism in bilingual lexical access, and whether this differs across language pairs of varying typological distance.

4.3.1 Investigating the Cognate Effect: Reaction Time

Before examining group-specific effects, it is useful to situate the RT findings within the broader pattern established in *Section 4.2*. As reported in Table 4.4, the Cognate condition produced intermediate reaction times for Group EF ($M = 1{,}107ms$) and the slowest reaction times of any condition for Group EM ($M = 1{,}224ms$). These descriptive values already hint at a group difference in how cognate stimuli are processed: whereas Group EF responded to cognates at a speed comparable to the NC baseline, Group EM showed a pronounced slowing relative to their own NC performance ($M = 1{,}036ms$). This divergence suggests that cognate status does not affect both bilingual populations in the same way, and motivates the group-specific analyses that follow.

To examine this more rigorously, reaction times in the Cognate condition were analysed by contrasting responses to cognate stimuli against responses to non-cognate real words within each language group. This comparison isolates the effect of cross-linguistic form overlap from the broader condition-level differences already reported. Separate linear mixed-effects models were fitted for Group EF and Group EM, each including cognate status (cognate vs. non-cognate) as the primary fixed effect, with random intercepts for participants and lexical items. This approach allows the Cognate Effect to be evaluated independently within each bilingual population, while accounting for variability attributable to individual participants and specific word stimuli.

Across both groups, the results point consistently in the same direction: cognate stimuli produced longer reaction times than non-cognate real words, indicating interference rather than facilitation. However, the magnitude of this effect differed substantially between groups, with Group EM exhibiting a considerably larger cognate cost than Group EF. The following subsections report these findings in detail.

4.3.1.1 Cognate Processing in English-French Bilinguals

To examine whether cognate status influenced lexical decision performance, reaction times were analyzed using linear mixed-effects models. Separate models were fit for each language group in order to evaluate whether the Cognate Effect differed across bilingual populations. Each model included Cognate Status (cognate vs. non-cognate) as a fixed effect, with random intercepts for participants and lexical items to account for variability across individuals and stimuli:

$$\text{RT} \sim \text{COGNATE\_STATUS} + (1 — \text{WORD}) + (1 — \text{PARTICIPANT})$$

Random slopes for Cognate Status were initially included for participants and items. However, the maximal model failed to converge. Following standard practice, the random-effects structure was simplified to include random intercepts only.

For Group EF, the model revealed a clear and robust cognate interference effect. Cognate status significantly predicted reaction times in the language discrimination task, with cognates producing response latencies that were $604.1 msec$ longer on average than non-cognates ($SE = 45.3$,

$t = 13.34$, $p < .001$). The effect size ($d = 0.89$) indicates a large and practically meaningful difference in performance between the two lexical categories. This pattern is illustrated in Figure 4.10, which shows the mean RT for cognate and non-cognate trials for Group EF.

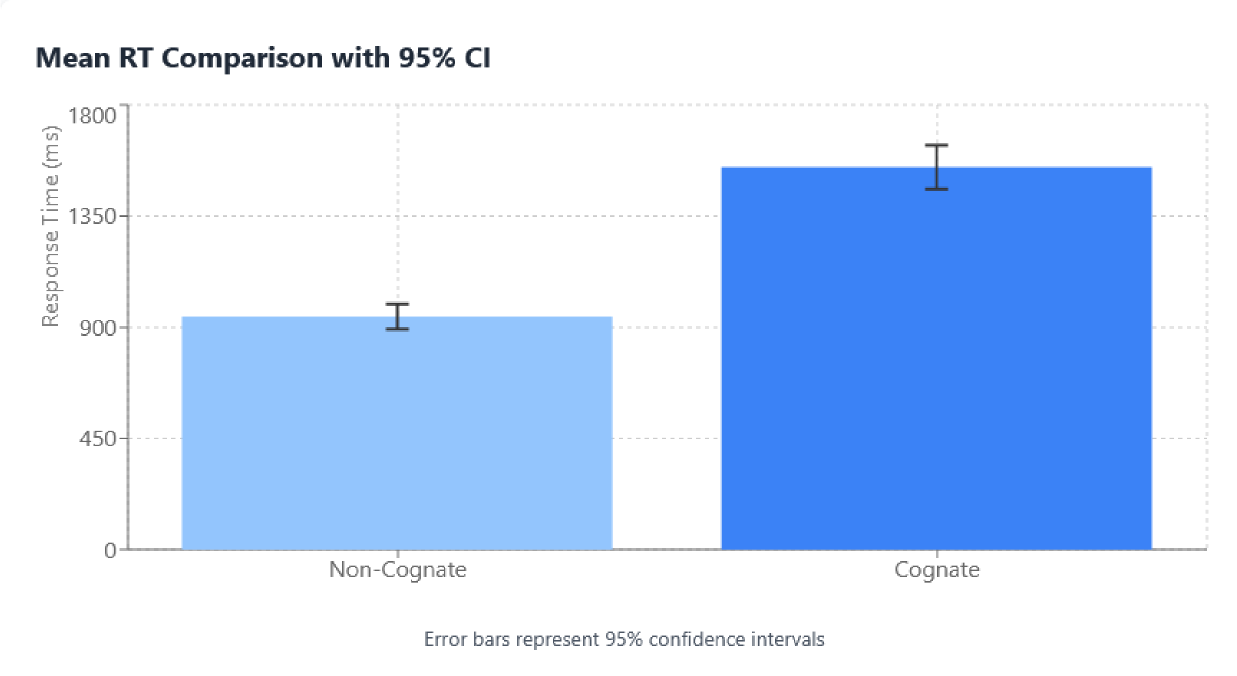


Figure 4.10: Mean RT comparison of non-cognate versus cognate trials for Group EF, indicating that cognates require longer RT—approximately double the mean values for non-cognate trials.

This result challenges the canonical cognate facilitation hypothesis, which would predict faster responses for cognates due to shared form–meaning mappings across languages (Dijkstra & Van Heuven, 2002 [1]). Instead, the data support a cognate interference interpretation, suggesting that during lexical recognition, bilinguals engage in simultaneous activation of both the English and French lexical representations, which subsequently must be resolved through inhibitory control mechanisms (Green, 1998 [58]; Green & Abutalebi, 2013 [73]).

The mixed-effects model provided strong statistical reliability and explanatory power, with a marginal R^2 of 0.34 (fixed effects only) and a conditional R^2 of 0.61 (including random effects).

This suggests that while the cognate manipulation accounted for a substantial portion of the variance in reaction times, individual differences among participants and word-specific variability also contributed meaningfully to performance. Random effect estimates revealed that 23% of the total variance was attributable to participant-level differences and 18% to item-level differences, consistent with prior bilingual research showing considerable variability in cross-linguistic activation strength and control efficiency (Poarch & Van Hell, 2012 [65]; Brenders et al., 2011 [66]).

The cognitive interpretation of these findings centers on cross-linguistic lexical competition. When a bilingual encounters a cognate, both lexical systems are activated due to the shared orthographic and semantic overlap. For English–French bilinguals, this overlap is particularly salient because both languages share a largely alphabetic writing system and substantial lexical borrowing. The participant's task, however, required language-selective processing (responding in one language context). This likely necessitated inhibition of the non-target lexical entry, creating additional cognitive demand and extending reaction time.

Although variability across participants was captured through the random-effects structure of the model, the present dataset does not allow us to determine the sources of this variation. Factors such as language dominance, proficiency balance, or individual bilingual experience may play a role, but these variables were not directly measured in the current study.

Interpretively, results from Group EF indicate that cognate recognition under comprehension demands elicits inhibitory rather than facilitative processing. This finding reinforces the theoretical distinction between language production (where co-activation yields facilitation) and language recognition (where co-activation leads to competition and delay).

4.3.1.2 Cognate Processing in English-Mandarin Bilinguals

The English–Mandarin bilingual group exhibited an even more pronounced cognate interference effect. The mixed-effects model estimated a cognate effect of $+899.1$ ms ($SE = 52.7$, $t = 17.06$, $p < .001$), corresponding to a large effect size (Cohen's $d = 1.12$). As with the Group EF, this direction of effect indicates slower reaction times for cognates, implying cross-linguistic interference rather than facilitation. This pattern is illustrated in Figure 4.11, which shows the mean RT for cognate and non-cognate trials for Group EM.

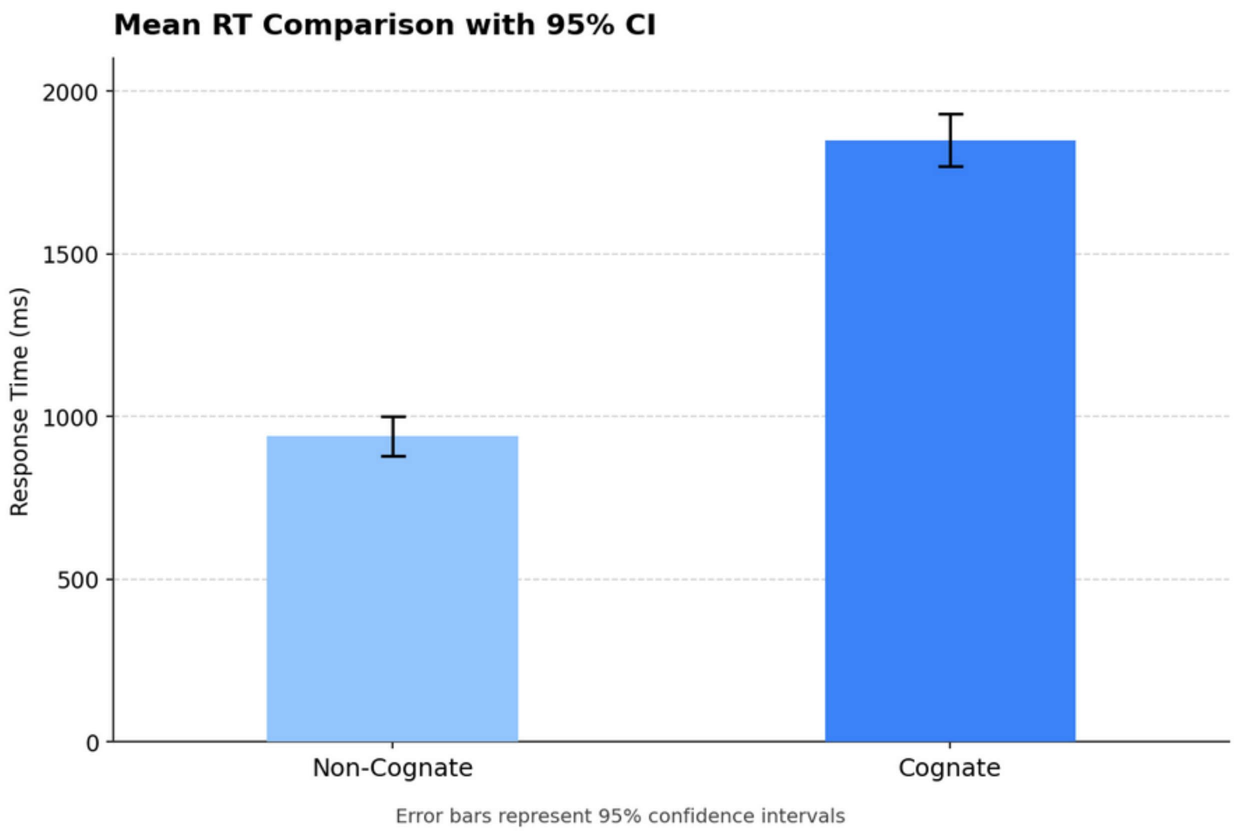


Figure 4.11: Mean RT comparison of non-cognate versus cognate trials for Group EM, indicating that cognates require longer RT—approximately double the mean values for non-cognate trials.

The model diagnostics revealed excellent reliability, with a marginal R^2 of 0.42 and a conditional R^2 of 0.68, indicating that both fixed and random effects together explained a substantial proportion of variance in lexical decision times. Variance partitioning showed 31% of total variance arising from subject-level differences and 22% from item-level variation, higher than in the Group EF data, suggesting greater individual and lexical variability in cognate processing among Group EM bilinguals.

The strength of the interference effect here is particularly noteworthy given the typological distance between English and Mandarin. Unlike English-French, the English-Mandarin pair does not contain true cognates in the traditional sense, but rather lexical borrowings that share semantic content and partial phonological overlap. At face value, the two languages share minimal orthographic overlap due to Mandarin's logographic writing system and tonal phonology, in contrast to English's alphabetic structure; however, this experiment was conducted using pinyin and employed borrowed forms that preserved phonological similarity across languages. As a result, these items occupy an ambiguous status: although presented in a Mandarin-consistent form, they may still be recognized as English-derived at some level of processing. This ambiguity may contribute to increased reaction times, as participants are required to resolve whether the stimulus should be treated as English or Mandarin despite its Mandarin-consistent surface form.

Consequently, the cognates in this context were primarily semantic or conceptual cognates, with phonological influences, rather than direct orthographic–phonological counterparts. Their processing therefore depended more heavily on semantic-level associations, where parallel activation may produce conceptual competition rather than form-based facilitation (Degani & Tokowicz, 2010 [75]). This effect may be further amplified by the fact that many borrowed Mandarin forms are multisyllabic, increasing processing demands relative to shorter, orthographically overlapping cognates in closely related language pairs.

This interpretation aligns with cross-language semantic interference effects observed in bilingual lexical decision and translation tasks (Duyck et al., 2007 [74]). When bilinguals encounter semantically related but form-dissimilar items—particularly borrowings that do not fully conform to either language's canonical lexical structure—both lexical networks may be activated via conceptual mediation, requiring the bilingual to suppress the non-target pathway. The reaction time delay of nearly $900msec$ in this group therefore reflects high-level inhibitory control costs associated with resolving between-language conceptual overlap under conditions of reduced form transparency.

The elevated subject-level variance (31%) likely reflects differences in Mandarin proficiency, pinyin familiarity, and sensitivity to borrowed lexical forms. Some participants may have relied primarily on semantic cues, while others engaged in phonological or transliterative processing, leading to heterogeneity in response patterns. The word-level variance (22%) further suggests that lexical characteristics, such as the degree of phonological transparency and syllabic complexity of borrowed forms, modulate interference strength.

Comparatively, the interference magnitude is greater for Group EM than for Group EF bilinguals, consistent with the hypothesis that linguistic distance, combined with the absence of true cognates, amplifies processing costs. As overlap between lexicons decreases and lexical items occupy an intermediate borrowing status, the bilingual must rely on more abstract semantic mapping processes that engage greater cognitive control and working memory resources (Green & Abutalebi, 2013 [73]).

Overall, these findings extend the literature on cognate effects by showing that interference is not limited to closely related language pairs. Even across typologically distant languages, and in the absence of true cognates, bilinguals exhibit parallel activation of both lexicons, supporting the non-selective nature of bilingual lexical access proposed by the BIA+ model (Dijkstra & Van Heuven, 2002 [1]).

4.3.2 Investigating the Cognate Effect: Pupil Dilation

To complement the behavioural findings, pupil dilation analyses were conducted to assess the cognitive load associated with cognate processing. Here, pupil dilation served as a sensitive index of mental effort and arousal, with larger dilations reflecting increased attentional allocation and working memory demands (Hyönä, Tommola, & Alaja, 1995 [155]; Kuchinsky et al., 2013 [76]).

To investigate whether cognate status modulates pupillary responses during word processing, the **pupil dilation differential**—defined as the difference between peak dilation at stimulus presentation and baseline dilation for each trial—was used as the dependent variable. Word type (cognate vs. non-cognate real words) was the primary fixed-effect predictor, and the following hierarchical mixed-effects model was employed using REML estimation:

$$\text{PupilDiff} \sim \text{word\_type} + (1 \;\text{—}\; \text{subject}) + (1 \;\text{—}\; \text{word})$$

Unlike the task-level pupillometry analyses reported in *Section 4.2*, where left and right pupil data were averaged to reduce noise and improve model parsimony, the present analyses examine left and right pupils separately. This decision was motivated by an environmental confound described in *Section 5.3* in *Chapter V: Discussion* the large window to the left of the eyetracker setup introduced differential ambient light exposure, which disproportionately affected left pupil measurements on sunnier days. Averaging the pupils under these conditions would risk propagating this asymmetric noise into the combined measure; separating them allows the confound to be isolated and its impact assessed independently.

4.3.2.1 Pupil Dilation: Observations in English-French Bilinguals

In this section, we explore the relationship between cognate status and pupil dilation in English-French bilinguals during the Cognate Condition, examining results from both left and right pupil dilation analyses.

The analysis of left pupil dilation revealed a statistically significant positive effect of cognate status. Specifically, cognates elicited pupil dilations that were $0.634mm$ larger than non-cognate real words ($SE = 0.199$, $z = 3.181$, $p = .001$, 95% CI $[0.243, 1.024]$), as shown in Table 4.18. This effect was robust and clearly distinguishable from zero, with the confidence interval excluding the null value by a substantial margin.

Table 4.18: Hierarchical Mixed-Effects Model Estimates with Crossed Random Effects for left pupil dilation in Group EF.

Parameter	**Coefficient**	**SE**	**z**	**p**	**CI Lower**	**CI Upper**
Fixed Effects						
Intercept	3.80	0.706	5.38	$< .001$	2.41	5.18
Word Type (Cognate)	0.63	0.20	3.18	**.001**	0.24	1.02
Random Effects (Variance Components)						
Group (Participants)	11.13					
Stimulus (Words)	4.38					
Residual	6.26					
Variance Proportions						
ICC (Groups): 51.1%	ICC (Stimuli): 20.1%	Residual: 28.8%				
Model Fit Statistics						
No. Observations: 1,172	No. Groups: 23 (mean size: 51)					
Method: REML	Converged: Yes	Log-Likelihood: -3093.38				

The model intercept of 3.797 mm ($SE = 0.706$, $z = 5.381$, $p < .001$) represents the estimated mean pupil dilation for non-cognate real words, providing a stable baseline reference. The variance components revealed substantial heterogeneity at both the group level ($Var = 11.13$) and stimulus level ($Var = 4.38$), with group variance being approximately 2.5 times larger than stimulus variance. This variance decomposition indicates that individual differences among participants accounted for a larger proportion of unexplained variability than differences among specific word stimuli, though both sources contributed meaningfully. The residual variance was estimated at 6.26.

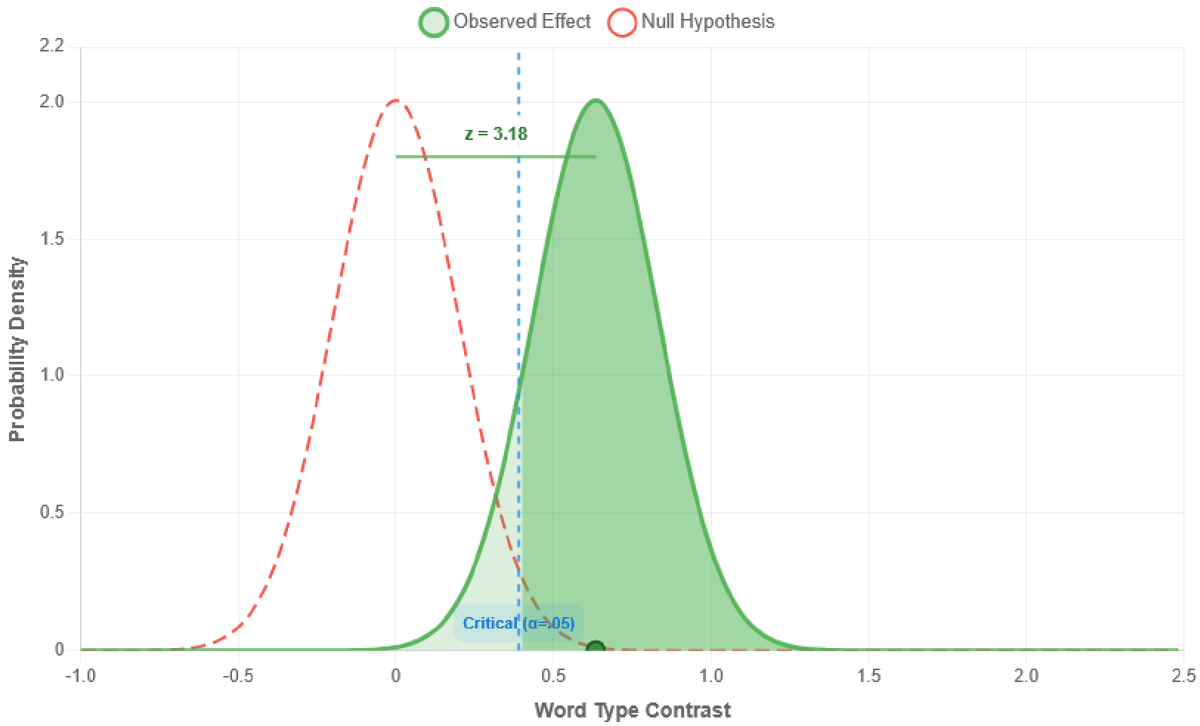


Figure 4.12: Sampling distribution of the cognate effect estimate (green curve) versus the null hypothesis distribution (red dashed curve). The observed effect ($\beta = 0.634$, marked with green circle) substantially exceeds the critical value for significance at $\alpha = .05$ (blue dotted line), yielding $z = 3.18$ and $p = .001$.

