Tracing the algorithm of bilingual language learning

Jose Armando Aguasvivas Manzano

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

2021

Tracing the algorithm of bilingual language learning

Doctoral dissertation by:

Jose Armando Aguasvivas Manzano

A thesis submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

Supervised by:

Prof. Manuel Carreiras

Basque Center on Cognition, Brain and Language (BCBL)

Universidad del País Vasco / Euskal Herriko Unibersitatea (UPV/EHU)

Donostia-San Sebastian, 2021

The project that gave rise to these results received the support of a fellowship from “la Caixa” Foundation (ID 100010434). The personal fellowship code is LCF/BQ/IN17/116200154004. This project has also received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie-Skłodowska-Curie grant agreement no. 713673.

BCBL Basque Center on Cognition, Brain and Language
Paseo Mikeletegi, 69, 2
Donsotia-San Sebastián, Spain
2021

To Faby, Sarah, and Polina,

my three pillars.

ACKNOWLEDGEMENTS

Esta tesis culmina una etapa relativamente larga y dura de mi vida. A lo largo de casi cuatro años, me he visto crecer como científico y como persona, acompañado de maravillosas personas que han guiado mi camino y brindado su apoyo. En primer lugar, agradezco a mi madre, Fabi, quien desde pequeño me ha enseñado la importancia de aprender, no por conseguir una meta o premio, sino por el valor intrínseco del aprendizaje. Su guía y palabras de apoyo han estado siempre a mi lado y, aunque mudarme a España ha significado separarnos, nunca he dejado de sentir su presencia y la de sus angelitos en todo lo que hago. En segundo lugar, mi hermana Sarah, quien desde lo lejos siempre ha tratado de mantener toda la familia unida y en contacto. Me enorgullezco mucho de que más que mi hermana es mi amiga, a quien puedo decirle lo que me preocupa, quien siempre tratará de comprenderme y ayudarme. Agradezco a mi esposa Polina, встреча с тобой - это изюминка моей жизни. Я тебя люблю. Gracias por todo el apoyo, las risas, los llantos, y sobre todo el amor. Finalmente, agradezco también a mi padre Luis José, a mis hermanos Javier y Arturo, a tío Fernando, tía Sandra y tío Ramiro. Todas las experiencias buenas y malas que hemos vivido juntos forman la base de quien hoy soy.

En mi larga estancia en España he conocido también grandes personas, tanto en el ámbito laboral como personal. Agradezco primero a Manolo. Su paciencia, guía y comprensión hacen que esta tesis sea posible. Gracias también a Jon Andoni, quien siempre me ha brindado su apoyo pese a cualquier circumstancia. Agradezco a todo el equipo del BCBL: Ana, Eider, Maider, Larraitz, Leire, Itziar, Amets, Marisa, Ainhoa, y todas las demás personas que dan sentido y pertenencia a los “BCBLianos”. Incluyo aquí también a mis compañeros de trabajo, que más que colegas les considero verdaderos amigos: Candice, Stefano, Merel, Jose (el otro), Alberto, Chiara, Karina, y Dani. En el ámbito personal, agradezco a Andrey, mi amigo de la infancia, a Fernando, mi compañero de andanzas, a Giancarlo, mi primo y guía filosófico, y a Aratz, que siempre será mi bestie Vasco. Gracias también a aquellas personas que, aunque nos hemos distanciado, siempre tienen un espacio en mi corazón: Ramiro, Jose (el otro otro), Roberto, Jean, Isa, Jhomy, Luis, Melody, Rosa, Aída y Laura. Finalmente, agradezco a Gisela y todo el equipo de “la Caixa” Foundation por haber creído en mi y financiado esta gran oportunidad.

ABSTRACT

Learning a new language is a challenging but highly rewarding experience. Learners need to acquire a massive vocabulary, and the rules to vary and combine this vocabulary to produce and understand sentences correctly. It is possible that learning new languages becomes easier once we already speak at least two. Based on this idea, in this thesis, I explore whether adults who already know two languages (bilinguals) are better at learning a foreign language than those who only know one (monolinguals). For this, I carried six behavioral experiments with three groups of young adult participants: Spanish monolinguals, Spanish-English bilinguals, and Spanish-Basque bilinguals. Together, these experiments targeted implicit and explicit foreign language learning using artificially constructed linguistic materials. Overall, the results from all experiments indicated that both bilingual groups outperformed their monolingual peers when implicitly and explicitly learning vocabulary but not in other aspects (e.g., sub-lexical phonology/orthography, morphology). To explain how the differences in vocabulary learning between monolinguals and bilinguals could emerge, I developed a computational model of the orthographic lexicon. The model could learn written vocabulary based on orthographic patterns (orthotactics) within words of one or two languages. This model revealed that it is easier to recognize and produce novel words when the model is trained on bilingual input than when it is trained with monolingual input. The totality of results from this thesis led me to conclude that monolingual and bilingual individuals differ fundamentally—and possibly only—in vocabulary learning. Exposure to distinct orthotactics within words in two languages could make bilinguals more flexible when integrating the orthographic form of novel vocabulary than monolinguals.

CONTENTS

Preface

We have all been there, sitting in a classroom, trying to learn several new words in Japanese, French, English, or any other foreign language. Does いす(Japanese word for chair romanized as *isu*) mean ***chair*** or ***dog*** (written as いぬ and romanized as *inu*)? The feeling, at first, is that learning a foreign language seems like an insurmountable challenge. One needs to acquire a novel vocabulary of several thousand words and learn how to pronounce, write—sometimes in an entirely different script—, and combine them correctly to form sentences and communicate proficiently. Foreign language learning is, undoubtedly, a challenging but rewarding experience. Humans are unique in their ability to learn not one but multiple languages throughout our lifespans. Furthermore, it is possible that learning new languages becomes easier the more languages we already know.

This observation, combined with several discussion sessions with my advisor, led me to the simple question that motivates this entire work: *Are adults who already know two languages (bilinguals) better at learning a foreign language than those who only know one (monolinguals)?* Intuitively, knowing two languages provides individuals with a more extensive pool of knowledge to exploit during foreign language learning than knowing a single language. For instance, coming back to the Japanese language, the romanized word for a part-time job is アルバイト (read as *arubaito*), which any German (monolingual or bilingual) speaker might recognize, as it stems from the word ***arbeit*** with similar meaning (part-time job). However, this is perhaps a trivial answer. What drove me to pursue this question extends beyond any specific similarities people could employ during foreign language learning. In other words, if all things as equal as possible, are bilinguals inherently better than monolinguals when learning a foreign language?

As with any research question, there is a long trek from its conception to the methodologies that can address it. For example, what does it mean to "learn a foreign language"? Which set of experiments can better address foreign language learning? These are only two of the questions I faced while conceptualizing, designing, and carrying out the research. Learning a foreign language is a complex activity comprised of multiple interconnected levels (e.g., phonology, vocabulary, morphology) acquired through implicit and explicit processes. Consequently, to answer whether bilinguals and monolinguals differ, it is essential to pinpoint at which level of foreign language learning, if any, are they different (RQ1), what these differences are (RQ2), and how these differences could emerge from the bilingual experience (RQ3). These questions justify the title of this thesis, as I was effectively *tracing* the where, what, and how of potential bilingual language learning advantages.

Overview of the chapters

This thesis is divided into five chapters. Chapter 1 introduces foreign language learning and its multiple domains, emphasizing the implicit and explicit learning mechanisms through which humans learn a foreign language. I then define bilingualism and review the extant literature investigating potential differences between monolingual and bilingual foreign language learning. Next, I offer some possible theoretical accounts rooted in experimental linguistic for some aspects of monolingual and bilingual foreign language learning. Finally, I outline the general methodology for the following chapters' experimental work. Chapter 1 establishes the conceptual and theoretical framework upon which the rest of the thesis develops.

Chapters 2 and 3 contain the main body of experimental work, addressing implicit and explicit foreign language learning with six behavioral experiments. Chapter 2 contains four implicit learning experiments targeting different levels of foreign language learning as a proxy: sub-lexical, morphological, word order (syntax), and vocabulary (lexical). Chapter 3 focuses on explicit language learning and presents two additional experiments addressing novel morphology and vocabulary learning. These experiments target my central research question from different angles. Together, these two chapters show "where" (RQ1) and "what" (RQ2) the differences between bilinguals and monolinguals are.

Chapter 4 takes a distinct perspective from the previous two chapters to address the "how" question (RQ3). In this chapter, I first introduce the framework for modeling cognitive systems using computational approaches. Then, I briefly cover some of the most well-known models of bilingual word recognition and learning, emphasizing the gap in models of language learning. Finally, I present a computational model that simulates monolingual and bilingual vocabulary learning to explain some of the findings from Chapter 3. The model proposes a novel view of vocabulary learning, whereby orthographic word forms are stored together using a distributed code that is dynamically modified by experience—and specifically by bilingual experience.

The final chapter summarizes the contributions from the previous three chapters and draws the main conclusions of the thesis. I then discuss the findings by highlighting their relevance in the broader context of the existing literature. Finally, I present arguments for how this work could open up exciting pathways for addressing outstanding questions on the behavioral, cognitive, neural, and computational mechanisms of monolingual and bilingual foreign language learning.

Chapter 1: Introduction

The human brain is remarkable in its ability to learn and generalize knowledge. There is ample evidence that different life experiences can serve as the basis for enhanced learning. To cite a few examples, taxi drivers are especially good at navigation and spatial learning tasks (Maguire et al., 2000), musicians are better at learning and discriminating lexical tone categories (T. C. Zhao & Kuhl, 2015), and action video games players have increased attentional control and meta-learning (learning to learn) skills (Green & Bavelier, 2012), compared to individuals without these life experiences. It is perhaps not a wild stretch to assume that linguistic experiences, such as experience with two languages (bilingual experience), could also foster language learning skills. This idea summarizes the primary goal of this thesis, which explores whether bilingual experience influences foreign language learning. The central research question is, are bilingual adults—those that know two languages—better than their monolingual peers when learning a foreign language?

There are two issues with the main research question as currently postulated. First, languages have different levels (e.g., phonology, vocabulary, morphology). The bilingual experience could influence none, one, multiple, or all levels of foreign language learning. Hence, concluding that there is a bilingual experience effect in foreign language learning from comparing bilinguals and monolinguals in a single aspect (e.g., vocabulary) would be erroneous. Second, and perhaps more importantly, there should be a definition for the bilingual experience. That is to say, which linguistic experiences comprise bilingualism as compared to monolingualism? What characteristics define a bilingual individual? And how could these experiences influence foreign language learning?

In what follows, I will address these issues in turn by reviewing the relevant theoretical and experimental work from multiple perspectives. In the initial section of the introduction, I will cover what it means to learn a foreign language. In this regard, the (psycho)linguistic literature offers valuable information regarding the different levels of a language and the mechanisms by which individuals acquire them. Next, I address the definition of bilingualism and bilingual experience, where most of the literature draws from psycholinguistics and experimental psychology. Then, I combine these two ideas and address cognitive psychology and experimental linguistics research on bilingual experience and foreign language learning. In all, this chapter provides the theoretical framework for the thesis's experimental part and highlights the gaps in the literature regarding bilingual and monolingual foreign language learning.

1.1. What exactly is foreign language learning?

One can think of a language as a mental rulebook. Within this book, there are multiple interconnected chapters. The information in each chapter is relevant to other sections of this book, and therefore one cannot read them in isolation. For instance, the first chapter could take the form of a dictionary, where a collection of words (in their written and spoken forms) pair with their concepts to form a language's vocabulary. This dictionary would naturally depend on other book sections detailing how to correctly write, pronounce, and use these words in different contexts. Other chapters in this rulebook would also show how to modify the words, combine them into sentences, and convey information that cannot be transmitted using a single word. Using this analogy, learning a foreign language could be akin to acquiring and memorizing a new rulebook. Of course, there are multiple problems associated with this oversimplified analogy, and I will touch on them throughout this thesis. A critical first step is to determine the title and contents of each chapter in this rulebook analogy, provided that they are highly interconnected.

Here, I present two—out of many—naming conventions to establish a common terminology for the rest of this thesis. These are also briefly summarized in Table 1. The first naming convention comes from the linguistic perspective. It contains six analytic levels (or chapters in our rulebook analogy): phonetics/phonology, morphology, syntax, semantics, and pragmatics (Hickey, 2005; Shopen, 2007). The levels range from more fine-grained to more coarse and abstract dimensions of language. According to this convention, one can describe a language as a combination of its sounds, including the set of all possible human sounds (phonetics) and a specific language's sounds (phonology). The vocabulary and its variations (e.g., different endings) comprise the morphology level, while the syntax level encompasses multiple-word sentences. The semantic level describes meaning in the broad and abstract sense, distinguishing it from individual words (morphology) and sentences (syntax). Finally, pragmatics covers language use in specific situations and is the most abstract language level.

The second naming convention for our rulebook comes from the psycholinguistic perspective, which draws from experimental psychology and linguistics. It originates from experiments investigating how humans recognize and produce individual words (e.g., Grainger & Ferrand, 1996; Rastle, 2015). As a result, this convention further subdivides the fine-grained linguistic levels into orthographic and phonological components. Specifically, there are three additional levels compared to the linguistic levels: sub-lexical phonology, sub-lexical orthography, and lexical levels. The sub-lexical levels target the elements that comprise words, such as individual sounds and letters and their combinations. The lexical level covers the orthography and phonology of words as units. Using this convention, the vocabulary of a language would entail the sub-lexical, lexical, and even the more abstract levels (González-Fernández & Schmitt, 2020).

Table 1. *Analytic levels of language.*

Linguistics	Description	Psycholinguistics	Description
Phonetics & Phonology	All sounds in a language	Sub-lexical phonology	Sounds & combinations
Morphology	Words and endings	Sub-lexical orthography	Letters & combinations
		Lexical	Orthography and phonology of words
		Morphology	Word variations
Syntax	Clauses and sentences	Syntax	Clauses and sentences
Semantics	Meanings	Semantic	Meanings
Pragmatics	Language use	Pragmatics	Language use

Note. Domains are ordered from more fine-grained (top) to more coarse aspects of language (bottom).

Given this terminology, foreign language learning refers to how individuals acquire, at least partially, constructions (or elements) from one or more analytic levels in a new language. A foreign language differs from a native language because it is indigenous to an individual's country (e.g., English in Spain). Typically, an individual's fluency in a foreign language—their proficiency—is a measure of their ability to comprehend and produce distinct constructions correctly. There are four main fluency domains in foreign language learning: reading, listening, writing, and speaking. It is easy to see how these fluency domains involve all analytic levels and fundamentally depend on vocabulary knowledge. To put this idea into perspective, the Council of Europe (Council of Europe, 2001) estimates that proficient speakers of a foreign language should control around 10,000 base words (without considering morphological variations), including words not used in everyday conversations. Consequently, although not explicitly an analytic level, vocabulary knowledge is an essential aspect of foreign language learning, tapping into multiple levels.

I will consider the analytic levels mentioned above as internal dimensions in the context of foreign language learning. Some constructions may be more or less challenging to learn, but their overall acquisition difficulty is invariant for a specific language. In contrast, external dimensions influence how a particular individual learns a foreign language but lie outside the internal analytic levels. These dimensions can include aspects such as instruction method, individual motivation, age, and, more importantly for this thesis, prior linguistic experiences such as bilingualism, among many others. Referring back to the rulebook analogy, how fast parts of each chapter are acquired could differ from individual to individual based on these external dimensions.

On a relevant side note, the literature conveniently refers to our rulebook analogy—or at least the part related to vocabulary—as the mental lexicon (Acha & Carreiras, 2014; Aitchison, 2012; Baxter et al., 2021). In other words, the mental lexicon is the brain's storage for words, meanings, morphology, and other aspects (e.g., collocations, pragmatics). This concept will become more relevant in Chapter 4, where I discuss the bilingual mental lexicon's theoretical and computational models. Before, however, I focus on the implicit and explicit mechanisms for language learning.

1.1.1. What mechanisms sustain language learning?

Logically, one cannot simply memorize all possible constructions in a new language. Instead, learning a foreign language requires acquiring a set of rules, abstracting and generalizing them to possibly infinite constructions. That is, there are many ways in which one can transmit the same meaning using different sentences. It does not make sense to memorize each possible sentence, but general rules to combine words and produce them correctly. This property, known as generativity (Corballis, 1992), is perhaps one of the defining characteristics of human languages.

In this regard, there are two related mechanisms through which humans learn constructions in a language. The first mechanism is implicit learning. It is defined as the unconscious acquisition of knowledge about structure in the environment (Ellis, 2002, 2015). Through this mechanism, individuals are sensitive to and can passively learn from frequent and repetitive patterns. Implicit learning seems to be a ubiquitous and domain-general mechanism that applies to different sensory modalities (Kirkham et al., 2002). That is, individuals can implicitly learn from multiple input types, such as music (Ponsford et al., 1999), motor sequences (Masters et al., 2020), visual patterns (Kirkham et al., 2002), and, critically, written and spoken language (Romberg & Saffran, 2010), among others.

A large body of research shows that both infants and adults can implicitly acquire information from structured linguistic material at all analytic levels (e.g., Ellis, 2002; Misyak & Christiansen, 2012; Romberg & Saffran, 2010; Saffran, 2003). The literature employs two interchangeable terms for this process, one is implicit learning, and the other one is statistical learning (Christiansen, 2019; Perruchet & Pacton, 2006). The "statistical" part refers to the statistics individuals unconsciously track in the input to learn from it. Language learners can acquire different aspects of a foreign language through statistical learning. These include, among others, sub-lexical units (Maye et al., 2002; McMurray et al., 2009), morphological rules (Rebecca Frost & Monaghan, 2016; Peña et al., 2002), and vocabulary (Smith & Yu, 2008; Yu & Smith, 2007). It is thought that a large portion of native and foreign language learning hinges on implicit learning (Cleeremans et al., 1998). However, learners cannot acquire the totality of a new language's constructions simply through implicit/statistical learning.

The second mechanism is explicit learning. This mechanism requires conscious processing from individuals to memorize and produce novel constructions in a foreign language and is usually related to explicit instruction (Ellis, 2015). In other words, there is an effort to allocate attentional and memory resources towards acquiring different constructions, including both rules and specific instances. Typically, explicit learning involves producing multiple erroneous constructions and correcting mistakes to internalize the rules of a foreign language (Varnosfadrani & Basturkmen, 2009). In this sense, some authors suggest that explicit learning requires attentional and hypothesis testing (trial-and-error) processes (Ellis, 2015; Perruchet & Pacton, 2006). In all, learners explicitly acquire constructions in a foreign language by deploying their attentional mechanisms and utilizing trial-and-error methods. The ample and rich literature on explicit learning covers different analytic levels (for reviews, see Ellis, 1993, 2015; Rosenshine, 1986), but an in-depth review is beyond the scope of this thesis, which concerns bilingual and monolingual foreign language learning.

Now, what is the limit between implicit and explicit learning? First, it is essential to note that language learning is not purely implicit or explicit. Some aspects might be learned beyond the realm of consciousness, with individuals not realizing they have acquired specific structures. On the other hand, there might be a conscious effort from individuals to acquire some knowledge. Nevertheless, what is learned through explicit methods can serve as a baseline for subsequent implicit learning and vice-versa (Ellis, 2015). Second, learners can acquire content simultaneously through both implicit and explicit mechanisms. For instance, an individual might learn the words ***unbreakable*** and ***untouchable*** separately and through explicit mechanisms. However, they might implicitly discover meaning in the affixes ***un-*** and ***-able***, subsequently understanding other words like ***undo***, ***doable***, ***undoable***, and even ***unable***. Importantly, these mechanisms have different relevance throughout life. Since the focus of this thesis is on adult language learners, it is necessary to highlight some of the differences between children and adult foreign language learning.

1.1.2. What is different for adults?

It is well-documented that age is an external dimension that impacts learning ability, in general, and language learning, in particular (DeKeyser, 2000; Scovel, 2000). It has been suggested that infants and children can acquire a foreign language implicitly and by imitation, using mostly memory-based processes (Nikolov & Djigunović, 2006). In contrast, adult learners often struggle to learn a foreign language, mainly relying on explicit rule-based learning (Muñoz, 2008). These contrasting findings seem to suggest the presence of a critical period (i.e., Critical Period Hypothesis; Scovel, 2000), whereby children are more successful in acquiring constructions in a foreign language than adults.

While the existence and exact age for this critical period is still a matter of debate, adults can still achieve high proficiency in a foreign language provided they have the opportunity, the willingness, and, most importantly, the time to learn (Chen & Hartshorne, 2021; Hartshorne et al., 2018). For instance, in an influential study, Harthorne et al. (2018) tested around 600,000 native and non-native English speakers using an online English sentence completion test. Their results indicated that performance in this test (i.e., English proficiency) was lower for non-native participants who reported starting to learn English after 17-18 years. This critical age is not a coincidence and might not be related to neuronal decay as previously thought (Birdsong & Molis, 2001). It is around this time that individuals transition into university or the workforce. Hence, there is usually less time and willingness from individuals at this stage to learn a new language.

Crucially, while infants can learn a foreign language and potentially develop native-like proficiency, older children and adults tend to filter the foreign language through their native language. This filter—better known as cross-linguistic transfer—can positively and negatively affect foreign language learning and is indirectly related to age (Alonso, 2016). For example, a Spanish speaker learning English might find the word ***vocabulary*** easier to learn than the word ***country***, even if they both occur frequently. This is because the former has a translation equivalent with similar meaning and form (***vocabulario***). These types of words, known as cognates, can be an excellent aid for individuals to jumpstart their vocabulary during the early stages of foreign language learning and are examples of positive cross-linguistic transfer (Hayakawa et al., 2020; Marian et al., 2021). Conversely, English speakers might struggle while learning to read aloud in Spanish, as English is an opaque language with one-to-many grapheme-phoneme mappings (Rafat, 2016). The difficulties caused by the differences between the native and foreign languages are known as negative cross-linguistic transfer.

Having defined foreign language learning, its multiple analytic levels, and the implicit and explicit learning mechanisms, I now focus on the definition and measurement of bilingual experiences.

1.2. Defining and measuring bilingual experiences

Bilingualism could be loosely defined as the ability to speak two languages, acquired simultaneously during childhood or by sequentially learning a foreign language later in life (Costa & Sebastián-Gallés, 2014). This definition often appears in juxtaposition to monolingualism, which refers to individuals that only speak one language. Although the definitions may vary depending on the author and the context where the term is used, bilingual individuals must possess specific characteristics. These include proficiency in both languages, the ability to switch between them in distinct situations (code-switching), and some contact or identification with the culture of both languages, among others (Moschkovich, 2007). Bilingualism can refer to individuals that learn a foreign language or a second native language as their second language. For example, a Spanish-English bilingual from Madrid speaks a native and a foreign language, but a Spanish-Basque bilingual from the Basque Country speaks two native languages.

There are differences in how bilingual individuals develop depending, among other factors, on their second language's age of acquisition and proficiency (Marian & Hayakawa, 2021; Place & Hoff, 2011). Table 2 shows four possible configurations for bilingual individuals depending solely on their age of acquisition and proficiency. Balanced bilingualism occurs if there is similar proficiency in the first and second languages (L1 and L2). Most commonly, however, an individual will develop one language more than the other, leading to unbalanced bilingualism (Place & Hoff, 2011). These definitions often intermix with the amount of exposure to each language. For example, balanced bilinguals might experience the two languages in similar proportions, while unbalanced bilingualism might use their L1 or L2 more in specific contexts.

Table 2. *Types of bilingualism according to the age of acquisition and proficiency in the L2.*

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">L2 Proficiency/Exposure</td></tr>
<tr><td>Low</td><td>High</td></tr>
<tr><td rowspan="2">L2 age of acquisition</td><td>Early (< 6)</td><td>Simultaneous unbalanced bilingual</td><td>Simultaneous balanced bilingual</td></tr>
<tr><td>Late (> 10)</td><td>Sequential unbalanced bilingual</td><td>Sequential balanced bilingual</td></tr>
</table>

Similarly, the age of acquisition of the L2 plays a fundamental role in the consolidation of bilingualism, distinguishing between those who are simultaneous (early) and sequential (late) bilinguals (Carlson & Meltzoff, 2008). As I mentioned earlier, some findings suggest that the linguistic performance of late bilinguals is limited by the decline of neural plasticity due to the neuronal maturation that occurs after adolescence (Birdsong & Molis, 2001). Still, individuals can achieve high proficiency in a foreign language even if they learn it late in life (Chen & Hartshorne, 2021; Hartshorne et al., 2018; Nikolov & Djigunović, 2006). However, the strategies for language acquisition in adulthood might differ due to how the two languages are acquired in infancy (DeKeyser, 2000). Because of this, there might be differences in the processing and control of the L2, with a shared system for both languages in early bilinguals and separate, L1-dependent systems for each spoken language in late bilinguals (Struys et al., 2015). I will further discuss whether there are one or two language systems in the bilingual mind in Chapter 4.

The two factors mentioned above are not necessarily the only ones that can distinguish between different bilingual experiences. For example, bilinguals could differ in how they acquired the two languages, distinguishing between formal and informal instruction. Additionally, bilinguals can vary in their cultural identification, code-switching, and many other factors (Marian & Hayakawa, 2021). Regardless, it is essential to consider that these are not dichotomous categories, and individuals can vary continuously alongside any of these factors. For example, a bilingual individual might have learned their second language primarily through informal instruction at a relatively early age (e.g., 7) and show intermediate proficiency in the L2.

1.2.1. Subjective and objective measures of bilingual experience

An equally important matter is how to measure bilingualism. According to the definitions offered above, to recruit bilingual participants for an experiment, one could simply ask whether they consider themselves bilingual. Although valid, this method would not necessarily yield information from other important factors such as language exposure, age of acquisition, or proficiency. Alternatively, besides their overall perceived bilingual experience, one could also ask specific questions about individual language proficiency, use, and other variables. The problem with this second approach is that measuring bilingualism using a comprehensive test would require additional time to capture all the nuanced aspects of the bilingual experience.

As a result, many researchers have developed short self-report questionnaires to measure specific aspects of the bilingual experience (for an overview; see Marian & Hayakawa, 2021). For example, the Language Experience and Proficiency Questionnaire is a short self-report test that has been used extensively in bilingualism research (Kaushanskaya et al., 2020; Marian et al., 2007). This questionnaire asks questions about L1 and L2 proficiency, age of acquisition, exposure, cultural identification, and even perceived accent, among many others. Tests such as these rely primarily on individuals remembering and reporting their language history accurately and without bias. Notably, some of these variables (e.g., language exposure, cultural identification) cannot be measured without self-report, as there is no objective way to obtain them. Nevertheless, some authors argue that the self-reported proficiency scores in these tests correlate with other—more objective—measures of bilingual proficiency (Gollan et al., 2012). However, self-report questionnaires should not replace these objective measures, and instead serve as a quick pre-screening of bilingual experiences before conducting more comprehensive tests.

Other authors have employed objective proficiency measures of participants' L1 and L2 (e.g., de Bruin et al., 2017; Gollan et al., 2012; Sheng et al., 2014). Usually, these objective tests include some receptive and productive vocabulary measures in individuals' L1 and L2, providing more precise estimates of their proficiency. For instance, some tests include tasks where individuals need to name pictures with increasing difficulty, offering an efficient way of measuring L1 and L2 productive vocabulary (de Bruin et al., 2017; Gollan et al., 2012). Still, unless these measures are used as a pre-screening, discarding participants before they complete any experiment is very likely. Therefore, in the experimental part of this thesis, I use a combination of self-reported and objective measures to measure bilingualism.

With these definitions as a basis, I now turn to the main idea of this thesis: how bilingualism could influence foreign language learning over monolingualism.

1.3. Bilingualism and foreign language learning

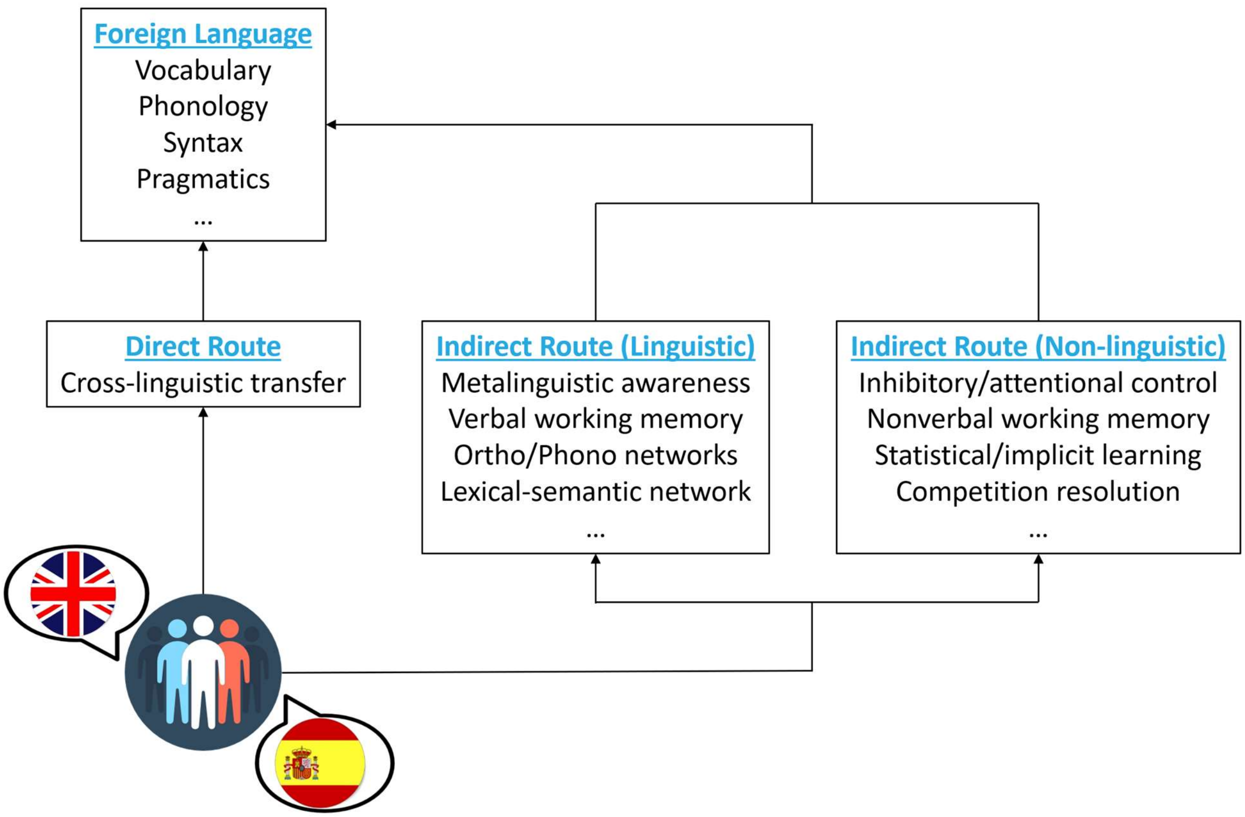
The idea that knowing more languages facilitates subsequent language learning spans decades (Cenoz, 2003; Festman, 2021; R. Nation & Mclaughlin, 1986). It is thought that the less predictable linguistic environment faced by multilinguals—a more general term for speakers of more than one language—might induce them to more efficiently explore and acquire new linguistic information than monolinguals (Festman, 2021; Filippi et al., 2019). In this regard, the experimental work on this idea generally falls into studies that target general or specific aspects of foreign language learning.

On the one hand, research focused on measuring the general proficiency (i.e., writing, reading, speaking, listening) highlights that bilinguals may use different strategies during foreign language learning than monolinguals (e.g., Cenoz, 2013; Tuncer, 2009). These findings hint towards an additive effect of bilingual experience on foreign language learning, whereby knowing more languages leads to an increased linguistic repertoire available during learning (Cenoz, 2003, 2013). These studies do not propose any additional mechanism by which bilingualism might foster language learning over monolingualism. Instead, they suggest that this additive effect relies primarily on cross-linguistic transfer from a more extensive knowledge pool in bilinguals.