Figure 4.12 provides a visual representation of the statistical evidence supporting the cognate effect. The observed cognate effect of $\beta = 0.634mm$ falls well into the tail of the null distribution, beyond the critical value threshold of $\alpha = 0.05$. The minimal overlap between distributions underscores why the 95% confidence interval $[0.243, 1.024]$ excludes zero by a substantial margin, providing further evidence against the null hypothesis of no cognate effect. This graphical evidence reinforces the interpretation that cognate processing genuinely elicits increased pupillary responses—potentially reflecting heightened cross-linguistic activation, lexical competition, or metalinguistic awareness—rather than the reduced cognitive effort predicted by traditional facilitation accounts.

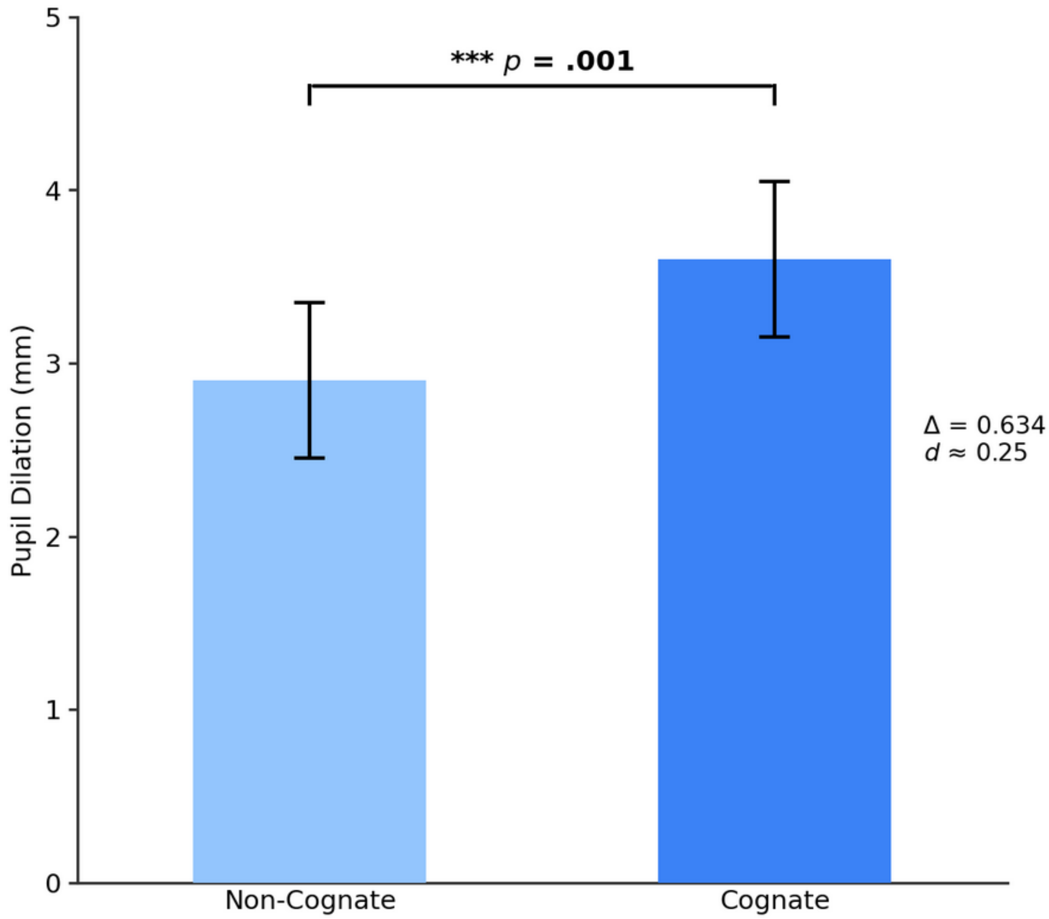


Figure 4.13: Estimated mean pupil dilation by word type with 95% confidence intervals. Cognate words show significantly higher pupil dilation than non-cognate words ($\beta = 0.634mm$, $p = .001$). Asterisks indicate statistical significance ($***p = .001$).

Figure 4.13 illustrates the estimated mean pupil dilation for both word types. These findings challenge traditional facilitation-based theories of cognate processing, which predict reduced cognitive effort due to form-meaning overlap between languages. Instead, the results suggest that recognizing cognates engages additional cognitive resources, likely due to cross-linguistic lexical activation or competition between translation equivalents. The small effect size ($d \approx 0.25$) indicates a modest but directionally consistent increase in cognitive load associated with cognate processing.

Table 4.19: REML Mixed-Effects Model Estimates for right pupil dilation by word type for Group EF.

Parameter	**Coefficient**	**SE**	**z**	**p**	**95%CI (Lower)**	**CI Upper**
Fixed Effects						
Intercept	3.63	0.16	22.33	$< .001$	3.31	3.95
Word Type (Cognate)	0.45	0.27	1.56	.12	-0.11	0.96
Random Effects (Variance Components)						
Group Variance	0.00	0.07				
Residual Variance	19.99					
Model Fit Statistics						
No. Observations: 1,172 No. Groups: 51 (range: 22–23 per group)						
Method: REML Converged: Yes Log-Likelihood: -3418.97						

The analysis of right pupil dilation revealed a positive but non-significant effect of cognate status (Table 4.19). Cognates elicited pupil dilations that were $0.43mm$ larger than non-cognate real words ($SE = 0.273$, $z = 1.555$, $p = .120$, 95% CI $[-0.111, 0.961]$). While this effect did not reach conventional significance at $\alpha = 0.05$, the p-value of $.120$ suggests a trend-level effect that warrants consideration given its directional consistency with the left pupil result. This is illustrated in Figure 4.14.

The marginal significance ($p = .120$) and confidence interval that narrowly includes zero (lower bound: -0.111) indicate that while the data trend toward a positive effect, the evidence remains equivocal. This pattern may reflect insufficient statistical power to detect a genuine small effect, particularly given the substantial residual variance relative to the effect magnitude. Figure 4.15 contextualises the observed effect within the framework of estimation uncertainty.

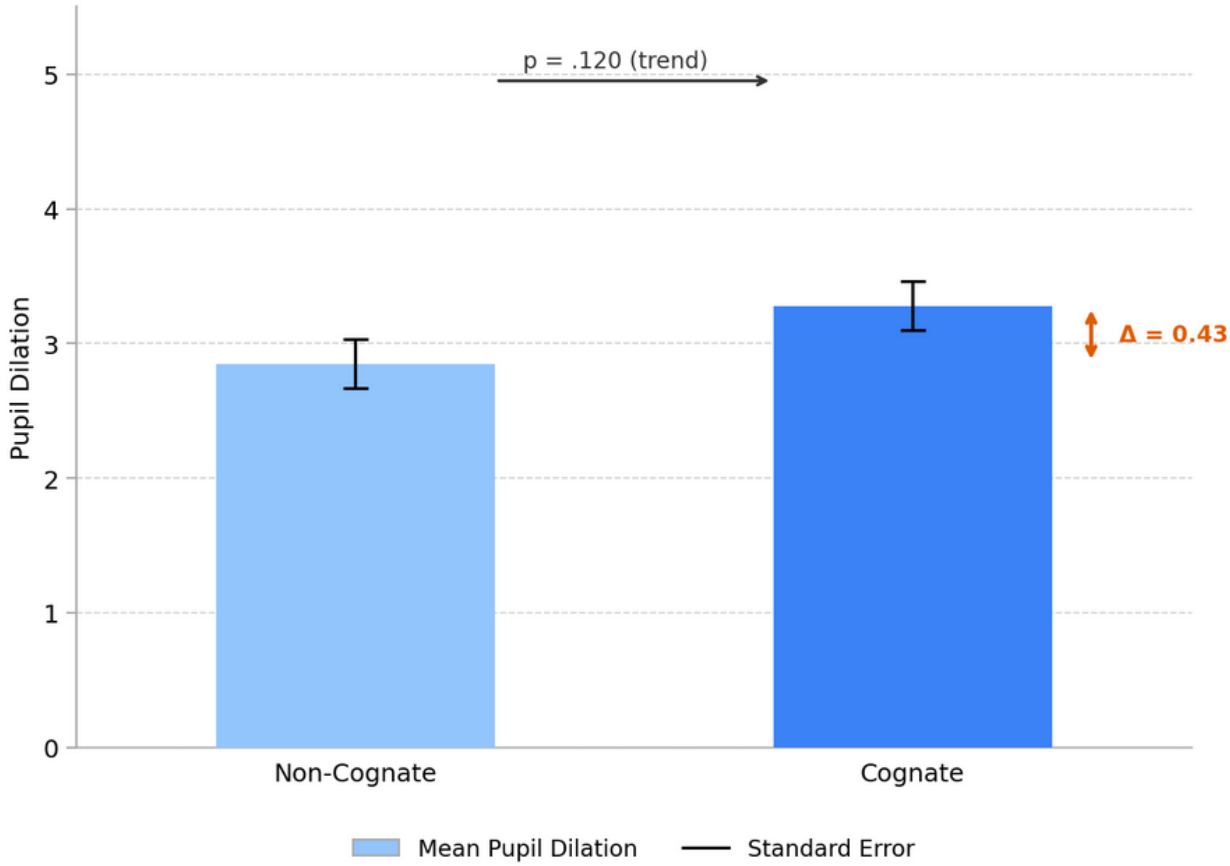


Figure 4.14: Estimated mean pupil dilation by word type with 95% confidence intervals. Cognate words show numerically higher pupil dilation ($\beta = 0.425$), though this difference is marginally non-significant ($p = .120$).

Taken together, the significant cognate effect in left pupil dilation, alongside the directionally consistent but non-significant right pupil result, suggests that bilingual word recognition involves distinct processing dynamics when orthographic and semantic overlap are present, potentially requiring additional cognitive resources to manage cross-linguistic activation and competition. This increased pupillary response may reflect greater attentional engagement, deeper lexical processing, or activation of metalinguistic awareness when cognate relationships are detected, consistent with a cognate interference interpretation for Group EF.

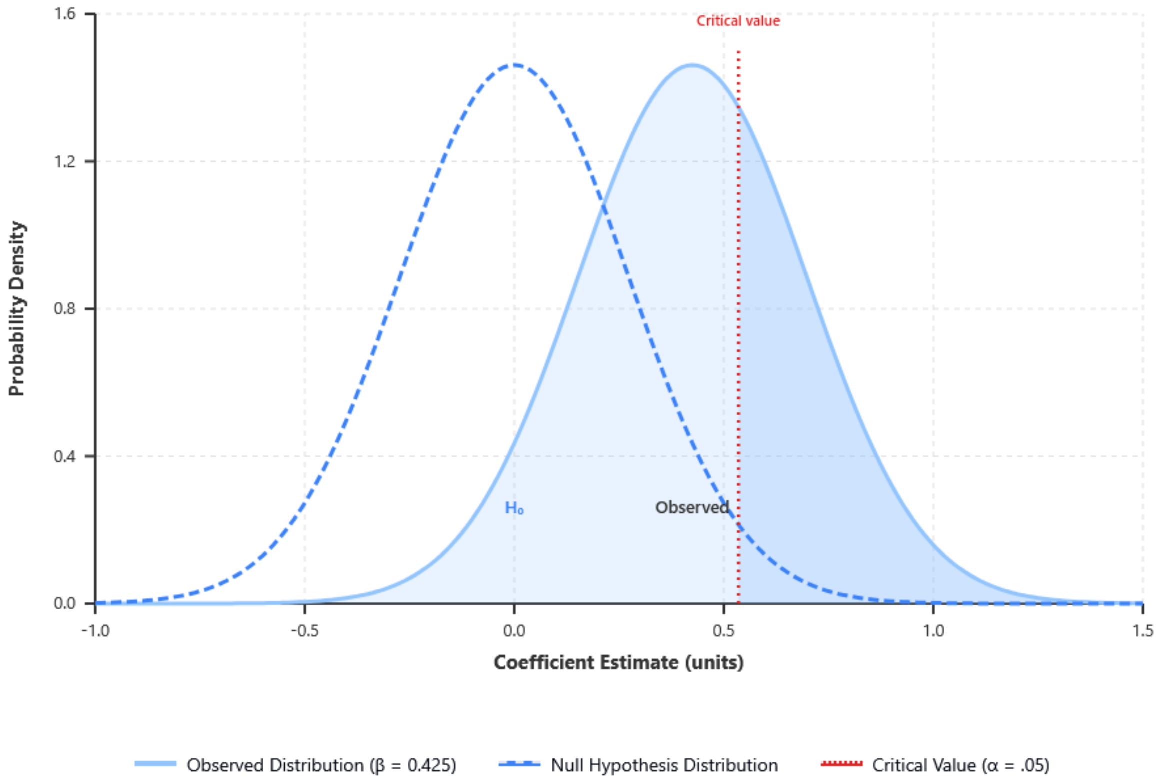


Figure 4.15: Distribution of the cognate effect estimate (blue solid curve) compared to the null hypothesis distribution (blue dashed curve). The observed effect ($\beta = 0.425$) is displaced from zero but does not exceed the critical threshold for significance at $\alpha = .05$ (red dotted line).

4.3.2.2 Pupil Dilation: Observations in English-Mandarin Bilinguals

The Group EF analyses revealed a significant increase in left pupil dilation for cognates relative to non-cognates, alongside a directionally consistent but non-significant right pupil effect. Whether a comparable pattern emerges for Group EM is less straightforward to predict: given the considerable typological distance between English and Mandarin, and the reliance on borrowed rather than true cognate forms, the mechanisms driving cross-linguistic activation may operate differently, or may not produce a measurable pupillary signature at all.

Table 4.20: REML-Based Mixed-Effects Model Parameter Estimates for left pupil dilation in Group EM.

Parameter	Coefficient	SE	z	p	95%CI (Lower)	CI Upper
Fixed Effects						
Intercept[a]	-0.114	6.48	~ 0	1.00	-12.71	12.71
Cognate Status	-0.19	0.25	-0.778	.44	-0.67	0.29
Model Fit Statistics						
No. Observations: 1,275 No. Groups: 25 (mean size: 51)						
Method: REML Converged: Yes Residual Variance: 17.85						
Group Variance: ~ 0 Log-Likelihood: undefined						

The analysis of left pupil dilation revealed no statistically significant effect of cognate status, as outlined in Table 4.20. The coefficient for cognate words was -0.190 units ($SE = 0.245$, $z = -0.778$, $p = .436$, 95% CI $[-0.670, 0.289]$), suggesting that cognates showed approximately $0.19mm$ lower pupil dilation compared to non-cognate words, though this difference was not statistically distinguishable from zero.

Figure 4.16 illustrates the estimated pupil dilation differential for both word types, with overlapping confidence intervals reinforcing the lack of statistical differentiation. The negligible effect size ($d \approx -0.04$) combined with the non-significant result suggests that cognate processing does not substantially alter cognitive load as indexed by pupillary metrics in this context.

The results for right pupil dilation in Group EM similarly revealed no statistically significant effect of word type (Table 4.21). The estimated coefficient for cognate status was -0.197 units ($SE = 0.290$, $z = -0.681$, $p = .496$, 95% CI $[-0.765, 0.371]$), as illustrated in Figure 4.17.

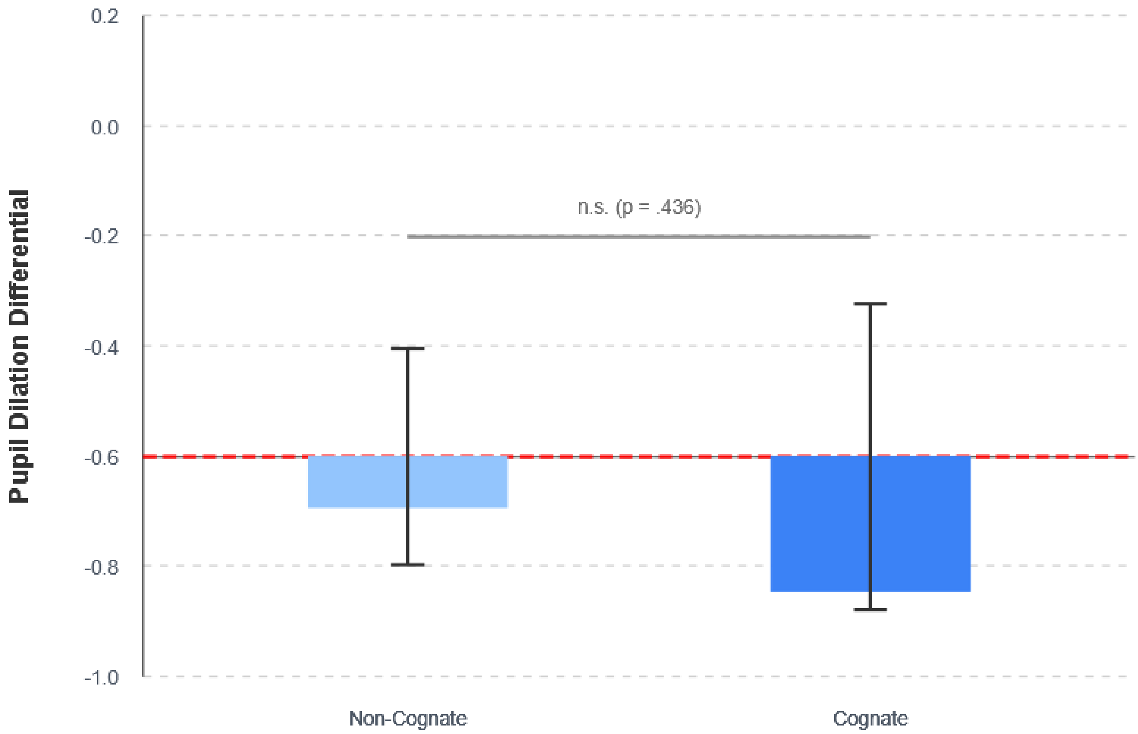


Figure 4.16: Estimated mean pupil dilation differential by word type with 95% confidence intervals for Group EM left pupil. Overlapping intervals indicate no significant difference between cognate and non-cognate conditions.

Table 4.21: Mixed-Effects Model Parameter Estimates for right pupil dilation in Group EM.

Parameter	**Coefficient**	**SE**	**z**	**p**	**95%CI (Lower)**	**CI Upper**
Fixed Effects						
Intercept	5.46	0.178	30.85	$< .001$	5.11	5.81
WordType (Cognate)	-0.120	0.29	-0.68	.50	-0.77	0.37
Model Fit Statistics						
No. Observations: 1,275	Log-Likelihood: -3861.36	Residual Variance: 25.01				
Method: Maximum Likelihood	Converged: Yes					

Taken as a whole, the findings on pupil dilation for Group EM suggest that cognate status does not modulate cognitive effort or arousal as indexed by pupillary response. The negligible effect sizes ($d \approx -0.04$) across both left and right pupil analyses are consistent with this interpretation. Several factors may account for this null finding. First, the considerable linguistic distance between English and Mandarin may mean that cognate-like stimuli do not meaningfully engage cross-linguistic activation in the same way they do for English-French bilinguals. Second, the use of Pinyin without tonal markings may have reduced the perceptual salience of any cognate relationship for Group EM participants. Finally, the task itself may have imposed a sufficiently high baseline cognitive demand on Group EM participants—as suggested by their consistently longer RTs—leaving less room for cognate status to produce a detectable additional load. These remain speculative interpretations, and future research with larger samples and more controlled stimuli would be needed to adjudicate between them.

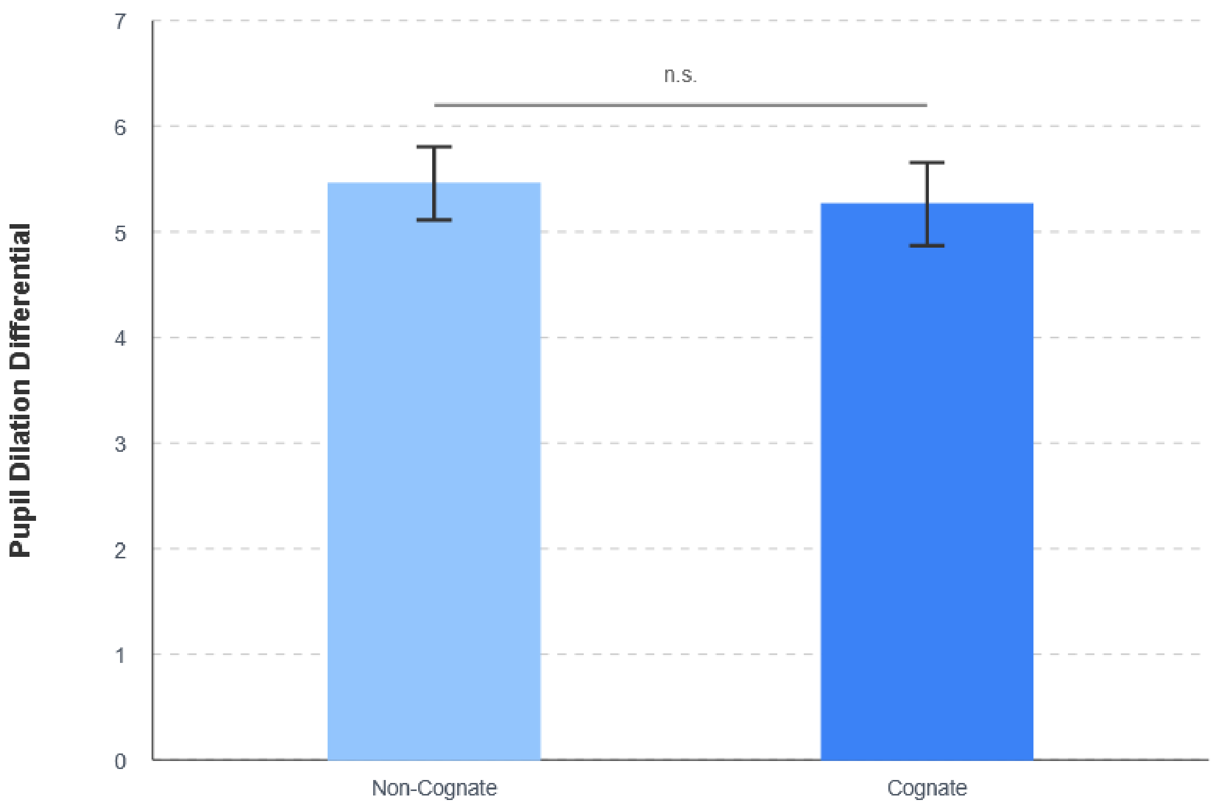


Figure 4.17: Estimated mean pupil dilation differential by word type with 95% confidence intervals for Group EM right pupil. The lack of separation between conditions indicates no statistically significant difference (n.s. = not significant, $p = .496$).