On the other hand, some researchers have examined the differences between monolinguals and bilinguals on specific analytic levels, the more predominant being vocabulary (e.g., Antoniou et al., 2015; Kaushanskaya & Marian, 2009b, 2009a). These studies suggest that bilingual experience improves vocabulary learning outcomes over monolingualism by either strengthening the phonological system, the lexical-semantic network, or inhibitory control mechanisms or making these more flexible to accommodate new information (Kaushanskaya & Marian, 2009b; Kaushanskaya & Rechtzigel, 2012; Yoshida et al., 2011). However, due to these typically ad hoc explanations, the mechanisms underlying the differences between monolingual and bilingual in foreign language learning remain largely understudied.

To review and integrate all these findings into a common framework, Hirosh and Degani (2018) proposed two broad and complementary routes through which multilingualism could facilitate foreign language learning. Figure 1 illustrates these two routes. The first route involves a ***direct*** transfer of linguistic information from any (or all) available languages to the foreign language. Studies show that bilinguals can transfer information from known languages to a foreign language, provided that they are similar in structure (Festman, 2021; Rothman, 2015). This cross-linguistic transfer is the central tenet of linguistic theories like the Linguistic Proximity Model, which proposes incremental learning of languages with positive and negative influences (Westergaard et al., 2017). In other words, bilinguals can use the knowledge in any of the languages they know during learning to find similarities to exploit (i.e., positive transfer), but these might also lead to negative transfer. I offer a brief overview of this and other linguistic theories of bilingual transfer in Section 1.4.

Figure 1. *Direct and indirect effects of multilingualism on foreign language learning.*



Note. Figure adapted from Hirosh & Degani (2018), p. 893. The bottom left depicts Spanish-English bilingual individuals. Ortho = orthographic; Phono = phonological.

The second route through which multilingualism affects foreign language learning is ***indirect***. Simply put, the multilingual experience might potentiate linguistic or non-linguistic cognitive abilities, improving foreign language learning outcomes in turn (Hirosh & Degani, 2018). The linguistic abilities include metalinguistic awareness, verbal working memory, and lexical-semantic network strength, among others. For example, a Spanish-English bilingual might possess enhanced metalinguistic awareness—the knowledge about language itself—due to how they learned their L2, compared to a Spanish monolingual.

This ability could allow them to acquire the constructions they are trying to learn in a foreign language faster (e.g., Rauch et al., 2012). With a few notable exceptions (i.e., verbal working memory), these linguistic abilities tend to be challenging to quantify as they are significantly language-specific (Bialystok et al., 2014; Serratrice et al., 2009). In other words, it would not be possible to measure metalinguistic awareness as a general ability without individuals already knowing a specific foreign language. As a result, studies have yet to explore (1) how to properly quantify these skills, (2) whether bilinguals and monolinguals differ in these skills, and (3) how these skills mediate foreign language learning.

Compared to the linguistic abilities mentioned above, the non-linguistic cognitive abilities are perhaps even more controversial. Many studies suggest that bilinguals might possess enhanced cognitive abilities over monolinguals (for reviews, see Adesope et al., 2010; Kroll & Bialystok, 2013; Schroeder & Marian, 2017). The typical narrative is that—due to their constant practice of inhibiting/selecting which language to use—bilinguals might show enhanced non-linguistic cognitive abilities over monolinguals, such as conflict monitoring, inhibitory or attentional control. These results have been under severe scrutiny recently due to theoretical and methodological issues (e.g., Blanco-Elorrieta & Pylkkänen, 2018; de Bruin et al., 2021; Dick et al., 2019). Thus, both the existence and relevance of these findings have been rendered moot. Moreover, as in the previous case, it would require substantial research to understand how bilingual experiences foster these non-linguistic cognitive abilities and how these abilities influence foreign language learning in turn.

In sum, the literature has established the direct (i.e., cross-linguistic transfer) effects of bilingualism on foreign language learning. Intuitively, knowing more languages can provide a more extensive pool of knowledge from where to draw similarities during learning. Unfortunately, though, the indirect effects remain understudied at best or questionable at worst. Critically, out of the non-linguistic abilities listed in Figure 1, statistical/implicit learning seems like a prime candidate to investigate in this thesis for three reasons. First, as a cognitive mechanism, statistical learning seems to be an essential method for implicit foreign language learning (Romberg & Saffran, 2010). Second, as an experimental, statistical learning task can facilitate the measurement of bilingual and monolingual implicit language learning, and only a handful of studies have compared monolingual and bilingual statistical learning performance. Third, statistical learning tasks can capture different analytic levels of a language as a proxy by modifying the stimuli used in each task. Hence, Chapter 2 focuses primarily on statistical language learning tasks to measure bilingual and monolingual implicit/statistical foreign language learning ability. Before, however, it is vital to examine the evidence in favor of and against bilingual experience influencing foreign language learning.

1.3.1. Empirical evidence

This subsection reviews the extant research targeting bilingual and monolingual foreign language learning at different analytic levels. Since this is currently an active field of research, this review is far from exhaustive (for more comprehensive recent reviews, see Festman, 2021; Hirosh & Degani, 2018; Montanari, 2019). Instead, I primarily cover the main experimental work addressing differences, or lack thereof, between bilingual and monolingual adults at each analytic level. For the sake of brevity, I combine certain psycholinguistic aspects to target vocabulary (lexical-semantics) and grammar (morpho-syntax) as knowledge domains in a foreign language. Moreover, I primarily focus on experimental studies with bilinguals and monolinguals but avoid, wherever possible, those involving multilingual participants. The research involving multilingual individuals, although informative, has additional confounding factors, such as the interactions between multiple languages and a general lack of control measures (i.e., age of acquisition, proficiency, exposure, general intelligence) for these participants.

1.3.1.1. Sub-lexical phonology: Phonology/Phonetics

Learning sub-lexical phonology can refer to acquiring new words that differ in phonology at the syllable level, learning to discriminate or produce new phonemes, or incidentally (implicitly) learning to segment words with foreign phonology from continuous speech. Here, I will discuss studies covering these three types of tasks. Naturally, if an individual already knows how to pronounce or discriminate specific phonemes—due to their prior linguistic knowledge—, they should be better overall in these three tasks. Beyond these direct effects, I also present studies showing differences between bilinguals and monolinguals and arguing in favor of indirect effects.

Antoniou et al. (2015) compared monolinguals (English) and bilinguals (Mandarin-English and Korean-English) across two vocabulary learning experiments. That is, participants learned eight new names (words) for eight objects. The critical manipulations in these experiments were the phonological patterns that comprised the words, which were only one syllable long or monosyllabic. Pairs of syllables in the vocabulary differed in one phoneme, and were either English-like (voiced fricative; e.g., /ɸ/ vs. /β/), Mandarin-like (i.e., retroflex; e.g., /t/ vs. /ʈ/), or Korean-like (i.e., lenition; e.g., /θ/ vs. /θˈ/) minimal pairs. In the first experiment, Mandarin-English bilinguals outperformed monolinguals in learning both the English-like and Mandarin-like minimal pairs. In the second experiment, both Mandarin-English and Korean-English bilinguals outperformed the monolinguals in the Mandarin-like minimal pairs. However, only the Korean-English group could distinguish between the more challenging Korean-like minimal pairs. Their results indicated that general bilingual experience (in terms of enhanced phonological network) and phonetic similarity (derived from experience with Korean) could influence learning new phonology over monolinguals.

Other studies have shown minimal or no differences between monolingual and bilingual children and adults when learning to discriminate new phonemic contrasts (Polka et al., 2001; Sundara et al., 2006; Tremblay & Sabourin, 2012). These tasks typically involve determining whether a heard sound belongs to a phonemic category (e.g., Hindi's retroflex t versus dental stop t). For instance, Tremblay & Sabourin (2012) tested the phonemic discrimination performance of English monolingual, English-French bilingual, and multilingual adults using pre-training and post-training tests. Before training, there were no differences between the groups. After training, only the multilinguals—but not the bilinguals—performed the discrimination task better than the monolinguals.

Wang and Saffran (2014) compared monolinguals (English and Mandarin) and bilinguals (Mandarin-English and English-Spanish) in a statistical learning task. The task combined syllables and tones to simulate the experience of learning a tonal language as a foreign language. The artificial language contained three trisyllabic words (i.e., words with three syllables) with varying tones. After familiarizing themselves with the artificial language, participants heard a series of word pairs (one target and one foil) and selected which one belonged to the artificial language they had previously heard. Their results indicated that both bilingual groups outperformed the monolingual groups, regardless of their experience with tonal languages. Again, the authors of this study pointed at bilingual experience conferring indirect advantages in terms of the phonological network (Wang & Saffran, 2014). In other words, experience with the different phonology of two languages might confer an advantage when segmenting words from the continuous speech in a foreign language.

To summarize, the results seem to be mixed in terms of the influence of bilingualism on explicit and implicit sub-lexical phonology learning, with some authors pointing to both direct transfer and indirect (phonological network) effects and others no differences between the groups. One crucial factor in all of these experiments is that the number of participants tends to be very low, with 24 participants per group at most but less than 15 participants on average. As a result, it is still unclear whether bilingualism could influence sub-lexical phonology learning beyond the cross-linguistic transfer effects.

1.3.1.2. Sub-lexical orthography: Literacy and Orthotactics

Under the sub-lexical orthography level, I consider studies where individuals learned to read and write in a foreign script (literacy) or vocabulary learning tasks where the words have varying orthographic patterns (orthotactics). Some of these studies target bilingual individuals who know two scripts, known as biliteracy. For example, a Spanish-Mandarin bilingual knows an alphabetic (Spanish) and a logographic (Mandarin) script.

Most studies on monolingual and bilingual literacy skills involve infants, children, and adolescents learning new scripts. In brief, some suggest that bilingualism or biliteracy can facilitate spelling and decoding in a new script over monolingualism (Clyne et al., 2004; Kahn-Horwitz et al., 2011; Schwartz et al., 2014; Trapman et al., 2014), whereas others have shown comparable performance between the groups (Van Gelderen et al., 2003). Although these findings imply that bilinguals might possess a more flexible orthographic system than monolinguals, it is essential to consider that other critical socio-cultural factors might be involved (Montanari, 2019). For instance, immigrant and immersion bilinguals might be more willing to communicate and have lower anxiety when facing a foreign language than local monolinguals (Mady, 2014; Schwartz et al., 2014). Thus, socio-cultural factors might have a more considerable effect on ultimate learning achievement than bilingual experience in itself.

The research on adult monolingual and bilingual literacy is surprisingly very scarce. Of note, Modirkhamene (2006) compared the English reading comprehension of adult bilingual (Turkish-Persian) and monolingual (Persian) speakers three times over two years of their classroom learning studies. Persian possesses a Perso-Arabic script, and the Turkish script contains 29 letters compared to the 26 in the English alphabet. They found that the bilingual participants outperformed their monolingual peers over time, suggesting that bilinguals could better acquire the decoding and comprehension skills to succeed in this test (Modirkhamene, 2006). Similarly, another study found that Turkish-Farsi bilinguals outperformed Farsi monolinguals in their English writing skills (Poorebrahim et al., 2020). Although other mechanisms might be involved, it is easy to see how Turkish knowledge might have directly benefited the outcomes in these studies because of its similar script to English.

A growing literature suggests that orthotactics—that is, the pattern of combinations between characters in words—can facilitate vocabulary acquisition (Hayakawa et al., 2020; Marian et al., 2021). These studies consistently show that words similar in their orthotactics to other known words can be acquired more quickly than dissimilar words (Bartolotti & Marian, 2017; Hayakawa et al., 2020; Marecka et al., 2021). For instance, the non-word ***londa*** would be easier to acquire than the non-word ***bupto*** for a Spanish speaker, as the former is more similar to other existing words (e.g., *lenta*, *linda*, *luna, lonja*) than the latter. In this regard, while it has been shown that bilinguals can exploit the similarity to any of their known languages (Bartolotti & Marian, 2017), studies have yet to explore whether bilinguals and monolinguals differ in learning these types of words. One study, however, has reported no differences between multilingual and monolingual participants using a similar vocabulary learning task (Hayakawa et al., 2020). Hence, in experiment 6 I compared the performance of monolingual and bilingual speakers while learning vocabulary with varying orthographic similarities to Spanish. This experiment targets both the sub-lexical orthography and the lexical-semantics levels.

In sum, adult monolingual and bilingual sub-lexical orthography learning remains largely unexplored. Research with children and adolescents suggests that other socio-cultural factors provide better indicators of learning achievement than bilingual experience. Moreover, the few studies targeting monolingual and bilingual adults point more towards an effect of cross-linguistic transfer (i.e., experience with the target script) than bilingual experience. Whether the experience with one or multiple orthographic systems could facilitate acquiring a novel script remains an open question.

1.3.1.3. Morpho-syntax: Grammar

Morpho-syntax encompasses the variations of words (morphology) and the rules by which multiple words combine to form clauses and sentences (syntax). These aspects are collectively referred to as ***grammar***. The research on monolingual and bilingual grammar learning is limited, partly because learning grammar requires prior knowledge of other fine-grained aspects at the sub-lexical, lexical, and even semantic levels (Hirosh & Degani, 2018; Montanari, 2019).

Nation and McLaughlin (1986) tested the grammar learning performance of English monolinguals, bilinguals, and multilinguals (with varying languages) using artificially generated character strings that followed two underlying sets of grammatical rules. A subset in each group was told that the character strings followed specific rules, which they had to discover (explicit condition). Conversely, the remaining participants did not receive any instructions and passively visualized the strings (implicit condition). After seeing the strings, all participants performed a grammaticality judgment test where they had to indicate whether a series of character strings were correct or not. Their results revealed no differences between the groups in the explicit condition. Multilinguals—and not bilinguals—outperformed the other two groups in the implicit condition. The authors suggested that the multilingual advantage was related to increased grammatical sensitivity (R. Nation & Mclaughlin, 1986). A similar study used English grammar instead of artificial character strings, also indicating a multilingual advantage for grammar learning in adolescents (Klein, 1995). Regardless, there have been no reported differences between monolinguals and bilinguals in these tasks.

A more recent study compared an older population (older than 60) of English monolinguals and English-Spanish bilinguals that explicitly or implicitly learned basic Latin grammar (Cox, 2017). They used four tasks to measure learning outcomes: written and auditory sentence interpretation, grammaticality judgment, and sentence production. Overall, their results showed that bilinguals outperformed monolinguals in both sentence interpretation tasks, irrespective of whether they received explicit or implicit instructions. However, there were no differences between the groups in the other two tasks. They attributed these results to bilinguals having an enhanced metalinguistic awareness than monolinguals. An alternative explanation is that knowledge of Spanish—a Latin-derived language—could have directly influenced the learning of Latin grammar.

Finally, another study compared Mandarin-English bilinguals against English monolinguals in an artificial grammar learning task (Grey et al., 2018). After receiving some instruction in the novel language, the participants performed comprehension and production tasks and a grammaticality judgment task. They then performed the comprehension and production tasks for 20 blocks to practice extensively in the language. Their results showed that bilinguals and monolinguals did not differ in any of the tasks. This result was consistent when testing the participants at a low or a high proficiency level in the artificial language and comparing them through the comprehension and production learning blocks. In other words, at least at the behavioral level, there was no evidence of bilingual experience conferring advantages over monolingual experience.

In all, these few studies point towards a benefit of multilingualism—but not necessarily bilingualism—on novel grammar learning. Substantial research is necessary to investigate other aspects of morpho-syntax and examine whether bilinguals and monolinguals genuinely differ. Therefore, in experiments 2, 3, and 5, I tested implicit and explicit learning of morpho-syntactic information.

1.3.1.4. Lexical-semantics: Vocabulary

In this subsection, I consider studies that target how individuals acquire the mapping between foreign words and their meanings, known as vocabulary learning. Vocabulary learning is the area that has received the most attention from the literature (Festman, 2021; Montanari, 2019). Admittedly, some of the studies I have reviewed thus far fall into the lexical-semantics level. However, their designs and purpose more appropriately addressed other analytic levels of language learning rather than vocabulary specifically.

In two seminal studies, Kaushanskaya & Marian (2009a; 2009b) compared the performance of bilinguals (English-Spanish English-Mandarin) and English monolinguals when learning foreign words paired with their English translations. The words were artificial and had unfamiliar phonology created for these experiments. Additionally, a subset of the participants only heard the new words and saw their English translation (unimodal condition). In contrast, others heard and saw the new words alongside their translation (bimodal condition). Their results indicated that the bilingual group outperformed the monolingual group when recognizing and recalling the words immediately after learning them in both conditions. After a delay, bilinguals were only better than monolinguals in the bimodal condition. To explain this advantage, the authors suggested that bilingual experience either provided a more flexible mapping between orthography and phonology or lead to increased phonological working memory (Kaushanskaya & Marian, 2009b, 2009a). An equally plausible interpretation offered by other studies is that bilinguals are better at reducing phonological interference of their known languages by recruiting inhibitory control mechanisms (Antoniou et al., 2015; Bartolotti et al., 2011; Yoshida et al., 2011).

This bilingual advantage in vocabulary learning seems consistent irrespective of whether bilinguals learned their languages since birth, through formal instruction or later in life, or when matching bilinguals and monolinguals on their phonological awareness and verbal working memory (Kan & Sadagopan, 2014; Kaushanskaya, 2012; Nair et al., 2016; Van Hell & Mahn, 1997). Additionally, related studies have shown that the differences between monolinguals and bilinguals are more prominent when learning concrete as compared to abstract meanings for the words (Kaushanskaya & Rechtzigel, 2012). Consequently, it would seem that bilingual experience improves language learning outcomes by either strengthening the phonological network, the lexical-semantic network, or inhibitory control mechanisms (Hirosh & Degani, 2018; Kaushanskaya & Rechtzigel, 2012; Yoshida et al., 2011). All of these comprise the indirect effects mentioned above.

Although the evidence seems overwhelmingly in favor of a bilingual vocabulary learning advantage, other studies employing different designs have shown no differences between monolinguals and bilinguals. For instance, Bakker-Marshal et al. (2021) compared English monolinguals and Spanish-English bilinguals when learning two sets of Swahili-English word pairs over two days. Participants performed four recall tasks with increasing difficulty and a primed lexical decision task where the Swahili words were used as primes (Bakker-Marshall et al., 2021). The authors included this latter task to measure the lexical integration of the new words. Overall, their results suggested that the groups performed equally on all the recall tasks, and only the monolingual group showed the priming effect indicative of lexical integration. Admittedly, these results could be biased since participants in the bilingual group were learning the words in their second language. Regardless, these results show that it might not always be the case that bilinguals outperform monolinguals.

Other studies have focused on lexical-semantics learning using implicit tasks. For instance, Poepsel and Weiss (2016) compared English monolinguals, Chinese-English, and English-Spanish late bilinguals across two experiments using an implicit vocabulary learning task. During an initial familiarization phase, participants saw a series of novel objects on the screen and heard non-words as the names for these objects in no particular order. Thus, participants implicitly learned the name of each object by implicitly tracking the co-occurrence of a specific object and its name through multiple scenes. After the familiarization phase, they performed a test where they heard one name and decided which of two objects corresponded to that name. In the first experiment, each object only had one unique name and were thus unique one-to-one mappings. Their results here indicated no differences between the monolingual and bilingual groups in this experiment. In the second experiment, they introduced two objects with one name (two-to-one mappings) alongside the one-to-one mappings. This experiment revealed that both bilingual groups outperformed the monolingual group only in these two-to-one mappings. They concluded that this advantage might emerge either because of enhanced implicit learning in bilinguals or due to indirect effects such as inhibitory control or phonological working memory. Importantly, studies using a similar design have found no differences between monolinguals and bilinguals or minimal differences only in one-to-one mappings but not in multiple mappings (Benitez et al., 2016; Chan & Monaghan, 2019; Escudero et al., 2016). Experiment 4 of this thesis addresses these discrepancies and targets monolingual and bilingual implicit vocabulary learning.

In essence, the literature suggests a somewhat robust bilingual advantage for vocabulary learning. There are multiple interrelated explanations for this effect, including enhanced verbal working memory, more flexible phonological and lexical-semantic systems, and improved cognitive control in bilinguals than monolinguals (Montanari, 2019). Additionally, direct transfer effects could explain why bilinguals are better at mapping words to their meanings, even when there are inconsistencies between their sub-lexical orthography and phonology (Hirosh & Degani, 2018). However, some recent evidence indicates no vocabulary learning differences between monolinguals and bilinguals. Considering these findings, experiments 4, 6 of this thesis targeted implicit and explicit learning at the lexical-semantic level.

1.3.1.5. Pragmatics

The final and more abstract analytic level is pragmatics, referring to language use and meaning in different contexts. There are virtually no studies concerning how bilinguals and monolinguals acquire this aspect.

A possibility is that this highly abstract level is hard to measure objectively and even harder to manipulate using experimental learning tasks. Additionally, as mentioned before, proficiency in the more fine-grained levels is a prerequisite for acquiring this abstract level (Ellis, 2015). Therefore, in the context of this thesis, I did not address the pragmatics learning differences between monolinguals and bilinguals. Future research could further develop measures for pragmatics knowledge and experimentally study how monolinguals and bilinguals vary in learning constructions at this level.

1.3.1.6. Summary

It seems like the evidence in favor of a bilingual advantage in foreign language learning vastly overpowers the evidence against this claim. This advantage seems to rely both on direct transfer and indirect enhancement of cognitive abilities (Hirosh & Degani, 2018). As I mentioned earlier, the literature on the indirect effects does not provide a reliable basis for these differences. Furthermore, it is essential to consider that the lack of negative evidence—that is, results against this bilingual foreign language learning advantage—might be due to publication bias, as has been the case in other areas of bilingual research like the bilingual cognitive advantage (de Bruin et al., 2015). Regardless, there is sufficient evidence to uphold that bilinguals should outperform monolinguals in all experiments in this thesis. This is my central hypothesis.

Before outlining the experimental part of this thesis, the last topic that remains to cover is how the linguistic theoretical models explain the direct effects in foreign language learning in more detail.

1.4. Cross-linguistic transfer theories

The linguistic perspective on adult language learning concurs that certain aspects of a known language can transfer to the second or third languages (Alonso, 2016; Cabrelli Amaro et al., 2012; VanPatten & Benati, 2015). A crucial concept in the context of foreign language learning is that of language *typology*, which refers to the structural features or parameters of a language at all analytic levels (VanPatten & Benati, 2015). All languages vary across these parameters, with languages oscillating according to how close or distant they are (i.e., their proximity). For instance, *word order* is a typological parameter that defines the positioning of subjects, objects, and verbs in sentences. Languages can be similar or dissimilar in their predominant word order—that is, which word order occurs the most in the sentences of a language. The most prevalent word orders are SOV (subject-object-verb; e.g., Basque, Japanese) and SVO (subject-verb-object; e.g., Spanish, English). Assuming that a foreign language has a similar predominant word order as the native language, learners would struggle less to learn and understand the order of words in that language than if the word order differed.

A recurring idea in applied and theoretical linguistics is *typological universals*, which are parameters present in all languages. For example, all languages have some form of nouns and verbs, but not all languages have articles and determiners (e.g., Russian). Accordingly, a core tenet in linguistic theories of language learning is that these universal features—albeit abstract—are always transferable (VanPatten & Benati, 2015), but the transfer of other features will depend on the proximity between the native and foreign language (Foote, 2009; Rothman, 2015). Notably, the typological similarity need not be objective, as learners can perceive idiosyncratic and subjective similarities—these are collectively known as psycho-typology (Hermas, 2014). In other words, the knowledge of nouns and verbs and their relation to one another automatically transfers to the foreign language. However, other specific aspects depend on the native and foreign languages' perceived or objective similarity.

The question for experimental and applied linguistics is how similar the languages need to be for positive or negative cross-linguistic transfer to occur. Naturally, there is only one source for transfer during second language acquisition: the native language (L1). In contrast, in acquiring a third language, the questions are whether transfer can occur from any or both languages and what factors modulate this transfer. Accordingly, several linguistic models have been put forward to account for these different factors (Cabrelli Amaro et al., 2012). Although empirically contrasting these models is beyond the scope of this work, here, I offer a brief overview with regards to my central hypothesis:

L1 Transfer Model. The L1 Transfer Model essentially posits that positive and negative cross-linguistic transfer of linguistic information occurs mainly from the first language (L1), at least during the initial stages of foreign language learning (Hermas, 2014). According to this model, transferring from the L1 incurs the least cognitive effort (i.e., is more efficient), thereby inducing learners to find perceived similarities to their L1 and not other known languages. Under this model, there should not be any differences between monolingual and bilingual foreign language learning so long as they share their L1.

L2 Status Factor Model. Contrary to the previous model, the L2 Status Factor Model suggests that the second language (L2) can take a more decisive role in cross-linguistic transfer, provided that it has a higher degree of psycho-typological similarity (or status factor) to the foreign language (Bardel & Falk, 2012). The reasoning for this model is that the learner does not actively select which language to use for cross-linguistic transfer. Hence, transfer effects can occur naturally from the L2 if the learner is sufficiently proficient and if the L2 and target foreign language have similar psycho-typology. In this case, bilinguals should outperform monolinguals only if they are sufficiently proficient in their L2 to find similarities and transfer knowledge.

Cumulative Enhancement Model. This model assumes both that language learning is cumulative—that is, all known languages contribute to subsequent foreign language learning—and that a specific language can enhance learning (Flynn et al., 2004). In other words, the L1 is not exceptional, and either the L1 or the L2 can be used as sources for direct cross-linguistic transfer. The Cumulative Enhancement Model would predict not only that bilinguals would outperform monolinguals but that trilinguals would, in turn, be better than bilinguals at learning a foreign language. That is, the more languages one knows, the easier it becomes to learn more.

Typological Primacy Model. The Typological Primacy Model extended the ideas from the previous model by suggesting that cross-linguistic transfer is selective (Rothman, 2011, 2015). In other words, it is not a question of either-or—as in the Cumulative Enhancement Model—but rather “when” do learners transfer from a specific language. In this model, the more psycho-typologically similar language has “primacy” during cross-linguistic transfer. Furthermore, the transfer can occur on a construction-by-construction basis. Therefore, learners can transfer from any known language so long as a specific construction is similar to any known language. The Typological Primacy Model would predict an additive effect of language knowledge on foreign language learning (i.e., the more, the merrier).

Linguistic Proximity Model. The last model I consider is the Linguistic Proximity Model (Westergaard et al., 2017). Compared to the prior models, this account further assumes that transfer could occur from all known languages simultaneously. For instance, there could be a positive transfer from the L1 and a negative influence from the L2 operating upon the same construction (Westergaard et al., 2017). That is, learners find similarities to any language to the foreign language. These similarities range from psycho-typology to more abstract and structural similarities. Therefore, this model predicts a bilingual advantage as long as both languages do not produce opposite cross-linguistic effects on the target construction to learn.

It is critical to note that this literature does not particularly address the differences between second and third language acquisition (i.e., monolingual versus bilingual foreign language learning). Instead, the predictions from these models fundamentally rely on whether bilinguals use their more extensive pool of knowledge to transfer linguistic information during learning. Although these models make no claims about the cognitive system, they provide valuable insights regarding the interplay of languages in the learner's mind. First, according to these models—particularly the last two—the information from both languages is active and available for cross-linguistic transfer in bilinguals. Second, there does not seem to be separate systems for each language an individual acquires, and all languages cumulatively interact during subsequent learning. I will elaborate on these ideas further in Chapter 4. Third, the cross-linguistic transfer occurs from any or both languages on a construction-by-construction basis. Finally, the similarities do not need to be objective, as learners can transfer based on perceived and subjective similarities.

With these theories and predictions, I now outline the purpose and general methodology of this thesis.

1.5. The current thesis

1.5.1. Purpose of this thesis

Throughout this first chapter, I have reviewed the experimental and theoretical literature underlying potential bilingual learning advantages at different foreign language analytic levels. Briefly, while the vocabulary level has been studied extensively, other analytic levels have received little to no attention from the literature. The findings are also mixed regarding whether differences between monolinguals and bilinguals genuinely exist. Moreover, there is no reliable consensus on how these differences might emerge. Some studies posit that the differences are fundamentally due to cross-linguistic transfer, while others propose indirect cognitive effects of bilingual experience.

Accordingly, the central research question of this thesis remains unchanged: **Are bilinguals better at learning a foreign language than monolinguals?** I further divide this research question into three specific research questions distributed throughout the following chapters:

- **RQ1. At which analytic level, if any, do bilinguals and monolinguals differ?** This represents the "where" question. Under this question, I explore multiple levels of foreign language learning through experiments in Chapters 2 and 3.
- **RQ2. What are the differences?** This is the "what" question. In other words, are bilinguals always better, or just under some specific conditions? Most of the experiments in this work contain multiple conditions to address specific questions regarding these potential differences.
- **RQ3. How could these differences emerge from the bilingual experience?** This is the "how" question. In this regard, Chapter 4 focuses on understanding some of the differences found throughout the experimental part of this thesis using cognitive computational modeling.

I compared two bilingual groups (Spanish-English and Spanish-Basque) to each other and against a Spanish monolingual group in all the experiments. The general hypothesis was that the two bilingual groups would outperform monolinguals in all foreign language learning experiments. Nevertheless, to adequately address these research questions, this work needs to fulfill several requirements. First, how to ensure that participants are learning linguistic information that is sufficiently novel for them? Second, which experiments can capture foreign language learning at different analytic levels? Three, how to properly recruit comparable groups of monolingual and bilingual participants? Finally, how to address the general and specific questions with the appropriate statistical analyses? Below, I describe how I addressed these desiderata in turn.

1.5.1.1. A brief note on the use of artificial languages in language learning research

In the experimental part of this thesis, I tested adult individuals' capacity to learn constructions in a foreign language. Given that these individuals were either monolingual (i.e., they know only one language) or bilingual (i.e., they already know two languages), many possible interactions and transfer effects could occur due to their prior linguistic knowledge. Thus, I used artificial languages rather than natural languages to account for their potential prior knowledge of the target constructions to learn. In other words, instead of using any existing foreign language (e.g., Japanese, Swahili), participants in each experiment learned artificial and, therefore, unfamiliar constructions. The only exception in this thesis is Experiment 5, where the participants learned artificial suffixes using existing Spanish words as roots.

Artificial languages provide a means to tap into various analytic levels of a foreign language while still maintaining sufficient experimental control. This premise has been the hallmark of psycholinguistic research for over a century (Weiss, 2020). Artificial languages have been used extensively to investigate language learning and processing at different analytic levels (e.g., Endress & Bonatti, 2016; Ettlinger et al., 2016; Hayakawa et al., 2019; Kaushanskaya & Marian, 2009b, 2009a; Morgan-Short et al., 2014). Participants can learn artificial linguistic materials in a few hours or days, making artificial languages helpful in investigating language learning trajectories (Grey et al., 2018; Morgan-Short et al., 2014). For example, artificially constructed linguistic materials have been used in language learning research to investigate how different sub-lexical, lexical, and other more abstract aspects can influence vocabulary acquisition (for a review, see Hayakawa et al., 2019). Researchers can also carefully control aspects such as sub-lexical similarity and the influence of prior knowledge in artificial linguistic materials. Finally, studies have shown that artificial language learning performance tasks positively correlates with natural language learning performance (Ettlinger et al., 2016). Consequently, artificial languages offer an excellent resource to level the field between monolinguals and bilinguals and test their foreign language learning abilities while simultaneously reducing or controlling for the influence of their prior knowledge.

1.5.2. Overview of experiments

Table 3 shows an overview of the experiments targeting different analytic levels of a language. The details about the design, materials, and procedure are left to each experiment's specific section. The main research question remains the same throughout all the experiments—i.e., are bilinguals better? There are different specific hypotheses and predictions within each experiment. Therefore, I briefly summarize the purpose, the specific hypotheses, and the predictions for each experiment.