4.3.3 Implications of Pupil Dilation for the Cognate Effect

Taken together, the behavioural and pupillometric findings speak directly to theoretical accounts of how bilinguals manage cross-linguistic competition during lexical access. The convergence of slower RTs and sustained pupil dilation in the Cognate condition—particularly for Group EF—is consistent with the Adaptive Control Hypothesis (Green & Abutalebi, 2013 [73]), which proposes that bilinguals continuously regulate cognitive control networks in response to cross-linguistic competition. Cognate trials, by jointly activating representations across languages, appear to require greater inhibitory engagement, reflected behaviourally in slower responses and physiologically in sustained pupil dilation across the decision window.

In summary, the RT and pupil-dilation findings are mutually reinforcing: both demonstrate that cognate processing imposes a cognitive cost, with the pupil data capturing the sustained nature of that cost and the RT data capturing its cumulative behavioural impact. Together, they show that bilingual lexical access is non-selective and cognitively taxing when cross-linguistic similarity is high, and that pupil dilation provides a sensitive, temporally precise index of the interference mechanism underlying RT-based cognate effects.

CHAPTER V

DISCUSSION

5.1 Summary of Results

To better understand language processing in a bilingual context and further investigate cognate effects, the data for Reaction Time (RT), Response Accuracy, and Pupil Dilation were considered and evaluated. In this chapter, I summarize the findings of this research before moving on to the implications of these findings.

5.1.1 Behavioural Data

5.1.1.1 Summary of Reaction Time (RT) Data

The RT data across all three conditions revealed a consistent pattern: responses were fastest in the NC condition, slowest in the Control condition, and intermediate in the Cognate condition; a pattern that held across both language groups and was confirmed by mixed-effects modelling.

With regard to RT data, RTs were shortest in the NC condition and longest in the Control condition, with the Cognate condition patterning between the two. At first glance, the Control condition appeared to elicit the longest RTs overall across both language groups, a pattern later corroborated by statistical analysis. These findings indicate that RT differences across conditions are largely driven by differences in task demands. Specifically, it was found that not only did CONDITION have a significant effect on RT, but lexical access was also influenced by LANGUAGE GROUP (i.e., linguistic similarity). In particular, while Group EF speakers exhibited the longest RTs in the Control condition, Group EM speakers showed consistently longer RTs across both the Control and Cognate conditions (see Table 4.4).

These condition-level differences are best understood in terms of task demands. The NC condition's short RTs reflect the relative simplicity of straightforward L1–L2 identification with no competing token types. The Control condition's pronounced slowdown—nearly $400msec$ longer than NC for both groups—is attributable to the additional verification required when non-word tokens are present: participants could not rely on language identity alone but had to first confirm wordhood, a process that appears to have introduced meaningful processing overhead. The Cognate condition sits between these two poles not because it is intermediate in difficulty per se, but because it shares the core task structure of NC (i.e., L1 vs. L2 identification) while adding the complication of cross-linguistic form overlap at the lexical level.

Focusing first on the Control condition, the longer RTs can be attributed to the increased cognitive demands associated with evaluating word status when non-word tokens are expected. Because participants were aware that non-words would appear, each trial required additional verification beyond simple language identification. Given that these non-words were constructed to conform to the phonotactic and morphological constraints of English, French, and Mandarin, it is plausible that participants experienced uncertainty regarding wordhood, thereby increasing processing time. This raises the possibility of priming effects: because tokens were presented in random order, exposure to a real word may have biased participants toward interpreting subsequent non-word tokens as real words.

A secondary observation concerns differences between Group EF and Group EM speakers. While both groups demonstrated relatively similar RTs in the Control condition, Group EF exhibited marginally longer RTs. This may be attributable to shared linguistic features between English and French, which could bias speakers toward falsely identifying non-words as real words based on familiar phonotactics or morphology. In contrast, interference in Group EM appeared less related to cross-linguistic similarity and more to uncertainty in distinguishing Mandarin words from non-words. Because Mandarin stimuli were presented in *Pinyin* without tonal markings, and non-words were generated to closely match Mandarin phonotactic constraints, participants may have experienced greater difficulty assessing word status in Mandarin than in English.

Turning finally to the Cognate condition, RTs were longer than in the NC condition but shorter than in the Control condition for both language groups. Crucially, the Cognate condition shared its core task demands with the NC condition: participants were required only to decide whether a word belonged to L1 or L2, with no requirement to assess word status. As such, the increased RTs observed relative to NC are not attributable to word/non-word interference, but rather to the additional processing demands introduced by cross-linguistic overlap at the lexical level. In this sense, the Cognate condition can be viewed as a special case of the NC condition; one in which shared linguistic features between languages modulate processing efficiency.

These condition-level RT differences are directly interpretable within the BIA+ framework. The model's task/decision system—introduced in part through Green's (1998) Inhibitory Control model—predicts that observable performance is not determined solely by lexical activation, but by the interaction between bottom-up lexical competition and top-down task schema constraints. Under this account, participants do not approach each condition naively; rather, they reconfigure their decision criteria in anticipation of the stimulus type they expect to encounter. In the NC condition, where only unambiguous L1 and L2 words are presented, the task schema imposes minimal inhibitory demands, consistent with the fastest observed RTs. In the CONT condition, the introduction of non-words raises participants' uncertainty about word status, leading to a more conservative decision criterion and the longest RTs observed across all conditions; a pattern fully consistent with the BIA+ prediction that greater task complexity increases the processing overhead imposed by the task/decision system. The COG condition sits between these two poles: participants must resolve language membership rather than word status, but the presence of cross-linguistically overlapping items increases lexical competition at the word identification level rather than the decision level per se. The intermediate RTs in this condition are therefore consistent with BIA+ predictions that cognate processing engages the word identification system more heavily than NC, while imposing less decision-level overhead than CONT. Taken together, the RT data across all three conditions are broadly consistent with the task dependency predictions of the BIA+ model, and provide empirical grounding for the third research question outlined in *III: Current Study*.

5.1.1.2 Summary of Response Accuracy Data

Moving to the Response Accuracy data, we observed clear effects of *Language Group* and *Condition* on overall accuracy. Accuracy was evaluated using two measures across all experimental conditions. The first, *net accuracy*, was calculated at the participant level as the proportion of correct responses out of total responses within each condition, without distinguishing between trial types (e.g., real word, non-word, cognate). This measure reflects overall performance but is less sensitive to variability within a condition. The second, *weighted accuracy*, treated each individual trial as a binary observation (correct = 1, incorrect = 0) regardless of trial type, weighting all trials equally and thus providing greater sensitivity to within-condition variability. As detailed in Chapter IV: Results and Analysis , these two measures sometimes diverged—most notably in the Cognate condition—and that divergence is itself informative.

Accuracy followed the same broad ordering as RT, though the two measures diverged in informative ways. NC yielded near-ceiling accuracy for both groups, consistent with its low task demands. The Control condition produced the steepest accuracy drop—particularly for Group EM, whose scores fell roughly 15 percentage points below Group EF—suggesting that non-word interference places a greater burden on groups whose L2 shares less orthographic and phonological structure with L1. The Cognate condition showed an accuracy pattern that depended heavily on how accuracy was measured: net scores were high (suggesting surface-level facilitation), but weighted scores—which treated cognate identification strictly as correct or incorrect—were the lowest of any condition. This divergence is itself informative: it indicates that cognates introduce a form of processing ambiguity that inflates apparent accuracy under lenient scoring but reveals real difficulty under stricter criteria.

These observations provide nuanced insight into how cognates are processed relative to non-cognates in bilingual word recognition (see Poort & Rodd, 2017 [156]; Costa et al., 2000 [138]; Sherkina-Lieber, 2004 [157]). Participants were more accurate in tasks involving only L1–L2 identification (NC and Cognate) than in tasks requiring word status evaluation (Control), highlighting the influence of task demands. Notably, Group EF showed slightly higher accuracy in the Cognate condition than in the NC condition, suggesting more efficient processing when cognates are present and the languages share substantial overlap. In contrast, Group EM exhibited reduced accuracy in the Cognate condition, indicating that the presence of cognates may introduce inhibitory effects when linguistic overlap is limited.

As shown in Table 4.9, weighted accuracy confirmed this ordering across both language groups: NC yielded the highest scores, followed by Control, with Cognate yielding the lowest. Although initially surprising, this reflects the greater sensitivity of weighted measures to within-condition variability, and is consistent with the interpretation that cognates introduce genuine processing difficulty that net accuracy alone does not fully capture. Mixed-effects modelling confirmed significant effects of CONDITION and LANGUAGE GROUP, reinforcing the conclusion that both task demands and linguistic similarity shape accuracy outcomes.

Importantly, the significant effect of *Condition* on accuracy should not be interpreted as direct evidence of cognate facilitation, since Cognate trials actually yielded the lowest weighted accuracy. Rather, these differences underscore how task demands and linguistic similarity jointly influence performance. Considering the observed effect of *Language Group* on accuracy scores, particularly in the Control and Cognate conditions, we can conclude with greater confidence that cognate effects depend on the similarity of the language pair. Specifically, when languages are highly similar, cognates may facilitate processing in terms of speed or ease, whereas when languages differ substantially in syntax, orthography, or phonology, the presence of cognates may instead introduce inhibitory effects.

5.1.2 Summary of Pupil Dilation Data

The pupillometry data converged with the behavioural findings: the Cognate condition imposed the highest cognitive load for Group EF bilinguals, while Group EM showed a comparatively flat profile across all three conditions (see Table 4.15). This pattern is taken up in the following section in relation to the cognate-specific analyses.

The secondary analysis of this same dataset—using mixed-effects modeling—yielded converging yet more nuanced results. While there were little to no overarching effects of language group on overall pupil dilation (i.e., Group EF and Group EM bilinguals did not differ substantially in baseline pupillary reactivity), each group responded differently to the experimental conditions. As discussed in *Chapter IV: Results and Analysis*, this reflects a reliable LANGUAGE_GROUP × CONDITION interaction, indicating that the impact of condition complexity depends, to some extent, on language background. In other words, while the overall magnitude of dilation may not differ between bilingual groups, the patterning of pupillary responses across NC, Control, and Cognate trials does.

The mixed-effects modelling also indicated that CONDITION had a significant effect on pupillary response. This finding is consistent with the preliminary averages, which showed that the Control and Cognate conditions presented more variable responses between groups, while the No Control condition yielded comparatively stable dilation patterns across both bilingual groups. A significant difference emerged between NC and Control conditions—mirroring both RT and accuracy findings—suggesting that the presence of non-words (in Control) introduced inhibitory demands that increased cognitive load. Interestingly, this same NC–Control difference did not emerge between NC–Cognate or Control–Cognate, which indicates that cognate processing introduces its own distinct cognitive dynamics.

5.1.2.1 Results of Pupil Dilation in Relation to the Cognate Effect

The mixed-effects models fitted for the Cognate condition refine and extend these earlier observations by directly examining whether cognate status modulates pupil dilation when bilinguals process real L1–L2 word pairs.

For Group EF, the hierarchical mixed-effects model revealed a significant and robust cognate effect on left-pupil dilation: cognates elicited $0.634mm$ larger dilations than non-cognate real words. This finding is consistent with the earlier suggestion that task complexity (such as increased lexical competition or activation) may increase cognitive load. Yet, it importantly extends the earlier interpretation by showing that even in the Cognate condition—where average dilation was lower than Control—within-condition cognate processing still increases cognitive load relative to matched non-cognates. In other words, the Cognate condition may overall appear less demanding than Control when averaged across word types, but at the level of individual word processing, cognates themselves elicit greater pupillary responses than non-cognates. This pattern parallels RT findings showing that cognates did not always facilitate responses, contrary to classical facilitation accounts, and instead sometimes induced processing difficulty.

Right-eye analyses in Group EF revealed a positive but non-significant trend ($\beta = 0.425$, $p = .120$) in the same direction, with overlapping confidence intervals. Although weaker in magnitude, this consistency across eyes reinforces that recognition of cognates engages additional cognitive and attentional resources. The findings suggest that the processing of cognates—even when familiar and semantically transparent—activates cross-linguistic lexical networks, eliciting increased cognitive work rather than reduced effort.

In contrast, English–Mandarin bilinguals (Group EM) showed no significant cognate effect for either left or right pupil dilation. While the direction of the effect was negative (approximately -0.19 to -0.20 mm), the confidence intervals were wide and included zero, and variance at the group level was negligible. This null effect suggests that cognate status does not meaningfully modulate cognitive load in English–Mandarin bilinguals, a result that is consistent across all model specifications. This aligns with the earlier observation that the Cognate and Control conditions elicited more variable responses in English–French bilinguals than English–Mandarin bilinguals, indicating that linguistic similarity between languages modulates the cognitive dynamics of word recognition. Because English and Mandarin share limited orthographic and morphological overlap, cognate-like stimuli may not meaningfully engage cross-linguistic activation processes in the same way they do for English–French bilinguals.

Taken together, the conclusions from the preliminary analyses and the detailed cognate modelling produce a coherent picture:

1. *NC trials show low variation across groups because the task is simple, involving straightforward word recognition with minimal interference.*

2. *Control and Cognate trials show greater cognitive load, consistent with both increased pupil dilation and slower RTs in conditions involving non-words or cross-linguistic activation.*

3. *Differences between NC and Control reflect the inhibitory cost of non-words, which raise cognitive load in both pupil dilation and behavioural performance measures.*

Building on those points, we note also that the group-specific modulation of cognate processing—present for Group EF but absent for Group EM—raises important questions for theoretical accounts of bilingual lexical access, which are taken up in the following section.

In sum, the combined data are broadly consistent with the broader research objective, offering preliminary support for the use of pupil dilation as a physiological index of cognitive load during bilingual word processing—one that may capture sensitivity to processing demands not fully reflected in reaction times or accuracy alone. The hierarchical mixed-effects analyses, especially for Group EF, suggest that pupillometry may be sensitive to cross-linguistic activation costs even when behavioural metrics show minimal differences—serving as a proof of concept for the use of pupillometry in bilingual word processing tasks. As such, pupil dilation represents a promising and complementary methodological tool for investigating the cognitive dynamics of bilingual lexical processing, pending replication with larger samples and more controlled stimuli.

5.2 Implications

The findings of this study are broadly consistent with an extension of the BIA+ model—one that more explicitly accounts for the interaction between L1 and L2 at the language node, and for the role of linguistic similarity in modulating lexical access. While the task employed here differs from a standard Lexical Decision Task, and therefore warrants caution in drawing direct theoretical conclusions, the findings of this study are nonetheless suggestive of a revision to the BIA+ model—or at minimum, are consistent with an extension of it to more explicitly account for the complex interactions of L1 and L2. To this end, I put forward the following revision:

The key modification to the BIA+ model—which is already a robust model to represent the task dependency of lexical processing via Green's (1998) [58] IC Model—is the incorporation of Kroll & Stewart's (1994) [54] Revised Hierarchical Model (see Figure 2.5) at the language node. The motivation for this modification is twofold. First, the between-group differences observed in reaction time and pupil dilation across conditions are consistent with the broader literature suggesting that relative language dominance shapes lexical access — a dimension not explicitly tested here, but one that the participant profiles (differing in acquisition history and self-reported language use) make relevant as a contextualising factor. Second, and more directly supported by the present data, typological similarity between the L1 and the L2 was shown to have a meaningful effect on lexical access, with Group EF exhibiting distinct behavioural and pupillometric patterns relative to Group EM across all three conditions.

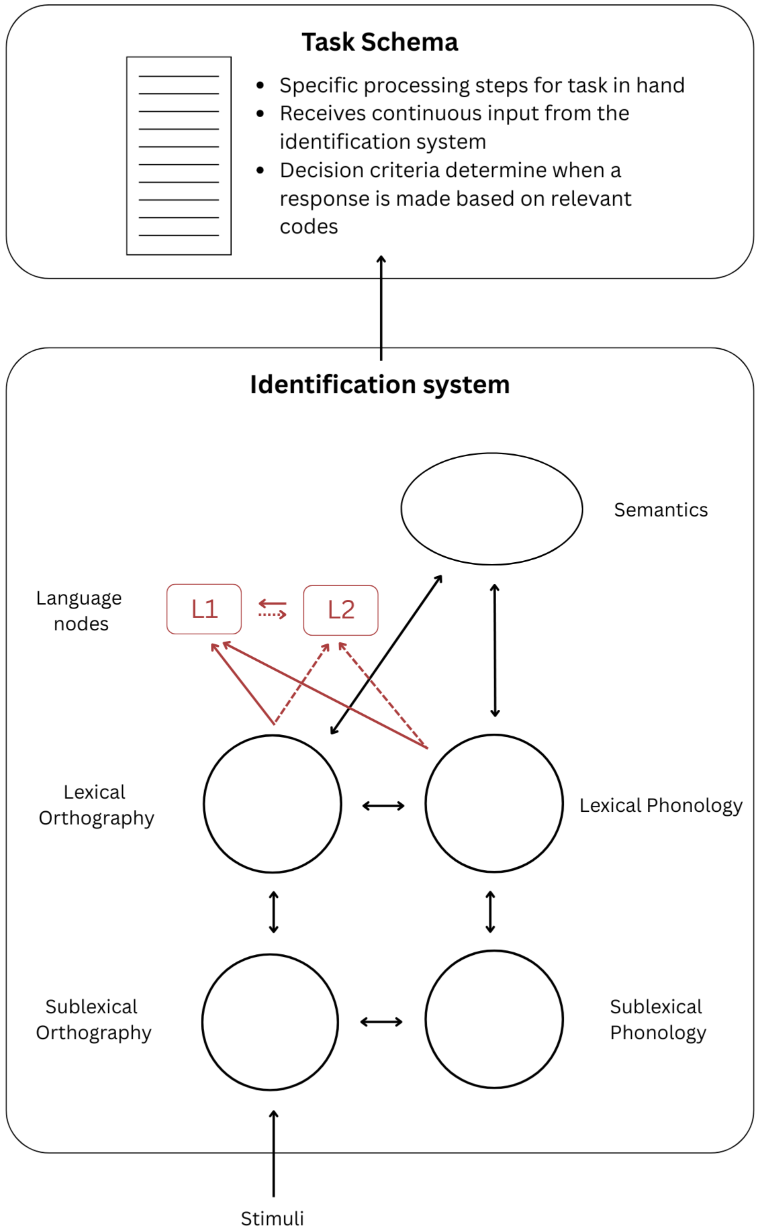


Figure 5.1: A revised version of the Dijkstra & van Heuven's (2002) BIA+ model, which incorporates Kroll & Stewart's (1994) Revised Hierarchical Model. This revised model is intended to more explicitly represent the complex interaction between L1, L2 and lexical nodes during processing, and to incorporate language dominance into the bilingual processing framework—a dimension not fully captured in the original BIA+ model.

In the proposed revised model (see Figure 5.1), we note the direct interaction of the lexical orthographic and phonological nodes with the language node (present already in BIA+), but in this case, they interact separately with the L1 and the L2 respectively. Furthermore, there is a direct interaction of L1 and L2, which was not previously accounted for in the BIA+ model. Finally, the use of solid lines is meant to indicate a robust interaction with the L1 (assuming that the L1 is the dominant language), and the dotted lines signal a weaker interaction with the L2 (assuming that L2 is the non-dominant language). Crucially, it is the addition of the direct interaction between L1 and L2, and between Lexical nodes with L1 and L2 respectively, that provides a more realistic representation of lexical access, and therefore language processing in a bilingual context, as demonstrated by this study.

On the topic of pupil dilation, it is known that pupil dilation is an involuntary and natural process in which the pupil adjusts to the amount of light entering the eye. To this end, it has been shown that pupils dilate more and for longer durations when a visual stimulus, or cue, requires greater cognitive processing or mental effort. In other words, there is evidence that pupil dilation correlates with cognitive load. This correlation was examined and further corroborated in this study; though interestingly, the findings for pupil dilation may contradict what is currently known about language processing.

It should be noted that previous literature is still generally on the fence about a cognate facilitation effect (see Poort & Rodd, 2017 [158]), however, most researchers agree that language similarities facilitate processing in a bilingual context (see van Hell & Dijkstra [159]). Given that the task employed here differs from a standard Lexical Decision Task, the pupillometry findings of this study are not straightforwardly comparable to those of the broader cognate processing literature. Rather, the pupillometry component of this study is best understood as a proof of concept, demonstrating that pupil dilation is sensitive to processing demands in bilingual word recognition tasks. With that caveat in mind, the data are broadly consistent with the suggestion that language similarity modulates cognitive load during processing, though stronger conclusions must await replication under more controlled conditions. This is evidenced by the fact that Group EF consistently demonstrated greater dilation in all experimental conditions. This is especially fascinating, as it suggests that language similarities may induce a form of inhibition, thereby requiring more reflection. Even so, more reflection does not entail slower processing—as evidenced by the findings of this study. To this end, I put forward the following suggestion: While processing is task dependent and performance is modulated by word status (i.e., word vs. non-word vs. cognate) and task demands (i.e., NC, Control and Cognate conditions), linguistic similarities between L1 and L2 will require more cognitive load (e.g., more reflection) to complete processing tasks. In contrast, if L1 and L2 do not share many or any linguistic features, they will be less likely to interact, therefore requiring less cognitive load to complete processing tasks, even when writing systems are modified to increase difficulty.