Table 3. *Overview of experiments by analytic level.*

<table>
<tr><th>Analytic Level</th><th>Implicit</th><th>Explicit</th></tr>
<tr><td>Sub-lexical phonology</td><td>Experiment 1</td><td>--</td></tr>
<tr><td rowspan="2">Morpho-syntax</td><td>Experiment 2</td><td rowspan="2">Experiment 5</td></tr>
<tr><td>Experiment 3</td></tr>
<tr><td>Sub-lexical orthography</td><td>--</td><td rowspan="2">Experiment 6 & 7</td></tr>
<tr><td>Lexical-semantics</td><td>Experiment 4</td></tr>
</table>

Note. The levels are grouped as in Section 1.3 and reorganized to facilitate comprehension.

Chapter 2: Implicit foreign language learning.

Chapter 2 targets implicit foreign language learning at four analytic levels using statistical learning experiments. I chose this paradigm because it allows experimentally testing multiple aspects of implicit learning using a similar design (Perruchet & Pacton, 2006; Romberg & Saffran, 2010). Experiment 1 compared the ability of bilinguals and monolinguals to segment words from three artificial speech streams. The critical manipulation was the sub-lexical phonological patterns that comprised the words in each stream, ranging from simple to complex. Therefore, the specific question was whether the groups would differ in all conditions or just in the complex conditions. The bilingual experience exposes participants to distinct phonological patterns (Antoniou et al., 2015). Therefore, I expected that bilinguals should show an advantage over monolinguals in this task.

Experiment 2 extends the findings from the previous experiment by testing participants' ability to generalize the knowledge to novel words. Specifically, I designed the artificial speech streams to mimic affixal morphology (e.g., **un**break**able**, **un**touch**able**), and participants had to generalize this knowledge to new items (e.g., **un**fill**able**). Experience with two languages might foster participants' ability to attend to language's structure and rules rather than explicit instances (Cenoz, 2013). Hence, as in the previous experiment, the overall question was whether the groups would differ in generalizing this knowledge.

Experiment 3 also targets the morpho-syntax analytic level. In this experiment, I tested participants' ability to segment words from an ambiguous speech stream. The stream targeted a specific syntactic property of languages: their word order. In this regard, participants in the Spanish-Basque group have experience with both SVO and SOV word orders. Therefore, I expected participants in this group to be sensitive to both segmentation patterns in the artificial language. The general idea was to test whether experience with two SVO languages—as is the case with Spanish-English bilinguals—would also confer advantages over experience with a single language. Hence, this experiment is vital to disentangle the overall effects of bilingual experience from the specific properties of known languages.

Finally, in Experiment 4, I targeted the lexical-semantics level using a statistical word-referent learning task. In this task, participants had to discover the names (words) of non-existing objects (referents) implicitly. They saw the objects across different scenes and heard the name of each object in no particular order. The manipulation in this experiment was the mappings between words and objects. Some objects only had one name—i.e., they were exclusive mappings—, but other objects could have two distinct names, or two distinct objects could have the same name. For example, in English, the word ***bat*** can refer to an animal or an object, and therefore this word is a homonym. Conversely, the words ***money*** and ***cash*** can refer to the same meaning, an example of synonyms. Consequently, the specific question was whether bilinguals and monolinguals would differ in learning any or all of these words. Intuitively, the bilingual experience exposes participants to multiple types of word-referent mappings, so the natural prediction is that bilinguals would outperform monolinguals in learning these three types of mappings.

Chapter 3: Explicit foreign language learning.

In Chapter 3, I addressed explicit foreign language learning using two additional experiments. These experiments jointly addressed the gap in knowledge about sub-lexical orthography and morphology learning and the discrepant findings regarding monolingual and bilingual foreign vocabulary learning. In these two experiments, participants had to learn linguistic material by using rehearsal and trial-and-error methods. The procedure involved a familiarization block followed by some tests of their recently acquired knowledge.

Experiment 5 compared monolinguals and bilinguals when learning new suffixes (e.g., break**able**). For this experiment, it was a necessary prerequisite that participants already knew the words to which these suffixes would append. Therefore, instead of using artificial words and having participants learn them beforehand, I used existing Spanish word roots and artificially constructed novel suffixes (e.g., laboral**suti**). Furthermore, since participants already knew the definitions for these words, it allowed me to focus only on how they learned the artificial suffixes' orthographic form. As in Experiment 2, the specific questions were whether bilinguals would better remember the suffixes than monolinguals. Additionally, since Basque is a postpositional and agglutinative language—i.e., information is appended to the end of words—this experiment also disentangled between overall bilingual experience and experience with a specific language. The predictions were that both bilingual groups would outperform monolinguals in an old versus new task where they saw different words with the new suffixes.

Experiment 6 combines the sub-lexical orthography and the lexical-semantics level. In this experiment, participants were tasked with learning an artificial vocabulary. The critical manipulation was that each words' sub-lexical orthography was either more similar or dissimilar to Spanish. Notably, all participants had Spanish as a common native language. Hence, the specific questions were (1) whether words more similar to Spanish would be learned better, and (2) whether orthographic similarity would interact with the bilingual status of participants. Beyond predicting that participants would learn similar words better than dissimilar words, I expected the bilingual experience effect to be independent of orthographic similarity. That is, I expected bilinguals to outperform monolinguals in learning both similar and dissimilar words.

Chapter 4: Tracing the algorithm of bilingual foreign vocabulary learning.

Chapters 2 and 3 jointly address "where" (RQ1) and "what" (RQ2) the differences between bilinguals and monolinguals are. In contrast, Chapter 4 focuses mainly on vocabulary learning and targets RQ3. In this chapter, I attempt to unify the direct and indirect accounts of bilingual and monolingual vocabulary learning. I first offer an overview of the cognitive computational modeling framework, reviewing some of the prior modeling work on the bilingual mental lexicon for word processing and learning. In short, the extant literature has mainly addressed (1) the processes involved in word processing using manually engineered computational models; or (2) how individuals acquire words in two languages using simple learning rules. Despite this, none of these models adequately accounts for the vocabulary learning differences between monolinguals and bilinguals.

Therefore, in Experiment 7, I propose and develop a new computational model specifically focused on the sub-lexical orthographic lexicon. As in Experiment 6, I employ this model to examine how orthographic similarity and bilingual experience influence foreign vocabulary learning. First, I simulate the development of bilingual and monolingual adults' orthographic lexicon. I then validate the model by replicating some well-established findings in the psycholinguistic and modeling literature. Finally, I contrast the model's performance with human behavior using an adapted version of the vocabulary learning task from Experiment 6. This experiment unifies the seemingly disparate findings of orthographic similarity and bilingual experience under a common computational framework, whereby distributed orthographic representations reside in a unified lexicon and are modified by learning experiences.

1.5.3. General Methodology

In this subsection, I describe the methodology that is common to all the experiments in this thesis. All experiments tested participants with similar linguistic profiles, namely bilinguals (Spanish-English and Spanish-Basque) and Spanish monolinguals. I start by offering a brief overview of the differences between the three languages under study (Spanish, Basque, English). Then I describe the measures employed to measure participants' linguistic experiences. Finally, I briefly cover the statistical analytic approach I employed throughout this thesis. Since my predictions and hypotheses are clearly defined—i.e., I expected bilinguals to outperform monolinguals in all the language learning tasks—, I disregarded the exploratory analyses used in prior studies in favor of confirmatory approaches, as supported by more recent literature (Schad et al., 2020).

1.5.3.1. Participants

Two bilingual groups were included in this work to isolate the effects of bilingual experience rather than specific language combinations. Spanish-English and Spanish-Basque bilinguals were compared to each other and against a Spanish monolingual group. All participants were from Spain, with Spanish-Basque bilinguals belonging to a different region (Basque Country) than the Spanish and Spanish-English groups (Murcia). The Spanish-English bilinguals knew a native and a foreign language, while the Spanish-Basque knew two native languages since Basque is a native language in the Basque Country region. The monolinguals knew one native language (Spanish). Although they might have experienced other languages, I controlled so that their knowledge and exposure to languages other than Spanish was minimal. The Spanish-English group had a more similar socio-cultural profile to the Spanish monolingual group, as they were all university students from the same university in the same region of Spain. All participants reported normal or corrected-to-normal vision and no history of hearing or other neurological disorders. The entire experimental protocol was approved by the Ethics Committee of the Basque Center for Cognition, Brain and Language (BCBL) and carried out following the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. Before their inclusion in each experiment, all subjects provided written informed consent. They received monetary compensation for their participation.

Participants needed a similar profile to make the results across experiments comparable. Therefore, before participating in the experiments, they completed a language background questionnaire, a picture-naming task, and a matrix reasoning task from the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2014) to measure non-verbal IQ. The linguistic questionnaire included information about the age of acquisition, total exposure to each language, self-rated proficiency scales for each known language, and demographic questions such as age and gender. This questionnaire is an abridged version of the LEAP-Q questionnaire (Marian et al., 2007). The picture naming task consisted of sixty-five images from the Basque, English, and Spanish Test (BEST) (de Bruin et al., 2017). Participants in the monolingual group named the images in Spanish and English, while the bilingual groups named them in both Spanish and their L2 (English or Basque). Additionally, participants in the bilingual groups completed the LexTALE test in Spanish and their respective L2 (Izura et al., 2014; Lemhöfer & Broersma, 2012). The combination of these tests provided information on receptive and productive L2 vocabulary in the bilingual groups.

I calculated the sample size a priori for all experiments using G*Power version 3.1 (Faul et al., 2009). A mixed within-between F-test was chosen as the statistic, with Group (monolingual vs. Spanish-English vs. Spanish-Basque) as between factors. The number of within contrasts oscillated between two and four to account for the different conditions in each experiment. The effect size was chosen to be conservative (Cohen's f of 0.20) in all experiments, and an expected power of 95% was selected. The alpha level was set to 0.05. On average, the estimated sample size to achieve this power level was 34 participants per group. Therefore, to err on the safe side, I collected data from around 40 participants per group in each experiment to account for any potential technical errors.

A final detail to emphasize is the differences between Spanish, Basque, and English. Table 4 summarizes the main characteristics of each language at different analytic levels that are relevant for this thesis. Naturally, a complete linguistic description of each language is beyond the scope of this work. While prior studies have emphasized the characteristics that differentiate Spanish and English (e.g., transparent vs. opaque grapheme to phoneme (G2P) mappings), it is essential to briefly comment on the commonalities and differences between them and the Basque language. For instance, Spanish and Basque do not share any common root, but Basque possesses many Spanish loan words due to their geographic proximity within Spain. Some language-specific bigrams differentiate Basque from Spanish and English (e.g., "tx", "tz"). Spanish and Basque are phonologically similar, differing fundamentally on this aspect from English. The Basque language also possesses a predominantly subject-object-verb (SOV) word order, which involves a postpositional and agglutinative morphology—i.e., morphemes and determiners are appended to the end of word roots (e.g., *eskolan* – *school the in*). These properties differentiate Basque from Spanish, English, and many other Indo-European languages (Bengtson, 2011).

Table 4. *Main differences between Spanish, Basque, and English.*

Analytic level	Spanish	Basque	English
Sub-lexical phonology	Transparent G2P mapping	Transparent G2P mapping	Opaque G2P mapping
Sub-lexical orthography	Alphabetic orthography	Alphabetic orthography	Alphabetic orthography
Morphology	Prepositional & Inflectional	Postpositional & Agglutinative	Prepositional & Inflectional
Syntax	SVO word order	SOV word order	SVO word order
Lexical-Semantics/Pragmatics	-	-	-

Note. G2P = grapheme to phoneme mapping; SVO = subject-verb-object; SOV = subject-object-verb.

1.5.3.2. Statistical analyses

The primary dependent variable of interest was participants' accuracy after learning the linguistic constructions in each experiment. I analyzed the accuracy in all experiments using generalized or linear mixed models (G/LMM). For the GLMMs, I assumed a binomial distribution and a logit link for the accuracy scores, with the group, any possible conditions, and their interaction as fixed effects of interest. This implies that the analysis modeled a function of the log-odds and not the aggregate percent correct. There are many limitations to using percentages instead of log-odds, the foremost being that percentages are sensitive to ceiling and floor effects, whereas log-odds are not (Hartshorne et al., 2018). In all experiments (except for Experiment 3), the group factor was reverse Helmert coded according to the central hypothesis. In other words, I first contrasted the two bilingual groups, expecting no differences between them. Then I contrasted both bilingual groups against the monolingual group. These factors were dummy-coded for each different contrast. I further included participants´ de-meaned Age and Non-verbal IQ and reported gender as covariates in each experiment´s model. Any other specific additional contrasts or approaches are described within each experiment's method section.

Crucially, this is a different approach from prior studies that focus on omnibus exploratory analytical methods (e.g., ANOVAs). I decided to adopt a confirmatory approach to evaluate the effects of bilingual experience in all experiments more robustly through a priori planned contrasts. This approach makes the G/LMM analyses directly interpretable without the need for corrected post-hoc tests (Schad et al., 2020). The G/LMM models were fitted using the *glmer* and *lmer* functions from the "lme4" package in R (Bates et al., 2015). Following standard practice, I tried to fit the maximal random-effects structure and reduced it to achieve convergence by eliminating the correlations between random slopes or the random slopes themselves (Barr et al., 2013). Each model's assumptions (e.g., variance inflation factor, collinearity, normality of residuals) were verified using the *performance* package for R (Lüdecke et al., 2021).

Finally, I report both the frequentist and Bayesian versions of the tests where possible. I opted to report the results from the Bayesian framework because these provide robust tests of the differences between the groups while simultaneously testing for the null hypothesis (Kruschke, 2010; van Doorn et al., 2020). Exact p-values are reported up to the 0.001 level for frequentist tests. For Bayesian tests, I report the exact Bayes Factor (BF_{10}) from 0.001 to 100. BF_{10} values below 1 indicate more support for the null hypothesis, and values above 1 support the alternative hypothesis. I used uninformative priors for all Bayesian analyses. Paired-samples, one-sample, independent-samples, and ANOVA frequentist and Bayesian tests were conducted in JASP, ensuring that the data met the corresponding assumptions (Love et al., 2019). To obtain an approximate BF_{10} from the G/LMM models, I contrasted nested models, including each fixed factor and interaction step-wise, against an intercept-only model with the same random-effects structure (Nakagawa et al., 2017; Wagenmakers, 2007). This approach is similar to the one implemented in other statistical software (e.g., JASP), except that the G/LMMs are not executed under a fully Bayesian framework. Instead, I calculated the BF_{10} using the Bayesian Information Criterion (BIC) of the intercept-only (m_0) and each fixed-effect model (m_1) with the following formula:

$$BF_{10} = \exp\left(\frac{BIC(m_0) - BIC(m_1)}{2}\right)$$

Chapter 2: Implicit foreign language learning

Theoretical motivation

The human brain is remarkably susceptible to structure in the environment. Both infants and adults can implicitly and quickly learn from patterns presented repetitively, a mechanism broadly known as statistical learning (Saffran, 2003). Statistical learning occurs irrespective of the input type (be it auditory, visual, or tactile), hinting towards a domain-general mechanism to implicitly track and learn from regularities in the environment (Kirkham et al., 2002). The field of language acquisition has taken a particular interest in *statistical language learning* (SL) as one of the primary mechanisms through which humans implicitly learn languages (Perruchet & Pacton, 2006; Romberg & Saffran, 2010). However, the extent to which prior linguistic experiences affect SL remains largely unexplored. In this chapter, I examined whether bilingualism influences performance in SL tasks. In particular, I compared monolingual and bilingual adults in four well-established SL tasks that targeted different analytic levels of foreign language learning as a proxy.

The idea of SL stems from observing the natural world. For example, given sufficient exposure to the words in the phrase *baby monkey*, infants and adults learn that the syllables within the words (i.e., *ba-by*) predict each other more reliably than the syllables at the boundary between the words (i.e., *by mo*) (Erickson & Thiessen, 2015). As an experimental task, SL consists of exposing participants to a continuous stream of artificial linguistic input and testing them on plausible—those that follow the statistics of the input—versus implausible items from the stream (Siegelman, Bogaerts, Christiansen, et al., 2017). Ideally, the participants should implicitly discover the boundaries between words solely from the statistics of the artificial language and mentally represent them as separate units. By manipulating the syllables or words themselves or the transitions between elements in an SL task, researchers have examined how individuals implicitly learn different aspects of a novel language, including segmenting words from speech (Saffran, 2003; Saffran et al., 1996), acquiring sub-lexical units (Maye et al., 2002; McMurray et al., 2009), discovering morphological rules (Rebecca Frost & Monaghan, 2016; Peña et al., 2002), and learning word-referent pairs (Smith & Yu, 2008; Yu & Smith, 2007).

Individual linguistic experiences seem to influence performance in SL tasks. For instance, studies have shown that the native language's stress pattern can interfere with SL performance in 9-month-old, but not in younger infants with less exposure to the native language (Jusczyk et al., 1999; Thiessen & Saffran, 2003). Adults also seem to struggle in SL tasks that contain syllables not plausible in their native language

(Finn & Hudson Kam, 2008). Furthermore, the SL performance of both infants and adults seems to be biased towards their native language's word order (Onnis et al., 2016; Onnis & Thiessen, 2013). Overall, it seems that the specific properties of a known language (e.g., stress patterns, word order) may facilitate or interfere with SL performance. It makes sense that prior linguistic experiences predispose the learning of novel linguistic material, as evidenced by cross-linguistic transfer studies (Alonso, 2016).

In bilingual adults—who consistently use two languages—their practice with conflicting statistics may influence their performance in SL tasks. To put this idea to the test, Wang and Saffran (2014) compared two monolingual (English and Mandarin) and two bilingual (Mandarin-English, and English-Spanish) groups in a challenging SL task that combined syllables and tones to emulate a foreign tonal language. They found that, while the monolingual groups could not perform above chance level in this task, both bilingual groups outperformed monolinguals. They concluded that this "bilingual advantage" was irrespective of experience with tonal languages and emerged from enhanced phonological working memory in bilinguals. Additional research has supported the idea of bilingual experience influencing implicit learning performance using other challenging SL tasks, such as learning a Morse Code language with interfering statistics (Bartolotti et al., 2011), simultaneous learning of two grammars (Onnis et al., 2018), and learning multiple word-referent pairs (Poepsel & Weiss, 2016). By using challenging SL tasks, all of these studies fail to disentangle the effects of SL task difficulty from those stemming from bilingual experience. Thus, there is no consensus on whether bilingual experience improves SL performance or if bilingual individuals are simply better at managing conflicting information.

Given that SL tasks can measure different aspects of foreign language learning, several questions arise from this overview. First, if bilingual experience facilitates SL in general, do bilinguals show an overall advantage regardless of the specific SL task? Second, do the effects emerge from a specific language combination or an overall bilingual advantage? Third, do these effects emerge at specific analytic levels of implicit foreign language learning? To address these questions, I compared monolingual and bilingual adults' performance across four SL experiments. The experiments targeted four analytic levels of foreign language learning as a proxy—i.e., sub-lexical phonology, morphology, word order (syntax), and vocabulary (lexical-semantics)—using well-established tasks reported in the literature.

Critically, two bilingual groups were included in this work to isolate the effects of bilingual experience rather than specific language combinations. Spanish-English and Spanish-Basque bilinguals were compared to each other and against a Spanish monolingual group. All participants were from Spain, with Spanish-Basque bilinguals belonging to a different region (Basque Country) than the Spanish and Spanish-

English groups (Murcia). While prior studies have emphasized the characteristics that differentiate Spanish and English, it is essential to briefly comment on the commonalities and differences between the Spanish and Basque languages. Spanish and Basque do not share any common root, but Basque possesses many Spanish loan words due to their geographic proximity within Spain. Some language-specific bigrams differentiate Basque from Spanish (e.g., "tx", "tz"). However, the two languages are phonologically similar. The Basque language also possesses a predominantly subject-object-verb (SOV) word order, which involves a postpositional and agglutinative morphology—i.e., morphemes and determiners are appended to the end of word roots (e.g., *eskolan* – *school the in*). These properties differentiate Basque from Spanish, English, and many other Indo-European languages (Bengtson, 2011).

Given that bilinguals experience different and potentially conflicting statistics in both known languages, the general hypothesis was that bilinguals should outperform their monolingual peers in the four tasks. However, within each experiment, there were specific questions and hypotheses. Experiment 1 compared the SL performance of monolinguals and bilinguals across three tasks. Each task contained different sub-lexical phonotactics that modulated its difficulty. Here, I expected the task's difficulty to interact with bilingual status, with bilinguals possibly outperforming monolinguals in the more complex conditions. Experiment 2 further tested participants' capacity to generalize from an SL stream. Experience with two languages—and especially with two distinct morphologies—could foster individuals' ability to discover and generalize rules from continuous speech.

Experiment 3 tested whether knowledge of different word orders could influence SL performance. In this case, I expected the Spanish-Basque group to perform differently from the other two groups due to their experience with different word orders in Spanish and Basque. I included this experiment to separate the overall effects of bilingual experience from the influence of specific properties of known languages. Finally, in Experiment 4, participants learned word-referent pairs. There were one-to-one (exclusive), two-to-one (synonym), and one-to-two (homonym) word-referent mappings, and participants learned them through a cross-situational SL task (Smith & Yu, 2008; Yu & Smith, 2007). Prior studies have suggested a bilingual advantage in explicit one-to-one word-referent learning (Kaushanskaya & Marian, 2009b). However, the defining characteristic of bilingual experience is learning multiple word-referent mappings (e.g., puerta – door/ate). Therefore, I expected three possibilities. First, bilinguals could better learn the exclusive mappings than monolinguals, but not the more challenging multiple mappings. Second, bilinguals could be better on the multiple, but not the exclusive mappings. Finally, bilinguals could overall be better than monolinguals regardless of the mapping type.

2.1. Experiment 1: Word segmentation from continuous speech

2.1.1. Rationale

Experiment 1 tested the word segmentation performance of monolinguals and bilinguals adults in three SL streams that differed in their sub-lexical phonotactic patterns. Phonotactics is an aspect of sub-lexical phonology that defines the allowed combinations of phonemes in a given language. Studies have shown that violating the phonotactics of the native language can hinder the word learning performance of infants and adults (Estes et al., 2016; Finn & Hudson Kam, 2008). For instance, exposing English-speaking adults to an SL stream containing syllables that do not exist in English (e.g., */tfobu/)* leads to lower performance than stimuli that only contain plausible English syllables (Finn & Hudson Kam, 2008). Thus, sub-lexical phonotactics—defined here as the plausibility of constituent syllables—provide an excellent way of adjusting the difficulty of SL tasks.

Bilinguals can potentially experience distinct phonotactics in both languages, affecting their learning performance (Antoniou et al., 2015; Kaushanskaya & Marian, 2009b). Consequently, I expected both bilingual groups to show an advantage compared to monolinguals in this task. Additionally, a specific question for this experiment was whether task difficulty could modulate this advantage. Therefore, participants in this experiment listened to three SL streams. The first one was constructed using simple consonant-vowel syllables, whereas the other two contained more complex consonant clusters that were either legal or illegal.

2.1.2. Methods

Participants

Forty Spanish monolinguals (M_{age} = 21.8, SD = 2.6; 37 females), forty Spanish-Basque bilinguals (M_{age} = 21.2, SD = 1.9; 35 females), and thirty-seven Spanish-English bilinguals (M_{age} = 20.9, SD = 2.3; 33 females) participated in this experiment. Before participating in the study, they completed a language background questionnaire, a picture-naming task, and a matrix reasoning (see Subsection 1.5.3). Additionally, participants in the bilingual groups completed the LexTALE test in Spanish and their respective L2. I compared the three groups using several ANOVAs with Helmert contrasts. In other words, I contrasted the Spanish monolingual group against the two bilingual groups in the first level and both bilingual groups against each other in the second level. A complete table with demographic information and statistical comparisons can be found in Appendix A1. The results suggested that the three groups did not differ in their age (p = 0.204, BF_{10} = 0.303), non-verbal IQ (p = 0.619, BF_{10} = 0.303), self-reported Spanish proficiency (p = 0.419, BF_{10} = 0.166) and Spanish picture naming (p = 0.245, BF_{10} = 0.260). As expected, the monolingual and bilingual groups differed in their exposure to Spanish, L2 age of acquisition, L2 exposure, L2 self-rated proficiency, and BEST picture naming in L2 (all $p < 0.001$). Crucially, the two bilingual groups did not differ in their L1 and L2 exposure, age of acquisition, self-rated proficiency, BEST picture naming, and LexTALE scores (all $p > 0.05$).

Materials

I constructed three auditory SL streams with eight trisyllabic words per stream. The syllables did not repeat within the streams, maintaining the transitional probabilities within the words at a constant value of 1. The eight words for each condition were randomly concatenated into continuous speech streams, with the constraint that there was no immediate word repetition and that each word appeared 80 times per stream.

The first of these streams, the *simple condition*, was only composed of consonant-vowel (CV) syllables and represented a version of the original statistical learning experiment (Saffran et al., 1996). An example of a word in this stream would be */motufi/*. In the second and third conditions—or complex conditions—, each word contained one consonant cluster (CCV) syllable, either at the beginning or at the end of a word. The difference between these conditions was the sub-lexical phonotactic patterns that defined the syllables. In the *complex legal* condition, the consonant cluster in the syllable (i.e., */fre/, /bla/, /fle/, /gli/, /gra/, /pre/, /pli/, /tre/*) were plausible at the beginning or end of a word without interfering with its segmentation (e.g., */beta<u>fre</u>/*). However, in the *complex illegal* condition, participants would hear words with consonant clusters that generally mark the boundary between syllables (i.e., */rnu/, /gma/, /lgi/, /nfu/, /rbu/, /rfo/, /sfe/, /bso/*) and were thus not plausible at the beginning or end of words (e.g., */teno<u>bso</u>/*). These consonant clusters were selected because they provided interfering cues—i.e., syllable boundaries—when inserted in an uninterrupted speech.

I synthesized the streams and each separate word using the MBROLA software (Dutoit et al., 1996), using the it3 voice, a constant pitch of 82.63Hz, and a duration of 200ms per syllable. No pauses were inserted between the words in any of the auditory streams. These constraints ensured minimal interference of segmentation cues due to varying pitch, amplitudes, or co-articulation of the syllables. Finally, following previous research, I introduced an amplitude ramp for the first and final 5 seconds of the stream (Onnis et al., 2016). It produced a fade-in and fade-out effect, giving the impression of an unbounded speech stream. The duration of each SL stream was around 7 minutes.

Procedure

Participants completed a familiarization phase followed by a 2-alternatives forced-choice (2AFC) test. They were asked to wear headphones and sit in a quiet room for the entire experiment. During the familiarization task, they listened to the artificial SL streams. As in prior studies, they were instructed to pay close attention to the made-up languages because later, they would answer questions about them (Saffran et al., 1999; Wang & Saffran, 2014). Each condition's order of presentation was counterbalanced for each participant. The 2AFC test immediately followed the familiarization phase in each condition. During this test, participants heard two words (one target and one foil) separated by 500 ms of silence and decided which word belonged to the previously heard language. The trials and the order of presentation of target and foil words were fully randomized. They responded using the keyboard ("f" and "j" keys) to indicate which word they thought was more similar to the previously heard language. The computer automatically recorded their accuracy during the experiment.

I included two types of foils during the test phase, synthesized in the same manner as the original words. The first type was non-words, created by inverting the syllables of each word in the stream. The second type of foils was part-words. I created these foils by combining the ending of one word and the beginning of another, maintaining the correspondence between syllables. For instance, at test, a word in the simple condition (*/bukoni/*) would be paired both with a non-word (*/nikobu/*) and with a part-word (*/konito/*) in separate trials. The inclusion of two types of foils allowed us to double the number of trials per condition to sixteen. With three familiarization and three test phases, the entire experiment lasted around 30 minutes. Participants were encouraged to take a small break between each condition. The tasks for this and the rest of the experiments were programmed in Psychopy (Peirce et al., 2019).

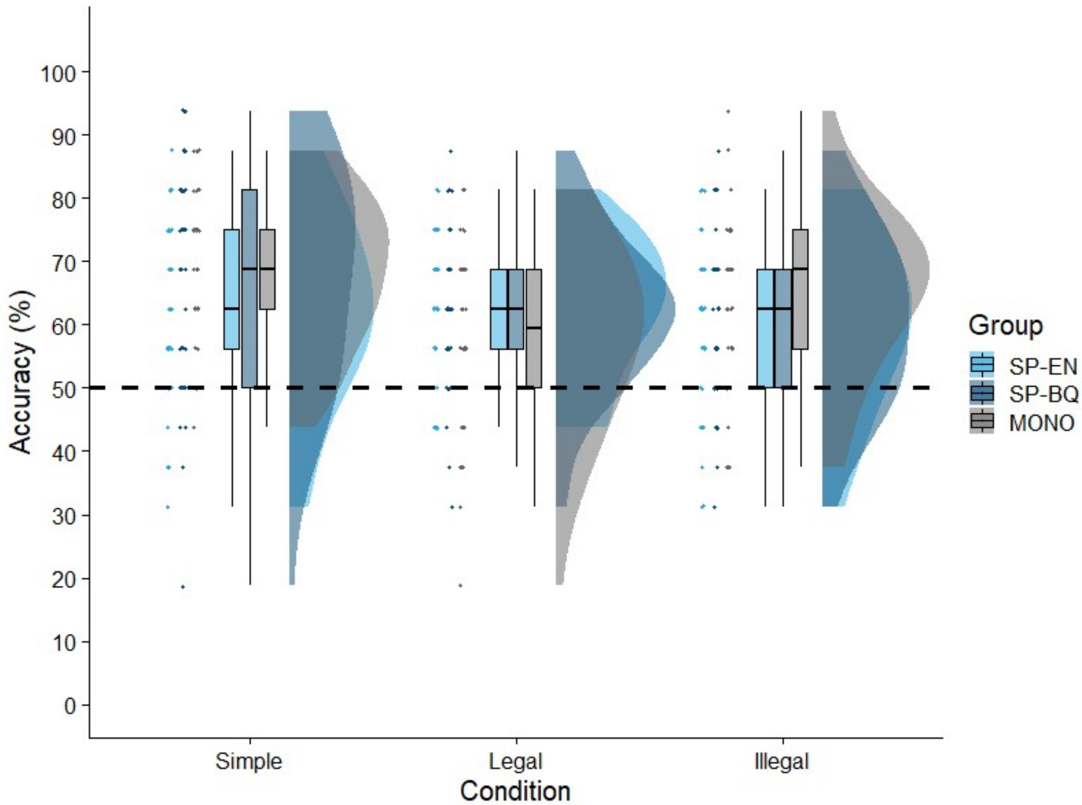
Data analysis

I analyzed participant's accuracy in the three SL streams using a GLMM. I assumed a binomial distribution and a logit link for the accuracy scores, with the Condition, Group, and their interaction as fixed effects of interest, and the Age, Non-verbal IQ, and Gender nuisance covariates. The Condition factor was Helmert coded to contrast the simple against the complex conditions first and then contrast the two complex conditions against each other. Similarly, the Group factor was reverse Helmert coded according to the hypothesis of differences between bilinguals and monolinguals but not between the bilingual groups. The final model achieved convergence using the by-participant and by-item intercepts with by-participant uncorrelated slopes for the Condition factor.

2.1.3. Results

I first tested whether the type of foil (i.e., non-words, part-words) influenced the SL tasks' results. Wilcoxon paired-samples tests over the aggregated accuracy suggested no differences between the non-word and part-word trials in the simple (p = 0.764, BF_{10} = 0.106), the complex legal (p = 0.567, BF_{10} = 0.112), or the complex illegal (p = 0.992, BF_{10} = 0.104) conditions. Thus, I did not include a type of foil factor in the GLMM analysis. As a second step, I tested whether participants' average accuracy score was above chance level (50%) in the three conditions using Wilcoxon one-sample tests—since these data were non-normal as suggested by a Shapiro-Wilk test (all $p < 0.05$). Participants in the Spanish monolingual group performed above chance level in the simple (M_{acc} = 68.4, SD = 12.1, $p < 0.001$, $BF_{10} > 100$), complex legal (M_{acc} = 58.3, SD = 14.4, $p < 0.001$, $BF_{10} > 100$), and complex illegal (M_{acc} = 65.0, SD = 12.8, $p < 0.001$, $BF_{10} > 100$) conditions. Participants in the Spanish-English bilingual group also performed above chance level in the simple (M_{acc} = 62.5, SD = 14.8, $p < 0.001$, $BF_{10} > 100$), complex legal (M_{acc} = 65.5, SD = 11.4, $p < 0.001$, $BF_{10} > 100$), and complex illegal (M_{acc} = 60.1, SD = 13.8, $p < 0.001$, $BF_{10} > 100$) conditions. Similarly, participants in the Spanish-Basque bilingual group performed above chance level in the simple (M_{acc} = 67.3, SD = 17.7, $p < 0.001$, $BF_{10} > 100$), legal (M_{acc} = 61.9, SD = 11.9, $p < 0.001$, $BF_{10} > 100$), and illegal (M_{acc} = 60.6, SD = 13.1, $p < 0.001$, $BF_{10} > 100$) conditions. Figure 2 presents the accuracy results by group and condition. Notably, participants' average performance on each task (around 60-70%) was in line with previous statistical learning studies reporting an average accuracy in the range of 55-70% (Erickson & Thiessen, 2015; Saffran, 2003).

Figure 2. *Average accuracy by condition and group in Experiment 1.*



Note. Raincloud plots showing probability density. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores (slightly jittered to minimize overlap). SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

The GLMM analysis (shown in Table 5) indicated a significant difference between the simple and complex conditions (p = 0.001, BF_{10} = 2.816), but no significant differences between the two complex conditions (p = 0.508, BF_{10} = 0.017). Specifically, participants performed better in the simple than the complex conditions, without any differences in performance between the legal and illegal complex conditions. As expected, there were no differences between Spanish-Basque and Spanish-English bilingual groups (p = 0.425, BF_{10} = 0.018). However, the monolingual and bilingual groups did not differ in their performance (p = 0.398, BF_{10} = 0.018). A significant two-way interaction indicated a difference between monolinguals and bilinguals in the complex conditions (p = 0.010, BF_{10} = 0.329). This interaction seemed to be driven by the monolingual group having a lower score in the complex legal compared to the complex illegal conditions and is not indicative of a bilingual experience effect since both bilingual groups had comparable performance in the two complex conditions. Only participants´ non-verbal IQ seemed to contribute to their overall scores in Experiment 1 (p = 0.003, BF_{10} = 0.751) out of all covariates (all p > 0.05, BF_{10} < 0.1).

Table 5. *Accuracy GLMM results of Experiment 1.*

Fixed Effects		Estimate	SE	z	*p*	BF_{10}
(Intercept)		**0.570**	**0.070**	**8.148**	**0.000**	**-**
Simple-Complex		**0.071**	**0.022**	**3.286**	**0.001**	**2.816**
Legal-Illegal		-0.023	0.034	-0.663	0.507	0.017
SPBQ-SPEN		0.032	0.043	0.739	0.460	0.018
MONO-BIL		0.016	0.025	0.626	0.531	0.018
Simple-Complex x ESEU-ESEN		0.038	0.027	1.415	0.157	0.022
Simple-Complex x MONO-BIL		0.016	0.015	1.022	0.307	0.036
Legal-Illegal x SPBQ-SPEN		-0.012	0.042	-0.286	0.775	0.329
Legal-Illegal x MONO-BIL		**-0.062**	**0.024**	**-2.564**	**0.010**	**0.014**
Covariates						
Age		-0.008	0.016	-0.497	0.619	0.017
Non-verbal IQ		**0.013**	**0.005**	**2.968**	**0.003**	**0.751**
Gender		-0.179	0.116	-1.539	0.124	0.027
Random Effects	*Group*	*Variance*	*SD*			
Item	Intercept	0.056	0.237			
Participant	Intercept	0.049	0.222			
	Simple-Complex	0.008	0.088			
	Legal-Illegal	0.002	0.043			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

2.1.4. Summary

Experiment 1 revealed that more complex sub-lexical phonotactic patterns could modulate the word segmentation difficulty in SL tasks. All three groups, on average, performed the tasks above chance level but their performance was lower in the complex than the simple condition, with no differences between the two complex conditions. More importantly, the SL performance of monolingual and bilingual participants did not significantly differ in any of the conditions. A possible explanation for these findings is that I constructed the stimuli using syllables that naturally occur in Spanish, English, or Basque. Even if the complex illegal condition included syllables that could have interfered with the statistics in the speech stream, these syllables also appear within words in all three languages. Therefore, this task may not have been challenging enough to elucidate the differences shown in previous studies (Wang & Saffran, 2014). In other words, the bilingual experience might still come into play when the task is sufficiently challenging—i.e., by using tones or completely illegal phonotactic patterns—or produces interference with previously learned materials. Additionally, the groups might differ when generalizing the acquired knowledge from an SL stream rather than individual words. With this in mind, in Experiment 2, I tested monolinguals and bilinguals when extracting rules from an SL task designed as a proxy for morphological rule learning.

2.2. Experiment 2: Morphological rule learning and generalization

2.2.1. Rationale

In the previous experiment, learning the words largely depended on the transitional probabilities of adjacent syllables. Specifically, the patterns presented in these SL streams could not be generalized outside of their respective artificial speech streams. While language learning requires identifying and learning the patterns within and across different words, it also encompasses learning the morphological rules that bind them together. Learning these rules allows generalization to instances where the words vary due to their dependency on other constituents (Endress & Bonatti, 2016; Peña et al., 2002). The perfect example is affixal morphology, where structures append to the beginning (prefix) or ending (suffix) of a word to change its class or meaning (e.g., ***un****touch****able****,* ***un****break****able***). These rules can span multiple elements, and in the SL literature, they are commonly known as non-adjacent dependencies (Misyak & Christiansen, 2007). Some authors posit similar SL mechanisms to those employed in word segmentation underly learning these non-adjacent dependencies (Rebecca Frost & Monaghan, 2016; Misyak & Christiansen, 2007). To my knowledge, no prior study has addressed whether bilingual experience can affect the implicit learning and generalization of non-adjacent dependencies.

In Experiment 2, I explored the capacity of monolinguals and bilinguals to learn non-adjacent dependencies from an SL stream and generalize them to novel items. Having observed no differences arising from bilingual experience in the first experiment, the purpose behind this manipulation was to test whether the groups would differ in another aspect of language learning by using this task as a proxy for affixal morphology rule generalization. Since this task tested participants' generalization ability and was more challenging than the previous two tasks, I adhered to the original hypothesis and expected to find differences between the groups.

2.2.2. Methods

Participants

Forty Spanish monolinguals (M_{age} = 21.7, SD = 2.4; 35 females), forty Spanish-Basque bilinguals (M_{age} = 21.8, SD = 2.2; 32 females), and forty Spanish-English bilinguals (M_{age} = 21.0, SD = 2.4; 36 females) participated in Experiment 2. Participants had a similar profile as those in Experiment 1. Appendix A2 shows the demographic information and statistical contrasts between groups.

Materials

I concatenated nine words with an AXC form, where A and C established a frame with constant syllables, and X was a filler syllable that could vary. Following previous work (Rebecca Frost & Monaghan, 2016), I used plosive syllables for the A_C frames (*/ke/, /po/, /bi/, /ga/, /du/, /ti/*) and continuants for the X fill syllables (*/mu/*, */fe/*, */li/*). The frame syllables were randomized to create five counterbalanced versions of the SL stream. There were three different frames combined with the three fill syllables, creating nine words for each version. I manually verified that none of the streams contained words or parts of words that existed in Spanish, English, or Basque. There were 100 repetitions of each word in every version of the stream. I synthetized the streams in the same manner as in Experiment 2, ensuring this time that there were no immediate repetitions of words with the same frame. The overall duration of the speech streams was about 12 minutes.

Procedure

The procedure was identical to the previous experiment except for the test trials. For the 2AFC test, I used three continuant syllables as generalization fills (*/se/, /ya/, /ni/*) and inserted them in the original frames to produce nine AYC generalization words. I tested participants on their ability to distinguish these words from three types of foils. The first two types were part-words created by concatenating the ending of the target word and the beginning of a different word from the stream. For example, a target word (*/kenipo/*) would be paired with a part-word starting at the fill syllable (*/nipoti/*) or ending with the fill syllable (*/gakeni/*). The last type of trials were non-words, created by inverting the word's syllables (e.g., */ponike/*). Due to the randomized nature of the different versions, a target word in one of the lists could potentially be a non-word foil in any other version. The inclusion of three types of tests increased the number of trials to 27, representing a substantial increase from the previous experiment.

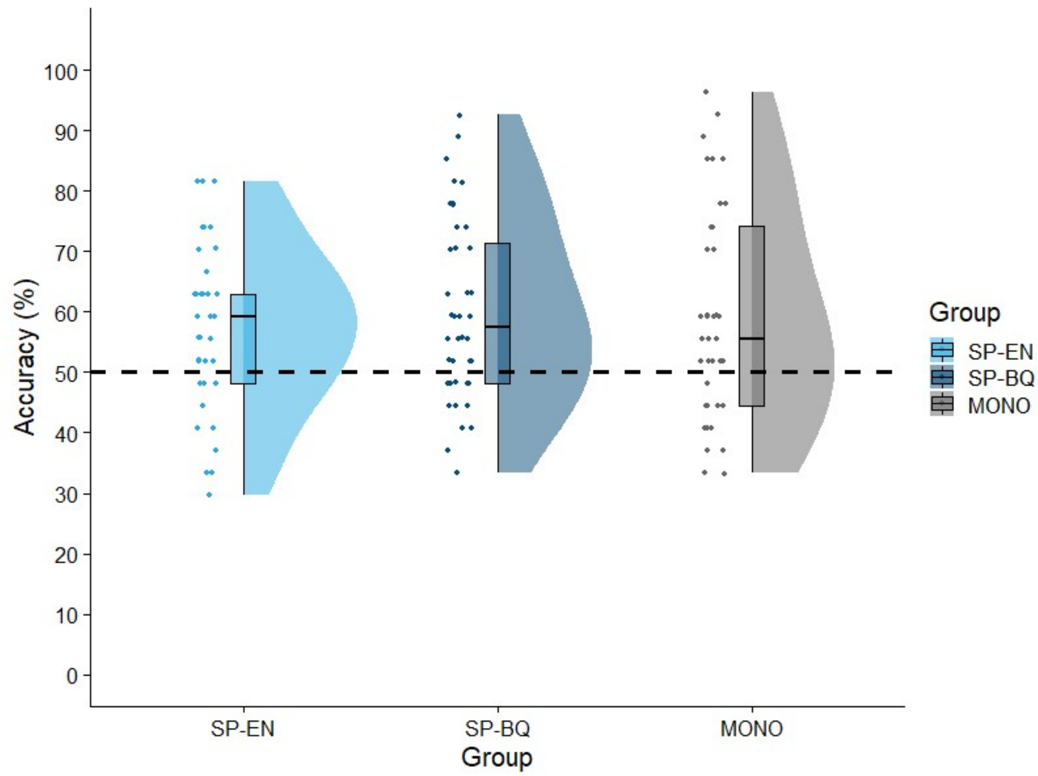
Data analysis

The data analysis closely followed Subsection 1.5.3.2 and Experiment 1, the only difference being that there were no conditions.

2.2.3. Results

One participant from the Spanish-English group was eliminated before the analysis due to technical errors during the experiment. As a first step, I tested whether the list or the type of foil influenced average performance in the task. An ANOVA suggested similar performance regardless of the stream's version ($p = 0.069$; $BF_{10} = 0.535$) and the type of foil ($p = 0.527$, $BF_{10} = 0.052$), and no interaction between the two ($p = 0.217$, $BF_{10} = 0.009$). Furthermore, Wilcoxon one-sample tests against chance-level indicated that participants in the monolingual (M_{acc} = 59.1, SD = 17.5, p = 0.011, BF_{10} > 100), Spanish-English (M_{acc} = 57.3, SD = 13.4, p = 0.005, BF_{10} > 100), Spanish-Basque (M_{acc} = 60.1, SD = 15.1, p < 0.001, BF_{10} > 100) groups performed above the 50% chance-level in the task. Figure 2 depicts the results by group. In this case, the GLMM revealed no differences between the two bilingual groups (Estimate = 0.066, SE = 0.077, z = 0.858, p = 0.391, BF_{10} = 0.025), or between the monolingual and bilingual groups (Estimate = 0.009, SE = 0.044, z = 0.216, p = 0.829, BF_{10} = 0.018). Additionally, none of the covariates reached significance (all p > 0.05, BF_{10} < 0.1).

Figure 3. *Average accuracy by Group in Experiment 2*



Note. Raincloud plots showing probability density. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores (slightly jittered to minimize overlap). SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

2.2.4. Summary

Overall, Experiment 2 extended the findings from the previous experiment by testing the generalization of non-adjacent dependencies learned during an SL task as a proxy for morphological rule learning. In line with previous studies, participants could generalize the frames extracted from the artificial language above chance-level, hinting towards similar implicit mechanisms for word and rule learning (Endress & Bonatti, 2016; Rebecca Frost & Monaghan, 2016). The focal point of Experiment 2 was to compare the performance of bilinguals and monolinguals in generalizing information from the SL task. The main challenge of this experiment was that participants needed to extract the non-adjacent frames from the stream—rather than specific words—and generalize this knowledge to novel items. Nevertheless, the results indicated similar performance between the three groups and thus no evidence of a bilingual advantage.

Spanish and Basque possess a rich morphology at the verb and noun levels (García Mayo & Villarreal Olaizola, 2011). Although English has a less rich morphology than Spanish and Basque, participants in the Spanish-English group might have benefited from their native knowledge of Spanish. In other words, the three languages may not differ sufficiently in their affixal morphology, such as to affect the non-adjacent dependencies generalization measured by this SL task. Notably, I did not target any specific differences that long-term speakers of two languages might exploit for a learning advantage. For instance, due to the Basque language´s word order and postpositional morphology, Basque speakers are more exposed to variations at the end of words (i.e., suffixes or other inflections) rather than at the beginning. Hence, in Experiment 3, I presented participants with an ambiguous SL stream that could be segmented based on participants’ known language’s word order.

2.3. Experiment 3: Statistical segmentation and word order

2.3.1. Rationale

In Experiment 3, I compared monolingual and bilingual participants' performance when exposed to an ambiguous artificial grammar that could be parsed based on their known languages' word order. As mentioned in the introduction, Basque is a language with a predominant subject-object-verb (SOV) word order instead of the subject-verb-object (SVO) word order of Spanish and English. In SVO languages, function words tend to precede content words (e.g., ***in the*** *school*), but this pattern is consistently the opposite in SOV languages such as Basque (e.g., *eskol**an*** – *school* ***the in***). Corpus studies have shown that function words (e.g., articles or similar inflections) are generally shorter and more frequent than content words (e.g., nouns), regardless of a language's word order (Gervain, Nespor, et al., 2008). Therefore, speakers of an SVO language experience more frequent and shorter words (i.e., determiners) preceding longer and less frequent tokens (i.e., nouns), whereas speakers of an SOV language experience the opposite pattern. Furthermore, prior studies have demonstrated that infants and adults prefer their native language's word order during SL (Gervain, Macagno, et al., 2008; Gervain, Nespor, et al., 2008; Onnis et al., 2016; Onnis & Thiessen, 2013). While the overall hypothesis concerns the bilingual advantage in foreign language learning tasks, in Experiment 3, I aimed to disentangle the overall effects of bilingual experience from those stemming from specific properties of known languages.

2.3.2. Methods

Participants

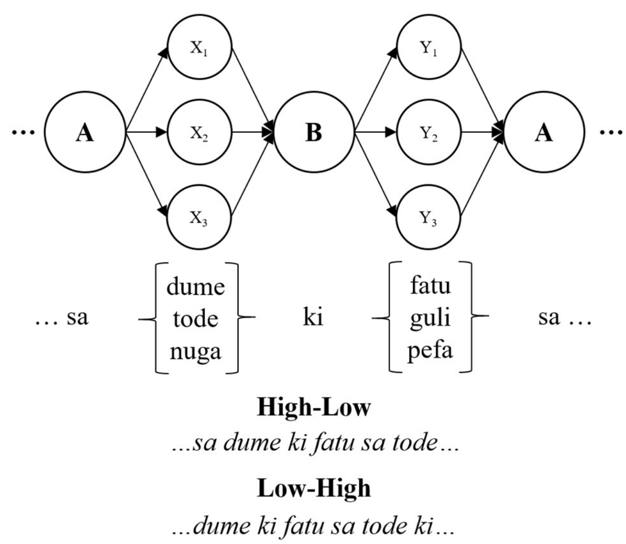
Participants were forty Spanish monolinguals (M_{age} = 21.7, SD = 2.5; 37 females), forty Spanish-Basque bilinguals (M_{age} = 21.9, SD = 1.9; 32 females), and forty Spanish-English bilinguals (M_{age} = 21.0, SD = 2.4; 36 females). They had a similar profile as those in Experiments 1 and 2. Appendix A3 shows tables with demographic information and statistical contrasts by groups.

Materials

I created an SL stream by concatenating two constant syllables with six variable disyllabic non-words (see Figure 4). Following previous work, the stream was instantiated as a Markov chain with AXBY form (Onnis & Thiessen, 2013; Perruchet & Desaulty, 2008). For the A and B slots, I used syllables with CV structure (i.e., */sa/* and */ki/*), whereas for the X and Y slots, I used CVCV words (e.g., */dume/* and */fatu/*). I diverged from previous studies in two ways. First, by introducing longer words in the X and Y slots, I more closely captured the relationship between frequency and length of function and content words. Second, while prior studies have only measured preference after familiarization with these types of streams, I included both preference and test trials during the test phase.

The SL stream for this experiment was generated by traversing the Markov chain depicted in Figure 4 for 100 steps. The final stream had 480 presentations of each A and B tokens and around 160 presentations of the X and Y words. Critically, the stream could be parsed in two ways, either as AX/BY (e.g., */sadume/*) or as XB/YA (e.g., */dumeki/*). For convenience and interpretability, I call the AX/BY pattern *High-Low* and the XB/YA pattern *Low-High* based on the frequency of their constituent tokens (see the bottom part of Figure 2). Finally, I synthesized the stimuli as in Experiment 1, except for each syllable duration, which I increased to 250ms following prior work (Onnis & Thiessen, 2013). Hence, the duration of the synthesized stream was around 12 minutes.

Figure 4. *Material creation for Experiment 3*



Note. The upper part of the figure depicts the Markov chain used to generate the artificial SL stream used in Experiment 3. The arrows represent the possible state transitions for each token. The circle sizes symbolize the frequency of presentation within the artificial language. Below is the instantiation of this artificial grammar with selected syllables and non-words. The bottom part shows possible ways of segmenting the stream depending on High-Low and Low-High patterns.

Procedure

The procedure was identical to the one in Experiments 1 and 2, except for the test phase. In the 2AFC test, I created three types of trials. The first two were the *segmentation* trials. These trials tested how well participants could segment the High-Low and Low-High patterns from the SL stream. In these trials, the participants heard words that followed the transitions of the stream against foils created by breaking the transitions in the stream. For example, a High-Low word */satode/* would be paired to its corresponding foil */tosade/*, whereas a Low-High word */todeki/* would be paired with */tokide/*. The third group of trials was the *preference* trials. In these trials, I directly contrasted High-Low and Low-High words to test whether participants preferred either of them. There were six trials per type to account for every possible combination generated by the Markov chain.

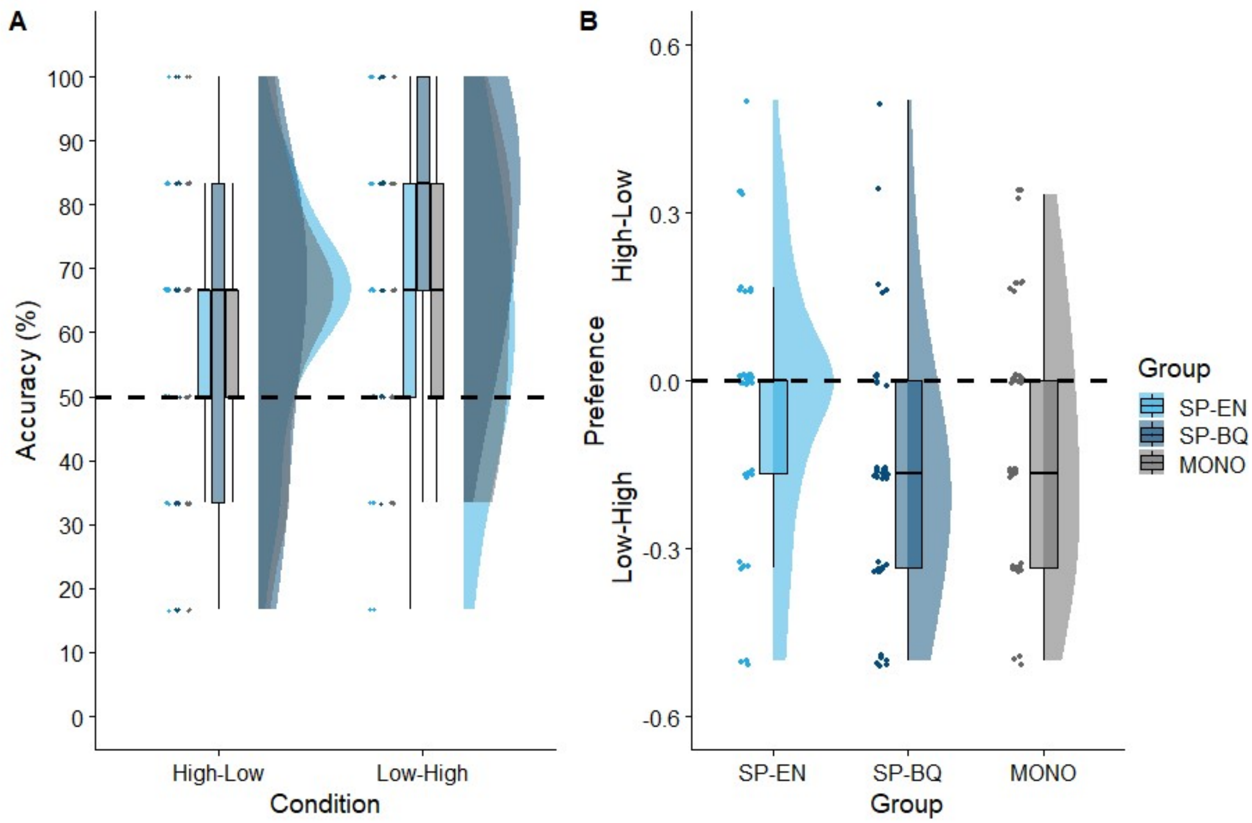
Data analysis

As in Experiments 1 and 2, I report both the frequentist and Bayesian versions of the analyses. In this case, however, I conducted two separate GLMMs for the segmentation and preference trials. Both models contained by-participant and by-item random intercepts. In addition, I -coded the Condition factor in the segmentation trials to compare the High-Low against the Low-High conditions (HL-LH). Furthermore, since I expected the Spanish-Basque bilinguals to perform differently due to their experience with an SOV language, I coded the group contrasts differently than in Experiments 1 and 2. On the first level, I contrasted the Spanish-Basque against the combined Spanish-English bilingual and Spanish monolingual groups (SOV-SVO). On the second level, I compared the Spanish-English against the Spanish monolingual group (SVO-SVO). I also used this group contrast for the preference trials' GLMM. The idea behind this coding was that if the experience with an SOV language modulates performance in this task, only the first contrast level should reach significance. Moreover, assuming the bilingual experience confers an advantage for tracking both patterns irrespective of experience with an SOV language, the second contrast should also be significant and in favor of the Spanish-English group.

2.3.3. Results

Before the analyses, I eliminated one participant from the monolingual group with zero scores in both segmentation conditions. I first compared the groups' overall performance in the segmentation trials against chance-level performance (50%). Non-parametric one-sample tests indicated that participants in the monolingual group performed above chance level in both the High-Low (M_{acc} = 61.5, SD = 21.3, p = 0.003, BF_{10} = 19.165) and Low-High conditions (M_{acc} = 65.4, SD = 20.7, p < 0.001, BF_{10} > 100). The Spanish-Basque bilingual group also performed above chance level on the High-Low (M_{acc} = 60.4, SD = 23.5, p = 0.011, BF_{10} = 5.056) and Low-High (M_{acc} = 77.5, SD = 18.7, p < 0.001, BF_{10} > 100) conditions. Finally, the Spanish-English participants also were above chance in the High-Low (M_{acc} = 63.5, SD = 17.5, p < 0.001, BF_{10} > 100) and Low-High (M_{acc} = 63.1, SD = 21.9, p = 0.002, BF_{10} = 34.791) conditions. The averaged accuracy scores for all groups and conditions are presented in Figure 5A.

Figure 5. *Average accuracy and preference scores in Experiment 3.*



Note. **(A)** Segmentation scores by Condition and Group. **(B)** Preference scores by group. The preference scores are centered at zero, with positive values suggesting a High-Low preference and negative values a Low-High preference. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores (slightly jittered to minimize overlap). SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

The GLMM results of the *segmentation* task (shown in Table 6) indicated a significant difference between the High-Low and Low-High conditions (p = 0.038, BF_{10} = 0.147). In addition, the difference between Spanish-Basque bilinguals and speakers of SVO languages (SOV-SVO) was significant (p = 0.033, BF_{10} = 0.203). However, the analysis revealed no difference between the Spanish monolinguals and Spanish-English bilinguals (SVO-SVO) on these tasks (p = 0.968, BF_{10} = 0.023). Moreover, there was a significant interaction term where speakers of an SOV language (Spanish-Basque bilinguals) showed a higher score for the Low-High condition than speakers of SVO languages (p = 0.002, BF_{10} = 1.966). This effect is also evidenced in Figure 5A. None of the covariates reached significance (p > 0.05, BF_{10} < 0.1).

Table 6. *Accuracy GLMM results of segmentation trials in Experiment 3.*

Fixed Effects		Estimate	SE	z	*p*	BF_{10}
(Intercept)		**0.657**	**0.086**	**7.626**	**< 0.001**	**-**
HL-LH		**-0.165**	**0.079**	**-2.079**	**0.038**	**0.147**
SOV-SVO		**0.096**	**0.045**	**2.133**	**0.033**	**0.203**
SVO-SVO		-0.003	0.075	-0.040	0.968	0.023
HL-LH x SOV-SVO		**-0.129**	**0.042**	**-3.097**	**0.002**	**1.966**
HL-LH x SVO-SVO		0.047	0.069	0.680	0.497	0.033
Covariates						
Age		0.007	0.027	0.268	0.789	0.056
Non-verbal IQ		-0.001	0.009	-0.155	0.877	0.054
Gender		-0.018	0.185	-0.096	0.923	0.058
Random Effects	*Group*	*Variance*	*SD*			
Item	Intercept	0.036	0.189			
Participant	Intercept	0.038	0.196			

Note. Significant terms are highlighted in bold. SE = standard error; SD = standard deviation; HL = High-Low; LH = Low-High, SOV = subject-object-verb; SVO = subject-verb-object.

For the *preference* trials, I centered the scores around zero by subtracting 50% from the aggregated scores (see Figure 5B). Wilcoxon one-sample tests against zero indicated that the monolingual (M_{pref} = -0.115, SD = 0.239, *p* = 0.006, BF_{10} = 8.031) and Spanish-Basque (M_{pref} = -0.175, SD = 0.233, *p* < 0.001, BF_{10} > 100) groups had preference scores significantly lower than zero. This was not the case for the Spanish-English group (M_{pref} = -0.041, SD = 0.237, *p* = 0.313, BF_{10} = 0.292). The GLMM for the preference trials also indicated a significant effect of knowledge of an SOV language (Estimate = -0.149, SE = 0.069, *z* = -2.147, *p* =0.032, BF_{10} = 0.358). Finally, the difference between the two SVO speaking groups was not significant (Estimate = -0.167, SE = 0.120, *z* = -1.391, *p* = 0.164, BF_{10} = 0.095).

2.3.4. Summary

Taken together, the results for Experiment 3 suggested that experience with an SOV language, rather than overall bilingual experience, influenced the tracking and learning of an ambiguous speech signal. Only the Spanish-Basque bilingual group segmented and preferred the Low-High patterns better than the other groups. These results align with research suggesting an effect of word order in segmentation of similar ambiguous streams (Gervain, Nespor, et al., 2008; Onnis & Thiessen, 2013). My results also extend these findings by suggesting that word order in a second language can influence segmentation in highly proficient Spanish-Basque bilinguals. Prior studies have also suggested that bilingual experience could influence simultaneous learning of similar—albeit interfering—artificial grammars (Onnis et al., 2018). My results do not seem to indicate an overall bilingual advantage and instead evidence that specific properties of the known languages (i.e., word order) influenced the task. Notably, I did not manipulate any conflict arising from learning two artificial grammars in this experiment. Hence, these results do not discard the possibility of bilingual experience influencing SL in contexts of interfering information (Bartolotti et al., 2011; Onnis et al., 2018).

A possibility for the lack of a general bilingual advantage in these three experiments is that bilingual experience might mainly influence learning at the lexical level (Kaushanskaya & Marian, 2009b; Poepsel & Weiss, 2016). In Experiment 4, I presented participants with an audio-visual SL task targeting word-referent learning.

2.4. Experiment 4: Implicit word referent learning

2.4.1. Rationale

Studies have shown that, through SL mechanisms, infants and adults can learn mappings between words and visual referents (objects) across multiple scenarios, an ability known as cross-situational (CS)SL (Smith & Yu, 2008; Yu & Smith, 2007). The CSSL task differs from other SL tasks because participants learn the names for different visual referents by aggregating information through multiple contexts. In other words, they implicitly "discover" each object's name by seeing it in conjunction with other distinct referents. This task was designed to mimic the crowded visual world where infants initially learn words and seems to target a crucial implicit mechanism for learning words in native and foreign languages (Benitez et al., 2016; Smith & Yu, 2008).

Perhaps one of the defining characteristics of bilingual experience is learning to map one referent to two potentially different words (e.g., *perro – dog/txakur*), but this phenomenon is not unique to bilingualism and happens within languages. For instance, there are words with multiple referents or *homonyms* (e.g., *bat* as the animal or the object) and concepts with multiple names or *synonyms* (e.g., *paper – sheet*). Bilinguals have the added difficulty of learning these two types of mappings within two languages and across languages, potentially influencing their ability to learn them. Prior work has shown that bilinguals can outperform monolinguals in a CSSL task when learning homonyms (i.e., one word to two referents), but not when learning exclusive one-to-one mappings (Poepsel & Weiss, 2016). Other authors only report differences between bilinguals and monolinguals in the one-to-one mappings (Escudero et al., 2016). Finally, some have only found minimal differences between bilinguals and monolinguals learning one-to-one and synonym (i.e., two words to one referent) mappings (Benitez et al., 2016).

To address these discrepant findings and target a different analytic level, in Experiment 4, I compared monolinguals and bilinguals in a CSSL task that entailed learning one-to-one, two-to-one (synonyms), and one-to-two (homonyms) word-referent pairs. If bilingual experience facilitates overall word learning, participants in the bilingual groups should outperform those in the monolingual group in the one-to-one mappings. However, if bilingual experience only potentiates learning of multiple word-referent mappings, they should perform better than monolinguals in the more challenging multiple mapping conditions.