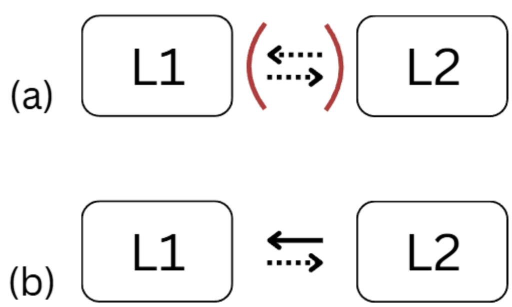


Figure 5.2: A representation of L1 and L2 interaction. In this figure, (a) illustrates the weak interaction between L1 and L2 in the case of Group EM. The red parentheses represent the optionality of this interaction which, I hypothesize, would not exist, had this experiment been conducted using Chinese characters rather than pinyin. In contrast, (b) illustrates the interaction between L1 and L2 for Group EF, taking into account language dominance.

Figure 5.2 illustrates this proposition, taking into account the optionality of the interaction between L1 and L2. The diagram in (a) can be taken to represent the interaction (or lack thereof) between two languages that are not linguistically similar. Here, the dotted arrows signal a weak interaction—which may occur if the languages share certain similarities (e.g., writing system, morphology, phonology, etc.) or in the case of cognates. The red parentheses signal the optionality principle I put forward; as demonstrated from the data, Group EM generally showed less interference of language on language, thereby supposing there was little to no overlap or interference regardless of task requirements.

The diagram in (b) shows the standard interaction, as put forward by Kroll & Stewart (1994) [54] in their Revised Hierarchical model. As previously mentioned, the dotted lines signal a weaker interaction/influence, while the solid line indicates a stronger interaction. This model is especially pertinent as it accounts for language dominance in a bilingual context and suggests that the more dominant language is more likely to have a stronger influence on processing and, in some cases, become the "point of reference" language, as illustrated in the revised BIA+ model I put forward (see figure 5.1). In this case, the more dominant language would likely be the first to be retrieved, as indicated by the solid arrow from the lexical level nodes. This weak/strong interaction was observed during this experiment, and can be summarized by highlighting the fact that in all experimental conditions, Group EF showed a higher degree of influence between L1 and L2.

Finally, I would like to clarify my point on the topic of language dominance and its role in processing. Notice that in figure 5.1, the solid line from the lexical level nodes point to L1 while the dotted lines point to L2. Notice also that in figure 5.2, I put forward the concept of "point of reference" language only in the context of mutual interaction between L1 and L2 (as modelled in (b)). I would like to clarify here that in this case, I am opting to use the term "point of reference" to reinforce the idea that the more dominant of the languages will become the foundation of processing as the speaker is more likely to reference that language first. This is not a feature which is particular to only linguistically similar languages, but is based on the notion that bilingualism is seldom balanced and one language will always hold more prestige, more importance, or more dominance over the other in the bilingual mind. As a result, it is crucial to remember that although languages may or may not directly interact at the language node, semantic representations are always shared and lexical-level features may overlap between L1 and L2; therefore, constant interaction at these levels should be presupposed.

In the following section, I will discuss in detail the limitations of the present research. However, I would like to close this section by putting forward that in future research, it would be important to further examine the limits of L1 and L2 interaction and the influence of L1 and L2 in linguistically similar and dissimilar language pairs. To this end, it would perhaps be interesting to observe such overlap by testing auditory cues, and visual cues (as was the case in this study) and perhaps capitalize on linguistic dissimilarities such as writing systems. Relatedly, future research should also consider how typological similarity itself is operationalized. In the present study, the distinction between Group EF and Group EM served as a broad proxy for linguistic similarity, but this conflates multiple dimensions—orthographic, phonological, morphological, and syntactic—that may contribute differentially to the observed effects. It remains an open question, for instance, whether the between-group differences reported here are primarily driven by orthographic overlap, phonological overlap, or some combination of the two, and whether the stimuli selected are truly representative of the broader English-French and English-Mandarin lexicons or are instead unique to the particular items chosen. Future research would benefit from developing more fine-grained, quantifiable measures of typological similarity—such as orthographic Levenshtein distance or phonological similarity metrics—and applying these systematically at the item level, so as to determine the extent to which observed processing differences generalize across the lexicon rather than reflecting properties specific to the stimulus set. This would allow researchers to move beyond broad language-pair comparisons and toward a more precise characterization of which dimensions of cross-linguistic similarity are most predictive of cognitive load and lexical access in bilingual word recognition.

5.3 Limitations and Challenges

As is the case with all research, the present project was not devoid of its fair share of challenges and limitations. This section covers those limitations and challenges, first discussing in detail the limitations and challenges around the conceptualization of this research, then touching on those regarding participants and recruitment, and finally discussing the challenges surrounding the experimental frame.

5.3.1 Limitations and Challenges Surrounding Conceptualization

While the intention was to evaluate two different language groups, practical considerations led to the selection of English-French and English-Mandarin bilinguals. Cantonese was initially considered as the L2 for the second group for two reasons: first, as a native Cantonese speaker, I would have been better positioned to independently control and validate token selection without relying on external consultation; and second, Cantonese has a well-documented history of phonetic borrowing from English, particularly in the Hong Kong context where English has long held official status and is widely taught in schools—properties that would have made it a linguistically interesting comparison to English-French. However, Cantonese was ultimately deemed unsuitable for the present study because it lacks a fully standardized romanization system, and literacy in written Cantonese among potential participants at York University was insufficient to support a visually-presented word recognition task of this kind. Mandarin was therefore selected as the L2 for Group EM, with token selection supported by consultation with a trilingual (English, French, Mandarin) colleague and a Mandarin language professor at York University to ensure appropriateness and validity for the experimental tasks.

Once all tokens were chosen and the general framework was drafted, I began setting up the experimental tasks. This process was particularly laborious because the originally intended software, PsychoPy, was not compatible with the Gazepoint GP3 eye tracker. As a result, and as detailed in *Subsection 3.4.2.2* of *Chapter III: Current Study*, all tasks had to be manually coded in 2 parts for the experiment to work with the eye tracking component—a component which is crucial for this research. As previously mentioned, each task was comprised of the main code, which was the base experimental code per task, and which also included script for tracking behavioural data such as RT and response, and a wrapper which was intended to sync the eye tracking data to the behavioural data. While the main code itself was relatively straightforward, being largely composed of a simple LDT in Pygame and script for tracking response and RT, the wrapper was another issue altogether. First, the intention was to get all the code working in a single file to avoid sync issues; however, most snippets of scripts used to accomplish this failed as they were either not compatible with the Gazepoint GP3 or it was not compatible with Pygame. The solution, a standalone PyOpenGaze Gazepoint wrapper, came after further research and a chance discussion with Dr Edwin Dalmaijer of the University of Bristol, who explained that the wrapper, written in Python, would be compatible with the Gazepoint GP3, and it would record and log all eyetracking data in the form of “events”. That said, it was also made clear that there would have to be some additional work to sync this data with that of the main code. Consequently, the solution for this issue was to develop the experiment codes (i.e., one task code and one wrapper per task, per language group) and later, develop the data processing code (see *Appendix B*).

5.3.2 Limitations and Challenges Surrounding Participants

During the recruitment and screening process for Group EF participants, a short interview was implemented to confirm that each participant met the minimum criteria for bilingual proficiency and exposure to French. The primary challenge with Group EF recruitment was therefore one of language verification: self-reported bilingualism did not always correspond to the level of French proficiency required for participation, as illustrated by the exclusion of one participant who, upon interview, was found not to meet the bilingualism criterion despite having self-identified as bilingual. As a result, recruiting sufficiently proficient French-English bilinguals required careful screening at the point of contact.

Interestingly, the issue for Group EM speakers was opposite to that of Group EF: whereas the central challenge for Group EF was one of verifying adequate proficiency, most Mandarin speakers who identified as bilingual were genuinely fully fluent, having learned the language as a heritage language or having moved to Canada during their childhood. Consequently, the issue was less of a language and proficiency problem, but more of a cultural one. Adding a personal observation here: a vast majority of the Chinese community is very reserved and will generally keep to themselves and their own circle. Furthermore, Chinese individuals who moved to Canada from Asia will often tell you their English is not great, which is typically not the case as these individuals are either highly educated and/or have lived in Canada or another English-speaking country long enough to have become fluent. Thus, to appeal to the Chinese community at York, it was crucial to network, and appeal to them via word of mouth. In this case, the opposite of the Group EF recruitment was true and I actually found myself struggling to recruit Group EM participants.

Thinking about this critically, I believe there are two other crucial factors that played into this struggle: the fact that I do not speak Mandarin, and the fact that I was born and raised in Canada. First, to not speak Mandarin already put me at a disadvantage in terms of interacting with the Mandarin speaking community so I had to rely heavily on peers and colleagues to help me spread the word. The second, and perhaps biggest factor, is my identity as a Chinese-Canadian individual; culturally, the Chinese community is generally more reserved—this statement is true across the board. However, within the Chinese community, those born and raised with the Western culture are seen as “others”—from the way they interact with the language, to the way they dress and act. Often times, I do feel like an ”other” when speaking with Chinese people who are from Asia, and this is only due to cultural nuances. But these cultural nuances are often enough to create a form of divide within the Chinese community. All this to say, as a Chinese-Canadian researcher who is Franco-Ontarian, a major challenge I faced with the recruitment process was appealing to the Chinese community and inviting them to participate in this study.

In sum, the main challenges at the conceptualization and execution stages involved difficulties in designing the study (e.g., creating appropriate word lists and implementing the experimental code) and in recruiting the “ideal” participant. As for the limitations at this stage, recruiting participants who met the minimum linguistic criteria to participate in this research was nothing short of difficult.

5.3.3 Limitations and Challenges Surrounding the Experiment

During the course of this research, three main limitations would have to be accounted for; first, that the brightness in the lab could only be controlled to a certain degree, second, that each task condition saw an inevitable priming effect, and third, that we had to put a cap on how many participants would be recruited for this study.

To address the first challenge, we must refer back to *Chapter III: Current Study* which discusses the methodology employed for this research. Recall that the experiments were held in a quiet, well lit room with two south facing windows. Note also that the blinds were kept closed during experiments to reduce excess lighting. While this setup did, generally speaking, help with controlling the lighting in the room, it did not negate the fact that the computer and eye tracker were set up next to one of the windows, and so on sunnier days or for those experiments done during the morning/afternoon, there was a bit more light in the lab at the time of the experiments. While not drastic, this did have the effect of the left pupil generally having smaller dilation than the right. For the purposes of this research, the amount of light let in by the blinds did not greatly affect the quality of the data, but should this research be replicated, it would be prudent to put more emphasis on controlling the amount of light in the room so as to ensure more accurate and consistent baseline dilation readings.

With regard to the priming effect, this is the unfortunate and inevitable nature of LDT and studies of language processing. On the one hand, if participants are asked to identify words as L1 or L2, should the participant see a word in L1 first, they may be more inclined to flag the next word as L1 as well. Furthermore, if the question was to identify words and non-words, a similar priming effect may be observable, wherein the participant may erroneously mark a word as a non-word simply because the previous token was a non-word. I observed what I can only speculate is a priming effect almost consistently across the Control and Cognate conditions across both language groups. Priming effects are inevitable and are therefore something to account for in any study of this nature.

A further interpretive limitation concerns the nature of the English-French cognate stimuli themselves. Cognates in English and French vary considerably along multiple dimensions, including degree of orthographic overlap (e.g., identical vs. near-identical spelling), phonological similarity, and semantic transparency. In the present study, this variability was not fully controlled for; cognate pairs were selected to be representative of the English-French lexicon, but no systematic manipulation or stratification of cognate type was implemented. As a result, it is difficult to determine whether the observed effects in Group EF are attributable to cognate status per se, or to specific sub-properties of the stimuli—such as particularly high orthographic overlap or salient semantic transparency. Future studies would benefit from explicitly stratifying cognate stimuli by degree of overlap across these dimensions, which would allow for a more fine-grained analysis of which properties of cognates drive the observed pupillary and behavioural effects. This limitation does not invalidate the present findings, but it does suggest that the cognate effect observed here should be interpreted as a broad index of cross-linguistic similarity rather than a precise measure of any single cognate property.

A related limitation concerns the extent to which the stimuli selected for this study are capable of fully capturing the nature of phonological and morphological overlap between language pairs. While the experimental conditions were designed to probe cross-linguistic activation through the use of cognates and non-words constructed to conform to the phonotactic and morphological constraints of each language, the stimuli themselves may not tell the whole story. Specifically, it remains unclear whether the behavioural and pupillometric effects observed here are driven primarily by phonological overlap, morphological overlap, or some interaction of the two; dimensions that were not independently manipulated in the present design. For instance, English-French cognates may share both phonological and morphological properties to varying degrees, and it is therefore difficult to disentangle the contribution of each to the observed effects. Similarly, the non-words used in the Control condition were constructed to approximate the phonotactics of each language, but this process does not guarantee that they are uniformly matched for morphological plausibility. Future research would benefit from independently manipulating phonological and morphological overlap in the stimulus set—for example, by including word pairs that share morphological structure but differ in phonological form, and vice versa—so as to better isolate the specific dimensions of cross-linguistic similarity that drive lexical access and cognitive load in bilingual word recognition tasks.

A further limitation concerns the potential influence of confounds inherent to the cognate stimuli themselves, particularly with respect to the English-French and English-Mandarin language pairs. Foremost among these is the issue of false friends—word pairs that share phonological or orthographic similarity across languages but diverge in meaning (e.g., English *library* and French *librairie*, meaning *bookstore*). While cognate pairs in the present study were selected with semantic transparency in mind, the possibility that some items functioned as false friends for certain participants cannot be ruled out, and no formal measure of perceived semantic overlap was administered. If participants encountered items they interpreted as false friends, this could have introduced additional processing difficulty that was not attributable to genuine cognate facilitation or inhibition, thereby confounding the interpretation of condition-level effects. A related concern applies to the English-Mandarin stimuli: because English-Mandarin cognates are far more limited in number and nature than English-French cognates—owing to the typological and orthographic distance between the two languages—the cognate-like stimuli used for Group EM may not constitute true cognates in the traditional sense, and may instead reflect borrowings, loanwords, or partial phonological correspondences that behave differently from the interlexical homographs and near-homophones typically studied in the cognate processing literature. Future research should implement formal screening procedures to verify semantic transparency and cognate status as perceived by participants, and should carefully distinguish between cognate types—such as true cognates, false friends, and loanwords—so as to ensure that observed effects can be attributed to cross-linguistic similarity rather than stimulus-level confounds.

A final limitation concerns sample size. With 25 participants per group, the present study was adequately powered to detect the large effects observed here; however, smaller or more nuanced effects—such as those associated with individual differences in language dominance or proficiency—may have gone undetected. Future replications with larger samples would strengthen the generalizability of these findings and allow for a more comprehensive examination of individual difference variables in bilingual lexical processing.

Notwithstanding these limitations, the present findings offer a coherent and internally consistent account of bilingual lexical processing across two typologically distinct language pairs, and establish a foundation for more controlled replication.

CHAPTER VI

CONCLUSION

Having examined the theoretical, methodological, and empirical dimensions of this research in the preceding chapters, this conclusion draws together the central findings and situates them within the broader aims of the study. The purpose of this study was to examine bilingual language processing via both behavioural and pupillometric measures. Specifically, the study was designed to address the following research questions:

1. *Can pupillometry serve as a reliable index of cognitive effort in bilingual language discrimination tasks, and does pupillary response to cognate stimuli differ across bilingual groups?*
2. *To what extent does typological similarity between a bilingual's two languages modulate behavioural and physiological indices of lexical processing, as measured by reaction time, accuracy, and pupil dilation?*
3. *To what extent do task demands — operationalized here as the three experimental conditions (NC, CONT, COG) — modulate behavioural and physiological indices of lexical processing, and are these modulations consistent with the task/decision system predictions of the BIA+ model?*

The findings suggest that bilingual language processing varies depending on the degree to which the languages in a pair share linguistic features. From this finding, I suggest that the notion that perhaps language similarity has an inhibitory effect, contrary to the popular belief that it should facilitate processing; that is to say, shared linguistic features impose a greater cognitive load for standard usage during processing, but not enough to fully halt processing. To this point and as previously mentioned, languages that share phonological and morphological features appear to require longer processing time, which is evidenced by the longer RT and larger pupil dilation observed in Group EF participants during this study. This is likely to manifest at the lemma selection phase or lexical level (e.g., the lexical orthography and lexical phonology as illustrated by the BIA+ model) when speakers make the initial contact from lexical level nodes (i.e., lemma selection) to language nodes. In contrast, this same inhibitory effect of language was not observed in Group EM—in fact, the data from the Group EM group suggests little to no overlap of L1 and L2 during processing. Here again, this was demonstrated by the shorter RT and reduced pupil dilation observed in Group EM participants during this study. Crucially, however, the results of this study also demonstrated little effect of language group on accuracy, therefore suggesting language overlap will only affect processing speed and will not hinder a speaker's ability to parse language.

Building on this, and to address the topic of whether pupil dilation can give us a clear indication as to how cognates are processed (i.e., was a cognate facilitation effect observed), the findings from this research are generally in favour of a cognate facilitation effect.

When the RT data are also taken into account, we note that despite participants requiring a marginally longer time to respond in the Cognate condition than in the NC condition, the overall RT for the Cognate condition was still significantly less than that of the Control condition. Couple this with the fact that participants performed with an average accuracy score of 95$\%$ in both language groups, it becomes clear that regardless of language similarities or dissimilarities, cognates facilitate processing more than they inhibit processing. Looking at pupil dilation data alone, we notice that pupil dilation appears to be generally stable across all experimental conditions for both language groups. However, there is consistently less dilation during the Cognate condition (with an average dilation increase of $0.3535mm$ from baseline to stimuli presentation, as opposed to the $0.49mm$ increase for NC condition and $0.425mm$ for the Control condition). The almost $0.1mm$ difference in dilation between the Cognate condition and the NC and Control conditions is enough to suggest that less cognitive activity was required to process language during the Cognate condition. In short and to answer the second question underlying this research, pupil dilation measurements provide a clear indication that less cognitive activity is required to process cognates, suggesting a facilitation effect regardless of language pair.

Lastly, to answer the third and final question, while pupil dilation did not conclusively tell us anything about lexical organization in the bilingual domain, it was instrumental in shedding light on how languages interact during processing. To recap, what the pupil dilation data showed was that languages which are very similar (e.g., Group EF) require more cognitive effort to process. This suggests that there is a high degree of interaction between the L1 and the L2 during lexical level processing. This interaction is mutual, in that the L1 and the L2 will activate in parallel, but we must also account here for language dominance in that the dominant language will have more influence over the other—this point being exemplified by the evidence from the behavioural data. In contrast, pupil dilation data showed that languages which are not similar (e.g., Group EM) require less cognitive effort, therefore suggesting less or no direct interaction between the L1 and the L2. To this end, I had presented an amendment to the BIA+ model which essentially incorporated the Revised Hierarchical Model at the language node. This amendment addresses a key gap in the original BIA+ framework by explicitly representing the asymmetric interaction between the L1 and the L2 at the language node—an interaction modulated by linguistic similarity and language dominance—thereby providing a more theoretically complete account of how bilingual speakers navigate two language systems during lexical access. As I had elaborated upon in *Chapter V: Discussion*—, *Section 5.2*, the purpose for this amendment was to account for and better represent the variation in how language selection occurs once we enter the lemma selection stage. This amendment not only provides a more detailed account of bilingual language processing, but also offers a more realistic representation of how languages are organized in the bilingual mind.

Finally, the findings of the present study serve to further strengthen our current knowledge of language processing. The findings of this study contribute additional information to our current understanding of language processing as they further shed light on the question of cognitive load on processing and effects of language on processing. In conducting this research, we were able to lay a solid foundation for further research in bilingual language processing using pupillometry as a gauge for cognitive load. Taken together, the findings of this study address the research questions outlined in *Chapter III* in that pupillometry was shown to be sensitive to cognate-related processing demands in ways that complement and extend behavioural measures, and bilingual language processing was found to vary systematically as a function of typological similarity, with measurable differences between Group EF and Group EM observed across both behavioural and pupillometric indices. In conclusion, future research of this nature—whether purely in the NLP and psycholinguistics space or purely in the cognitive science space—can and should leverage pupillometry as a method to gauge attention, cognitive activity, and memory.

REFERENCES

[1] T. Dijkstra and W. J. Van Heuven, "The bilingual interactive activation (bia+) model and other models of bilingual word recognition," *Acta Psychologica*, vol. 109, no. 2, pp. 179–196, 2002.

[2] M. Davies, "The corpus of contemporary american english (COCA)," 2008–2020. Available online at `https://www.english-corpora.org/coca/`.

[3] B. New, C. Pallier, L. Ferrand, and R. Matos, "Une base de données lexicales du français contemporain sur Internet: LEXIQUE," *L'Année Psychologique*, vol. 101, no. 3, pp. 447–462, 2001.

[4] R. Speer, J. Chin, and C. Havasi, "Conceptnet 5.5: An open multilingual graph of general knowledge," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 4444–4451, 2018. wordfreq Python library. Data includes SUBTLEX-CH (Cai & Brysbaert, 2010) with permission. Available: `https://github.com/rspeer/wordfreq`.

[5] Q. Cai and M. Brysbaert, "Subtlex-ch: Chinese word and character frequencies based on film subtitles," *PLOS ONE*, vol. 5, no. 6, p. e10729, 2010.

[6] P. Kay, R. Lesser, and M. Coltheart, "The linguistic sign: An empirical investigation," *Cognition*, vol. 42, no. 3, pp. 245–276, 1992.

[7] W. L. Heij, "Selection processes in monolingual and bilingual lexical access," in *Handbook of Bilingualism: Psycholinguistic Approaches* (J. F. Kroll and A. M. B. de Groot, eds.), pp. 289–307, New York: Oxford University Press, 2005.

[8] T. Dijkstra, W. J. B. van Heuven, and J. Grainger, “Simulating cross-language competition with the bilingual interactive activation model,” *Psychologica Belgica*, vol. 38, pp. 177–196, 1998.

[9] P. Fuhrmeister and A. Bürki, “Distributional properties of semantic interference in picture naming: Bayesian meta-analyses,” *Psychonomic Bulletin & Review*, vol. 29, no. 3, pp. 635–647, 2022.

[10] M. Celce-Murcia and E. Olshtain, *Discourse and Context in Language Teaching: A Guide for Language Teachers*. Cambridge: Cambridge University Press, 2001.

[11] T. Roxbury, K. McMahon, and D. Copland, “An fmri study of concreteness effects in spoken word recognition,” *Behavioral and brain functions : BBF*, vol. 10, p. 34, 09 2014.

[12] J. S. Magnuson, D. Mirman, and H. D. Harris, *Computational Models of Spoken Word Recognition*, p. 76–103. Cambridge Handbooks in Psychology, Cambridge University Press, 2012.

[13] L. Jones, L. Wurm, R. Calcaterra, and N. Ofen, “Integrative priming of compositional and locative relations,” *Frontiers in Psychology*, vol. 8, 03 2017.

[14] T. Hannagan, J. Magnuson, and J. Grainger, “Spoken word recognition without a trace,” *Frontiers in psychology*, vol. 4, p. 563, 09 2013.

[15] F. Grosjean and P. Li, *The Psycholinguistics of Bilingualism*. Hoboken, NJ: John Wiley Sons, 2013.

[16] E. Bialystok, “Cognitive complexity and attentional control in the bilingual mind,” *Child Development*, vol. 70, no. 3, pp. 636–644, 2003.

[17] E. Bialystok and M. M. Martin, "Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task," *Developmental Science*, vol. 7, no. 3, pp. 325–339, 2004.

[18] J. B. Morton and S. N. Harper, "What did simon say? revisiting the bilingual advantage," *Developmental Science*, vol. 10, no. 6, pp. 719–726, 2007.

[19] M. van den Noort, E. Struys, P. Bosch, L. Jaswetz, B. Perriard, S. Yeo, P. Barisch, K. Vermeire, S.-H. Lee, and S. Lim, "Does the bilingual advantage in cognitive control exist and if so, what are its modulating factors? a systematic review," *Behavioral Sciences*, vol. 9, no. 3, 2019.

[20] K. R. Paap, R. Anders-Jefferson, L. Mason, K. Alvarado, and B. Zimiga, "Bilingual advantages in inhibition or selective attention: More challenges," *Frontiers in Psychology*, vol. 9, p. 1409, 2018.

[21] K. R. Paap, "Meta-analysis reveals a bilingual advantage that is dependent on task and age," *Frontiers in Psychology*, vol. 11, p. 1458, 2020.

[22] A. T. Ware, M. Kirkovski, and J. A. G. Lum, "Meta-analysis reveals a bilingual advantage that is dependent on task and age," *Frontiers in Psychology*, vol. 11, p. 1458, 2020.

[23] E. Bialystok, F. I. M. Craik, R. Klein, and M. Viswanathan, "Bilingualism, aging, and cognitive control: Evidence from the simon task," *Psychology and Aging*, vol. 19, no. 2, pp. 290–303, 2004.

[24] A. Costa, M. Hernández, and N. Sebastián-Gallés, "Bilingualism aids conflict resolution: Evidence from the ant task," *Cognition*, vol. 106, no. 1, pp. 59–86, 2008.

[25] J. F. Kroll and E. Bialystok, “Understanding the consequences of bilingualism for language processing and cognition,” *Journal of Cognitive Psychology*, vol. 25, no. 5, pp. 497–514, 2013.

[26] T. Dijkstra and W. J. B. van Heuven, “Modeling bilingual word recognition: Past, present and future,” *Bilingualism: Language and Cognition*, vol. 5, no. 3, pp. 219–224, 2002.

[27] W. J. B. van Heuven, T. Dijkstra, and J. Grainger, “Orthographic neighborhood effects in bilingual word recognition,” *Psychologica Belgica*, vol. 38, pp. 177–196, 1998.

[28] V. Marian and M. Spivey, “Competing activation in bilingual language processing: Within- and between-language competition,” *Bilingualism: Language and Cognition*, vol. 6, no. 2, pp. 97–115, 2003.

[29] M. Murujosa, C. Gattei, D. Shalom, and Y. Sevilla, “Cognitive effort during the processing of relative clauses with psychological predicates in spanish: a pupillometric study,” 2020. Poster presented at AMLaP 2020.

[30] E. H. Hess and J. M. Polt, “Pupil size in relation to mental activity during simple problem-solving,” *Science*, vol. 143, no. 3611, pp. 1190–1192, 1964.

[31] M. P. Janisse, “Pupil size and affect: A critical review of the literature since 1960,” *Canadian Psychologist / Psychologie canadienne*, vol. 14, no. 4, pp. 311–329, 1973.

[32] J. C. de Winter, S. M. Petermeijer, L. Kooijman, and D. Dodou, “Replicating five pupillometry studies of eckhard hess,” *International Journal of Psychophysiology*, vol. 165, pp. 145–205, 2021.

[33] U. Weinreich, *Languages in Contact: Findings and Problems*. The Hague: Mouton, 1968.

[34] L. Bloomfield, *Language*. New York: Holt, Rinehart & Winston, 1933.

[35] C. Hoffmann, *Introduction to Bilingualism*. Routledge, 1st ed., 1991.

[36] W. F. Mackey, "The description of bilingualism," *Canadian Journal of Linguistics / Revue canadienne de linguistique*, vol. 7, no. 2, pp. 51–85, 1962.

[37] W. F. Mackey, "A typology of bilingual education," in *Language, Culture, and Cognition: Essays in Honor of Wallace E. Lambert* (P. A. Hornby, ed.), pp. 333–347, Montreal: International Centre for Research on Bilingualism, 1970.

[38] W. Q. Yow and X. Li, "Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: why variations in bilingual experiences matter," *Frontiers in Psychology*, vol. 6, p. 164, 2015.

[39] M. Brysbaert and T. Dijkstra, "Changing views on word recognition in bilinguals," in *Bilingualism and Second Language Acquisition* (J. Morais and G. d'Ydewalle, eds.), pp. 51–85, Royal Academies for Science and the Arts of Belgium, 2006.

[40] S. Montrul, "Dominance and proficiency in early and late bilingualism," in *Language Dominance in Bilinguals: Issues of Measurement and Operationalization* (C. Silva-Corvalán and J. Treffers-Daller, eds.), pp. 15–35, Cambridge University Press, 2015.

[41] A. Caramazza, "How many levels of processing are there in lexical access?," *Cognitive Neuropsychology*, vol. 14, no. 1, pp. 177–208, 1997.

[42] J. K. Bock, "Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation," *Psychological Review*, vol. 89, no. 1, pp. 1–47, 1982.

[43] B. Butterworth, "Lexical access in speech production," in *Lexical Representation and Process* (W. D. Marslen-Wilson, ed.), pp. 240–287, Cambridge, MA: MIT Press, 1989.

[44] D. M. Burke, R. S. MacKay, J. M. Worthley, and R. Wade, "On the tip-of-the-tongue phenomenon: Evidence for a lexical retrieval deficit," *Journal of Memory and Language*, vol. 30, no. 5, pp. 542–559, 1991.

[45] J. K. Bock and W. J. M. Levelt, “Language production: Grammatical encoding,” in *Handbook of Psycholinguistics* (M. A. Gernsbacher, ed.), pp. 945–984, San Diego, CA: Academic Press, 1994.

[46] M. J. Spivey and V. Marian, “Cross talk between native and second languages: Partial activation of an irrelevant lexicon,” *Psychological Science*, vol. 10, no. 3, pp. 281–284, 1999.

[47] X. Wang, B. Hui, and S. Chen, “Language selective or non-selective in bilingual lexical access? it depends on lexical tones!,” *PLoS ONE*, vol. 15, no. 3, p. e0230412, 2020.

[48] T. Dijkstra, H. Van Jaarsveld, and S. Brinke, “Interlingual homograph recognition: Effects of task demands and language intermixing,” *Bilingualism: Language and Cognition*, vol. 1, no. 1, pp. 51–66, 1998.

[49] V. Marian and H. K. Blumenfeld, “Visual word recognition in bilinguals: Evidence for nonselective lexical access,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 33, no. 2, pp. 347–364, 2007.

[50] M. R. Libben and D. A. Titone, “Bilingual lexical access in context: Evidence from eye movements during reading,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 35, no. 2, pp. 381–390, 2009.

[51] H. K. Blumenfeld and V. Marian, “Parallel language activation and cognitive control during spoken word recognition in bilinguals,” *Journal of Cognitive Psychology*, vol. 25, no. 5, pp. 547–567, 2013.

[52] T. Dijkstra and W. J. B. Van Heuven, “The architecture of the bilingual word recognition system: From identification to decision,” *Bilingualism: Language and Cognition*, vol. 5, no. 3, pp. 175–197, 2002.

[53] D. E. Rumelhart and D. A. Norman, "Representation of knowledge and memory," *The Psychology of Learning and Motivation*, vol. 19, pp. 99–135, 1985.

[54] J. F. Kroll and E. Stewart, "Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations," *Journal of Memory and Language*, vol. 33, no. 2, pp. 149–174, 1994.

[55] J. F. Kroll and A. M. B. de Groot, "Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages," in *Tutorials in Bilingualism: Psycholinguistic Perspectives* (A. M. B. de Groot and J. F. Kroll, eds.), pp. 169–199, Hillsdale, NJ: Erlbaum, 1997.

[56] M. Finkbeiner, K. I. Forster, J. Nicol, and K. Nakamura, "The role of polysemy in masked semantic and translation priming," *Journal of Memory and Language*, vol. 51, no. 1, pp. 1–22, 2004.

[57] A. M. B. de Groot, "Bilingual lexical representation: A closer look at conceptual representations," in *Orthography, Phonology, Morphology, and Meaning* (R. Frost and L. Katz, eds.), pp. 389–412, Amsterdam: Elsevier, 1992.

[58] D. W. Green, "Mental control of the bilingual lexico-semantic system," *Bilingualism: Language and Cognition*, vol. 1, no. 2, pp. 67–81, 1998.

[59] P. A. Starreveld, A. M. B. de Groot, B. M. M. Rossmark, and J. G. van Hell, "Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context," *Bilingualism: Language and Cognition*, vol. 17, no. 2, pp. 258–276, 2013.

[60] L. W. Barsalou, "Context-independent and context-dependent information in concepts," *Memory & Cognition*, vol. 10, no. 1, pp. 82–93, 1982.

[61] T. Dijkstra, W. J. B. van Heuven, and J. Grainger, "Recognition of cognates and interlingual homographs: The neglected role of phonology," *Journal of Memory and Language*, vol. 41, no. 4, pp. 496–518, 1999.

[62] T. H. Gollan, K. I. Forster, and R. Frost, "Translation priming with different scripts: Masked priming with cognates and noncognates in hebrew-english bilinguals," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 23, no. 5, pp. 1122–1139, 1997.

[63] A. I. Schwartz, J. F. Kroll, and M. Diaz, "Reading words in spanish and english: Mapping orthography to phonology in two languages," *Language and Cognitive Processes*, vol. 22, no. 4, pp. 449–468, 2007.

[64] J. G. V. Hell and T. Dijkstra, "Foreign language knowledge can influence native language performance in exclusively native contexts," *Psychonomic Bulletin & Review*, vol. 9, no. 4, pp. 780–789, 2002.

[65] G. J. Poarch and J. G. Van Hell, "Cross-language activation in children's speech production: Evidence from cognates," *Bilingualism: Language and Cognition*, vol. 15, no. 2, pp. 329–344, 2012.

[66] P. Brenders, J. G. Van Hell, and T. Dijkstra, "Cognate effects in bilingual visual word recognition: The role of morphological structure," *Bilingualism: Language and Cognition*, vol. 14, no. 2, pp. 196–205, 2011.

[67] P. Cristoffanini, K. Kirsner, and D. Milech, "Bilingual lexical representation: The status of spanish–english cognates," *Quarterly Journal of Experimental Psychology*, vol. 38A, no. 3, pp. 367–393, 1986.

[68] A. Costa, M. Santesteban, and A. Caño, "Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals," *Journal of Memory and Language*, vol. 43, no. 4, pp. 527–546, 2000.

[69] K. Lemhöfer and T. Dijkstra, "Recognizing cognates and interlingual homographs: Effects of code similarity in bilingual word recognition," *Journal of Memory and Language*, vol. 51, no. 2, pp. 118–147, 2004.

[70] A. M. De Groot and G. L. Nas, "Cross-language differences in semantic representation," *Bilingualism: Language and Cognition*, vol. 33, no. 5, pp. 439–472, 1991.

[71] J. G. Van Hell and T. Dijkstra, "Sentence context modulates lexical competition in bilingual word recognition," *Bilingualism: Language and Cognition*, vol. 5, no. 2, pp. 95–116, 2002.

[72] K. J. Midgley, P. J. Holcomb, and J. Grainger, "Early parallel activation of lexical representations in bilinguals: Evidence from electrophysiology (erp)," *Brain Research*, vol. 1187, pp. 150–158, 2008.

[73] D. W. Green and J. Abutalebi, "Language control in bilinguals: The adaptive control hypothesis," *Journal of Cognitive Psychology*, vol. 25, no. 5, pp. 515–530, 2013.

[74] W. Duyck, E. Van Assche, D. Drieghe, and R. J. Hartsuiker, "Visual word recognition in bilinguals: Evidence from eye movements during reading," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 33, no. 4, pp. 673–690, 2007.

[75] T. Degani and N. Tokowicz, "Learning mechanisms in the bilingual lexicon: Evidence from translation priming with semantically ambiguous words," *Bilingualism: Language and Cognition*, vol. 13, no. 1, pp. 81–98, 2010.

[76] S. E. Kuchinsky, J. B. Ahlstrom, K. I. Vaden, S. L. Cute, L. E. Humes, J. R. Dubno, and M. A. Eckert, “Pupil size varies with word listening and response selection difficulty in older adults with hearing loss,” *Psychophysiology*, vol. 50, no. 1, pp. 23–34, 2013.

[77] R. R. Rosinski, “Picture-word interference is semantically based.,” *Child Development*, vol. 48, pp. 643–647, 1977.

[78] P. A. Starreveld and W. La Heij, “Semantic interference, orthographic facilitation, and their interaction in naming tasks,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, no. 3, pp. 686–698, 1995.

[79] T. T. Schnur, M. F. Schwartz, A. Brecher, and C. Hodgson, “Semantic interference during blocked-cyclic naming: Evidence from aphasia,” *Journal of Memory and Language*, vol. 54, no. 2, pp. 199–227, 2006.

[80] V. Piai, A. Roelofs, and H. Schriefers, “Task choice and semantic interference in picture naming,” *Acta Psychologica*, vol. 157, pp. 13–22, 2015.

[81] P. Starreveld and W. La Heij, “Picture-word interference is a stroop effect: A theoretical analysis and new empirical findings,” *Psychonomic Bulletin & Review*, vol. 24, pp. 721–733, Jun 2017.

[82] C. Mulatti and M. Coltheart, “Picture-word interference and the response exclusion hypothesis,” *Cortex; a journal devoted to the study of the nervous system and behavior*, vol. 48, pp. 363–72, 05 2011.

[83] B. Z. Mahon, F. E. Garcea, and E. Navarrete, “Picture-word interference and the response-exclusion hypothesis: a response to mulatti and coltheart,” *Cortex*, vol. 48, pp. 373–377, Mar. 2012. Epub 2011 Oct 29.

[84] M. L. Hall, “Bilingual picture-word studies constrain theories of lexical selection,” *Frontiers in Psychology*, vol. 2, p. 381, Dec. 2011. Published online 2011 Dec 29.

[85] S. Aristei, P. Zwitserlood, and R. Abdel Rahman, “Picture-induced semantic interference reflects lexical competition during object naming,” *Frontiers in Psychology*, vol. 3, p. 28, Feb. 2012. Erratum in: Front Psychol. 2017 Jun 12;8:1002. doi: 10.3389/fpsyg.2017.01002.

[86] M. Lowry, C. Dubé, and E. Schotter, “Evaluating theories of bilingual language control using computational models,” *Journal of Memory and Language*, vol. 117, p. 104195, 2021.

[87] Y. J. Wu and G. Thierry, “Event-related brain potential investigation of preparation for speech production in late bilinguals,” *Frontiers in Psychology*, vol. 2, p. 114, May 2011.

[88] A. Melinger, “Distinguishing languages from dialects: A litmus test using the picture-word interference task,” *Cognition*, vol. 172, pp. 73–88, March 2018. Epub 2017 Dec 9.

[89] G. Morini and R. S. Newman, “Monolingual and bilingual word recognition and word learning in background noise,” *Language and Speech*, vol. 63, no. 2, pp. 381–403, 2020. Originally published online in 2019.

[90] V. Marian, H. K. Blumenfeld, and O. V. Boukrina, “Sensitivity to phonological similarity within and across languages,” *Journal of Psycholinguistic Research*, vol. 37, no. 3, pp. 141–170, 2008.

[91] E. Lagrou, R. Hartsuiker, and W. Duyck, “Knowledge of a second language influences auditory word recognition in the native language,” *Journal of experimental psychology. Learning, memory, and cognition*, vol. 37, pp. 952–65, 04 2011.

[92] W. D. Marslen-Wilson and A. Welsh, “Processing interactions and lexical access during word recognition in continuous speech,” *Cognitive Psychology*, vol. 10, no. 1, pp. 29–63, 1978.

[93] S. Goldinger, "Auditory lexical decision," *Language and Cognitive Processes*, vol. 11, pp. 559–567, 12 1996.

[94] H. K. Blumenfeld, S. R. Schroeder, S. C. Bobb, M. R. Freeman, and V. Marian, "Auditory word recognition across the lifespan: Links between linguistic and nonlinguistic inhibitory control in bilinguals and monolinguals," *Linguistic Approaches to Bilingualism*, vol. 6, no. 1-2, pp. 119–146, 2016. Epub 2016 Feb 19.

[95] D. E. Meyer and R. W. Schvaneveldt, "Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations," *Journal of Experimental Psychology*, vol. 90, no. 2, pp. 227–234, 1971.

[96] J. F. Kroll, S. C. Bobb, and N. Hoshino, "Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain," *Current Directions in Psychological Science*, vol. 23, no. 3, pp. 159–163, 2014.

[97] J. A. Linck, N. Hoshino, and J. F. Kroll, "Cross-language lexical processes and inhibitory control," *Mental Lexicon*, vol. 3, no. 3, pp. 349–374, 2008.

[98] M. Nakayama, K. Ida, and S. J. Lupker, "Cross-script l2-l1 noncognate translation priming in lexical decision depends on l2 proficiency: Evidence from japanese–english bilinguals," *Bilingualism: Language and Cognition*, vol. 19, no. 5, pp. 1001–1022, 2016.

[99] R. A. López Zunini, M. Baart, A. G. Samuel, and B. C. Armstrong, "Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures," *Neuropsychologia*, vol. 137, p. 107305, 2020. Epub 2019 Dec 13.

[100] C. Rojas, B. Riffo, M. San Martín, and E. Guerra, "Lexical access restrictions after the age of 80," *Brain Sciences*, vol. 13, no. 9, p. 1343, 2023.

[101] D. C. Friesen, A. Chung-Fat-Yim, and E. Bialystok, "Lexical selection differences between monolingual and bilingual listeners," *Brain and Language*, vol. 152, pp. 1–13, 2016. Epub 2015 Dec 9.

[102] E. Bialystok, F. Craik, and G. Luk, "Cognitive control and lexical access in younger and older bilinguals," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 34, no. 4, pp. 859–873, 2008. Erratum in: J Exp Psychol Learn Mem Cogn. 2009 May;35(3):828.

[103] A. Nikolaev, E. Higby, J. Hyun, and S. Ashaie, "Lexical decision task for studying written word recognition in adults with and without dementia or mild cognitive impairment," *Journal of Visualized Experiments*, 06 2019.

[104] M. J. Yap, D. E. Sibley, D. A. Balota, R. Ratcliff, and J. Rueckl, "Responding to nonwords in the lexical decision task: Insights from the english lexicon project," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 41, no. 3, pp. 597–613, 2015. Epub 2014 Oct 20.

[105] H. S. Berberyan, H. van Rijn, and J. P. Borst, "Discovering the brain stages of lexical decision: Behavioral effects originate from a single neural decision process," *Brain and Cognition*, vol. 153, p. 105786, 2021. Epub 2021 Aug 9.