2.4.2. Methods

Participants

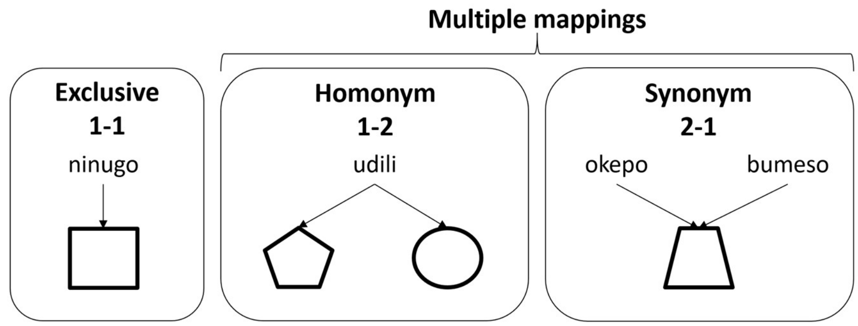
Forty Spanish monolinguals (M_{age} = 21.7, SD = 2.5; 37 females), forty Spanish-Basque bilinguals (M_{age} = 21.9, SD = 1.9; 31 females), and thirty-seven Spanish-English bilinguals (M_{age} = 21.0, SD = 2.4; 32 females) participated in Experiment 4. Participants had a similar profile as those in previous experiments. Appendix A4 shows the demographic information and statistical contrasts between groups.

Materials

I created 30 words by randomly concatenating letters from a pool of consonants (t, l, k, m, n, b, s, p, g, d) and vowels (a, i, u, e, o) in an alternating manner. The words had either a CVCVCV (e.g., */ninugo/*) or a VCVCV (e.g., */udili/*) form and were synthesized using the Mac OS X system's Text-to-Speech software with the Spanish female voice "Monica". Since these words were presented in isolation, I did not enforce as much control (e.g., constant pitch, vowel duration) as in the previous three experiments, thus creating more natural-sounding stimuli. In addition, I eliminated words that existed or sounded like actual Spanish, English, or Basque words from an initial list. The final 30 words were paired with 30 color depictions of non-existing objects manually selected from the NOUN database (Horst & Hout, 2016) based on their color saliency and visual complexity scores.

The critical manipulation in this task involved creating the different word-referent mappings. As shown in Figure 3, twelve words were randomly paired with twelve objects to form the *exclusive* (one-to-one) mappings. For the *homonym* (one-to-two) mappings, I randomly paired six words and twelve objects, producing words with two distinct referents. Conversely, for the *synonym* (two-to-one) mappings, twelve words were randomly paired with six objects, such that each of these objects had two different names. I manually verified that the words and objects in the multiple mapping conditions were not very similar to avoid artificially increasing the task's difficulty.

Figure 6. *Word-referent mappings in Experiment 4.*



Note. The leftmost panel depicts an example of the Exclusive one-to-one mappings, where one word was consistently presented with one visual referent. The middle panel illustrates an example of a Homonym one-to-two mapping, where one word was paired with two distinct visual referents. Finally, the rightmost panel shows an example of a Synonym two-to-one mapping, where two words referred to the same visual referent.

Procedure

Participants completed three blocks of alternating familiarization and test phases. As in previous experiments, they were asked to wear headphones and sit in a quiet room. In each familiarization phase, they saw thirty-two scenes comprised of three objects horizontally aligned and at the center of the screen over a white background. They also heard the three names for the depicted objects in each scene, starting after one second and with a one-second pause between them. Participants were instructed to pay attention to the different scenes because they would later perform a test based on them. The objects' position and the order of the audios were fully randomized in each scene so that there was no particular match between words and referents. Following previous studies (Poepsel & Weiss, 2016), only one side of the multiple mappings was presented during the first half of the scenes. During the second half, the other side of the multiple mappings was shown, simulating a change of context. This manipulation allowed us to test the effects of primacy and recency on the multiple mappings. I created four counterbalanced lists by varying which side of the two multiple mapping conditions was presented during the first and second halves of the familiarization phase. Thus, the scenes' order of presentation was pseudo-randomized for each familiarization phase so that no word or referent, including two sides of the multiple mappings, would appear in contiguous scenes.

Following each familiarization phase, participants completed a 2AFC test. In each trial, two horizontally aligned objects (one target and one foil) were presented, and the target object's name was played. The trials' order of presentation and the position of target and foil objects were fully randomized. Participants were instructed to select the object they thought corresponded to the heard name using the keyboard (f and j keys). The one-to-one mappings were tested during the three test blocks. However, during the experiment's first two blocks, only one side of the multiple mappings was presented, corresponding to the order of presentation during the familiarization phase (Poepsel & Weiss, 2016). The complete word-referent list was used for the last block, controlling that the foil object did not correspond to the multiple mapping condition. Participants were encouraged to take small breaks between each block.

Data analysis

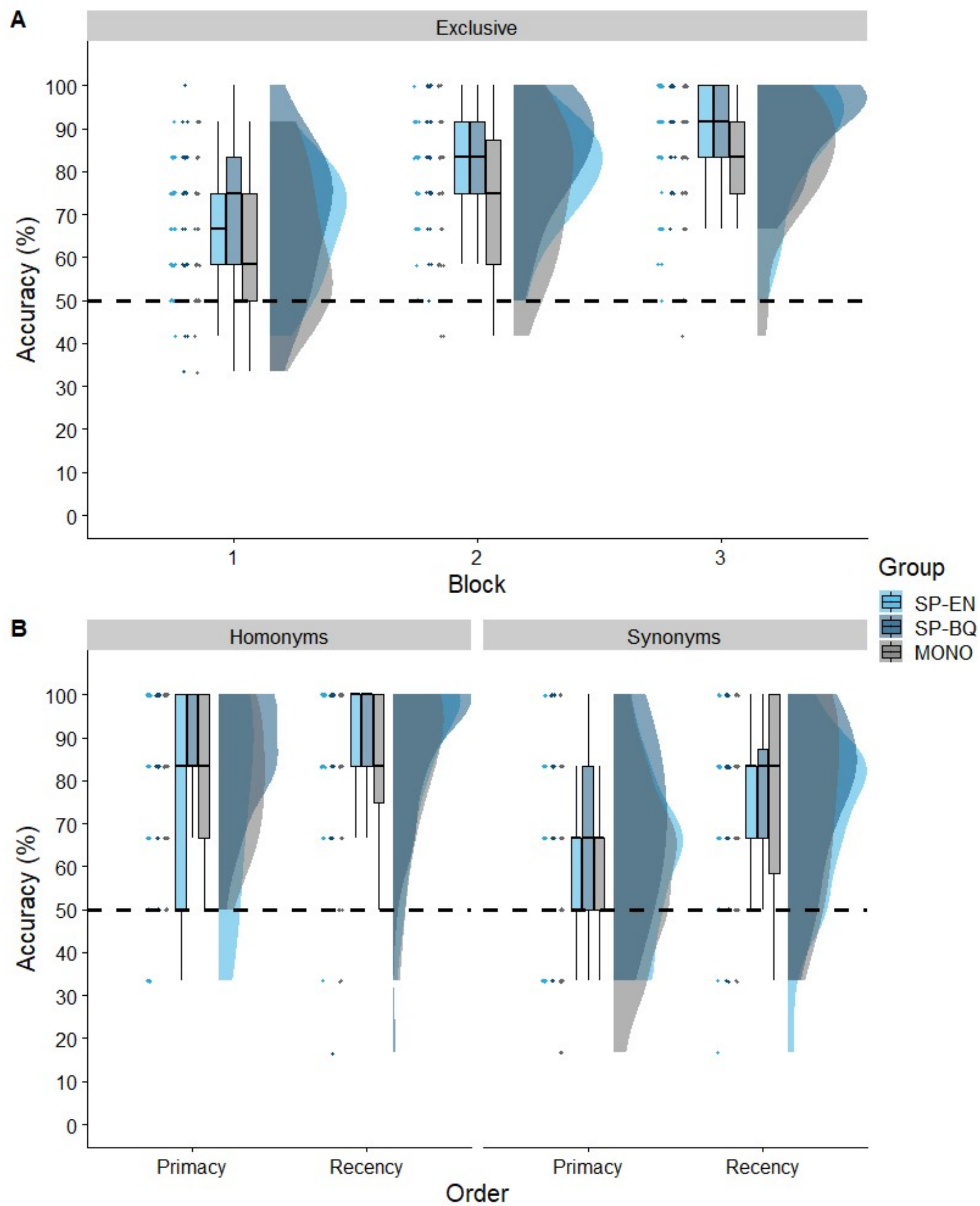
I modeled the exclusive and multiple (synonyms/homonyms) mappings data separately using GLMMs, similar to the previous two experiments. I treated the block as a continuous linear factor for the exclusive mappings and compared the reverse Helmert-coded groups and their interaction with the blocks as fixed effects. In this case, the GLMM was performed over the proportion correct data to avoid inflating the estimates for the linearized block (Mirman, 2017; J. D. Singer & Willett, 2009). As in the previous experiments, I contrasted the Spanish-Basque against the Spanish-English group first, and then both bilingual groups against the monolinguals. I included the group, block, and interactions as fixed effects and the by-participant intercept and uncorrelated block slope as random effects.

For the more challenging multiple mappings, since participants were tested on different mapping sides during the first and second blocks, I only contrasted the groups on the final test at block three. I created a factor to account for the order of presentation in the CSSL task. During the familiarization phase, the mappings presented first represented the *primacy* mappings, and those presented later the *recency* mappings. The group contrasts were the same as for the exclusive mappings. I tested the order-of-presentation factor and its interactions with the group as fixed effects. I included the by-participant and by-item intercepts and the by-participant slope for order-of-presentation as random effects in the final GLMMs.

2.4.3. Results

Before the analysis, I removed one participant from the monolingual group with chance-level performance in all three exclusive blocks. I first compared that there were no differences between the lists across the three blocks on the exclusive mappings, as the counterbalanced lists only targeted the multiple mappings. A series of ANOVAs revealed that there were no differences due to the list on the first (p =0.901, BF_{10} = 0.058), second (p = 0.586, BF_{10} = 0.096), or third blocks (p = 0.815, BF_{10} = 0.066). Next, I compared participant's accuracy on the Exclusive mappings to chance level (50%) in the last block. Wilcoxon one-sample tests indicated that Spanish monolinguals (M_{acc} = 83.1, SD = 14.6, p < 0.001, BF_{10} > 100), Spanish-English bilinguals (M_{acc} = 86.9, SD = 14.2, p < 0.001, BF_{10} > 100), and Spanish-Basque bilinguals (M_{acc} = 91.0, SD = 10.2, $p < 0.001$, $BF_{10} > 100$) performed significantly above chance level. Figure 7A depicts the accuracy for each group across the three blocks. Accuracy was higher for this task than in the previous experiments.

Figure 7. *Average accuracy by group and block in Experiment 4.*



Note. **(A)** Accuracy by group and block in the one-to-one (Exclusive) mappings. **(B)** Accuracy by group, mapping type, and order of presentation for the multiple (Homonyms and Synonyms) mappings. Raincloud plots showing probability density. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

Table 7 shows the results for the exclusive mappings' GLMM. Unsurprisingly, the analysis revealed a significant effect of the block ($p < 0.001$, $BF_{10} > 100$), indicating that participant's accuracy increased across the three blocks. The results also indicated no significant differences between the two bilingual groups ($p = 0.061$, $BF_{10} = 0.114$). However, there were significant differences between monolinguals and bilinguals, wherein both bilingual groups outperformed the monolingual group ($p = 0.008$, $BF_{10} = 1.071$). The interactions between block and group did not reach statistical significance ($p = 0.075$, $BF_{10} = 0.055$; $p = 0.147$, $BF_{10} = 0.073$; respectively). None of the covariates significantly influenced the scores (all $p > 0.05$, $BF_{10} < 0.1$).

Table 7. *Accuracy GLMM results of exclusive condition in Experiment 4.*

Fixed Effects		Estimate	SE	z	*p*	BF_{10}
(Intercept)		**1.453**	**0.074**	**19.596**	**< 0.001**	**-**
Block		**1.043**	**0.088**	**11.815**	**< 0.001**	**> 100**
SPBQ-SPEN		0.160	0.085	1.870	0.061	0.114
MONO-BIL		**-0.126**	**0.047**	**-2.662**	**0.008**	**1.071**
Block x SPBQ-SPEN		0.183	0.103	1.778	0.075	0.055
Block x MONO-BIL		-0.081	0.056	-1.451	0.147	0.073
Covariates						
Age		-0.006	0.025	-0.255	0.799	0.056
Non-verbal IQ		< 0.001	0.009	0.016	0.987	0.054
Gender		-0.011	0.176	-0.061	0.951	0.058
Random Effects	*Group*	*Variance*	*SD*			
Participant	Intercept	0.314	0.561			
	Linear	0.153	0.391			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

In the homonym condition (Figure 7B, left), Wilcoxon one-sample tests indicated that participants in the monolingual ($M_{acc} = 82.3$, SD = 14.6, $p < 0.001$, $BF_{10} > 100$), Spanish-Basque ($M_{acc} = 87.1$, SD = 12.1, $p < 0.001$, $BF_{10} > 100$), and Spanish-English ($M_{acc} = 80.6$, SD = 17.9, $p < 0.001$, $BF_{10} > 100$) groups performed above the 50% chance level on average. The GLMM results for the homonym mappings indicated a significant effect of the order of presentation (Estimate = -0.364, SE = 0.113, z = -3.219, $p = 0.001$, $BF_{10} = 4.874$), where performance in the recency mappings was on average better than in the primacy mappings. There were no differences between the two bilingual groups (Estimate = 0.275, SE = 0.143, z = 1.921, $p = 0.055$, $BF_{10} = 0.148$), or between the monolingual and bilingual groups (Estimate = -0.042, SE = 0.081, z =

-0.521, $p = 0.603$, $BF_{10} = 0.034$). There were no significant interactions between the order of presentation and the group contrasts (all $p > 0.05$, $BF_{10} < 0.1$).

In the synonym condition, (Figure 7B, right), Wilcoxon one-sample tests indicated that participants in the monolingual ($M_{acc} = 67.7$, SD = 16.0, $p < 0.001$, $BF_{10} > 100$), Spanish-Basque ($M_{acc} = 73.5$, SD = 14.2, $p < 0.001$, $BF_{10} > 100$), and Spanish-English ($M_{acc} = 66.9$, SD = 13.5, $p < 0.001$, $BF_{10} > 100$) groups performed above the 50% chance level on average. The GLMM for this condition revealed an effect of order of presentation in the same direction as for the homonym mappings (Estimate = -0.284, SE = 0.068, z = -4.179, $p < 0.001$, $BF_{10} > 100$). However, in this case, the analysis revealed that the Spanish-Basque group performed slightly better than the Spanish-English group (Estimate = 0.173, SE = 0.087, z = 1.991, $p = 0.046$, $BF_{10} = 0.211$), but there were difference between monolinguals and bilinguals (Estimate = -0.046, SE = 0.049, z = -0.924, $p = 0.355$, $BF_{10} = 0.037$). As in the homonym condition, there were no significant interactions between the order of presentation and the group contrasts (all $p > 0.05$, $BF_{10} < 0.1$). Additionally, a Wilcoxon paired-samples test confirmed that accuracy was overall higher in the homonym than the synonym mappings ($p < 0.001$, $BF_{10} > 100$).

2.4.4. Summary

Overall, Experiment 4 indicated that participants could learn the exclusive and the two types of multiple mappings from the same audio-visual CSSL task. Contrary to the previous experiments, this design also allowed us to test participants' learning at three stages, showing a progressive increase in their performance for the exclusive mappings. In this regard, both bilingual groups outperformed the monolingual group on the exclusive one-to-one mappings. Thus, the results align with prior findings suggesting bilingual experience effects for these types of mappings in CSSL and other explicit word learning tasks (Escudero et al., 2016; Kaushanskaya & Marian, 2009b). These findings should be interpreted in the context of the more challenging multiple mappings. The task itself was more demanding than the previous experiments—i.e., participants had to learn thirty novel items and words with different mappings in an implicit manner. Therefore, it is unclear whether the observed bilingual experience effect is constrained to the exclusive mappings or emerged due to the conflicting multiple mapping conditions.

Despite a similar design, the results diverge from prior work reporting bilingual experience effects in the homonym but not in the exclusive mappings (Poepsel & Weiss, 2016). The homonym mappings were easier to acquire for the participants than the synonym mappings, and there were no differences between monolinguals and bilinguals on these multiple mapping conditions. A possible reason for these findings is that pure synonyms—especially those referring to concrete objects with the same meaning—are significantly rarer than homonyms in most natural languages (Hurford, 2003). Moreover, it is often necessary to introduce additional cues, such as speaker identity or spacing between presentations, to avoid direct competition from multiple mappings during learning (Benitez et al., 2016). Out of these two factors, I only spaced the multiple mappings' presentations during the learning phase, leading to the observed difference between primacy and recency mappings extensively reported in associative learning research (Pineño & Miller, 2005).

Discussion

The present chapter compared monolingual and bilingual implicit learning across four statistical language learning (SL) experiments. I opted to use four well-established SL tasks to approximate different aspects of foreign language learning as a proxy, including word segmentation from continuous speech with varying sub-lexical phonology (Experiment 1), morphological rule generalization (Experiment 2), syntactic segmentation based on word order (Experiment 3), and lexical-semantics learning (Experiment 4). The inclusion of two bilingual groups and different manipulations permitted disentangling the overall effects of bilingual experience from those stemming from task difficulty or specific language pairs. Additionally, the confirmatory analytical approach and sample size—of almost three times the size reported in the literature—facilitated the straightforward evaluation of my central hypothesis.

Overall, Experiment 1 showed that, while manipulating the sub-lexical phonology of the SL stream affected participants' performance—thus making the task more challenging—, there were no bilingual experience effects for word segmentation. In Experiment 2, participants generalized the learned non-adjacent dependencies during the test phase, a slightly more difficult task than Experiment 1. However, again, I did not find any bilingual experience effects. In Experiment 3, only the knowledge of an SOV language—and not overall bilingual experience—produced an advantage when segmenting from an ambiguous SL stream. Lastly, in Experiment 4, I tested participants' ability to learn exclusive and multiple word-referent pairs. The results revealed that bilingual participants outperformed their monolingual peers when learning the exclusive but not the multiple word-referent mappings. In all experiments, regardless of the manipulation, the average performance of all groups was significantly above chance and in the range reported in the literature (Erickson & Thiessen, 2015; Perruchet & Pacton, 2006; Saffran, 2003), which points to the robustness of SL as an implicit foreign language learning mechanism.

It is essential to distinguish SL as a cognitive mechanism, an experimental task, and a proxy for foreign language learning. As a cognitive mechanism, other individual differences could perhaps account for auditory SL performance better than the bilingual experience, such as spontaneous synchronization to speech (Assaneo et al., 2019). The bilingual experience could operate over these abilities and indirectly influence SL. Prior studies seem to support this idea, showing bilingual experience effects only when participants learn from interfering SL streams (Bartolotti et al., 2011; Onnis et al., 2018). Similarly, Experiments 1 through 3 of this chapter showed no overall bilingual advantage, but the differences in Experiment 4 could have emerged in the context of the more difficult multiple mapping conditions. Nevertheless, the purpose of this study was to isolate the effects of bilingual experience from other factors (e.g., task difficulty, interference) previously shown in the literature (Bartolotti et al., 2011; Onnis et al., 2018; Wang & Saffran, 2014). In doing so, I purposefully abstained from introducing additional variables that could more likely benefit from the bilingual experience. For instance, other studies have shown that bilingual experience could influence non-native phonetic learning and conflict resolution (Antoniou et al., 2015; Donnelly et al., 2015). These abilities could indirectly affect SL as a cognitive ability, but whether they account for bilingual experience effects in SL tasks remains to be explored.

Considering SL as an experimental task, participants are sensitive to specific and controlled manipulations of the probabilities between and within words that might target properties of their known languages. I tested this idea in Experiment 3, where only bilinguals with knowledge of an SOV language outperformed the other two groups. Similar interfering or facilitatory effects have been reported in the literature using sub-lexical phonotactics (Finn & Hudson Kam, 2008) or stress patterns (Thiessen & Saffran, 2003) in monolingual adults and infants. Consequently, any bilingual advantage may primarily rely on the properties of participants' known languages and not the overall bilingual experience.

In addition, bilingual experience effects might emerge progressively through learning, and one-shot 2AFC tests are not sensitive enough to capture them. Most SL studies base their results on a single familiarization and test phase, constraining their findings to what is learned upon first exposure to a foreign language (Romberg & Saffran, 2010). A common criticism with this approach is that performance, as measured by a one-shot 2AFC test, is highly variable and noisy, partly due to the low number of trials generated from these artificial streams (Siegelman & Frost, 2015). In other words, there usually are only a handful of target and foil words presented during the test phase, and using a single test introduces additional variability in participants' scores. Artificially increasing the number of trials by repeating the target words multiple times only leads to participants learning during the test phase (Siegelman, Bogaerts, & Frost, 2017). Thus, while these experimental tasks could be adequate to measure learning at the group level, they might not be sensitive enough to detect individual differences (Siegelman & Frost, 2015). I partially addressed these limitations in Experiment 4 by adding three familiarization and test phases for the twelve exclusive mappings and found differences between monolinguals and bilinguals. Observing any effects in other SL tasks might be unlikely due to the variability in responses from one-shot 2AFC tests. Future research could further address these gaps by using continuous measures to focus on the process rather than the learning outcome.

Lastly, considering SL as a proxy for foreign language learning, a possibility is that bilingual experience effects might arise mainly at the vocabulary level. Most of the literature supporting a bilingual advantage primarily focuses on explicit vocabulary learning tasks, stressing that bilinguals might more easily establish links between new word forms (orthographic or phonological) and their meanings within the mental lexicon than monolinguals (e.g., Antoniou et al., 2015; Kaushanskaya & Marian, 2009b). In this regard, the present study showed no bilingual experience effects using SL tasks as proxies for word segmentation and morpho-syntactic levels. However, significant differences were observed at the lexical-semantics level. Notably, experimental work seldom targets other aspects of language (e.g., pragmatics, grammar at the sentence level). It would be interesting for future implicit learning studies to explore whether bilingual experience effects are constrained to the lexical-semantics level or extend to more higher-level aspects or broader contexts (i.e., sentence or paragraph structures) during foreign language.

Chapter 2 Conclusion

Are bilinguals better than monolinguals at implicit (statistical) language learning? My answer is that it depends. Statistical learning is a robust cognitive process supported by a plethora of experimental findings. At the same time, decisions about stimuli selection, experiment design, participant selection, data collection, and statistical analysis can significantly influence the outcome and interpretation of these experiments. This chapter explores only a handful of manipulations that could have potentially elicited a bilingual advantage, such as experience with different phonotactic patterns, the ability to generalize learned linguistic knowledge, knowledge of different word orders, and the capacity to learn exclusive and multiple word-referent pairs. The question of whether bilingualism can influence SL as a cognitive mechanism is perhaps ill-posed. Instead, to make progress in this field, I believe studies should focus on how linguistic and life experiences affect the strategies and filters through which individuals implicitly process and learn from linguistic and non-linguistic materials.

Moreover, SL only accounts for a minimal part of what learning a foreign language entails. Adult individuals acquire a significant portion of a foreign language through explicit processes—and implicit learning might play a more prominent role after individuals have achieved a certain proficiency level in a new language (Ellis, 2015). Therefore, in Chapter 3, I explored bilingual and monolingual language learning using explicit tasks.

Chapter 3: Explicit foreign language learning

Theoretical motivation

One of the most challenging aspects of foreign language learning is acquiring the substantial vocabulary necessary to perform proficiently in that language (Baxter et al., 2021; Schmitt, 2019). Language learners require around 10,000 words to be considered proficient speakers of a foreign language (Council of Europe, 2001). This number refers only to the base vocabulary and does not consider morphological variations. The majority of words in a language involve affixal (i.e., suffix, prefix, or similar inflections) morphology (Merkx et al., 2011). For instance, considering the words ***break***, ***breaks***, and ***breakable*** as separate entries in a language's vocabulary leads to the number of necessary words growing exponentially (Brysbaert et al., 2016a). Previous research has highlighted that bilingual experience can facilitate explicit vocabulary learning (e.g., Antoniou et al., 2015; Kaushanskaya, 2012; Kaushanskaya & Marian, 2009b, 2009a). However, it is still unclear whether bilinguals are better at learning morphological variations of known words or if bilinguals are only better at learning dissimilar—foreign-sounding—words to their known languages. This chapter addresses these two issues in turn by comparing monolinguals and bilinguals when explicitly learning the form of novel suffixes for known Spanish words as a proxy for the morphology level (Experiment 5) and novel words with different degrees of orthographic similarity to Spanish to target the sub-lexical orthography and the lexical-semantics levels (Experiment 6).

At a bare minimum, foreign vocabulary knowledge entails learning to recognize or produce the written or spoken forms of words when prompted (González-Fernández & Schmitt, 2020). The sub-lexical, lexical-semantic, and morphological information (among others) about vocabulary are theorized to lie in the so-called *mental lexicon*, the brain's storage for known words (Acha & Carreiras, 2014; Aitchison, 2012; Dijkstra, 2012). There are two fundamental representations—or neural traces—within the mental lexicon. The first is *word forms*, referring to the written (orthographic) or spoken (phonological) word representations. The second is *meaning representations*, which is the knowledge or information arbitrarily mapped to one or many word forms (Aitchison, 2012). This relationship can go from meaning to form during production or from form to meaning during recognition (Baxter et al., 2021). Thus, learning foreign vocabulary essentially involves establishing new form or meaning representations in the mental lexicon or validating existing form and meaning relationships (Marecka et al., 2021; T. Singer et al., 2003; Yang et al., 2015). Notably, during the early stages of foreign language learning, learners require conscious effort and explicit content regarding the mappings between form and meaning of foreign words (Ellis, 2015).

Concerning multilinguals—who know more than one language—, the literature has converged into the idea that the multilingual mental lexicon is language-nonselective. In other words, there is an *unified* mental storage for the word forms and meanings in all languages (Baxter et al., 2021; de Groot, 1992; Dijkstra & van Heuven, 2002). When individuals (monolinguals or bilinguals) learn foreign vocabulary, the novel word forms or meanings integrate into the pool of already known words and start taking part in processes such as competition (Baxter et al., 2021; Gaskell & Dumay, 2003). Based on this idea, the reported differences between monolinguals and bilinguals in foreign vocabulary learning and the results from Experiment 4 could be due to two reasons. First, bilinguals might possess a more extensive pool of knowledge—i.e., word forms or meanings in two languages—, allowing them to find perceived (or objective) similarities between form or meaning in any language (Bartolotti & Marian, 2017). Second, due to their experience mapping multiple words and meanings, they might be more efficient in establishing form or meaning representations within their mental lexicon than monolinguals (Kaushanskaya & Marian, 2009b). The former could be considered a direct effect, while the latter would be an indirect effect (Hirosh & Degani, 2018).

The present chapter explores whether monolingual and bilingual individuals differ when explicitly learning vocabulary-related information. Are bilinguals better than monolinguals at learning variations of known words? Experiment 5 compared the participants' performance when learning to recognize the orthographic form of novel suffixes appended to existing Spanish word stems (e.g., *laboral**suti***) without additional semantic information. Do bilinguals and monolinguals differ when learning similar- or only foreign-looking vocabulary? Experiment 6 tested participants' ability to learn foreign vocabulary paired with black-and-white drawings of existing objects. The critical manipulation in Experiment 6 was the sub-lexical orthographic similarity of this vocabulary to existing words in Spanish. Some words were more similar to Spanish, and other words were more foreign-looking when compared to Spanish orthography. Therefore, this experiment targeted both the sub-lexical orthography and the lexical-semantics levels.

3.1. Experiment 5: Morphological suffix learning

3.1.1. Rationale

Prefixes and suffixes (known together as affixes) occur in a combinatorial manner in a language (Rastle & Davis, 2008). For example, a specific suffix (e.g., -ness) can appear in conjunction with many word roots (e.g., kind<u>ness</u>, fit<u>ness</u>). In this regard, there are three dominant theories regarding the acquisition of affixal morphology (Merkx et al., 2011). First, according to the *morpheme boundary detection* view, learners implicitly track the sequential probabilities to detect boundaries and segment affixes. The extensive statistical learning literature backs this view (e.g., Misyak & Christiansen, 2007; Romberg & Saffran, 2010). In the context of bilingualism, the experience with tracking multiple sets of—often conflicting—statistics could provide an advantage for bilinguals over monolinguals when implicitly integrating this morphological information into the mental lexicon (Wang & Saffran, 2014). Indeed, Experiment 2 of this thesis showed that, despite the groups on average successfully generalizing the information from the statistical learning task, there were no differences between monolinguals and bilinguals.

The second theory is *morpheme chunking*, referring to how individuals learn affixes due to their combinatorial nature (Rastle & Davis, 2008). In other words, because affixes appear in conjunction with different known words, learners can efficiently segment them based on their pre-existing knowledge. For instance, knowing the words ***kind*** and ***fit*** a priori can facilitate the acquisition of ***kindness*** and ***fitness***, and more importantly, the acquisition of ***-ness*** as an affix that modifies the meaning of a word. Rastle and Davis (2008) suggested that, in this case, top-down influences of word form and meaning influence the bottom-up learning and lexicalization of affixes. This theory suggests that learners can acquire suffixal information regardless of the meaning of the unit as a whole—i.e., ***kindness*** does not mean the same as ***kind***. Thus, this view places a greater emphasis on an affixes' form rather than its meaning when accompanying a known word. Contextualizing this theory to the goal of this thesis, the experience with learning different affixal forms in two languages could potentially benefit bilinguals over monolinguals. However, no study has targeted bilingual and monolingual suffix learning.

The third and final view is that affixal learning is *semantically driven* (Merkx et al., 2011). According to this view, the co-occurrence of affixes and the variation in meanings drive the acquisition of morphological information. To compare the morpheme chunking and semantically driven views, Merks et al. (2011) tested two groups of English monolinguals when learning novel suffixes for known words (e.g., *sleepnept*). One group of participants only saw the form of the novel affixes accompanying a series of known words (form learning condition). The other group received a semantic description accompanying the novel words and suffixes (semantic learning condition). Their results suggested a comparable performance from both groups of learners when recognizing the novel suffixes, but successful lexicalization required a period of offline consolidation. Notably, while consolidation was further boosted in the semantic learning experiment, learning occurred regardless of the accompanying meaning.

To my knowledge, no study to date has compared monolinguals and bilinguals when learning affixal information. In Experiment 5, I compared Spanish monolinguals, Spanish-Basque bilinguals, and Spanish-English when learning novel suffixes for known Spanish words. Critically, the participants did not receive any semantic information accompanying these novel suffixes. Thus, I tested their learning of each suffix's form rather than its meaning. Basque speakers know a postpositional and agglutinative language—i.e., determiners and other morphological inflections append to the end of words, although there are some prefixes. In contrast, Spanish or English are prepositional languages where both prefixes and suffixes are used in affixal morphology. Corpus studies suggest that suffixes are more common than prefixes, even in prepositional languages (Ramscar, 2013). However, it is unclear whether bilinguals would outperform monolinguals due to the specific properties of their known languages. Conversely, experience with two languages may allow bilinguals to more flexibly integrate the form of novel suffixes, as supported by artificial vocabulary learning studies (Bartolotti & Marian, 2017). Hence, I adhered to the original hypothesis and expected bilinguals to outperform monolinguals when learning the form of novel suffixes.

3.1.2. Methods

Participants

Participants were forty Spanish monolinguals (M_{age} = 21.7, SD = 2.4; 35 females), forty Spanish-Basque bilinguals (M_{age} = 21.8, SD = 2.2; 32 females), and forty Spanish-English bilinguals (M_{age} = 21.0, SD = 2.4; 36 females). These participants were the same as in Experiment 2 and thus had the same profile as in the other experiments. Appendix A2 shows the demographic information and statistical contrasts between groups.

Materials

I combined sixty-four exiting Spanish word stems with sixteen artificial suffixes. The Spanish words could all be suffixed without changing the root (e.g., total/total-idad, color/color-ido)—also known as independent stems. This step was necessary so the words could be randomly combined with the artificial suffixes without breaking the Spanish rules for morphological inflection. For example, the word ***comer*** (to eat) cannot be suffixed without changing it (e.g., comedor, comiendo) and therefore would not represent a good candidate for this experiment. The words were 4 to 8 characters in length, were highly frequent (Zipf larger than 2.9), and were either nouns or adjectives. The information for each word was obtained from the EsPal database (Duchon et al., 2013).

I generated the artificial suffixes by modifying the letters of existing suffixes in Spanish. The selected existing suffixes had CVCV (e.g., -dado), VCVC (e.g., -idad), CVC (e.g., -tud), or VCV (e.g, -ido) configurations, where C is a consonant and V is a vowel. I modified the consonants and vowels to produce inexistent, and thus, artificial suffixes. These artificial suffixes are presented in Table 8. There were four suffixes in each configuration divided into two counterbalance lists (A and B). The Spanish word stems were then randomly paired with the artificial suffixes in each list at a ratio of eight words per suffix. Finally, the words were synthesized using the Mac OS X system's Text-to-Speech software with the Spanish female voice "Monica". The final lists of words are presented in Appendix B1.

Table 8. *Artificial suffixes in Experiment 5.*

List	CVCV	VCVC	CVC	VCV
A	-boru, -suti	-omed, -utet	-bur, -pol	-ibe, -une
B	-sotu, -bire	-isos, -atut	-pel, -ter	-odi, -ule

Additionally, I created three sets of foils to test participants' recognition performance after learning. The first type was comprised of thirty-two additional Spanish stems with similar properties as the original stimuli, randomly paired with the learned artificial suffixes (novel–learned). The second set of foils contained thirty-two of the original stems but paired with the suffixes from the other list (learned–novel). These two types of foils were *non-words* compared to the original stems and suffixes. Finally, the last type of foils was sixty-four *recombinant* pairs, created by randomly pairing the original stems with distinct learned suffixes. Naturally, I expected the recombinant condition to be more challenging than the non-word conditions.

Procedure

Participants were instructed to wear headphones and sit in a quiet room for the duration of this experiment. They performed a learning phase followed by an immediate recognition memory test and a delayed recognition test after one night of sleep (a minimum of 12 hours). Participants performed eight learning blocks during the learning phase. In each block, there were two presentations of each word (stem + suffix). Each trial presented a word at the screen center followed by an audio recording after 250 milliseconds from stimulus onset. Participants had to type the word using the keyboard and press enter to continue with the next trial. I adopted this procedure to ensure that the participants acquired the novel suffixes by learning them in multiple modalities (visual and auditory) and producing them. In total, there were 16 (2x8) presentations of each stem and 128 (16x8) presentations of each suffix throughout the learning phase. After the learning phase, participants performed an immediate recognition memory test. The test was an old versus new task where participants had to decide whether a word presented on the screen was learned or not. There were 192 trials in this test, and participants were encouraged to take a break after every 64 trials. They responded using the keyboard ("f" or "j" keys), and the computer automatically computed their accuracy. Notably, they also performed the old versus new task on a second day after at least one night of sleep.

Data Analysis

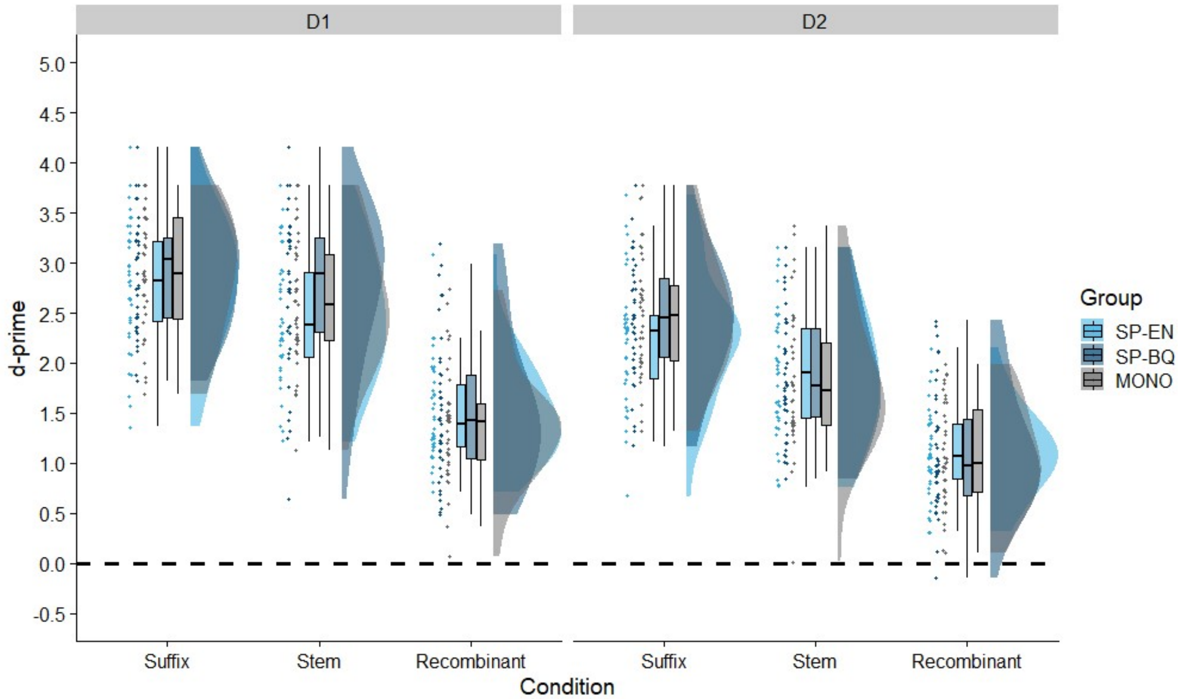
To correct response biases in the recognition memory test, I calculated d-prime to measure participants' sensitivity to the learned stems and suffixes against the three types of foils. The d-prime is a sensitivity metric extensively used in signal detection theory. It is computed as the differences between the Z-scores of hit and false alarm rates (Stanislaw & Todorov, 1999). This measure typically ranges from -1 to 2.5, with values significantly above 0 indicating above-chance discrimination of the stimuli. Following previous studies (Merkx et al., 2011), I obtained the d-prime scores for each participant with regards to each foil type, using the *dprime* function from the "psycho" package for R (Makowski, 2018). In other words, I calculated the false alarm rate for the novel-learned condition, trained-novel, and recombinant conditions separately. Then, I computed the d-prime of each condition against the hits in the original words (trained stems and suffixes). Thus, the d-prime for the novel-learned condition tested the participants' recognition of the trained stems (Stem condition). The scores for the learned-novel condition tested their sensitivity to the trained suffixes (Suffix condition). Finally, the d-prime for the recombinant condition tested participants' whole-word recognition memory (Recombinant condition).

I analyzed these scores using a linear mixed-effects model (LMM). The model included the main effects of day-of-testing (Day 1, Day 2), group (Spanish monolinguals, Spanish-Basque bilinguals, and Spanish-English bilinguals), condition (Stem, Suffix, Recombinant), and their two-way interactions. The day-of-testing factor was deviation coded as -1 and 1. The group factor was reverse Helmert coded to contrast the Spanish-Basque and Spanish-English groups on the first level and the bilinguals against the monolinguals on the second level. Finally, I also reverse Helmert coded the condition factor. In this case, I contrasted the Stem and Suffix conditions on the first level and both non-word conditions against the recombinant condition on the second level. The final model included the by-participant random intercepts and the random slope of day-of-testing and uncorrelated slopes for the condition contrasts. Critically, since d-prime is calculated over the by-participant aggregate data, I did not include by-item random effects. The degrees of freedom were approximated in the LMM using the Satterthwaite method as implemented by the *lmerTest* package in R (Kuznetsova et al., 2017).

3.1.3. Results

Before the analysis, I eliminated four participants. One participant from the Spanish-English group did not perform the recognition test on the second day. Furthermore, I eliminated one participant from each group with lower than chance scores in three out of the four conditions. Therefore, thirty-eight Spanish-English bilinguals, thirty-nine Spanish-Basque bilinguals, and thirty-nine Spanish monolinguals were included in the final analysis. As an initial step, I tested whether the list influenced participants' d-prime scores on the immediate test after the learning phase. A series of Mann-Whitney U-tests indicated no differences between the lists on the Stem (p = 0.719, BF_{10} = 0.209), Suffix (p = 0.330, BF_{10} = 0.268), or Recombinant (p = 0.546, BF_{10} = 0.227) conditions. Additionally, Wilcoxon one-sample tests revealed that participants, overall, discriminated above chance-level in the Stem (M_d = 2.611, SD = 0.702, p < 0.001, BF_{10} > 100), Suffix (M_d = 2.857, SD = 0.606, p < 0.001, BF_{10} > 100), and Recombinant (M_d = 1.472, SD = 0.597, p < 0.001, BF_{10} > 100) conditions on Day 1. The pattern was similar for Day 2 (all p < 0.001, BF_{10} > 100). The d-prime scores of each group and day-of-test are shown in Figure 8.

Figure 8. *Results of Experiment 5.*



Note. Raincloud plots showing the probability density of the d-prime scores by condition, group, and day-of-test. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

Table 9 shows the LMM results of Experiment 5. The results indicated that, as expected, the discrimination performance declined on the second day-of-test ($p < 0.001$, $BF_{10} > 100$). There was a significant difference between Suffix and Stem discrimination ($p < 0.001$, $BF_{10} > 100$), and between the Recombinant and Nonword discrimination performance ($p < 0.001$, $BF_{10} > 100$). Participants were better at discriminating the words when the target contained a completely new suffix than in the other two conditions. However, they performed the worst when discriminating new combinations of the learned stems and suffixes. These two contrasts interacted with the day-of-test factor. In brief, the gap in performance between the Suffix and Stem conditions was larger on Day 2 ($p < 0.001$, $BF_{10} = 8.004$). Similarly, the difference between the combined Nonword conditions and the Recombinant condition was also larger on Day 2 ($p < 0.001$, $BF_{10} = 8.004$). Furthermore, the analysis revealed no differences between the Spanish-Basque and Spanish-English bilinguals ($p = 0.367$, $BF_{10} = 0.007$). There were no differences between the monolingual and bilingual groups (p = 0.949, BF_{10} = 0.003). None of the other two-way interactions were significant (all p > 0.05, BF_{10} < 0.1). Additionally, none of the covariates reached significance (all p > 0.05, BF_{10} < 0.1)

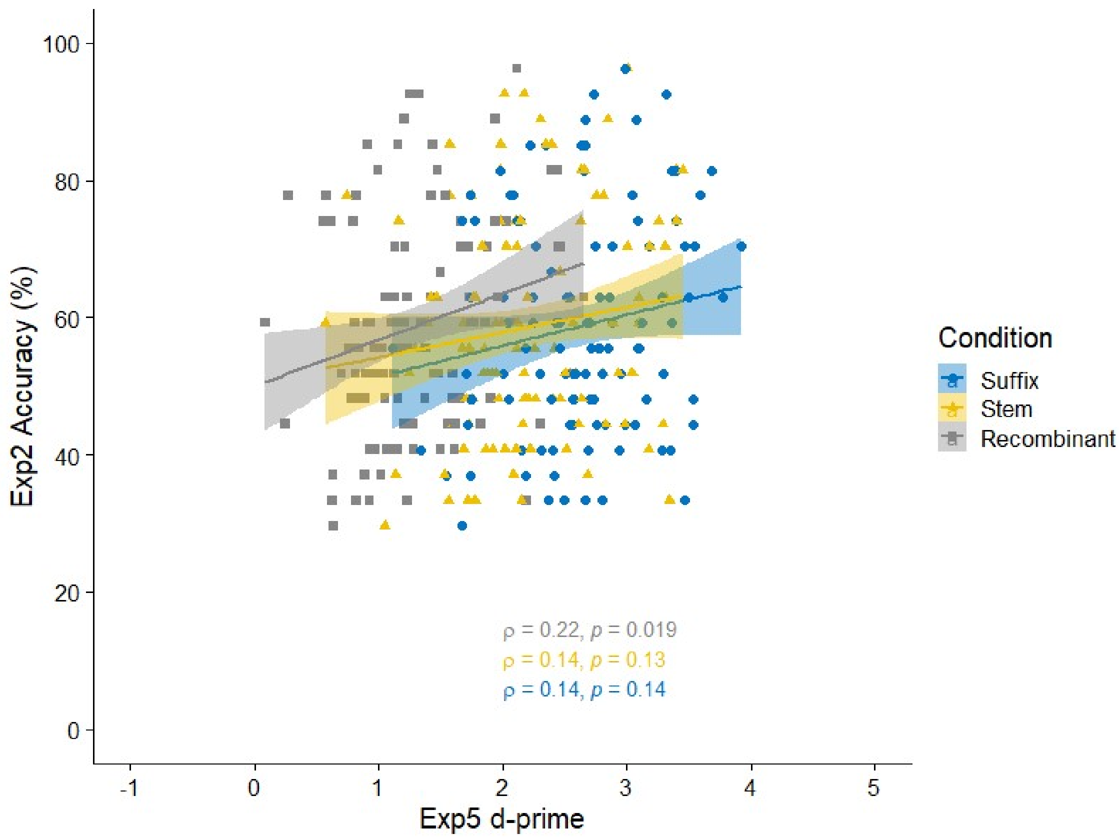
Table 9. *LMM results of Experiment 5.*

Fixed Effects		Estimate	SE	df	t	*p*	BF_{10}
(Intercept)		**2.049**	**0.051**	**110.6**	**40.223**	**< 0.001**	**-**
Day		**-0.265**	**0.016**	**113.0**	**-16.278**	**< 0.001**	**> 100**
Stem-Suffix		**-0.195**	**0.025**	**113.0**	**-7.803**	**< 0.001**	**> 100**
Recombinant-Nonword		**-0.380**	**0.009**	**113.0**	**-40.606**	**< 0.001**	**> 100**
SPBQ-SPEN		0.058	0.059	110.1	0.980	0.329	0.007
MONO-BIL		0.004	0.034	110.0	0.109	0.914	0.003
Day x Stem-Suffix		**-0.072**	**0.012**	**230.0**	**-5.949**	**< 0.001**	**8.004**
Day x Recombinant-Nonword		**0.040**	**0.007**	**230.0**	**5.690**	**< 0.001**	**9.209**
Day x SPBQ-SPEN		-0.016	0.020	113.0	-0.806	0.422	< 0.001
Day x MONO-BIL		0.007	0.012	113.0	0.648	0.519	< 0.001
Stem-Suffix x SPBQ-SPEN		-0.004	0.031	113.0	-0.141	0.888	0.004
Stem-Suffix x MONO-BIL		-0.021	0.018	113.0	-1.204	0.231	0.003
Recombinant-Nonword x SPBQ-SPEN		-0.019	0.011	113.0	-1.669	0.098	0.004
Recombinant-Nonword x MONO-BIL		-0.011	0.007	113.0	-1.731	0.086	0.003
Covariates							
Age		-0.022	0.020	110.0	-1.088	0.279	0.003
Non-verbal IQ		0.009	0.006	110.0	1.354	0.179	0.002
Gender		-0.026	0.142	110.0	-0.183	0.855	0.013
Random Effects	*Group*	*Variance*	*SD*	*Correlation*			
Participant	(Intercept)	0.251	0.501				
	Day	0.019	0.139	-0.22			
	Stem-Suffix	0.055	0.236				
	Recombinant-Nonword	0.004	0.067				
	Residual	0.068	0.261				

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

As a final exploratory step, I computed the correlation between Experiment 2's accuracy scores and the d-prime from Experiment 5 using Spearman rank correlations. Since I observed no differences between the groups in Experiments 2 and 5, I did not perform this analysis by each group. Instead, I computed a correlation score for each condition in Experiment 5, pooling the scores of the two days. The scatterplot and correlation results are shown in Figure 9. This analysis indicated no significant correlations between the Stem and Suffix conditions and the scores in Experiment 2 (all p > 0.05). However, the Recombinant condition d-prime scores and the statistical rule generalization scores from Experiment 2 showed a significant positive correlation (ρ = 0.22, *p* = 0.019). Notably, the Recombinant condition was the most challenging condition to discriminate during the recognition test in Experiment 5, as these were stems and suffixes seen during the learning phase. These results suggest that participants who were better at generalizing the "morphological" frames in Experiment 2 also performed better in this condition but not in the Stem or Suffix conditions.

Figure 9. *Spearman correlations of Experiment 2 and Experiment 5 scores.*



Note. Shading indicates the 95% confidence interval. Each shape is an individual observation.

3.1.4. Summary

Experiment 5 showed that, while participants could learn to recognize the novel artificial suffixes, there were no differences between monolinguals and bilinguals in this artificial suffix recognition task. Additionally, in this experiment, I tested participants' ability to retain their knowledge over a short delay and one night of sleep. While performance decreased on the second day, the groups were comparable regardless of the test delay. The three languages under study (Spanish, English, and Basque) in this work possess more suffixes than prefixes (Ramscar, 2013). Even in bilingual participants, there might not be sufficient differences to affect suffix learning in particular. Another possibility for this lack of differences is that, by using Spanish words as stems, this experiment might not have targeted foreign language learning but learning of novel linguistic information within a language. In this regard, the experiment's instructions and content might have implicitly primed bilingual participants to operate in Spanish and disregard any contribution from their respective L2s.

There is extensive research suggesting an essential role of sleep in consolidating knowledge, in general, and linguistic information, in particular (e.g., Batterink et al., 2017; Mirkovic & Gaskell, 2016; Stickgold, 2005; Walker & Stickgold, 2004). While this experiment did not target consolidation as a process, the results showed that participants could retain the information after just a few exposures to the new suffixes—particularly their form—and over at least 12 hours. Naturally, the performance declined on the second day of testing, but participants could still successfully discriminate the learned stem and suffixes from the foil conditions above chance level. Previous studies have shown that consolidation and semantic information might primarily influence the lexicalization of novel morphological information rather than recognizing suffixes' form (Merkx et al., 2011). It is unclear whether adding meaning to these novel suffixes—or testing their lexicalization—would have led to observed differences between the monolingual and bilingual groups.

Given the scarce literature on monolingual and bilingual morphology learning (Hirosh & Degani, 2018), there is ample room for improvement at the theoretical, methodological, and practical (i.e., developing new instruction methods) levels. I can claim there were no differences between monolinguals and bilinguals in this suffix form recognition task, but this claim does not extend to the entirety of morphological learning. Future studies could investigate suffix learning using other methods to assess knowledge of the learned materials. In addition, constructing an entirely foreign artificial vocabulary might also help disentangle the effects of language mode—i.e., participants performing the entire task in a "Spanish" setting—from the influence of the bilingual experience. With this in mind, in Experiment 6, I tested monolingual and bilingual foreign vocabulary learning with varying orthographic similarity to Spanish to address the sub-lexical orthography and lexical-semantics level.

3.2. Experiment 6: Foreign language vocabulary learning

3.2.1. Rationale

Individuals often rely on words similar in form (orthographic or phonological) to the native language to jump-start their foreign vocabulary knowledge (Bartolotti & Marian, 2017; Hayakawa et al., 2020). For instance, an English speaker learning Spanish should find it easier to acquire the words “pera” (pear; a cognate) and “carta” (letter and not card; a false friend) over the word “perro” (dog; a non-cognate). Indeed, artificial vocabulary learning studies have shown that both cognates and false friends are acquired faster than non-cognate words (Marecka et al., 2021; Otwinowska et al., 2020). Even without complete form overlap, similarities in how groups of letters and sounds combine—sub-lexical orthotactic or phonotactic probabilities—within a word can lead to faster vocabulary acquisition (Meyer & Schmitt, 2002; Storkel et al., 2006). Learners seem to easily integrate similar words into the mental lexicon regardless of their meaning, suggesting a reliance on the form representations within the mental lexicon (Ecke, 2015; Marecka et al., 2021; Ringbom, 2006). In other words, as long as a word is similar in at least its sub-lexical orthographic or phonological form to a known language, individuals should acquire them faster.

Compared to monolinguals learning words in their second language, bilinguals could exploit knowledge from their two languages during third language vocabulary learning. Several studies have highlighted that bilinguals could transfer knowledge from any known language on a construction-by-construction basis (Jarvis & Pavlenko, 2007; Mihi, 2016; Rothman, 2015; Westergaard et al., 2017). Prior research also suggests that bilinguals can incorporate novel words similar in orthography to at least one of their languages faster than non-cognate words when learning foreign vocabulary (Bartolotti & Marian, 2017). Furthermore, as covered in Chapter 1, studies that have directly compared monolingual and bilingual vocabulary learning abilities have shown that bilinguals outperform their monolinguals peers when learning foreign-sounding unfamiliar words (e.g., Antoniou et al., 2015; Kaushanskaya & Marian, 2009b). Together, these findings suggest that experience with two languages could provide a vocabulary learning advantage over experience with a single language.

Despite all these findings, several questions remain unanswered. Most studies measure recognition and recall of foreign-sounding—and hence entirely unfamiliar—vocabulary at the end-state of the learning process. However, it is essential to establish whether this advantage emerges from different learning trajectories in monolinguals and bilinguals or just during the test phase. In other words, do bilinguals differ from monolinguals in the process of learning or only in the outcome? Furthermore, studies that have compared monolingual and bilingual learning have only employed foreign-sounding or looking words. Bilinguals may exploit the similarity to any known languages while maintaining flexibility to integrate these foreign-looking or sounding words. Do bilinguals only outperform monolinguals in learning foreign-sounding/looking words or also in learning similar words?

To my knowledge, no prior study has compared monolinguals and bilinguals when learning novel vocabulary with varying sub-lexical orthographic similarity to their common native language. To test the effects of sub-lexical orthographic similarity, I constructed a small (48 words) artificial written vocabulary—dubbed *Flavian* from the acronym for foreign language vocabulary (FLV)—as a benchmark for monolingual and bilingual foreign vocabulary learning performance. Half of the vocabulary consisted of orthographically similar words to Spanish (ES+ words), while the other half were orthographically dissimilar or non-cognate words (ES- words). Participants learned this vocabulary through an active and explicit vocabulary learning task that involved recognizing and producing Flavian words *with feedback*. The hypothesis in Experiment 6 was that the ES+ words, due to their more similar sub-lexical orthography to Spanish, would be learned better than the ES- words. Critically, I expected bilinguals to outperform monolinguals in learning both types of words (i.e., a general bilingual vocabulary learning advantage). Alternatively, these two variables could interact, and therefore bilinguals could be better only at the dissimilar but not the similar words.

3.2.2. Methods

Participants

Forty Spanish monolinguals (M_{age} = 20.4, SD = 2.1; 36 females), forty Spanish-English bilinguals (M_{age} = 21.0, SD = 2.5; 36 females), and forty Spanish-Basque bilinguals (M_{age} = 21.5, SD = 2.6; 29 females) participated in Experiment 6. Participants had a similar profile as in all the previous experiments. Appendix B2 shows the demographic information and statistical contrasts between groups.

Materials

I created an artificial vocabulary by randomly producing nonwords with CVC-CV structure—where each C is a consonant and each V a vowel, and the hyphen marks the syllabic boundary. I used all possible consonants for the initial and second consonants, excluding the letters h, k, q, v, w, x, y, z. The letters l and r were excluded from the third consonant pool to avoid fluid consonant clusters and thus maintain the syllabic boundary (e.g., fl, fr, cr, cl). Initially, I generated 1,000 nonwords, from which I removed those that already existed in Spanish, English, or Basque. For each nonword, I calculated the gram and bigram frequency sum in Spanish and English using CLEARPOND (Marian et al., 2012) and in Basque using E-HITZ

(Perea et al., 2006). I also obtained each nonword's orthographic Levenshtein distance (old20) to existing Spanish, English, and Basque words with the *vwr* package in R (Keuleers, 2013). Using these scores, I calculated a global composite sub-lexical orthographic similarity score by averaging the Z-scores for each variable; and then discarded nonwords with a composite score between -1 and 1. From the resulting list of 248 nonwords, I manually removed those nonwords that were very similar to each other and recalculated the sub-lexical orthographic similarity scores with respect to Spanish, reducing it to a pool of 180 nonwords.

These 180 nonwords were rated by ten native Spanish-speakers on a scale from 1—the nonword does not look at all like Spanish—to 4—the nonword is very similar to Spanish. I included categories for whether the nonword resembled an existing Spanish word (due to its very similar phonology or orthography) or any other known language. Twenty real Spanish words were also included to control for each rater's attention during the task. I used the intra-class correlation coefficient to measure the agreement between raters of the nonwords (Koo & Li, 2016). There was a moderate agreement between the ten raters, using the two-way random effect model and the "average rater" unit (kappa = 0.74, $F_{(199, 1791)} = 6.9$, $p < 0.001$). Thus, I included the Z-score of the average ratings per nonword into the composite score, using only the Spanish old20 and gram/bigram scores. Nonwords with scores between -0.5 and 0.5 were removed to create two dichotomous categories based on the sub-lexical orthographic patterns. Finally, I manually selected 24 nonwords orthographically similar to Spanish from this list, called ES+ words, and 24 nonwords that were orthographically dissimilar to Spanish or ES- words[1]. The final Flavian word list can be found in Appendix B3.

Procedure

Participants performed an active vocabulary learning task. At the beginning of the experiment, they were told they would learn vocabulary from a language named Flavian. This instruction was included based on prior experiments and to engage participants in the task (Bartolotti & Marian, 2017). Each Flavian word was paired with a black and white depiction of a real object selected from the MultiPic dataset (Duñabeitia et al., 2018). The images were selected based on their low visual complexity, the gender of the depicted object and had an H-index of 100 (the percentage of agreement across different participants). The

[1] The original list contained 28 ES+ and 28 ES- words, but was reduced after piloting the experiment. Pilot participants performed poorly in the task due to the large number of items. Using 48 items also diminished the total duration of the experiment.

depicted objects were all concrete nouns (e.g., button, chair, pear). There were four counterbalance lists to ensure that the effects were not due to any specific pairings of Flavian words and images.

The vocabulary learning task involved recognizing and producing the Flavian words *with feedback*. I chose this procedure to maximize comparability with similar vocabulary learning studies (Bartolotti & Marian, 2017; Hayakawa et al., 2020; Marecka et al., 2021). The participants first performed one block of familiarization with the Flavian words, where the pairings between the images and words were shown once for 2 seconds each. After the familiarization phase, they completed five blocks of alternating recognition and production tasks as a learning phase. The order of the trials in all these tasks was fully randomized.

In the *recognition task*, participants visualized a Flavian word at the center of the screen with four possible images (one target and three foils) presented at each corner of the screen. Therefore, this was a 4-alternative forced-choice (4AFC). I controlled that each image appeared once as a target and three times as a foil. Participants used the keyboard (”d”, “c”, “k”, and “m” keys) to select which image corresponded with the presented word without any time pressure. After submitting their responses, the correct image remained on the screen alongside the word for 2 seconds, and a feedback message (correct or incorrect) appeared on the screen. The computer automatically recorded their accuracy. There was a variable intertrial interval sampled from a uniform distribution between 0.5 and 1 second.

In the *production task*, participants saw an image at the center of the screen, and they had to type the correct Flavian name of the depicted object using the keyboard. They could correct the typed word if needed and pressed enter to submit their final response. The computer automatically calculated the typed words’ accuracy by assigning a score of 0.2 for each letter in its correct position for a maximum score of 1. Each additional letter past the maximum length of each Flavian word (5) received a 0.2 penalty, with a minimum score of 0. After submitting their responses, the correct Flavian name appeared below the typed word, alongside a feedback message (correct, partially correct, or incorrect) for 2 seconds, followed by a variable intertrial interval sampled from a uniform distribution between 0.5 and 1 second. After 3 seconds, if participants had not still submitted a response, a message appeared on the screen encouraging them to type the word as best as possible or just press enter to see the response.

After the learning phase, participants performed a computerized version of the *Operation Span* task (Unsworth et al., 2005) as a distractor task. This task lasted about 25 minutes and measured participants’ working memory capacity. In this task, participants verified a series of simple mathematical equations (e.g., (5 * 2) + 5 = 15 ?) and decided whether they were correct or incorrect using the keyboard (“f” and “j” keys). After submitting the response for each equation, they visualized a letter from the alphabet for 2 seconds, which they were instructed to memorize. After a fixed number of equations, participants were instructed to type the letters presented during the equation verification task in the order in which they saw them. The task's difficulty increased gradually, starting with two equations and two letters trials as practice and increasing up to seven equations and letters. Each difficulty level was repeated three times with different letters.

Finally, participants performed a *test phase* that consisted of two blocks of production and recognition trials: one after the short delay (D1) and another after one night of sleep (D2). In these blocks, participants did not receive any feedback for the typed or recognized words. However, to avoid participants rehearsing the Flavian words during the recognition task, they completed the production task first, followed by the recognition task. Other than this, the trials in each block were precisely the same as in the learning phase.

Data analysis

The learning phase’s recognition and production data were modeled separately using logistic and linear Growth Curve Analyses (GCA), respectively, in R (Mirman, 2017). GCA is a multi-level mixed model perfectly suited to capture the nested structure of time-course data while simultaneously quantifying group-level and individual-level patterns (Mirman, 2017; Mirman et al., 2008; J. D. Singer & Willett, 2009). I included second-order orthogonal polynomials to reflect linear and quadratic changes across the blocks, using the group (i.e., SP-EN vs. SP-BQ vs. MONO), condition (ES+ vs. ES-) factors, and their interactions as fixed effects on all time-terms. Participants’ age, non-verbal IQ, gender, and Operation Span scores were also introduced as covariates in the models.

I contrast coded the condition and group factors according to my hypothesis of a facilitatory effect of orthographically similar words and overall bilingual advantage. The condition factor was deviation coded as -1 (ES-) and +1 (ES+). The group factor was reverse Helmert coded first to contrast the Spanish-English and Spanish-Basque groups (coded as -1 and 1), then contrasting the bilingual (BIL) groups against the MONO group (coded as -2 and 2, respectively). I included the main effects and interactions of the time, condition, and group factors into the fixed effects but excluded any three-way interactions with the time-terms from the GCA models. The degrees of freedom for the linear GCA were approximated in the LMM using the Satterthwaite method as implemented by the *lmerTest* package in R (Kuznetsova et al., 2017). In contrast, the logistic GCA was estimated as an exact Z-test—since the mean and standard deviation are known in the logistic model. I performed similar G/LMM analyses for the test phase, using the recognition and production data, respectively. However, in this case, the only additional factor was the test day instead of the learning block, coded as -1 (Day 1) or +1 (Day 2).

3.2.3. Results

As an initial step, I compared whether the list influenced participants' performance during the learning and testing phases. For the learning phase, an ANOVA indicated no significant effect of the list in the last block of the recognition (p = 0.499, BF_{10} = 0.113) and production (p = 0.404, BF_{10} = 0.136) tasks. Similarly, there were no differences due to the list on the recognition test on Day 1 (p = 0.336, BF_{10} = 0.162) and Day 2 (p = 0.366, BF_{10} = 0.131), or on the production test on Day 1 (p = 0.440, BF_{10} = 0.125) and Day 2 (p = 0.435, BF_{10} = 0.127). Therefore, I did not include the list as a factor in the final analyses.

I also calculated the score obtained by each participant in the Operation Span task as the sum of each fully correct letter recall, with a maximum score of 75 (Unsworth et al., 2005). I compared the three groups using an ANOVA with Helmert coding for the group factor. The results indicated no differences between the three groups ($F_{(2, 117)}$ = 2.527, p = 0.084, BF_{10} = 0.633). Participants in the monolingual group had an average score of 55.1 (SD = 15.3), while those in the Spanish-English scored 56.7 (SD = 15.9), and the Spanish-Basque on average scored 61.9 (SD = 10.7). The Helmert contrasts also indicated no differences between the monolingual and bilingual groups (p = 0.128) and no differences between the two bilingual groups (p = 0.103). Thus, overall, the groups had comparable working memory capacity as measured by this task.