[106] L. Ferrand, M. Brysbaert, E. Keuleers, B. New, P. Bonin, A. Méot, M. Augustinova, and C. Pallier, "Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from chronolex," *Frontiers in Psychology*, vol. 2, p. 306, 2011. Published 2011 Nov 1.

[107] R. E. von Studnitz and D. W. Green, “Lexical decision and language switching,” *International Journal of Bilingualism*, vol. 1, no. 1, pp. 3–24, 1997. Original work published 1997.

[108] M. Broersma, D. Carter, K. Donnelly, and A. Konopka, “Triggered codeswitching: Lexical processing and conversational dynamics,” *Bilingualism: Language and Cognition*, vol. 23, no. 2, pp. 295–308, 2020.

[109] C. Soares and F. Grosjean, “Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access,” *Memory & Cognition*, vol. 12, no. 4, pp. 380–386, 1984.

[110] J. L. McClelland and J. L. Elman, “The trace model of speech perception,” *Cognitive Psychology*, vol. 18, no. 1, pp. 1–86, 1986.

[111] R. M. Warren, “Perceptual restoration of missing speech sounds,” *Science*, vol. 167, no. 3917, pp. 392–393, 1970.

[112] A. Colomé, “Lexical activation in bilinguals’ speech production: Language-specific or language-independent?,” *Journal of Memory and Language*, vol. 45, no. 4, pp. 721–736, 2001.

[113] W. J. B. Van Heuven and T. Dijkstra, “Cognate effects in lexical decision and semantic priming: A review of the literature,” *Bilingualism: Language and Cognition*, vol. 13, no. 3, pp. 42–50, 2010.

[114] E. I. Alshanskaia, G. V. Portnova, K. Liaukovich, and O. V. Martynova, “Pupillometry and autonomic nervous system responses to cognitive load and false feedback: an unsupervised machine learning approach,” *Frontiers in Neuroscience*, vol. 18, p. 1445697, 2024.

[115] I. E. Loewenfeld, *The Pupil: Anatomy, Physiology, and Clinical Applications*. Iowa State University Press, 1993. Two volumes.

[116] H. M. Simpson and S. M. Hale, "Pupillary changes during a decision-making task," *Perceptual and Motor Skills*, vol. 29, no. 3, pp. 1111–1114, 1969.

[117] D. Kahneman and J. Beatty, "Pupil diameter and load on memory," *Science*, vol. 154, no. 3756, pp. 1583–1585, 1966.

[118] A. Kafkas and D. Montaldi, "Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity," *The Quarterly Journal of Experimental Psychology*, vol. 64, no. 10, pp. 1971–1989, 2011.

[119] H. E. Fitzgerald, "Autonomic pupillary reflex activity during early infancy and its relation to social and nonsocial visual stimuli," *Journal of Experimental Child Psychology*, vol. 6, no. 3, pp. 470–482, 1968.

[120] D. Kahneman and A. Treisman, "Pupil diameter and load on memory," *Science*, vol. 165, no. 3897, pp. 1123–1124, 1969.

[121] E. H. Hess and J. M. Polt, "Pupil size as related to interest value of visual stimuli," *Science*, vol. 132, no. 3436, pp. 349–350, 1960.

[122] R. P. R. D. van der Wel and H. van Steenbergen, "Pupillometry as a tool to study the link between mental effort and cognitive control," *Psychophysiology*, vol. 55, no. 3, p. e13039, 2018.

[123] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Personality and Social Psychology*, vol. 61, no. 5, pp. 210–220, 2000.

[124] S. E. Samuels and J. D. Hamilton, "The effect of cognitive load on pupil diameter in a memory retrieval task," *Cognitive Psychology*, vol. 22, no. 2, pp. 242–257, 1990.

[125] J. Beatty, “Phasic pupil response and the operationalization of mental workload,” *Human Factors*, vol. 24, no. 5, pp. 631–642, 1982.

[126] M. W. O’Boyle and D. M. O’Hara, “The effect of cognitive workload on pupil diameter during visual search,” *Perception Psychophysics*, vol. 51, no. 2, pp. 177–182, 1992.

[127] J. Beatty and B. Lucero-Wagoner, “The pupillary system,” in *Handbook of Psychophysiology* (J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, eds.), pp. 142–162, Cambridge University Press, 2 ed., 2000.

[128] J. Schmidtke, “Pupillometry in linguistic research: An introduction and review for second language researchers,” *Studies in Second Language Acquisition*, vol. 40, no. 3, pp. 529–549, 2017.

[129] J. Beatty, “Task-evoked pupillary responses, processing load, and the structure of processing resources,” *Psychological Bulletin*, vol. 91, no. 2, pp. 276–292, 1982.

[130] M. A. Just and P. A. Carpenter, “The intensity dimension of thought: Pupillometric indices of sentence processing,” *Canadian Journal of Experimental Psychology*, vol. 47, no. 2, pp. 310–339, 1993.

[131] C. Scheepers, M. D. Demberg, and R. K. K. Keller, “Pupillometry in linguistic processing: A review,” *Journal of Psycholinguistic Research*, vol. 42, no. 6, pp. 519–531, 2013.

[132] J. Sauppe, “Pupillometry in language processing: A review,” *Language and Linguistics Compass*, vol. 11, no. 5, p. e12216, 2017.

[133] J. Tommola and J. Hyönä, “Pupillary responses during processing of spoken and written target languages,” *Journal of Psycholinguistic Research*, vol. 19, no. 6, pp. 485–494, 1990.

[134] P. E. Engelhardt, F. Ferreira, and E. G. Patsenko, “Pupillometry reveals processing load during spoken sentence comprehension,” *The Quarterly Journal of Experimental Psychology*, vol. 63, no. 4, pp. 639–645, 2010.

[135] J. M. Schneider, A. F. Healy, and L. E. Bourne, “Pupillometric evidence for the use of executive control in language switching,” *Bilingualism: Language and Cognition*, vol. 19, no. 5, pp. 1029–1044, 2016.

[136] T. Brothers, T. Y. Swaab, and M. J. Traxler, “Goals and strategies influence lexical prediction during sentence comprehension,” *Journal of Memory and Language*, vol. 93, pp. 203–216, 2017.

[137] B. R. Payne and E. A. L. Stine-Morrow, “Aging, parafoveal preview, and semantic integration in sentence processing: Testing the cognitive workload of wrap-up,” *Psychology and Aging*, vol. 27, no. 3, pp. 638–649, 2012.

[138] A. Costa, A. Caramazza, and N. Sebastian-Galles, “The cognate facilitation effect: Implications for models of lexical access,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 26, pp. 1283–1296, Sep 2000.

[139] P. Li, N. Jiang, and K. Gor, “Lexical processing in second language learners and bilinguals: The roles of lexical representation and processing mechanisms,” *Bilingualism: Language and Cognition*, vol. 20, no. 4, pp. 682–692, 2017.

[140] J. Schmidtke, “Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry,” *Frontiers in Psychology*, vol. 5, p. 137, 2014.

[141] H. Ershaid, D. McLaughlin, and M. Lallier, “Highly proficient bilinguals compensate language dominance effects with differential attentional resource allocation: Insights from pupillometry,” *Bilingualism: Language and Cognition*, 2025.

[142] Y. Yao, K. Connell, and S. Politzer-Ahles, "Hearing emotion in two languages: A pupillometry study of cantonese–mandarin bilinguals' perception of affective cognates in l1 and l2," *Bilingualism: Language and Cognition*, vol. 26, no. 4, pp. 795–808, 2023.

[143] J. A. E. Anderson, L. Mak, A. Keyvani Chahi, and E. Bialystok, "The language and social background questionnaire: Assessing degree of bilingualism in a diverse population," *Behavior Research Methods*, vol. 50, no. 1, pp. 250–263, 2018.

[144] E. Dalmaijer, S. Mathôt, and S. Van der Stigchel, "Pygaze: an open-source, cross-platform toolbox for minimal-effort programming of eye tracking experiments. behavior research methods," *Handbook of Clinical Neurology*, vol. 46, pp. 913–921, 2014.

[145] M. Brysbaert and B. New, "Moving beyond kucera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english," *Behavior Research Methods*, vol. 41, pp. 977–990, 2009.

[146] B. New, C. Pallier, M. Brysbaert, and L. Ferrand, "Lexique 2: A new french lexical database," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 3, pp. 516–524, 2004.

[147] S. Mathôt, "Pupillometry: Psychology, physiology, and function," *Journal of Cognition*, vol. 1, no. 1, p. 16, 2018.

[148] R. S. Hoedemaker and P. C. Gordon, "The onset and time course of semantic priming during rapid recognition of visual words," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 43, pp. 881–902, May 2017.

[149] D. A. Balota, M. J. Yap, M. J. Cortese, and J. M. Watson, "Beyond mean response latency: Response time distributional analyses of semantic priming," *Journal of Memory and Language*, vol. 59, no. 5, pp. 495–523, 2008.

[150] S. Kinoshita, B. de Wit, M. Aji, and D. Norris, "Evidence accumulation in the integrated and primed stroop tasks," *Memory & Cognition*, vol. 45, no. 5, pp. 824–836, 2017.

[151] N. V. Valtakari, I. T. C. Hooge, C. Viktorsson, P. Nyström, T. Falck-Ytter, and R. S. Hessels, "Eye tracking in human interaction: Possibilities and limitations," *Behavior Research Methods*, vol. 53, pp. 1592–1608, Aug 2021.

[152] H. C. Cuve, J. Stojanov, X. Roberts-Gaal, C. Catmur, and G. Bird, "Validation of gazepoint low-cost eye-tracking and psychophysiology bundle," *Behavior Research Methods*, vol. 54, pp. 1027–1049, Apr 2022.

[153] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[154] S. Mathôt and A. Vilotijević, "Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis," *bioRxiv*, 2022. Review article noting that pupil dilation typically peaks 1 s after stimulus onset.

[155] J. Hyönä, J. Tommola, and A.-M. Alaja, "Pupillary response as an indicator of processing load in reading: Evidence from eye movements," *Quarterly Journal of Experimental Psychology*, vol. 48A, no. 3, pp. 598–612, 1995.

[156] E. D. Poort and J. M. Rodd, "The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition," *Acta Psychologica*, vol. 180, pp. 52–63, 2017.

[157] M. Sherkina-Lieber, "The cognate facilitation effect in bilingual speech processing: The case of russian-english bilingualism," *Cahiers linguistics d'Ottawa*, vol. 32, pp. 108–121, 2004.

[158] E. D. Poort and J. M. Rodd, "The role of cognates in bilingual lexical processing: List composition effects in language decision," *Cognition*, vol. 164, pp. 1–13, 2017.

[159] J. G. Van Hell and T. Dijkstra, "Foreign language dominance and lexical access: Evidence from cognate and false friend processing in bilinguals," *Bilingualism: Language and Cognition*, vol. 5, no. 3, pp. 203–222, 2002.

Appendix A

PARTICIPANT PROFILE

Table A.1: Summary table of participant profile for Group EF, including age at the date of participation, country of birth, Speaker status with regard to French (i.e., Native vs. Non-Native speaker) and AoA.[6] To protect participant identity and in compliance with York University's Research Ethics Board (REB) guidelines, all participants are identified with a code EF, for language group, followed by their participant number (1-25).

Participant	**Age**	**Country of Birth**	**Speaker Status**	**AoA**
EF1	23	Canada	Native	-
EF2	18	Canada	Heritage	4
EF3	19	Canada	Heritage	6
EF4	19	Ivory Coast	Native	-
EF5	24	Burundi	Native	-
EF6	20	Canada	Native	-
EF7	25	Canada	Native	-
EF8	19	Chile	Native	-
EF9	17	Mauritius	Native	-
EF10	-	Canada	Native	-
EF12	19	Lebanon	Native	-
EF13	18	Canada	Heritage	5
EF14	25	Canada	Native	-
EF15	34	Belgium	Heritage	6
EF16	20	Canada	Native	-
EF17	18	Canada	Native	-
EF18	18	Canada	Heritage	12
EF19	27	Canada	Native	-
EF20	30	Canada	Native	-
EF21	19	Canada	Heritage	8
EF22	26	Canada	Native	-
EF23	27	Canada	Native	-
EF24	25	Canada	Native	-

[6] Note that based questionnaire responses and follow-up screening questions, the AoA listed by participants did not necessarily correspond to age at which they were exposed to French. For instance, with participant EF18, they highlighted that they grew up in a francophone environment but did not really ”feel fluent” until they were about 12 years of age.

Table A.2: Summary table of participant profile for Group EM, including age at the date of participation, country of birth, Speaker status with regard to Mandarin (i.e., Native vs. Non-Native speaker) and AoA.[7] To protect participant identity and in compliance with York University's Research Ethics Board (REB) guidelines, all participants are identified with a code EM, for language group, followed by their participant number (1-25).

Participant	Age	Country of Birth	Speaker Status	AoA
EM1	19	Malaysia	Native	-
EM2	18	China	Native	-
EM3	21	China	Native	-
EM4	23	China	Native	-
EM5	26	Hong Kong	Native	-
EM6	25	China	Native	-
EM7	29	China	Native	-
EM8	-	China	Native	-
EM9	31	China	Native	-
EM10	20	Canada	Heritage	3
EM12	27	Canada	Heritage	6
EM13	26	China	Native	-
EM14	23	China	Native	-
EM15	22	China	Native	-
EM16	20	China	Native	-
EM17	30	China	Native	-
EM18	25	China	Native	-
EM19	-	China	Native	-
EM20	24	Taiwan	Native	-
EM21	27	China	Native	-
EM22	60	China	Native	-
EM23	19	Hong Kong	Heritage	3
EM24	20	China	Native	-
EM25	21	China	Native	-

[7] Note that based questionnaire responses and follow-up screening questions, the AoA listed by participants did not necessarily correspond to age at which they were exposed to Mandarin. For the self-reported heritage speakers, this age often corresponded with when they started attending Mandarin language classes or when they travelled back to Asia to be with family.

Appendix B

DATA PREPARATION: METHODS AND NOTES

Data Cleaning

Recalling *Section 3.4.2.2: Coding* in *Chapter III: Current Study*, it was established that for each task condition, two result files would be generated for each task; RT, response and token timestamps would be captured into a single excel spreadsheet (see Figure B.1), and all eye-tracking data would be saved into a CSV. In total, each participant generated 6 files: one LDT response Excel and one debug CSV per condition multiplied by 3 test conditions.

	Stimulus	Response	ReactionTime(ms)	Timestamp
CrossTimestamp	ABEILLE	RIGHT	727	2024-10-30 11:52:18.244326
WordTimestamp	ABEILLE	RIGHT	727	2024-10-30 11:52:28.248135
KeyInputTimestamp	ABEILLE	RIGHT	727	2024-10-30 11:52:28.980319
CrossTimestamp	AUTILLER	RIGHT	2072	2024-10-30 11:47:35.152411
WordTimestamp	AUTILLER	RIGHT	2072	2024-10-30 11:47:45.155309
KeyInputTimestamp	AUTILLER	RIGHT	2072	2024-10-30 11:47:47.233017
CrossTimestamp	AVOIR	RIGHT	839	2024-10-30 11:47:13.651165
WordTimestamp	AVOIR	RIGHT	839	2024-10-30 11:47:23.654377
KeyInputTimestamp	AVOIR	RIGHT	839	2024-10-30 11:47:24.498235
CrossTimestamp	BARBE	RIGHT	963	2024-10-30 11:50:50.377188
WordTimestamp	BARBE	RIGHT	963	2024-10-30 11:51:00.380955
KeyInputTimestamp	BARBE	RIGHT	963	2024-10-30 11:51:01.348221
CrossTimestamp	BE	LEFT	696	2024-10-30 11:47:47.233017
WordTimestamp	BE	LEFT	696	2024-10-30 11:47:57.234703
KeyInputTimestamp	BE	LEFT	696	2024-10-30 11:47:57.934591
CrossTimestamp	BEC	LEFT	713	2024-10-30 11:53:57.983043
WordTimestamp	BEC	LEFT	713	2024-10-30 11:54:07.988052
KeyInputTimestamp	BEC	LEFT	713	2024-10-30 11:54:08.704744
CrossTimestamp	BERGER	RIGHT	1321	2024-10-30 11:45:59.186556
WordTimestamp	BERGER	RIGHT	1321	2024-10-30 11:46:09.190471
KeyInputTimestamp	BERGER	RIGHT	1321	2024-10-30 11:46:10.516471
CrossTimestamp	BESOIN	RIGHT	889	2024-10-30 11:54:40.381572
WordTimestamp	BESOIN	RIGHT	889	2024-10-30 11:54:50.385691

Figure B.1: Language Discrimination Task Response data as captured by the LDT task codes. This particular example comes from a Control task for Group EF. From left to right, we see the three timestamp conditions, the token, the response, the reaction time and the timestamp.

To avoid confusion and to maintain a comprehensive filing system, all files were renamed as follows: ***[Participant Initials]_[Language Group]_LDT [test condition]_[Date]_[Researcher Initials]*** for the LDT response excel and ***[Participant Initials]_[Language Group]_LDT [test condition]_[DEBUG]_[Date]_[Researcher Initials]*** for the Debug CSV file.

Crucially, the eye-tracking data and response data are captured in a separate Excel file. In order to consolidate the data into a single working file for each test condition, three additional scripts were written in Python to aid in the cleanup and review of said data: a Timestamps modification script, a Debug extraction script, and a final extraction script.

The purpose of the Timestamp modification script was to take the LDT response spreadsheet and extract each time stamp so we could separate the date from the hour, the minutes, and the seconds (see Figure B.2). After running the script, the changes would automatically populate into the original LDT response file (see Figure B.3).

```
    # Process the 'Timestamp' column
    timestamps = df['Timestamp']
    dates, hours, minutes, seconds = process_timestamp(timestamps)

    # Add new columns to the DataFrame
    df['date'] = dates
    df['hour'] = hours
    df['minutes'] = minutes
    df['seconds'] = seconds

    # Save the DataFrame back to the same Excel file
    try:
        df.to_excel(input_file, index=False)
        print(f"Processed data saved to {input_file}")
    except Exception as e:
        print(f"Error: {e}")
```

Figure B.2: Snippet of Timestamp Modification python script

This first script worked by first identifying the *Timestamp* column in the original LDT response file, then created four (4) new columns for *date*, *hour*, *minutes*, and *seconds*. It would then return to the *Timestamp* column and extract the corresponding information into the specified column.

Innamed:	Stimulus	Response	ctionTime(	Timestamp	date	hour	minutes	seconds	minutes	seconds
CrossTime	ABEILLE	RIGHT	727	2024-10-30 11:52:18.244326	2024-10-30	11	52	18.244326	52	18.24433
WordTime	ABEILLE	RIGHT	727	2024-10-30 11:52:28.248135	2024-10-30	11	52	28.248135	52	28.24814
KeyInputT	ABEILLE	RIGHT	727	2024-10-30 11:52:28.980319	2024-10-30	11	52	28.980319	52	28.98032
CrossTime	AUTILLER	RIGHT	2072	2024-10-30 11:47:35.152411	2024-10-30	11	47	35.152411	47	35.15241
WordTime	AUTILLER	RIGHT	2072	2024-10-30 11:47:45.155309	2024-10-30	11	47	45.155309	47	45.15531
KeyInputT	AUTILLER	RIGHT	2072	2024-10-30 11:47:47.233017	2024-10-30	11	47	47.233017	47	47.23302
CrossTime	AVOIR	RIGHT	839	2024-10-30 11:47:13.651165	2024-10-30	11	47	13.651165	47	13.65117
WordTime	AVOIR	RIGHT	839	2024-10-30 11:47:23.654377	2024-10-30	11	47	23.654377	47	23.65438
KeyInputT	AVOIR	RIGHT	839	2024-10-30 11:47:24.498235	2024-10-30	11	47	24.498235	47	24.49824
CrossTime	BARBE	RIGHT	963	2024-10-30 11:50:50.377188	2024-10-30	11	50	50.377188	50	50.37719
WordTime	BARBE	RIGHT	963	2024-10-30 11:51:00.380955	2024-10-30	11	51	0.380955	51	0.380955
KeyInputT	BARBE	RIGHT	963	2024-10-30 11:51:01.348221	2024-10-30	11	51	1.348221	51	1.348221
CrossTime	BE	LEFT	696	2024-10-30 11:47:47.233017	2024-10-30	11	47	47.233017	47	47.23302
WordTime	BE	LEFT	696	2024-10-30 11:47:57.234703	2024-10-30	11	47	57.234703	47	57.2347
KeyInputT	BE	LEFT	696	2024-10-30 11:47:57.934591	2024-10-30	11	47	57.934591	47	57.93459
CrossTime	BEC	LEFT	713	2024-10-30 11:53:57.983043	2024-10-30	11	53	57.983043	53	57.98304
WordTime	BEC	LEFT	713	2024-10-30 11:54:07.988052	2024-10-30	11	54	7.988052	54	7.988052
KeyInputT	BEC	LEFT	713	2024-10-30 11:54:08.704744	2024-10-30	11	54	8.704744	54	8.704744
CrossTime	BERGER	RIGHT	1321	2024-10-30 11:45:59.186556	2024-10-30	11	45	59.186556	45	59.18656
WordTime	BERGER	RIGHT	1321	2024-10-30 11:46:09.190471	2024-10-30	11	46	9.190471	46	9.190471
KeyInputT	BERGER	RIGHT	1321	2024-10-30 11:46:10.516471	2024-10-30	11	46	10.516471	46	10.51647
CrossTime	BESOIN	RIGHT	889	2024-10-30 11:54:40.381572	2024-10-30	11	54	40.381572	54	40.38157
WordTime	BESOIN	RIGHT	889	2024-10-30 11:54:50.385691	2024-10-30	11	54	50.385691	54	50.38569
KeyInputT	BESOIN	RIGHT	889	2024-10-30 11:54:51.279483	2024-10-30	11	54	51.279483	54	51.27948

Figure B.3: Language Discrimination Task Response data from the same LDT Group EF Control task after running the Timestamp modification script.

The second script was designed to extract the debug file such that specific data points were easy to recall for final analysis. As seen in figure B.4, what the Opengaze script generates is a full log of the eyetracker data from the Gazepoint GP3, which recorded data at 17msec intervals for the entire duration of the task.

```
11:44:37.024159: socket recv timeout
11:44:37.025166: Outqueue sent: '<SET ID="ENABLE_SEND_DATA" STATE="1" />\r\n'
11:44:37.027390: Raw instring: b'<ACK ID="ENABLE_SEND_DATA" STATE="1" />\r\n'
11:44:37.027390: Incoming: '<ACK ID="ENABLE_SEND_DATA" STATE="1" />'
11:44:37.027390: Raw instring: b'<REC CNT="61172" TIME="1009.69946" TIME_TICK="786600180172" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0"
RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431" BPOGV="0" LPCX="0.37550" LPCY="0.65798" LPD="12.08470" LPS="3.04082" LPV="1" RPCX="0.68598" RPCY="0.68257" RPD="9.63714" RPS="3.04082" RPV="1" LEYEX="-0.03750"
LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00418" LPUPILV="1" REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00408" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />\r\n'
11:44:37.027390: Incoming: '<REC CNT="61172" TIME="1009.69946" TIME_TICK="786600180172" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0"
RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431" BPOGV="0" LPCX="0.37550" LPCY="0.65798" LPD="12.08470" LPS="3.04082" LPV="1" RPCX="0.68598" RPCY="0.68257" RPD="9.63714" RPS="3.04082" RPV="1" LEYEX="-0.03750"
LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00418" LPUPILV="1" REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00408" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />'
11:44:37.028397: Outqueue acknowledged: '<SET ID="ENABLE_SEND_DATA" STATE="1" />\r\n'
11:44:37.060258: Raw instring: b'<REC CNT="61173" TIME="1009.71619" TIME_TICK="786600347177" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0"
RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431" BPOGV="0" LPCX="0.37478" LPCY="0.65959" LPD="11.85370" LPS="3.04082" LPV="1" RPCX="0.68511" RPCY="0.68360" RPD="9.95932" RPS="3.04082" RPV="1" LEYEX="-0.03750"
LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00416" LPUPILV="1" REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00426" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />\r\n<REC CNT="61174" TIME="1009.73285"
TIME_TICK="786600513876" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0" RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431"
BPOGV="0" LPCX="0.37253" LPCY="0.66185" LPD="12.32223" LPS="3.04082" LPV="1" RPCX="0.68350" RPCY="0.68501" RPD="10.21000" RPS="3.04082" RPV="1" LEYEX="-0.03750" LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00411" LPUPILV="1"
REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00416" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />\r\n'
11:44:37.060258: Incoming: '<REC CNT="61173" TIME="1009.71619" TIME_TICK="786600347177" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0"
RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431" BPOGV="0" LPCX="0.37478" LPCY="0.65959" LPD="11.85370" LPS="3.04082" LPV="1" RPCX="0.68511" RPCY="0.68360" RPD="9.95932" RPS="3.04082" RPV="1" LEYEX="-0.03750"
LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00416" LPUPILV="1" REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00426" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />'
11:44:37.060258: Incoming: '<REC CNT="61174" TIME="1009.73285" TIME_TICK="786600513876" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0"
RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431" BPOGV="0" LPCX="0.37253" LPCY="0.66185" LPD="12.32223" LPS="3.04082" LPV="1" RPCX="0.68350" RPCY="0.68501" RPD="10.21000" RPS="3.04082" RPV="1" LEYEX="-0.03750"
LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00411" LPUPILV="1" REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00416" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />'
11:44:37.088995: Raw instring: b'<REC CNT="61175" TIME="1009.74915" TIME_TICK="786600676977" FPOGX="0.00000" FPOGY="0.00000" FPOGS="1006.12256" FPOGD="0.23120" FPOGID="1311" FPOGV="0" LPOGX="0.60884" LPOGY="-1.32431" LPOGV="0"
RPOGX="0.50148" RPOGY="0.44719" RPOGV="0" BPOGX="0.60884" BPOGY="-1.32431" BPOGV="0" LPCX="0.37171" LPCY="0.66230" LPD="12.06222" LPS="3.04082" LPV="1" RPCX="0.68199" RPCY="0.68511" RPD="9.79370" RPS="3.04082" RPV="1" LEYEX="-0.03750"
LEYEY="-0.04039" LEYEZ="0.89450" LPUPILD="0.00416" LPUPILV="1" REYEX="0.02043" REYEY="-0.01091" REYEZ="0.83579" RPUPILD="0.00418" RPUPILV="1" CX="0.99948" CY="0.09722" CS="0" USER="0" />\r\n'
```


Figure B.4: Eyetracking data as captured by the Opengaze.py wrapper for the same Control English-French task. In this file, we see timestamps for each entry, and eyetracking data as outlined by the Open Gaze API by Gazepoint.

In order to process the Debug file and run the second script, I first had to import the contents into an Excel file (see figure B.5), then run the Debug extraction script. Much like the Timestamp Modification script, this script extracted information such as the timestamps for each entry, and separated each entry by its predetermined label (e.g., *CNT* for count, *LPUPILD* for left pupil dilation, etc.) into new columns.

hour	minutes	seconds	_2	raw data
11	44	37.02739	Incoming	'<REC CNT="61172" TIME="1009.69946" TIME_TICK="786600180172" FPOGX="0.00000" FPOGY="
11	44	37.060258	Incoming	'<REC CNT="61173" TIME="1009.71619" TIME_TICK="786600347177" FPOGX="0.00000" FPOGY="
11	44	37.060258	Incoming	'<REC CNT="61174" TIME="1009.73285" TIME_TICK="786600513876" FPOGX="0.00000" FPOGY="
11	44	37.088995	Incoming	'<REC CNT="61175" TIME="1009.74915" TIME_TICK="786600676977" FPOGX="0.00000" FPOGY="
11	44	37.11331	Incoming	'<REC CNT="61176" TIME="1009.76562" TIME_TICK="786600841778" FPOGX="0.00000" FPOGY="
11	44	37.114305	Incoming	'<REC CNT="61177" TIME="1009.78198" TIME_TICK="786601005193" FPOGX="0.00000" FPOGY="
11	44	37.124051	Incoming	'<REC CNT="61178" TIME="1009.79846" TIME_TICK="786601169716" FPOGX="0.00000" FPOGY="
11	44	37.140467	Incoming	'<REC CNT="61179" TIME="1009.81482" TIME_TICK="786601333758" FPOGX="0.00000" FPOGY="
11	44	37.158444	Incoming	'<REC CNT="61180" TIME="1009.83148" TIME_TICK="786601500164" FPOGX="0.00000" FPOGY="
11	44	37.174402	Incoming	'<REC CNT="61181" TIME="1009.84814" TIME_TICK="786601666494" FPOGX="0.00000" FPOGY="
11	44	37.191398	Incoming	'<REC CNT="61182" TIME="1009.86462" TIME_TICK="786601831547" FPOGX="0.00000" FPOGY="
11	44	37.20851	Incoming	'<REC CNT="61183" TIME="1009.88123" TIME_TICK="786601997496" FPOGX="0.00000" FPOGY="
11	44	37.225073	Incoming	'<REC CNT="61184" TIME="1009.89758" TIME_TICK="786602161243" FPOGX="0.00000" FPOGY="
11	44	37.239221	Incoming	'<REC CNT="61185" TIME="1009.91388" TIME_TICK="786602323987" FPOGX="0.00000" FPOGY="
11	44	37.256569	Incoming	'<REC CNT="61186" TIME="1009.93042" TIME_TICK="786602489371" FPOGX="0.00000" FPOGY="
11	44	37.273559	Incoming	'<REC CNT="61187" TIME="1009.94690" TIME_TICK="786602654061" FPOGX="0.00000" FPOGY="
11	44	37.292567	Incoming	'<REC CNT="61188" TIME="1009.96344" TIME_TICK="786602819921" FPOGX="0.00000" FPOGY="
11	44	37.312606	Incoming	'<REC CNT="61189" TIME="1009.98004" TIME_TICK="786602985696" FPOGX="0.00000" FPOGY="
11	44	37.329616	Incoming	'<REC CNT="61190" TIME="1009.99658" TIME_TICK="786603151309" FPOGX="0.00000" FPOGY="
11	44	37.352619	Incoming	'<REC CNT="61191" TIME="1010.01294" TIME_TICK="786603314515" FPOGX="0.00000" FPOGY="

Figure B.5: Debug data imported from CSV into excel.

In addition to separating eyetracking values, this script pulled only pupil-related data as that is the primary focus of this research. As a result, other data, such as gaze and fixation, were excluded by the extraction code. The excerpt of the code in figure B.6 provides a visual on what was specifically extracted. After running the script, a new Excel file with the suffix “updated” was generated into the results folder.

```
# Define the tags we are interested in
tags = ['CNT', 'LPCX', 'LPCY', 'LPD', 'LPUPILD', 'RPCX', 'RPCY',
 ↪  'RPD', 'RPUPILD']

# Check if 'raw data' column exists
if 'raw data' not in df.columns:
    print("Error: The Excel file must contain a column labelled 'raw
     ↪  data'")
    return

# Initialize new columns in the DataFrame
for tag in tags:
    df[tag] = None

# Process each row in the 'raw data' column
for index, row in df.iterrows():
    raw_data = row['raw data']
    values = extract_values(raw_data, tags)

    # Update the DataFrame with extracted values
    for tag in tags:
        df.at[index, tag] = values[tag]
```

Figure B.6: Snippet of Timestamp Modification python script, where CNT corresponds to token count, LPCX and RPCX corresponds to left and right pupil position on the X axis, LPCY and RPCY corresponds to left and right pupil position on the Y axis), LPD and RPD corresponds to left and right pupil dilation in pixels, and LPUPILD and RPUPILD corresponds to left and right pupil dilation in millimeters

At this stage, the updated debug file still contained all recorded eye-tracking events from start to end of the task, however only the lines corresponding to the recorded timestamps in the LDT results file would be pertinent. Consequently, a third and final script was run which compared the new debug file against the LDT results file, and matched and extracted only those entry lines which corresponded with the logged timestamps. In other words, this script combed through the entire debug file to look for only those entries taken at the same time as the Fixation, Token and Response timestamps. This was made possible as the code was designed to match the minutes and seconds of each timestamp in the LDT results file to an entry in the debug file (see Figure B.7).

Jnnamed:	Stimulus	Response	tionTime(	hour	minutes	seconds	CNT	LPCX	LPCY	LPD	LPUPILD	RPCX	RPCY	RPD	RPUPILD
CrossTime	ABEILLE	RIGHT	727	11	52	18.244326	89142	0.27946	0.50384	17.24916	0.00429	0.63547	0.49536	18.79289	0.00469
WordTime	ABEILLE	RIGHT	727	11	52	28.248135	89749	0.27208	0.506	18.49757	0.00449	0.62894	0.49515	19.33957	0.00474
KeyInputT	ABEILLE	RIGHT	727	11	52	28.980319	89793	0.27325	0.50602	16.64617	0.00424	0.62956	0.49625	18.06113	0.00448
CrossTime	AUTILLER	RIGHT	2072	11	47	35.152411	71973	0.29505	0.49466	17.77706	0.00433	0.65132	0.48684	18.24216	0.00461
WordTime	AUTILLER	RIGHT	2072	11	47	45.155309	72579	0.29173	0.50191	18.5254	0.00445	0.65038	0.50255	19.42775	0.00488
KeyInputT	AUTILLER	RIGHT	2072	11	47	47.233017	72705	0.28688	0.49573	20.08993	0.00489	0.64017	0.48852	20.55083	0.00504
CrossTime	AVOIR	RIGHT	839	11	47	13.651165	70669	0.29092	0.50307	20.76514	0.00503	0.64705	0.4974	20.46464	0.00503
WordTime	AVOIR	RIGHT	839	11	47	23.654377	71275	0.30963	0.48715	21.74311	0.00521	0.66592	0.48658	20.82961	0.00523
KeyInputT	AVOIR	RIGHT	839	11	47	24.498235	71327	0.30861	0.48885	21.66393	0.00516	0.66426	0.48822	21.34446	0.00533
CrossTime	BARBE	RIGHT	963	11	50	50.377188	83813	0.29424	0.50267	18.84776	0.00452	0.65104	0.49844	20.14074	0.00501
WordTime	BARBE	RIGHT	963	11	51	0.380955	84420	0.29896	0.50626	17.93451	0.00475	0.65509	0.50179	18.31484	0.00471
KeyInputT	BARBE	RIGHT	963	11	51	1.348221	84479	0.29439	0.5047	19.30134	0.00464	0.65137	0.49937	19.01844	0.00469
CrossTime	BE	LEFT	696	11	47	47.233017	72706	0.28609	0.49493	20.3824	0.00488	0.63956	0.48739	20.30117	0.00515
WordTime	BE	LEFT	696	11	47	57.234703	73312	0.29636	0.50155	18.11778	0.00474	0.65141	0.49519	16.51651	0.0047
KeyInputT	BE	LEFT	696	11	47	57.934591	73354	0.27404	0.50401	20.47466	0.00503	0.62893	0.48445	20.19475	0.00501
CrossTime	BEC	LEFT	713	11	53	57.983043	95191	0.27816	0.51221	17.63772	0.0043	0.63399	0.50621	19.20704	0.0048
WordTime	BEC	LEFT	713	11	54	7.988052	95798	0.27446	0.49332	19.4904	0.00487	0.63359	0.47805	20.43906	0.00512
KeyInputT	BEC	LEFT	713	11	54	8.704744	95842	0.262	0.48633	18.69092	0.00451	0.6181	0.46185	19.53481	0.00478
CrossTime	BERGER	RIGHT	1321	11	45	59.186556	66152	0.29224	0.49927	18.59456	0.00454	0.64016	0.49298	18.99939	0.00483
WordTime	BERGER	RIGHT	1321	11	46	9.190471	66759	0.36558	0.3773	14.81443	0.00448	0.71903	0.42678	11.39911	0.00523
KeyInputT	BERGER	RIGHT	1321	11	46	10.516471	66839	0.31983	0.45467	17.76425	0.00463	0.67112	0.46655	17.42365	0.00499
CrossTime	BESOIN	RIGHT	889	11	54	40.381572	97763	0.29463	0.50444	15.9128	0.00413	0.6506	0.50506	18.09872	0.00473
WordTime	BESOIN	RIGHT	889	11	54	50.385691	98370	0.28849	0.50649	18.48301	0.00458	0.64332	0.50338	19.15233	0.00486
KeyInputT	BESOIN	RIGHT	889	11	54	51.279483	98424	0.28976	0.50664	17.66939	0.00432	0.64391	0.5039	18.20433	0.00465

Figure B.7: Final results file with all eye-tracking data extracted and consolidated into a single Excel file.

These steps were repeated for each of the two files (LDT response and Debug CSV) for each test conditions for each participant. In sum, using these extraction scripts 294 files (2 files $\times$ 3 conditions $\times$ 49 participants) became 147 data files (1 results file $\times$ 3 conditions $\times$ 49 participants) for final consideration.

Appendix C

WORD LISTS

Table C.1: Complete word list for the English x French NC task condition, with number of syllables and word frequency (occurrences per million words). English frequencies were sourced from the Corpus of Contemporary American English (COCA; Davies & Gardner, 2010 [2]). French frequencies were sourced from the Lexique 3.83 film subtitles corpus (*freqlemfilms2*; New et al., 2001 [3]).

Token	Type	Syll.	Freq	Token	Type	Syll.	Freq
BARK	EN Word	1	45.59	ABEILLE	FR Word	2	9.11
BE	EN Word	1	32622.71	ABONDANT	FR Word	3	1.36
CALL	EN Word	1	2338.17	AFFRONT	FR Word	2	3.36
CLOSE	EN Word	1	798.54	BARBE	FR Word	1	24.15
CRY	EN Word	1	236.19	BATEAU	FR Word	2	124.82
EAT	EN Word	1	404.32	BEC	FR Word	1	7.39
FIND	EN Word	1	2552.64	BERGER	FR Word	2	11.84
FRIEND	EN Word	1	572.87	BESOIN	FR Word	2	771.24
GO	EN Word	1	6928.74	BEURRE	FR Word	1	15.14
MAKE	EN Word	1	4679.80	CANICHE	FR Word	2	2.56
MAN	EN Word	1	1516.38	CHANDAIL	FR Word	2	0.82
MONEY	EN Word	2	660.90	CHANGER	FR Word	2	411.99
PEOPLE	EN Word	2	3132.22	CRECHE	FR Word	1	3.91
PROBE	EN Word	1	41.99	CUIRE	FR Word	1	21.65
RIDE	EN Word	1	201.28	DONNER	FR Word	2	1209.67
RUN	EN Word	1	942.83	DORTOIR	FR Word	2	5.29
SAY	EN Word	1	4125.20	ECRIRE	FR Word	2	305.92
SPEAK	EN Word	1	297.60	FAIRE	FR Word	1	8813.48
STUFF	EN Word	1	325.49	FINIR	FR Word	2	557.61
TRIP	EN Word	1	246.56	FLOCON	FR Word	2	2.61
WAIT	EN Word	1	468.96	PRENDRE	FR Word	1	1913.83
WANT	EN Word	1	2385.54	REGARDER	FR Word	3	1197.29
WOMAN	EN Word	2	1393.51	RENDRE	FR Word	1	508.81
WORD	EN Word	1	541.61	VERITE	FR Word	3	193.24
WRITE	EN Word	1	686.10	VOIR	FR Word	1	4119.43

Table C.2: EN and FR word list, including Non-Words, for the English x French CONT task condition, including length (syllable count), and frequency as determined by the Corpus of Contemporary American English (COCA; Davies & Gardner, 2010 [2]) and the Lexique 3.83 film subtitles corpus (*freqlemfilms2*; New et al., 2001 [3]).

Token	Type	Syll.	Freq	Token	Type	Syll.	Freq
BE	EN Word	1	32622.71	BERGER	FR Word	2	11.84
CLOSE	EN Word	1	798.54	ABEILLE	FR Word	2	9.11
CRY	EN Word	1	236.19	AVOIR	FR Word	2	3.13
DO	EN Word	1	8244.70	BARBE	FR Word	1	24.15
EAT	EN Word	1	404.32	BEC	FR Word	1	7.39
FIND	EN Word	1	2552.64	BESOIN	FR Word	2	771.24
GO	EN Word	1	3571.65	CUIRE	FR Word	1	21.65
MAKE	EN Word	1	4679.80	ECRIRE	FR Word	2	305.92
PUT	EN Word	1	2100.52	ETRE	FR Word	1	100.69
RIDE	EN Word	1	201.28	FAIRE	FR Word	1	8813.48
RUN	EN Word	1	942.83	FINIR	FR Word	2	557.61
SAY	EN Word	1	4125.20	JURER	FR Word	2	148.41
SEE	EN Word	1	3270.45	MANGER	FR Word	2	5.62
SPEAK	EN Word	1	297.60	PARLER	FR Word	2	11.58
TRY	EN Word	1	1698.61	PRENDRE	FR Word	1	1913.83
WANT	EN Word	1	2385.54	RENDRE	FR Word	1	508.81
WRITE	EN Word	1	686.10	VERITE	FR Word	3	193.24

Token	Type	Syll.	Freq
AUTILLER	Non Word	3	N/A
CLANCELIR	Non Word	3	N/A
CORVENE	Non Word	2	N/A
FIEGER	Non Word	2	N/A
FLURIR	Non Word	2	N/A
HOLEATH	Non Word	2	N/A
JAWEL	Non Word	2	N/A
MANITHE	Non Word	2	N/A
ORTHONINTHE	Non Word	3	N/A
PALMER	Non Word	2	N/A
REWINAL	Non Word	3	N/A
ROSION	Non Word	2	N/A
ROUBILLE	Non Word	2	N/A
SERPINE	Non Word	2	N/A
VIRLOING	Non Word	3	N/A
YUZZ	Non Word	1	N/A

Table C.3: EN and FR word list, including Cognates, for the English x French COG task condition, including length (syllable count), and frequency as determined by the Corpus of Contemporary American English (COCA; Davies & Gardner, 2010 [2]) and the Lexique 3.83 film subtitles corpus (*freqlemfilms2*; New et al., 2001 [3]).