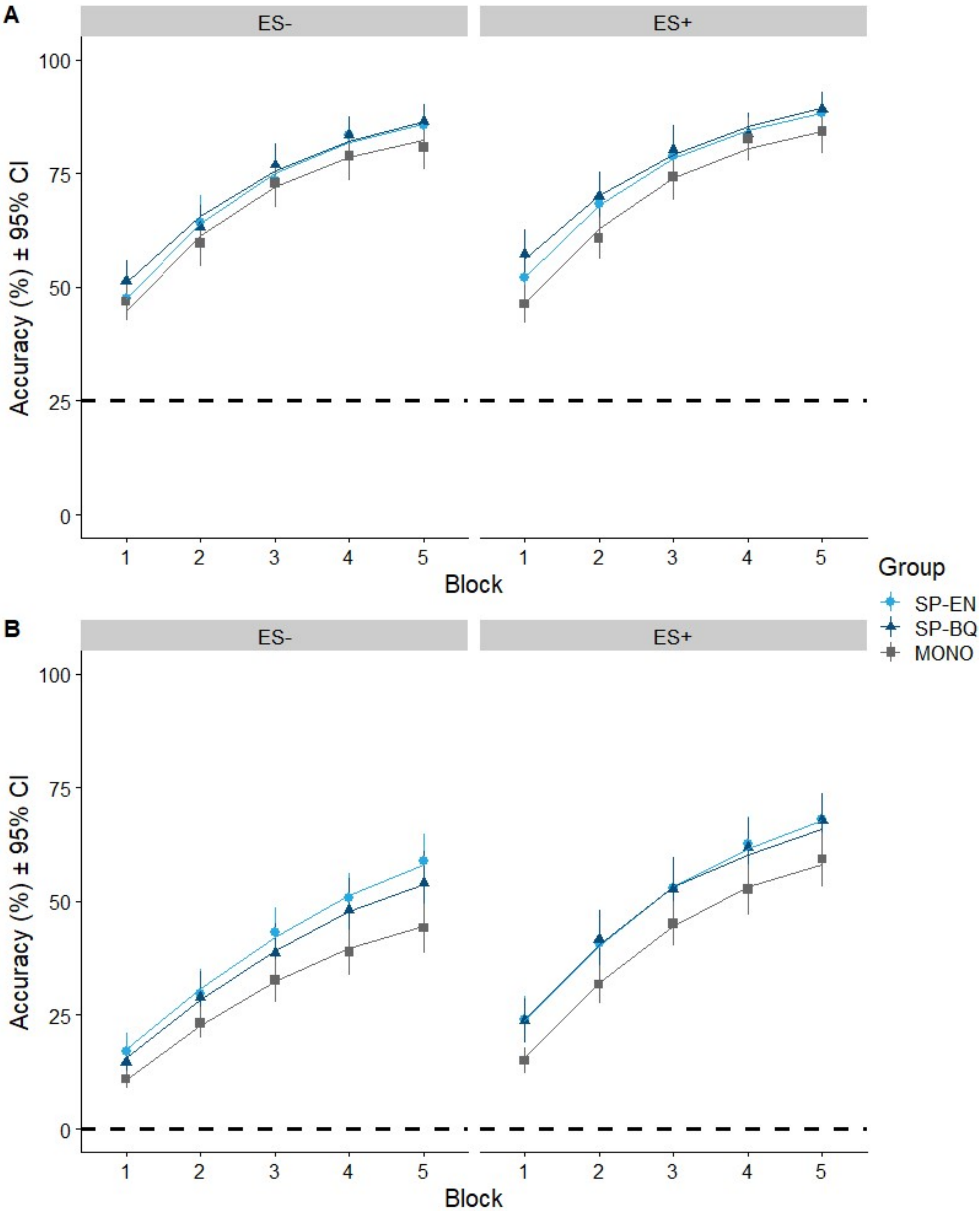
Learning phase. I performed the GCA analysis on the count of hits and misses using a binomial distribution and a logit link for the recognition task. The final model contained the by-participant random intercept and slopes for the Linear and Quadratic terms and a participant-by-condition random intercept. The results are shown in Table 10. The analysis revealed significant effects of the Linear (p < 0.001, BF_{10} > 100) and Quadratic (p < 0.001, BF_{10} > 100) terms, suggesting that individuals learned the words throughout the blocks in the recognition task. As expected, there was also a significant main effect of the condition (p < 0.001, BF_{10} > 100), where the ES+ words overall were recognized better than the ES- words. Moreover, the difference between Spanish-Basque and Spanish-English bilinguals was not significant (p = 0.628, BF_{10} = 0.088). Critically, bilinguals were better at recognizing the Flavian words than monolinguals throughout the learning phase (p = 0.030, BF_{10} = 0.254). Interestingly, there was a marginally significant interaction between the Condition and the monolingual versus bilingual contrast (p = 0.044, BF_{10} = 0.029). It seems that the differences between monolinguals and bilinguals were larger for the ES+ condition than the ES- condition. The remaining two-way interactions between the group, condition, time terms, or any covariate reached significance (all $p > 0.05$, $BF_{10} < 0.2$). The learning trajectories for each group in the recognition task are depicted in Figure 10A. By the fifth block, participants in the Spanish monolingual group on average reached 82.5% (SD = 14.3) accuracy; those in the Spanish-English group averaged 86.9% (SD = 13.3), and finally, the Spanish-Basque bilinguals averaged 87.8% (SD = 10.7).

Table 10. *Logistic GCA of the recognition learning task.*

Fixed Effects		**Estimate**	**SE**	**z**	***P***	**BF_{10}**
(Intercept)		**1.203**	**0.079**	**15.317**	**< 0.001**	**-**
Linear		**1.767**	**0.075**	**23.464**	**< 0.001**	**> 100**
Quadratic		**-0.201**	**0.041**	**-4.891**	**< 0.001**	**> 100**
Condition		**0.094**	**0.015**	**6.286**	**< 0.001**	**> 100**
SPBQ-SPEN		0.045	0.093	0.485	0.628	0.088
MONO-BIL		**-0.117**	**0.054**	**-2.168**	**0.030**	**0.254**
Linear x Condition		0.031	0.034	0.910	0.363	0.008
Quadratic x Condition		0.008	0.033	0.232	0.816	0.007
Linear x SPBQ-SPEN		-0.024	0.090	-0.268	0.789	0.078
Linear x MONO-BIL		-0.059	0.051	-1.169	0.243	0.124
Quadratic x SPBQ-SPEN		0.017	0.047	0.353	0.724	0.041
Quadratic x MONO-BIL		-0.020	0.026	-0.739	0.460	0.128
Condition x SPBQ-SPEN		0.008	0.018	0.437	0.662	0.003
Condition x MONO-BIL		**-0.020**	**0.010**	**-2.012**	**0.044**	**0.029**
Covariates						
Age		-0.021	0.021	-1.009	0.313	0.029
Non-verbal IQ		-0.005	0.004	-1.203	0.229	0.031
Gender		0.101	0.135	0.751	0.453	0.046
OSPAN score		0.004	0.003	1.245	0.213	0.080
Random Effects	*Group*	*Variance*	*SD*	*Correlation*		
Participant	(Intercept)	0.642	0.801			
	Linear	0.460	0.678	0.85		
	Quadratic	0.038	0.194	-0.67	-0.40	

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

Figure 10. *Recognition and production accuracy during the learning phase in Experiment 6.*



Note. Average observed accuracy (symbols, vertical lines indicate ± 95% confidence intervals) by group and condition for the recognition (A) and production (B) tasks. The solid lines depict the average GCA model predicted values. The conditions are plotted separately to avoid cluttering. The dashed line in indicates chance-level accuracy. SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

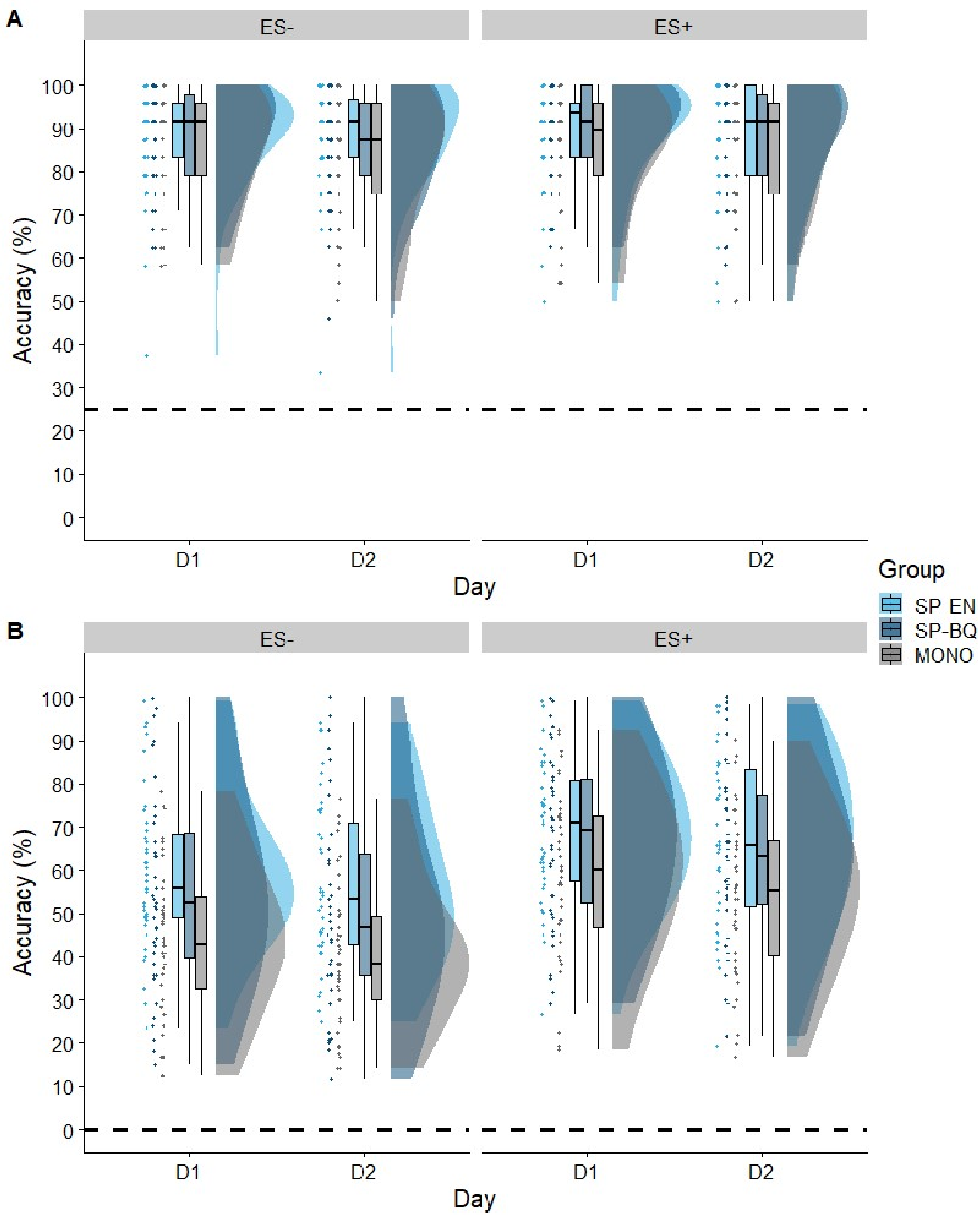
Table 11 shows the results for the linear GCA on the production task accuracy. In this case, the analysis was performed on the summed partial accuracy data by participant, block, and condition. The final model converged with by-participant random intercept and random slopes for the Linear and Quadratic terms, and participant-by-condition intercept, and Linear and Quadratic random slopes. As in the recognition task, the results showed significant effects of the Linear ($p < 0.001$, $BF_{10} > 100$) and ($p < 0.001$, $BF_{10} > 100$) time-terms. The accuracy in the ES+ condition was also significantly higher than in the ES- condition ($p < 0.001$, $BF_{10} > 100$). There were no differences between the two bilingual groups ($p = 0.398$, $BF_{10} = 0.025$), but both bilingual groups outperformed the monolingual group ($p = 0.005$, $BF_{10} = 10.122$). Additionally, there were two significant two-way interactions. First, the Linear time-term by Condition interaction suggested that participants had a steeper learning slope for the ES+ than the ES- condition ($p < 0.001$, $BF_{10} > 100$). Second, the Quadratic by Condition interaction suggested that the curvature through the learning blocks was different for the ES+ and the ES- conditions ($p < 0.001$, $BF_{10} > 100$). The rest of the interactions were not significant (all $p > 0.05$, $BF_{10} < 0.5$). Finally, out of the covariates, there was a significant effect of participants' gender ($p = 0.025$, $BF_{10} = 0.596$) where it seemed that males performed better than females in this task. I refrain from discussing any gender differences in vocabulary learning, as these are outside the scope of this thesis. Participants' performance through the production learning blocks is depicted in Figure 10B. By the fifth block, participants in the Spanish monolingual group reached an average score of 51.7% (SD = 17.0), participants in the Spanish-English group a score of 63.4% (SD = 17.2), and those in the Spanish-Basque bilingual group 61.0% (SD = 19.6). This implies that participants could correctly write around half of the words in the experiment.

Table 11. *Linear GCA of the production learning task.*

Fixed Effects		Estimate	SE	df	t	*p*	BF_{10}
(Intercept)		**9.627**	**0.342**	**124.1**	**28.183**	**< 0.001**	**-**
Linear		**7.759**	**0.241**	**117.3**	**32.207**	**< 0.001**	**> 100**
Quadratic		**-1.357**	**0.128**	**117.3**	**-10.571**	**< 0.001**	**> 100**
Condition		**1.329**	**0.091**	**116.4**	**14.619**	**< 0.001**	**> 100**
SPBQ-SPEN		-0.343	0.405	117.0	-0.848	0.398	0.025
MONO-BIL		**-0.679**	**0.235**	**118.7**	**-2.884**	**0.005**	**10.122**
Linear x Condition		**0.550**	**0.112**	**118.9**	**4.913**	**< 0.001**	**> 100**
Quadratic x Condition		**-0.295**	**0.086**	**119.0**	**-3.439**	**0.001**	**> 100**
Linear x SPBQ-SPEN		-0.166	0.295	117.5	-0.564	0.574	0.033
Linear x MONO-BIL		-0.246	0.170	117.1	-1.443	0.152	0.066
Quadratic x SPBQ-SPEN		-0.031	0.157	117.5	-0.200	0.842	0.011
Quadratic x MONO-BIL		-0.019	0.091	117.1	-0.209	0.835	0.159
Condition x SPBQ-SPEN		0.164	0.099	116.3	1.650	0.102	0.001
Condition x MONO-BIL		-0.071	0.057	115.8	-1.232	0.220	0.486
Covariates							
Age		-0.134	0.103	113.4	-1.301	0.196	0.008
Non-verbal IQ		-0.017	0.020	112.8	-0.847	0.399	0.001
Gender		**1.505**	**0.662**	**112.6**	**2.275**	**0.025**	**0.596**
OSPAN score		0.021	0.017	112.6	1.243	0.216	0.007
Random Effects	*Group*	*Variance*	*SD*	*Correlation*			
Participant by Condition	(Intercept)	1.697	1.303				
	Linear	1.573	1.254	0.49			
	Quadratic	0.339	0.582	-0.90	-0.56		
Participant	(Intercept)	11.690	3.419				
	Linear	5.450	2.335	0.54			
	Quadratic	1.089	1.044	-0.86	-0.66		

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

Figure 11. *Recognition and production accuracy during the test phase in Experiment 6.*



Note. Raincloud plots showing the probability density of the accuracy scores by condition, group, and day-of-test. The center of the boxplot indicates the median, and the limits of the box define the interquartile range (IQR = middle 50% of the data) for each group. Dots reflect individual participant scores. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; D1 = Day 1; D2 = Day 2.

Test phase. A GLMM on the recognition test scores only revealed a significant effect of the condition (p = 0.033, BF_{10} = 0.439). The rest of main effects and interactions were not significant (all $p > 0.05$, $BF_{10} < 0.3$). For brevity, I report the full table of results in Appendix B4. In all, the recognition test performance did not decrease from one day to the next, but the differences between monolinguals and bilinguals observed during the learning phase disappeared. The scores by condition, group, and day are presented in Figure 11A. On average, participants in the monolingual group scored 85.1% (SD = 11.8) on Day 1, and 84.8% (SD = 12.7) on Day 2. Similarly, participants in the Spanish-English bilingual group scored 88.5% (SD = 11.8) and 88.5% (SD = 11.9) on days 1 and 2, respectively. Finally, the Spanish-Basque group averaged 88.8% (SD = 10.6) on Day 1 and 87.1% (SD = 11.5) on Day 2.

An LMM of the production test scores (shown in Appendix B5) indicated that participants' performance in the production task decreased on Day 2 ($p < 0.001$, $BF_{10} > 100$). Additionally, as in the recognition task, participants were better at the ES+ words than the ES- words on both days ($p < 0.001$, $BF_{10} > 100$). The results revealed no differences between the Spanish-Basque and Spanish-English bilinguals (p = 0.357, BF_{10} = 0.075), but both bilingual groups outperformed their monolingual peers (p = 0.002, BF_{10} = 3.033). There were no significant two-way interactions between the day, condition, or group contrasts (all $p > 0.05$, $BF_{10} < 0.2$). The production scores by condition and day are depicted in Figure 11B. Participants in the Spanish monolingual group averaged 51.7% (SD = 17.7) and 47.6% (SD = 16.7) on days 1 and 2, respectively. Individuals in the Spanish-English bilingual group scored 64.1% (SD = 16.4) on Day 1 and 61.6% (SD = 18.1). Finally, the Spanish-Basque bilingual group averaged 61.0% (SD = 20.3) on Day 1 and 57.0% (SD = 21.6) on Day 2, respectively. Notably, these scores are based on participants' partial accuracy, which means that, on average, they could partly produce around half of the words (i.e., 24 words) in the experiment [2].

[2] Indeed, considering responses with a score of 0.8 and above as correct and performing a GLMM analysis using absolute accuracy instead of partial accuracy led to similar results, with lower scores across the board and a global average of around 40% in the ES- condition and 50% in the ES+ condition.

3.2.4. Summary

In Experiment 6, I tested the role of sub-lexical orthographic similarity and bilingual experience vocabulary learning. This experiment measured both the learning trajectory and the outcome after learning an artificial vocabulary (Flavian) with varying degrees of sub-lexical orthographic similarity to Spanish. Furthermore, I tested participants' receptive and productive vocabulary using two distinct tasks. Overall, the results consistently showed an effect of sub-lexical orthographic similarity during the learning and testing phase. In other words, words that were more similar to Spanish (ES+ condition) were learned faster and remembered better during the test by all groups in the recognition and production tasks than less similar (ES- condition) words. These findings are consistent with the growing literature suggesting a role of orthographic and phonological similarity during foreign vocabulary learning (e.g., Bartolotti & Marian, 2017; Hayakawa et al., 2020; Marecka et al., 2021). However, even though similarity aids vocabulary learning at early stages, it might hinder the acquisition of less similar vocabulary later on (Marian et al., 2021). Therefore, it would be interesting to see how sub-lexical orthographic or phonological similarity affects vocabulary learning using different word lists and across longer time-spans.

More importantly, the results revealed a difference between monolinguals and bilinguals, but no differences between the bilingual groups. Both bilingual groups outperformed their monolingual peers during the learning phase in their recognition and production accuracy. Bilinguals were also better at producing but not recognizing the words during the test phase. These two findings align with previous literature showing a bilingual vocabulary learning advantage (e.g., Antoniou et al., 2015; Kaushanskaya & Marian, 2009b, 2009a). The findings of this experiment extend the existing literature in two important ways:

First, the results of this experiment show how the differences emerge throughout the time-course of learning, with bilinguals outperforming monolinguals during the learning blocks in both recognition and production tasks. Curiously, bilinguals outperformed monolinguals in the ES+ condition (but not the ES- condition) after just the familiarization phase, particularly in the production task. It is possible that, compared to monolinguals, bilinguals were more willing to produce the words after having just seen them once. Nevertheless, the group differences disappeared during the recognition but not the production tests. In other words, measuring only the outcome of learning would have led to a lack of differences at the receptive vocabulary level. Admittedly, the recognition task was less challenging than the production task. The number of items (48) and extensive practice with the recognition task might have nudged participants' accuracy towards a ceiling, thus eliminating any differences between the groups. It would be helpful to explore whether the strategies employed by monolingual and bilingual participants differ during these types of active learning experiments.

Second, these differences were evident in both the ES- and ES+ conditions—with only a marginally significant interaction between condition and group in the recognition learning task—supporting the idea that the bilingual vocabulary learning advantage is present for similar and unfamiliar novel words. In other words, bilingual individuals seem to be better at learning vocabulary, not just with foreign-looking or sounding words. As suggested by prior studies, this bilingual advantage could be related to mechanisms such as enhanced orthographic or lexical-semantic networks, as well as enhanced working memory capacity in bilinguals (Hirosh & Degani, 2018; Kaushanskaya & Marian, 2009b). However, there were no differences between the groups on their verbal working memory as measured by the Operation Span task, nor did this variable significantly account for the results. Therefore, it is still unclear how experience with two languages can foster novel vocabulary acquisition, irrespective of the specific language pairs.

Discussion

In this chapter, I compared monolingual and bilingual language learning performance across two explicit learning experiments. Experiment 5 tested participants' ability to integrate the orthographic form of novel suffixes using existing Spanish words as stems to target the morphology level. This experiment indicated that participants could discriminate the learned stems and suffixes above chance and over two days. However, monolingual and bilingual participants' performance did not differ in this task. Experiment 6 addressed both the sub-lexical orthography and lexical-semantics level. In this experiment, participants learned words that varied in their sub-lexical orthographic similarity to the language they all had in common (Spanish). They learned the words' meaning by seeing them in association with depictions of existing objects. In this case, the results of this experiment revealed that sub-lexical orthographic similarity and bilingual status facilitated the learning of this novel vocabulary, with no interaction between these factors. In brief, words that were more similar in their orthography to Spanish were acquired faster during the learning blocks and remembered better during the test. Moreover, bilinguals showed an overall advantage over monolinguals at recognizing and producing these words during learning, regardless of their sub-lexical orthographic similarity. However, bilinguals were only better at producing the words during the test phase.

The results from Experiment 5 generally support the conclusions from its implicit learning counterpart: Experiment 2. That is, participants can learn to recognize and generalize morphological knowledge through implicit and explicit mechanisms, but there are no differences between bilinguals and monolinguals in their morphology learning—at least as measured by these tasks. In Experiments 2 and 5, I respectively tested the morpheme boundary detection and morpheme chunking theoretical views of affixal morphology learning (Merkx et al., 2011; Misyak & Christiansen, 2007; Rastle & Davis, 2008). The former argues that affixes are learned through statistical learning mechanisms, while the latter argues that affixes are integrated into the mental lexicon because of their combinatorial nature. While some studies argue that statistical language learning correlates with natural language comprehension (Misyak & Christiansen, 2012), whether this ability correlates specifically with explicit language learning is still a matter of debate (Ram Frost et al., 2015). Critically, an exploratory analysis revealed that participants' scores in Experiment 2 and their Recombinant discrimination ability in Experiment 5 were positively correlated. It seems that participants who were better at statistical rule learning also discriminated between the target stems and suffixes and the recombinant items better. These results open exciting questions about the nature and relation of implicit and explicit memory traces during learning, but these are beyond the scope of this work.

Another interpretation of Experiments 5 and 6 is as paired-associates tasks. In Experiment 5, each suffix was paired with eight different words. Therefore, participants learned many-to-one word-to-suffix pairings. In this regard, the lack of differences between monolinguals and bilinguals corroborates the results from the multiple mapping conditions in Experiment 4. Conversely, in Experiment 6, each Flavian word was uniquely paired with a different object. The pooled results from experiments 4 and 6 indicate that bilingual participants are better at implicit and explicit learning of novel vocabulary at the lexical-semantics level, provided there are one-to-one mappings during the learning session. These ideas lead to further questions regarding the role of non-exclusive mappings during the initial (and later) stages of implicit and explicit foreign vocabulary learning. However, they remain beyond the scope of this thesis.

Experiment 6 also revealed that the effects of bilingual experience did not interact with sub-lexical orthographic similarity—except during the recognition learning tasks. The bilingual participants were generally better at learning the vocabulary, regardless of how similar or dissimilar the Flavian words were to Spanish. Furthermore, neither participants' working memory capacity (as measured by the Operation Span task) nor non-verbal intelligence seemed to explain individual learning trajectories. Thus, these findings are consistent with studies suggesting a bilingual advantage even when participants are matched on certain variables such as phonological awareness or working memory (Kan & Sadagopan, 2014; Kaushanskaya, 2012; Nair et al., 2016; Van Hell & Mahn, 1997). Some authors have suggested that the differences between monolinguals and bilinguals are more significant for concrete than abstract referents (Kaushanskaya & Rechtzigel, 2012). Since both experiments 4 and 6 used visual depictions of concrete objects, whether concreteness further interacts with similarity and the bilingual status during learning remains to be seen. In addition, using translation pairs rather than visual depictions of objects—albeit more similar to the adult classroom learning environment—could prime participants to operate in their L1, which is a significant confound in experiments using translation pairs.

Chapter 3 Conclusion

The experiments I presented in Chapter 3 targeted different aspects of vocabulary learning in monolinguals and bilinguals. The results are consistent with Chapter 2, suggesting that bilinguals outperform their monolingual peers when learning vocabulary, but not in other aspects of foreign language learning. The experiments presented so far represent a handful of the manipulations and tasks to target the different analytic levels during initial exposure to a foreign language. Moreover, since these experiments tested participants only immediately and one night after learning the materials, the results cannot be generalized to long-term foreign language learning. Future studies could explore how the differences observed in the initial stages extend through the course of becoming highly proficient in a foreign language.

I can claim that, at least during the initial exposure to novel linguistic material, the bilingual participants in these experiments are consistently (and possibly only) better than monolinguals at learning vocabulary. Thus, Chapters 2 and 3 suggest that bilinguals outperform monolinguals in learning at the lexical-semantics level (Axis A). Furthermore, these differences are primarily observed on the one-to-one mappings between word forms and their referents, and regardless of the novel words' similarity to a known language (Axis B). Nevertheless, how could experience with two languages facilitate integrating novel word forms into the lexicon over experience with a single language? This is the central question of Chapter 4 (Axis C).

Chapter 4: Tracing the algorithm of bilingual vocabulary learning

Theoretical motivation

So far, I have examined and presented evidence favoring bilingual experience influencing foreign language vocabulary learning but not other analytic levels of foreign language learning. While the literature covered in Chapter 1 and the findings from Chapters 2 and 3 could generally fall into the direct and indirect effects framework (Hirosh & Degani, 2018), it is essential to establish the theoretical work explaining how these differences could emerge specifically at the vocabulary level. On the one hand, the theories of cross-linguistic (direct) transfer suggest that bilinguals are better than monolinguals insofar as they can exploit similarities from their known languages (e.g., Cabrelli Amaro et al., 2012; Rothman, 2015; Westergaard et al., 2017). However, I have shown that bilinguals might outperform monolinguals regardless of word form similarity.

On the other hand, there is scarce evidence in favor of bilingual experience affecting other non-linguistic cognitive abilities that could indirectly favor foreign language learning (e.g., Blanco-Elorrieta & Pylkkänen, 2018; de Bruin et al., 2021; Dick et al., 2019). Indeed, these results might be due to indirect cognitive linguistic abilities (e.g., metalinguistic awareness, enhanced phono/ortho networks) as suggested by the direct and indirect effects framework (Hirosh & Degani, 2018). However, these abilities are either challenging to assess in artificial languages or near impossible to measure using behavioral tasks. Consequently, the theoretical framework presented thus far seems insufficient to explain the observed differences at the lexical-semantics level.

In this chapter, I attempt to unify the findings from the previous sections revising some theoretical and computational models of word processing and learning in bilinguals. I start by briefly introducing some of the most prominent theoretical and computational models of the bilingual mental lexicon. These models help understand the nature and organization of word form and meaning representations in multiple languages. Nevertheless, I will argue that they do not provide a broad enough account of vocabulary learning to explain the differences between monolinguals and bilinguals. I then present a simulation study, proposing an emergentist account of foreign vocabulary learning in a second and a third language. This simulation study provides an initial approach towards understanding the role of orthographic regularities in monolingual and bilingual vocabulary learning. It is, naturally, an incomplete account of the entire mental lexicon, as it focuses primarily on sub-lexical orthography. I acknowledge that the literature, and this work, are still far from a holistic account of human language learning—or learning in general—and its relation to cognitive and neural processes.

4.1. Cognitive models of word processing and learning

The field of cognitive science has focused on the mental processes underlying learning. A general view of the learning process comes from information-processing theory (Schunk, 2012). In simple words, learners first integrate the information in short-term memory, which is then passed to long-term memory using consolidation processes (Walker & Stickgold, 2004). Thus, the knowledge is codified, stored, and retrieved from these systems during learning, production, comprehension. The codification process involves creating a mental (or neural) representation that can be efficiently stored and retrieved. In this regard, the central questions in the cognitive science literature concern the nature of these representations and the mental processes underlying the codification, storage, and retrieval stages. The literature devoted to language learning and processing has employed computational models to address some of these questions (Ellis, 1998; Seidenberg, 2005). These models force researchers to think about the adequate manner to represent both the information and the rules through which individuals learn, recognize, and produce language (Carreiras et al., 2014).

The majority of the language-related cognitive science literature has focused on the mental lexicon, which, as I mentioned before, is the brain's storage for word forms (orthography and phonology), their meanings, and other aspects (Acha & Carreiras, 2014; Aitchison, 2012). These mental lexicon's theoretical and computational models concur that word form representations are distinct from meaning representations (Baxter et al., 2021). However, some researchers suggest that these representations are *localist*, meaning that both forms and meanings have unique identifiers in the brain (Holman & Spivey, 2016). This would be akin to assigning a unique mental identification number (e.g., a number plate) to each possible word form and meaning. Conversely, other authors assume that the representations are *distributed,* and thus, word forms and meanings are compositions of many features (Holman & Spivey, 2016). In this case, one could think of a word's form or meaning as having different dimensions (e.g., color, shape, concreteness), but these dimensions are usually not directly interpretable.

Specifically, the second language processing literature concurs that meaning representations are common to the native and second languages, but there are two contrasting views regarding word form representations (de Groot, 1992; French & Jacquet, 2004). Proponents of the *language-selective* view argue that form representations are unique for each language or that additional mechanisms (e.g., bottom-up or top-down language tags) modulate which set of representations is more active for a specific word (Jacquet & French, 2002; Kroll et al., 2010). Conversely, the *language-nonselective* view posits that word form representations are also common to the two languages (De Groot et al., 2000). As I mentioned in Chapter 3, this debate seems to have settled in recent years, with the language-nonselective view becoming the predominant view of the multilingual mental lexicon (Baxter et al., 2021). Therefore, most models assume that word form and meaning representations from all known languages are unified in the mental lexicon.

For the sake of clarity, I will categorize these computational and theoretical models into two types: (1) those that address bilingual word processing using localist representations and (2) those that address word learning (and sometimes processing) in one or more languages using distributed representations. I briefly cover the main properties of each model with regards to the purpose of this thesis, avoiding an exhaustive review (for a more extensive overview, see Holman & Spivey, 2016). Both types of models provide valuable information as to how the differences between monolinguals and bilinguals might originate. Notably, while these models differ in their purpose and implementation, most of the models originate from what is known as the *connectionist framework* (Seidenberg, 2005). This framework employs artificially simulated connections between nodes in a network, a process loosely analogous to the functioning of neural networks in the brain.