Token	Type	Syll.	Freq	Token	Type	Syll.	Freq
BARK	EN Word	1	45.59	AVOIR	FR Word	2	3.13
BOY	EN Word	1	761.42	BARBE	FR Word	1	24.15
CALL	EN Word	1	773.51	BEC	FR Word	1	7.39
CRY	EN Word	1	236.19	BERGER	FR Word	2	11.84
DAY	EN Word	1	1724.87	BESOIN	FR Word	2	771.24
EAT	EN Word	1	404.32	BEURRE	FR Word	1	15.14
FIND	EN Word	1	2552.64	DONNER	FR Word	2	1209.67
MAKE	EN Word	1	4679.80	DORTOIR	FR Word	2	5.29
NIGHT	EN Word	1	764.39	ECRIRE	FR Word	2	305.92
PUT	EN Word	1	2100.52	ETRE	FR Word	1	100.69
RIDE	EN Word	1	201.28	FINIR	FR Word	2	557.61
SAY	EN Word	1	4125.20	FLOCON	FR Word	2	2.61
SEE	EN Word	1	3270.45	MANGER	FR Word	2	5.62
TRIP	EN Word	1	246.56	PARLER	FR Word	2	11.58
TRY	EN Word	1	1698.61	PRENDRE	FR Word	1	1913.83
				RENDRE	FR Word	1	508.81
				VOIR	FR Word	1	4119.43

Token	Type	Syll	EN Freq	FR Freq
CAUTION	Cognate	2	84.53	10.88
CAVE	Cognate	1	65.21	1.59
CHAIR	Cognate	1	290.18	1.72
CHAT	Cognate	1	75.44	93.00
COIN	Cognate	1	57.39	99.51
FIGURE	Cognate	2	569.23	25.18
GRAND	Cognate	1	224.87	638.72
LECTURE	Cognate	2	78.62	15.80
MAIN	Cognate	1	440.71	499.60
PAIN	Cognate	1	266.94	67.58
PAR	Cognate	1	33.18	10.58
PIECE	Cognate	1	411.87	151.10
PLUS	Cognate	1	263.51	607.21
POINT	Cognate	1	1163.42	5.28
POUR	Cognate	1	14.82	27.69
SALE	Cognate	1	197.63	146.86

Table C.4: Complete word list for the English x Mandarin NC task condition, with number of syllables and word frequency (occurrences per million words). English frequencies were sourced from the Corpus of Contemporary American English (COCA; Davies & Gardner, 2010 [2]). Mandarin frequencies were sourced from the `wordfreq` Python library (Speer et al., 2018 [4]), which incorporates SUBTLEX-CH (Cai & Brysbaert, 2010 [5]) as a primary Chinese data source (permission to redistribute granted by M. Brysbaert).

Token	Type	Syll	Freq	Token	Type	Syll	Freq
BARK	EN Word	1	45.59	CHE 车	CN Word	1	186.00
CALL	EN Word	1	773.51	CHUSHI 厨师	CN Word	2	11.00
CRY	EN Word	1	236.19	DAXUE 大学	CN Word	2	417.00
DAY	EN Word	1	1724.87	DE 得	CN Word	1	1100.00
DO	EN Word	1	8244.70	DI 睇	CN Word	1	1450.00
EAT	EN Word	1	404.32	DIANSHI 电视	CN Word	2	178.00
FIND	EN Word	1	2552.64	DU 肚	CN Word	1	2.19
GO	EN Word	1	3571.65	JI 机	CN Word	1	112.00
JOB	EN Word	1	618.43	JIAO 脚	CN Word	1	72.40
MAKE	EN Word	1	4679.80	JIU 酒	CN Word	1	58.90
MONEY	EN Word	2	660.90	LI 来	CN Word	1	2820.00
PEOPLE	EN Word	2	3132.22	LU 炉	CN Word	1	8.32
PUT	EN Word	1	2100.52	NU 女	CN Word	1	302.00
SAY	EN Word	1	4125.20	QU 去	CN Word	1	1450.00
SPEAK	EN Word	1	297.60	ROU 肉	CN Word	1	64.60
TRAIN	EN Word	1	452.18	SHI 食	CN Word	1	447.00
TRIP	EN Word	1	246.56	SHU 书	CN Word	1	240.00
TRY	EN Word	1	1698.61	SHUI 水	CN Word	1	200.00
WAIT	EN Word	1	468.96	XIAO 笑	CN Word	1	158.00
WANT	EN Word	1	2385.54	XING 行	CN Word	1	269.00
WOMAN	EN Word	2	1393.51	YIN 饮	CN Word	1	100.00
WORD	EN Word	1	541.61	ZI 仔	CN Word	1	141.00
WRITE	EN Word	1	686.10	ZOU 走	CN Word	1	457.00
				ZUO 做	CN Word	1	1200.00

Table C.5: English and Mandarin, including Non-Words, word list for the English x Mandarin CONT task condition, with number of syllables and word frequency (occurrences per million words). English frequencies were sourced from the Corpus of Contemporary American English (COCA; Davies & Gardner, 2010 [2]). Mandarin frequencies were sourced from the `wordfreq` Python library (Speer et al., 2018 [4]), which incorporates SUBTLEX-CH (Cai & Brysbaert, 2010 [5]) as a primary Chinese data source.

Token	Type	Syll	Freq	Token	Type	Syll	Freq
CLOSE	EN Word	1	798.54	BI 畀	CN word	1	603.00
CRANE	EN Word	1	18.34	CHA 茶	CN word	1	85.10
CRY	EN Word	1	236.19	DENG 等	CN word	1	2140.00
EAT	EN Word	1	404.32	JIU 酒	CN word	1	58.90
GO	EN Word	1	3571.65	LIN 拎	CN word	1	61.70
JOB	EN Word	1	618.43	MAI 买	CN word	1	380.00
MORNING	EN Word	2	260.95	NIAN 年	CN word	1	4370.00
PLACE	EN Word	1	1101.84	PAO 跑	CN word	1	170.00
PLATE	EN Word	1	58.23	RI 日	CN word	1	2000.00
RIDE	EN Word	1	201.28	SHUI 水	CN word	1	200.00
RUN	EN Word	1	942.83	TING 听	CN word	1	490.00
SAY	EN Word	1	4125.20	WAN 玩	CN word	1	316.00
SPEAK	EN Word	1	297.60	XIAOBA 小巴	CN word	2	1.29
STUFF	EN Word	1	325.49	XING 行	CN Word	1	269.00
TRAIN	EN Word	1	452.18	YONG 用	CN word	1	1170.00
WANT	EN Word	1	2385.54	ZOU 走	CN word	1	457.00
WRITE	EN Word	1	686.10				

Token	Type	Syll	Freq
AO	Non Word	1	N/A
BAIO	Non Word	1	N/A
DERSIALS	Non Word	2	N/A
ETHERAN	Non Word	3	N/A
JAWEL	Non Word	2	N/A
LAO	Non Word	1	N/A
MICIOUSLY	Non Word	3	N/A
NI	Non Word	1	N/A
OPNER	Non Word	2	N/A
PU	Non Word	1	N/A
ROSION	Non Word	2	N/A
SAN	Non Word	1	N/A
SPRIOTS	Non Word	2	N/A
WA	Non Word	1	N/A
ZHU	Non Word	1	N/A

Table C.6: English and Mandarin word list, including Cognates, for the English × Mandarin COG task condition, with number of syllables and word frequency (occurrences per million words). English frequencies were sourced from COCA (Davies & Gardner, 2010 [2]). Mandarin frequencies were sourced from `wordfreq` (Speer et al., 2018 [4]), incorporating SUBTLEX-CH (Cai & Brysbaert, 2010 [5]).

Token	Type	Syll	Freq	Token	Type	Syll	Freq
BARK	EN Word	1	45.59	CHUSHI 厨师	CN word	2	11.00
BOY	EN Word	1	761.42	DI 睇	CN word	1	1450.00
CALL	EN Word	1	773.51	DIANSHI 电视	CN word	2	178.00
CRY	EN Word	1	236.19	DU 肚	CN word	1	2.19
DAY	EN Word	1	1724.87	LI 来	CN Word	1	2820.00
EAT	EN Word	1	404.32	LU 炉	CN word	1	8.32
FIND	EN Word	1	2552.64	MAI买	CN word	1	380.00
MAKE	EN Word	1	4679.80	MAO 猫	CN word	1	53.70
NIGHT	EN Word	1	764.39	NU 女	CN word	1	302.00
PUT	EN Word	1	2100.52	PAO 跑	CN word	1	170.00
RIDE	EN Word	1	201.28	QU 去	CN word	1	1450.00
SAY	EN Word	1	4125.20	SHI 食	CN word	1	447.00
SEE	EN Word	1	3270.45	SHU 书	CN word	1	240.00
TRAIN	EN Word	1	452.18	XIAO 笑	CN word	1	158.00
TRIP	EN Word	1	246.56	XING 行	CN word	1	269.00
TRY	EN Word	1	1698.61	YIN 饮	CN Word	1	100.00
WANT	EN Word	1	2385.54	ZI 仔	CN Word	1	141.00
WRITE	EN Word	1	686.10	ZOU 走	CN word	1	457.00
				ZUO 做	CN Word	1	1200.00

Token	Type	Syll	EN Freq	ZH Freq
BA 吧	Cognate	1	N/A	955.00
BITE 比特	Cognate	1	20.40	10.70
CHAN 禅	Cognate	1	10.00	4.17
HANI 哈尼	Cognate	2	N/A	2.69
KAFEI 咖啡	Cognate	2	12.30	45.70
KAOLA 考拉	Cognate	2	1.10	3.27
KEKE 可可	Cognate	2	N/A	6.76
LAKE 腊克	Cognate	1	67.60	30.90
MAMA 妈妈	Cognate	2	17.40	174.00
MINI 迷你	Cognate	2	27.50	17.40
PIKA 皮卡	Cognate	2	0.37	2.32
PUKE 扑克	Cognate	1	1.91	5.01
QIPAO 旗袍	Cognate	2	0.03	2.24
WATE 瓦特	Cognate	2	N/A	3.47

Appendix D

FULL EXPERIMENTAL RESULTS

Table D.1: A summary of the average Reaction Time (RT) in msec for all experimental conditions

	No Control	Control	Cognate
English	1044.28		
French	1105.05		
English	915.13		
Mandarin	1147.87		
English		953.23	
French		1124.43	
FR-Non-Word		2312.97	
English		1014.87	
Mandarin		1363.60	
CH-Non-Word		1800.62	
English			975.63
French			1019.65
FR-Cognate			1348.91
English			916.54
Mandarin			1127.53
CH-Cognate			1750.88

Table D.2: A summary of the general distribution of responses for all experimental conditions

	No Control	Control	Cognate
English	50.42%		
French	49.25%		
No Response	0.33%		
English	48.17%		
Mandarin	50.83%		
No Response	1.00%		
English		34.92%	
French		37.42%	
FR-Non-Word		25.67%	
No Response		2.00%	
English		36.25%	
Mandarin		46.50%	
CH-Non-Word		15.58%	
No Response		1.67%	
English			50.00%
French			50.00%
No Response			0.00%
English			46.67%
Mandarin			48.31%
No Response			5.02%

Table D.3: Summary of response accuracy for the No Control condition.

		No Control
ENxFR	ENG Correct RESP	98.17%
	ENG Incorrect RESP	1.33%
	No Response	0.50%
	FR Correct RESP	49.25%
	FR Incorrect RESP	49.25%
	No Response	0.33%
ENxCH	ENG Correct RESP	98.09%
	ENG Incorrect RESP	1.57%
	No Response	0.35%
	CH Correct RESP	96.16%
	CH Incorrect RESP	2.24%
	No Response	1.60%

Table D.4: Summary of response accuracy for the Control condition.

		Control
ENxFR	ENG Correct RESP	98.28%
	ENG Incorrect RESP	1.47%
	NW Incorrect RESP	0.25%
	No Response	0.00%
	FR Correct RESP	97.30%
	FR Incorrect RESP	0.74%
	NW Incorrect RESP	1.47%
	No Response	0.49%
	NW Correct RESP	78.39%
	ENG Incorrect RESP	3.91%
	FR Incorrect RESP	11.98%
	No Response	0.49%
ENxCH	ENG Correct RESP	96.71%
	ENG Incorrect RESP	1.65%
	NW Incorrect RESP	1.18%
	No Response	0.47%
	CH Correct RESP	93.75%
	CH Incorrect RESP	0.75%
	NW Incorrect RESP	3.75%
	No Response	1.75%
	NW Correct RESP	44.53%
	ENG Incorrect RESP	5.60%
	CH Incorrect RESP	46.93%
	No Response	2.93%

Table D.5: Summary of response accuracy for the Cognate condition.

		Cognate
ENxFR	ENG Correct RESP	99.31%
	ENG Incorrect RESP	0.69%
	No Response	0.00%
	FR Correct RESP	97.55%
	FR Incorrect RESP	2.45%
	No Response	0.00%
	COG ENG RESP	45.05%
	COG FR RESP	54.95%
	No Response	0.00%
ENxCH	ENG Correct RESP	95.56%
	ENG Incorrect RESP	0.44%
	No Response	4.00%
	CH Correct RESP	87.58%
	CH Incorrect RESP	8.00%
	No Response	4.42%
	COG ENG RESP	36.29%
	COG CH RESP	56.57%
	No Response	7.14%

Appendix E

ADDITIONAL DOCUMENTATION AND MATERIALS

YORK U UNIVERSITÉ UNIVERSITY

Certificate #:	**STU 2024-012**
Approval Period:	**03/04/24-03/04/25**

OFFICE OF RESEARCH ETHICS (ORE)
Kaneff Tower

4700 Keele St.
Toronto ON
Canada M3J 1P3
Tel 416 736 5914
www.research.yorku.ca

ETHICS APPROVAL

To: **Konstantine Pho**
Graduate Student of Linguistics & Applied Linguistics
phokonst@my.yorku.ca

From: Alison M. Collins-Mrakas, Director, Research Ethics
(on behalf of You-ta Chuang, Chair, Human Participants Review Committee)

Date: Monday, March 4, 2024

Title: **Investigation of bilingual language processing via pupillometry**

Risk Level: ☒ Minimal Risk ☐ More than Minimal Risk

Level of Review: ☒ Delegated Review ☐ Full Committee Review

I am writing to inform you that this research project, **"Investigation of bilingual language processing via pupillometry"** has received ethics review and approval by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines.

Note that approval is granted for one year. Ongoing research – research that extends beyond one year – must be renewed prior to the expiry date.

Any changes to the approved protocol must be reviewed and approved through the amendment process by submission of an amendment application to the HPRC prior to its implementation.

Any adverse or unanticipated events in the research should be reported to the Office of Research ethics (ore@yorku.ca) as soon as possible.

For further information on researcher responsibilities as it pertains to this approved research ethics protocol, please refer to the attached document, **"RESEARCH ETHICS: PROCEDURES to ENSURE ONGOING COMPLIANCE".**

Should you have any questions, please feel free to contact me at acollins@yorku.ca.

Yours sincerely,

Alison M. Collins-Mrakas M.Sc., LLM
Director, Office of Research Ethics

Figure E.1: Ethics approval for March 2024-2025 as obtained by York University Research Ethics Board (REB)

YORK UNIVERSITÉ UNIVERSITY

Certificate #:	**STU 2024-012**
Initial Approval:	**03/04/24-03/04/25**
Amendments:	
Renewals:	**02/25/25-02/25/26**
Current Approval Period:	**02/25/25-02/25/26**

OFFICE OF RESEARCH ETHICS (ORE)
Kaneff Tower

4700 Keele St.
Toronto ON
Canada M3J 1P3
ore@yorku.ca
www.research.yorku.ca

ETHICS RENEWAL

To: **Konstantine Pho - Graduate Student**
Department of Linguistics & Applied Linguistics
Faculty of Liberal Arts & Professional Studies
phokonst@my.yorku.ca

From: Alison M. Collins-Mrakas, Director, Research Ethics
(on behalf of Gillian Parekh, Chair, Human Participants Review Committee)

Date: Tuesday, January 28, 2025

Title: **Investigation of bilingual language processing via pupillometry**

Risk Level: ☒ Minimal Risk ☐ More than Minimal Risk

Level of Review: ☒ Delegated Review ☐ Full Committee Review

I am writing to inform you that this research project, **"Investigation of bilingual language processing via pupillometry"** has received ethics review and renewal by the Human Participants Review Sub-Committee, York University's Ethics Review Board and conforms to the standards of the Canadian Tri-Council Research Ethics guidelines.

Note that renewal is granted for one year. Ongoing research – research that extends beyond one year – must be renewed prior to the expiry date.

Any changes to the approved protocol must be reviewed and approved through the amendment process by submission of an amendment application to the HPRC prior to its implementation.

Any adverse or unanticipated events in the research should be reported to the Office of Research ethics (ore@yorku.ca) as soon as possible.

For further information on researcher responsibilities as it pertains to this approved research ethics protocol, please refer to the attached document, **"RESEARCH ETHICS: PROCEDURES to ENSURE ONGOING COMPLIANCE".**

Should you have any questions, please feel free to contact me at: 416-736-5914 or via email at: acollins@yorku.ca.

Yours sincerely,

Alison M. Collins-Mrakas M.Sc., LLM
Director, Office of Research Ethics

Figure E.2: Ethics approval for February 2025-2026 as obtained by York University Research Ethics Board (REB)

Figure E.3: French English recruitment poster

Figure E.4: French French recruitment poster

Figure E.5: Mandarin English recruitment poster

Figure E.6: Mandarin Mandarin recruitment poster

YAV Reference Code ____________________

YORK UNIVERSITÉ UNIVERSITY

Speech and Psycholinguistics Lab

Konstantine K. Pho, Principal Investigator
York University

Language and Social Background Questionnaire

1. Today's date: ______________________________
day month year

2. Sex: male ☐ female ☐ **3.** Occupation/University Major: ____________________

4. Handedness: left ☐ right ☐ **5.** Age:

6. What is the highest year of school you have completed? ____________________

7. What is the highest degree you have earned? ____________________

8. On average, how many hours do you use a computer per day? ____________________

9. On average, how many hours do you play video/computer games per day? ____________________

10. Do you have hearing problems? **yes** ☐ **no** ☐

If **yes**, do you wear a hearing aid? **yes** ☐ **no** ☐

11. Do you have vision problems? **yes** ☐ **no** ☐

If **yes**, do you wear glasses/contacts? **yes** ☐ **no** ☐

Is your vision corrected to normal with glasses/contacts? **yes** ☐ **no** ☐

12. Are you colour blind? **yes** ☐ **no** ☐

If **yes**, what type? ____________________

13. Have you ever had a head injury? **yes** ☐ **no** ☐

14. Do you have any known neurological impairments (e.g., epilepsy etc) **yes** ☐ **no** ☐

If yes, please indicate: ____________________

15. Are you currently taking any psychoactive medications? **yes** ☐ **no** ☐

If yes, please indicate: ____________________

Reference Code ___________________________

16. Please indicate the highest level of education and occupation for each parent:

MOTHER	FATHER
1.______No high school diploma	1.______No high school diploma
2.______High school graduate	2.______High school graduate
3.______Some college or college diploma	3.______Some college or college diploma
4.______Bachelor's degree	4.______Bachelor's degree
5.______Graduate or professional degree	5.______Graduate or professional degree
Occupation: ____________________________	Occupation: ____________________________
Native language: ________________________	Native language: ________________________
Second language: _______________________	Second language: _______________________
Other language: _________________________	Other language: _________________________

17. Were you born in Canada? **yes** ☐ **no** ☐

If **no**, where were you born? ____

When did you move to Canada? ____________________________________
month **year**

18. Have you ever lived in a place where English is not the dominant communicating language? **yes** ☐ **no** ☐

		From		To	
If **yes**, where and for how long?	1				
	2				
	3				
		month	**year**	**month**	**year**

Language Background

19. List all the languages and dialects you can speak including English, *in order of fluency*:

Language	Where did you learn it? (Home, School, Community)	Where do you use it? (Home, School, Friends, Travel, Other)	At what age did you learn it?
1 __________________			
2 __________________			
3 __________________			
4 __________________			
5 __________________			

Reference Code ___________________________

20. For the languages indicated above, do you use them regularly? Please fill in the corresponding column.

Language	**If yes:** Were there any periods in your life when you did not use it? Indicate duration in months/years.	**If no:** How long (months/years) and up to which age did you use it?
1		
2		
3		
4		
5		

21. Please indicate the proportion of use for English and your other language in **daily life**. One end of the scale indicates that the activity in that environment is carried out only in English (All English). The other end of the scale indicates that only the other language is used (No English). You can mark anywhere on the scale, so please be as precise as possible.

	All English — No English
Language spoken to family members	●——|——|——|——●
Language spoken to friends	●——|——|——|——●
Language for watching TV/video	●——|——|——|——●
Language for reading books/magazines etc.	●——|——|——|——●
Language used for writing (e.g., shopping list, notes etc.)	●——|——|——|——●
Language used in the community and cultural activities	●——|——|——|——●

22. Relative to a native speaker's performance, rate your proficiency level in a scale of 'No Proficiency' to 'Fully Fluent' for the following activities conducted in English and your other language(s).

English

	No Proficiency — Fully fluent
Speaking	●——|——|——|——●
Understanding (Comprehension)	●——|——|——|——●
Reading	●——|——|——|——●
Writing	●——|——|——|——●

Reference Code ______________________

Other Language: ______________________ (please indicate)

	No Proficiency	Fully fluent
Speaking		
Understanding/Listening		
Reading		
Writing		

How often do you do you perform each of the following in this language? Be specific.

speak: ______________, listen ______________, read ______________, write______________?

Other Language: ______________________ (please indicate)

	No Proficiency	Fully fluent
Speaking		
Understanding/Listening		
Reading		
Writing		

How often do you do you perform each of the following in this language? Be specific.

speak: ______________, listen ______________, read ______________, write______________?

Other Language: ______________________ (please indicate)

	No Proficiency	Fully fluent
Speaking		
Understanding/Listening		
Reading		
Writing		

How often do you do you perform each of the following in this language? Be specific.

speak: ______________, listen ______________, read ______________, write______________?

24. Global self-assessment:
Overall, how would you describe your level of bilingualism or multilingualism? Please indicate.

Monolingual — Bilingual/Multilingual

Page 4 of 4

Figure E.7: Language and Social Background Questionnaire.