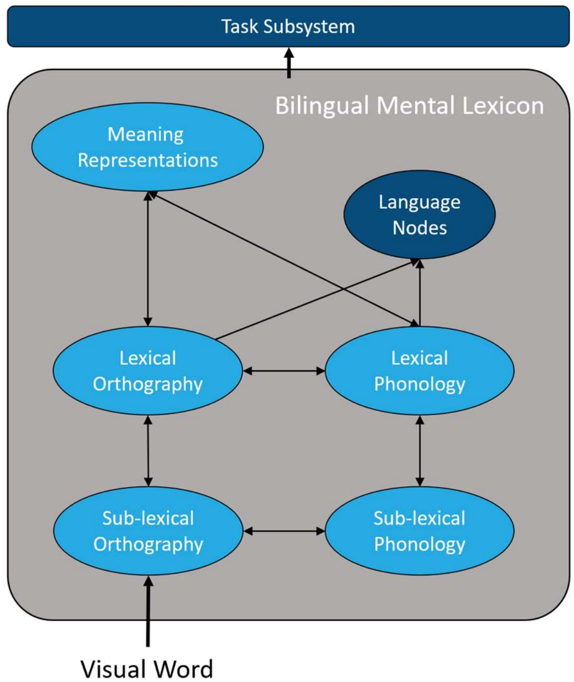
4.1.1. Localist models of bilingual word recognition

The dominant models of bilingual word recognition stem from the Interactive Activation (IA) connectionist account (McClelland & Rumelhart, 1981). This model contains different representations of visual and acoustic features, letter and sound units, and lexical levels, with bottom-up (features to words), top-down (word to features), and lateral (within the same level) activation and inhibition during word recognition. It was developed to account for context effects in visual letter recognition. For example, individuals are faster at recognizing whether a letter is present in a string if it is a known word (i.e., the R in MARKET) rather than a non-word (i.e., the R in MKRAET). Therefore, the IA model was first to computationally demonstrate the influence of top-down effects—i.e., knowledge of a word influencing low-level letter recognition—by capturing these findings. This model opened the field for the exploration of language representation in the human mind, leading to several bilingual models based on these principles:

Bilingual Interactive Activation (BIA) model. The BIA model is a direct extension of the IA model to the bilingual context. In essence, the BIA assumes a language-nonselective view of the bilingual mental lexicon at the initial stages of word recognition, with both word forms and meanings shared between the two languages (Grainger & Dijkstra, 1992). The model also uses a language node that modulates which representations are active in this shared lexicon via top-down connections, emulating an L1 or L2 "language context" (Van Heuven et al., 1998). This mechanism implements a sort of inhibitory control over which set of representations are active at any given time. While this model has been highly regarded for its ability to account for multiple phenomena in bilingual word recognition, the BIA model does not provide an account for how words are integrated into the mental lexicon. Regardless, this model was one of the first to operate using a shared and unified bilingual mental lexicon.

Bilingual Interactive Activation Plus (BIA+) model. The BIA+ model is an extension and improvement of the BIA model (Dijkstra & van Heuven, 2002). The architecture for the BIA+ model is depicted in Figure 12. According to this model, a visual word will non-selectively trigger the sub-lexical orthographic and phonological representations, subsequently activating the lexical representations. Contrary to its predecessor, the language nodes activate at a later stage and in a bottom-up manner. That is, the language nodes determine a word's language after receiving the lexical representations. Another addition is a task subsystem, which represents the cognitive processes involved in the task at hand. For example, if the task is determining whether a word exists or not (i.e., lexical decision), the task subsystem handles the cognitive, decision, and motor processes necessary to perform this task (Dijkstra & van Heuven, 2002). This model has been criticized because of its lack of learning mechanisms, the use of localist instead of distributed representations, and the language-dependent word representations at the lexical level, which violate the language-nonselective presupposition (Jacquet & French, 2002). In particular, these properties do not allow this model to explain monolingual and bilingual learning of orthographically similar and dissimilar vocabulary. Regardless, this is arguably the most influential model in the bilingual word processing literature (Holman & Spivey, 2016).

Figure 12. *The Bilingual Interactive Activation Plus (BIA+) model.*



Note. Adapted from Dijkstra & van Heuven (2002). Arrows indicate the direction of information flow.

Bilingual Interactive Activation developmental (BIA-d) model. This model represents yet another extension to the BIA+ model, but it is more theoretical than an implemented model (Grainger et al., 2010). This extension draws from a theoretical account of bilingual word learning and processing, known as the *Revised Hierarchical* (RHM) model (Kroll et al., 2010). Briefly, both of these models propose that, during second language learning, individuals initially establish a link between form representations in the L2 their corresponding form in the L1, without a direct link to meaning representations. With increased exposure to the L2, the connection between L2 words and meaning strengthens, while the original link between L1 and L2 word forms weakens. The difference between the BIA-d and the RHM models is that the RHM assumes that the representations are unique for each language (i.e., language-selective). Nevertheless, both models fail to explain learning effects such as orthographic similarity because the representations in these models are localists (Marecka et al., 2021; Otwinowska et al., 2020). Furthermore, it is hard to explain how monolingual and bilingual learners could differentially learn vocabulary in a foreign language using these models.

Multilink. A more recent development in the bilingual word processing literature is Multilink (Dijkstra et al., 2019b). Multilink is a computational model that can perform word recognition, spoken word production, and translation. It maintains the core ideas of the BIA+ model, with language-nonselective bottom-up activation and language nodes. Researchers can adjust some parameters in the model to address balanced and unbalanced bilingualism and even monolingualism. Additionally, it can be used to explain phenomena from more tasks than word recognition. Nevertheless, learning in these models remains largely unexplored.

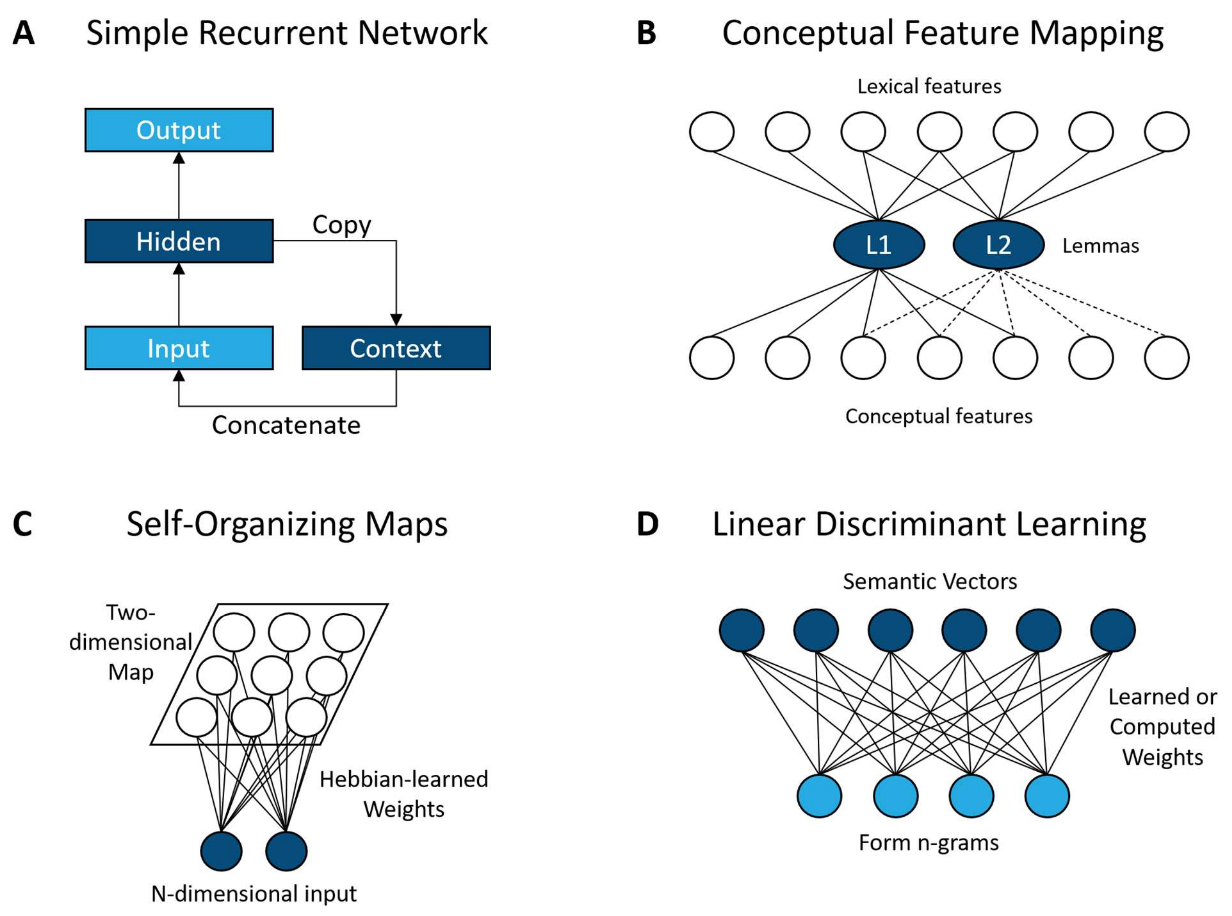
4.1.2. Distributed models of bilingual word learning

A premise of the emergentist view of language is that learning depends fundamentally on input co-occurrence—akin to statistical learning—and that the input determines both the units and the rules of a language (Christiansen & Chater, 2001; Ellis, 1998). In this regard, using basic learning rules and network architectures, the distributed models of word learning can explain how the organizational principles of the mental lexicon *emerge* from the input. There is an extensive tradition of using these distributed connectionist models in the cognitive science literature (e.g., Hawkins et al., 2019; Thomas & McClelland, 2012; Touretzky & Hinton, 1988). However, here I focus on those models that could potentially account for monolingual and bilingual word learning:

Bilingual Simple Recurrent Network model. The Bilingual Simple Recurrent Network was one of the first models to address bilingual language representation using a distributed approach (French, 1998). The architecture was a recurrent connectionist network that receives an input and a copy of the previous state (a vector of zeros at the beginning) and produced a prediction of the following input (see Figure 13A)—also known as Simple Recurrent (Elman) Network (Elman, 1991). This model was trained using simple three-word sentences from two languages with no indication of the boundary between languages (e.g., boy lift toy femme touche ballon). Notably, this was the first to show that lexical representations clustered according to each language without the need for explicit language tags, using a completely language-nonselective distributed lexicon. However, since the representations processed by the model were words, it did not address the learning of vocabulary but rather of sentence structure in two languages.

Conceptual Feature Mapping model. The Conceptual Feature Mapping model (Figure 13B) is a theoretical distributed connectionist model of word learning (Lee et al., 1999). This theoretical model provides a verbal explanation for the facilitatory effects of word form similarity on foreign vocabulary learning. According to this model, the bilingual mental lexicon is unified and distributed at both the word form (conceptual) and meaning (lexical) levels. An intermediate localist "lemma" layer maps a word form to its corresponding meaning in a specific language. Thus, learning a foreign word requires modifying the connections from the form level to the lemma level or from the lemma level to the meaning level, depending on whether the target word's form or meaning overlaps with its native language counterpart. This model further proposes that some features might not be shared between the languages but does not propose a mechanism through which these could be learned. Nevertheless, due to a lack of implementation, this model cannot be empirically contrasted to explain the differences between monolinguals and bilinguals.

Figure 13. *The Bilingual Interactive Activation Plus (BIA+) model.*



Note. (A) Adapted from Elman (1991). (B) Adapted from Lee et al. (1999). L1 = first language; L2 = second language. (C) Adapted from Kohonen (1998). (D) Adapted from Baayen et al. (2019).

Self-organizing maps. These are a collection of models rather than a single unified account of bilingual word learning. They were developed under the framework of self-organizing maps (Kohonen, 1998; T., 1982). In essence, self-organizing maps (SOM) are artificial neural networks that map a set of multidimensional inputs onto a two-dimensional topological map (Figure 13C) using weights learned through simple rules (e.g., Hebbian learning). The Self-Organizing Map of Bilingual Processing (SOMBIP) was one of the first models to show the differences in the organization of phonological word forms emerging from early versus late learning of the second language (Li & Farkas, 2002). More recent developments of this model, called Dev-LEX (Li et al., 2007; X. Zhao & Li, 2010, 2013), have further argued that distinct language-dependent form representations emerge from these models without any label. Also, these representations are structured depending on the onset of L2 learning (X. Zhao & Li, 2007). Simulations using these models require substantial manual labor to code the input correctly. Thus the models learn from highly distilled and typically very scarce input representations rather than natural language (Li & Grant, 2019). In all, while these models help visualize the structure of representations in the (multilingual) mental lexicon, they require substantial work to compare their output to actual human performance.

Linear Discriminant Learning models. The final group of mental lexicon models belongs to a recent development in word and paradigm morphology (Harald Baayen et al., 2019). These models establish a fully connected two-layer structure between word form and meaning (for word recognition; Figure 13D) or meaning and form representations (for word production). Notably, these are implemented as separate networks. Moreover, these models require a set of pre-determined form representations (typically 3-grams or 3-phones) and a corresponding set of meaning representations obtained from distributional semantic models such as word2vec (Church, 2017). The mappings of form to meaning (or vice-versa) can either be learned using simple learning rules (e.g., Rescorla-Wagner) or estimated using Linear Discriminant Learning, a form of multiple linear regression (Harald Baayen et al., 2019). Hence, these models can simulate both the learning trajectory and end-state of learning one or multiple languages. A recent study examined the learning trajectory of bilinguals and trilinguals using a small vocabulary (405 words), showing how different factors such as order of acquisition and number of inter-lingual homophones can give rise to errors in comprehension and production (Chuang et al., 2021). However, these models are highly susceptible to out-of-vocabulary errors and require extensive engineering to make the results comparable with human behavior. Regardless, future developments could prove the usefulness of this model for language acquisition, processing, and production.

4.1.3. Summary

Localist models suggest essential aspects regarding the organization of languages in the bilingual mind. First, languages are stored together and co-activated according to the language-nonselective view. Second, there is a distinction between the language subsystem and other cognitive (e.g., task, motor, decision) subsystems. Finally, these models reiterate the distinction between sub-lexical, lexical, and semantic analytic levels of a language from a computational perspective. Furthermore, these models are highly interpretable and can be directly compared to human behavior. However, they require extensive manual engineering of connections and activations, which complicates explaining their learning mechanisms (Holman & Spivey, 2016). Additionally, localist representations force researchers to find alternative ways to implement word form organization due to their similarity, a theorized organizing principle of the mental lexicon (Baxter et al., 2021).

On the other hand, although the distributed approaches seem promising, several issues remain to be addressed. First, there is no account for how meaning is acquired at present, and the models that do implement semantic representations employ pre-trained semantic vectors or localist labels (Chuang et al., 2021; French & Jacquet, 2004; Laszlo & Plaut, 2012). The meaning representations pose an additional challenge in more than one language—for instance, is the meaning of two words precisely the same across languages? Third, while distributed models can learn representations using simple learning mechanisms, these models have yet to be compared to human behavior directly (Holman & Spivey, 2016). Fourth, all models mentioned above only present small-scale simulations using carefully selected subsets of words (Li & Grant, 2019), hindering their generalizability. Finally, no model provides a general account for how bilinguals and monolinguals might differ in their vocabulary learning performance or even how words might be differentially learned due to their sub-lexical orthographic similarity.

To address the last point, I developed a recurrent neural network model of the orthographic lexicon. The model implements a language-nonselective view employed by many other models and learns distributed word form representations from the input in a self-supervised manner. Critically, this model addresses the effects of orthographic similarity and bilingual experience reported in Chapter 3. Although an argument could be made against the simplicity of my approach—namely, this is just a model of the orthographic lexicon without meaning—, there are benefits to starting with a small architecture. I validate this model using an adapted version of the lexical decision task, which has been the primary paradigm to investigate word recognition and processing (e.g., Grainger & Jacobs, 1996; Norris, 2013; Ziegler et al., 2000). The model processes individual words as input in a character-by-character manner, allowing me to reference the sub-lexical orthographic patterns of the input rather than the architecture itself. Third, I dissect the model's internal representations to provide information about the organization of the multilingual orthographic lexicon in the absence of meaning. Finally, my model is ideally suited to explore the effects of sub-lexical orthographic similarity and bilingual experience on foreign vocabulary learning.

4.2. Experiment 7: Simulating bilingual and monolingual vocabulary learning[3]

4.2.1. Rationale

Intuitively, a word that looks or sounds more similar to the native language should be easier to learn, irrespective of its meaning. The knowledge of two languages—bilingualism—might provide an advantage in terms of sources from where to draw similarities. Recent research corroborates these ideas, highlighting word form (written or spoken) similarity (Hayakawa et al., 2020; Marecka et al., 2021; Otwinowska et al., 2020) and experience with two (or more) languages (Festman, 2021; Hirosh & Degani, 2018; Montanari, 2019) as factors that facilitate foreign vocabulary learning. Experiment 6 of this thesis further indicated that bilinguals are better at learning similar and dissimilar written words than monolinguals.

Due to the orthographic nature of Experiment 6, these findings provide a testable hypothesis regarding the role of bilingual input on the organization of word forms in the mental lexicon. Because bilinguals experience potentially distinct orthographic patterns (orthotactics) in two languages during learning and recognizing different words, the bilingual orthographic lexicon might more flexibly integrate similar and dissimilar novel word forms than a lexicon that operates in only one language. Moreover, if a word is more similar in its orthography to other instances in the lexicon, it should be acquired faster than dissimilar words. Still, despite numerous efforts to formalize bilingual memory and language processing with computational approaches (e.g., Dijkstra et al., 2019a; French & Jacquet, 2004; Li & Grant, 2019), the basis for these seemingly disparate learning effects remains largely unexplored.

Here, I propose a computational model to account for the behavioral evidence in Experiment 6. The model addresses written word learning in a second and third language. It is instantiated as a recurrent neural network that can learn written vocabulary by implementing a unified, distributed, and dynamic view of the orthographic lexicon. In doing so, I address two theoretical questions. (a) Can a unified orthographic lexicon learn vocabulary with varying degrees of sub-lexical orthographic similarity? (b) Does exposure to the orthography of two versus one language benefit subsequent vocabulary learning?

Recurrent neural network models can approximate the probabilities of natural languages (French & Jacquet, 2004; Mikolov et al., 2010). They have been used extensively in natural language processing literature, where they are known as language models (e.g., Mikolov et al., 2010; Sundermeyer et al., 2013;

[3] This work was carried out in collaboration with Prof. Marco Zorzi and Prof. Alberto Testolin at the Computational Cognitive Neuroscience Lab, University of Padova, Italy. The work was supported by an EMBO Short-Term Fellowship grant (Fellowship code 8742). I also thank Prof. James Magnuson for his valuable insight regarding this work.

Williams et al., 2015). Since recurrent models keep track of current and past information, they can form representations of the co-occurrence patterns in the input (Testolin et al., 2016). Taking sentences as an example, a word-level model contains representations of word co-occurrence. For instance, a word-level model could output a higher probability of “house” following “my” than “die” following “my”. Similarly, a character-level model contains information about certain letters following others within words in a language.

To my knowledge, no study has compared the differences between monolingual and bilingual character-level models in regards to the acquisition and organization of words in the mental lexicon. Crucially, several criteria need to be met for these models to be helpful. First, the monolingual and bilingual models should have an architecture simple enough so that the relation between input and output is at least partially interpretable. Second, the monolingual and bilingual models should only differ in their input but not in their architecture. Third, there should be an appropriate metric for assessing the performance of these models. Fourth, in comparing monolingual and bilingual models, their performance on this metric should be matched. Finally, it is essential to adapt the task used in Experiment 6 for the results to be comparable. Below, I explain how I addressed these desiderata.

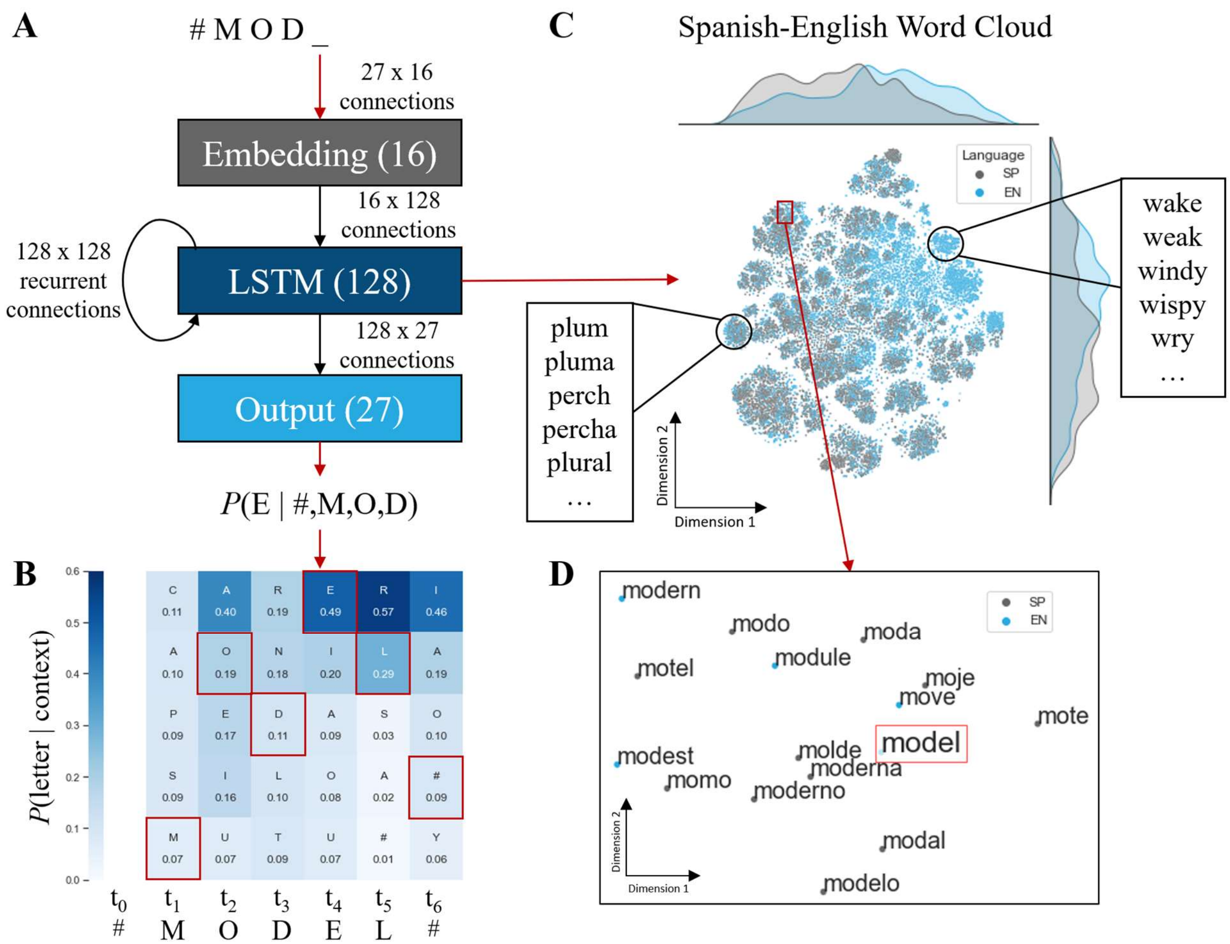
4.2.2. The CLOUD model

I developed a character-level model—termed CLOUD for *Constrained Learner of Orthography: Unified, Distributed, and Dynamic*—with three constraints. First, following the language-nonselective view predominant in the literature (Baxter et al., 2021), the model stores the word form representations of one or two languages in a *unified* structure. This constraint forces the model to learn to recognize the language category on its own—assuming the two languages have highly different orthotactics. Second, to test for the effects of orthographic similarity during foreign language learning, the model contains *distributed* representations of letters and word forms implemented as numeric vectors. Studies have previously used distributed word-level representations to investigate the organization of words in the bilingual lexicon, showing separation between the languages without any language-specific nodes (French, 1998). However, these have not been explored at the character level. Third, the model adjusts the distributed representations *dynamically* during vocabulary learning. This constraint implies that learning new words changes the representations of existing words in the lexicon. The dynamic constraint is also backed by recent research suggesting that mere exposure to words in a different language changes orthographic processing and word learning in the native language, hinting towards a dynamic lexicon (Bice & Kroll, 2015, 2019).

I structured the CLOUD model as a three-layer recurrent neural network that learned its representations by predicting the following character in multiple words—a mechanism loosely analogous to the predictive coding framework of the nervous system (Huang & Rao, 2011; Rao & Ballard, 1999). Figure 1 exemplifies the inner working of the CLOUD model using the word "model" as an example. In essence, at each time step, the model processes a letter as input and tries to predict the succeeding character (Figure 14A). The first layer, called the Embedding layer, transforms each letter into a distributed 16-dimensional numeric vector. While admittedly an over-reductionist approach, I chose to add this layer to exemplify how the printed letter becomes an internalized representation (Finkbeiner & Coltheart, 2009; Rapp & Caramazza, 1997). After this, 128 Long Short-Term Memory (LSTM) hidden nodes store the distributed orthographic word form representations (Hochreiter & Schmidhuber, 1997). These recurrent nodes allow the model to develop sensitivity to long- and short-term orthographic dependencies within words. The criteria for selecting the number of dimensions in each layer are detailed in the Methods section and Appendix C1. Using the self-supervised procedure of predicting the next character in a word, the model can learn the orthotactic patterns underlying words in multiple languages. Lastly, the Output layer encodes the successor character's probability distribution given a sequence of preceding characters (Figure 14B). The model then received an error backpropagation depending on whether the predicted character matches the expected character in a word (Rumelhart et al., 1995; Werbos, 1990). Hence, the CLOUD architecture can capture the orthographic dependencies in different words using fully distributed representations.

The second reason I named the model CLOUD is its capacity to generate *word clouds* using the representations learned by the LSTM nodes (Figure 14C). To obtain the representation of a word in the CLOUD, I first average the LSTM outputs at each time step. Then, I project each word's 128-dimensional vector onto a 2-dimensional space for visualization using the t-distributed stochastic neighbor embedding (t-SNE) (Van Der Maaten & Hinton, 2008). These word clouds help understand how words cluster according to their orthotactic patterns in different languages. One can also compare the representation of any word (or nonword) to other words in the dataset to identify similar words, further boosting CLOUD architecture's interpretability. For instance, as expected, the word "model" should be closer to orthographically similar words in the dataset, including its Spanish translation equivalent, "modelo" (Figure 14D). Since the model learns from individual letters and not whole words, its distributed representations (and outputs) are not limited to specific words or languages, making the CLOUD architecture ideal for probing questions about monolingual and bilingual written vocabulary learning.

Figure 14. *The CLOUD model architecture.*



Note. **(A)** Depiction of the CLOUD model architecture. The model processes each character sequentially, using the current character and an internal (context) state as inputs, and predicting the next character in the sequence in a recurrent and self-supervised manner. The model selects the next character from a pool of 27 characters, including 26 letters of the alphabet plus a space token (here marked as #). After training, the model approximates the next character's probability distribution given the previous sequence of characters. All connections are learnable parameters in the CLOUD architecture. **(B)** An example plot of the top five successor character's probability distribution at every time-step of processing the word "model" as input. At time-step 4 (t_4), after processing the sequence "#MOD", the model predicts a 49% probability that the next character in the sequence is "E". **(C)** A word cloud created using the LSTM hidden representations of the entire Spanish-English word dataset. The marginal distributions suggest a slight separation between the two languages. Hand-picked examples on the right suggest that some English words cluster based on patterns that do not exist in Spanish, such as words beginning with "w", ending with "y", or containing "k". The examples on the left part are a cluster of Spanish and English words that begin with "p", including perfect cognates (e.g., plural) and orthographic neighbors (e.g., plum, pluma). **(D)** The top 15 most similar English and Spanish words to the word "model". Pairwise similarity scores were calculated using cosine similarity, the dot product of the L2-normalized vector representations. The dimensions are jittered to minimize word overlap. The top words include those with similar orthography, beginning with "mo" (e.g., motel, modal, molde).

In developing this model, I aimed to answer three specific questions. (1) Can a unified model learn the orthotactic patterns of two languages without language-specific nodes? (2) Do the internal representations of the model contain information about the language of a word? Moreover, (3) Could this model emulate the effects of orthographic similarity and bilingual experience observed in Experiment 6?

4.2.3. Methods

Model architecture

As depicted in Figure 14A, the CLOUD architecture contains three fully connected layers. At each time step, the model receives one character as a categorical index from 0 to 26 for a total of 27 characters. The characters include 26 letters in the alphabet, plus a space character that indicated the beginning or end of a word (here depicted as #). After receiving the input, the Embedding layer converts each character index to a 16-dimensional distributed vector using a lookup table. I chose this layer's dimensions as the power of 2 closest but smaller than the alphabet size to obtain a distributed representation of each character. These distributed character representations can be individually interpreted. However, the organization of individual characters in monolingual and bilingual input is beyond the scope of this experiment, which concerns combinations of letters (i.e., sub-lexical orthotactics) rather than individual letters.

The second layer contains 128 recurrent LSTM nodes. I selected the final number of nodes for the LSTM layer after exploring the model's performance with varying nodes (see Appendix C1). The LSTM nodes capture long- and short-term dependencies in the input using two 128-dimensional distributed representations. These representations are respectively called cell state and hidden state. The cell and hidden states are combined with the Embedding layer's output through a series of gating mechanisms at each time step. First, the *forget gate* combines the information from the current input and the previous hidden state to decide which information to remove from the cell state. Parallelly, the *input gate* decides which information to add to the cell state using the combined input and hidden state. The LSTM layer uses these two gates to produce the next cell state. Then, it combines this new cell state with the *output gate* to calculate what should be propagated to the next hidden state. This process can be summarized using the forward pass equations shown in Appendix C2.

The Output layer implements a multinomial logistic regression. This layer uses the hidden state produced by the LSTM layer to output the logits for each possible character, implemented with the equation

$$logits_t = W_{ol} \cdot h_t + b_{ol} \quad ; \quad W_{ol} \in \mathbb{R}^{27 \, x \, 128}; \; b_{ol} \in \mathbb{R}^{27} \tag{1}$$

where $logits_t$ is a 27-dimensional vector of logits for each possible character output, W_{ol} and b_{ol} are the output layer's weight and bias parameters and h_t is the 128-dimensional hidden output from the LSTM layer. The weight parameters are initialized following standard practice (Glorot & Bengio, 2010), sampling random from a uniform distribution ranging from $-1/27^{0.5}$ to $1/27^{0.5}$, where 27 is the number of dimensions.

The model learns by predicting each word's next character and receiving a prediction error signal that updates the Embedding, LSTM, and Output layers' weight and bias terms. This mechanism is called backpropagation (Rumelhart et al., 1995; Werbos, 1990). I computed the prediction error using the cross-entropy loss function

$$loss(logits, character_{index}) = -\mathrm{logits}_{character_{index}} + \log\left(\sum_{i=0}^{27} \exp(logits_i)\right) \tag{2}$$

that takes the negative of the expected character's logit and adds the logarithm of the sum of all logits' exponent in the output vector.

During learning, the backpropagation algorithm adjusted every learnable parameter in the model. This algorithm requires calculating the partial derivative of Eq. 2 with respect to each parameter to determine which direction to nudge them. I implemented the CLOUD architecture using PyTorch (Paszke et al., 2017), a Python-based framework that provides automatic differentiation capabilities and already contains optimized Embedding, LSTM, and Output layers described above. The final model had 78,812 learnable parameters [4], and the objective was to minimize the prediction error (loss function) by adjusting these parameters.

Materials

Pre-training datasets. I started with 44,853 Spanish words from the SPALEX dataset (Aguasvivas et al., 2018), 61,851 English words from the English Crowdsourcing Project (L2 version) (Mandera et al., 2020), and 74,490 Basque words from the EHME dataset (Acha et al., 2014). I first reduced the number of words in each dataset by discarding words with a frequency lower than the median value and limiting the length between 3 and 10 characters. These filters reduced the Spanish words to 19,592, the English words to 28,174, and the Basque words to 37,978. I replaced every letter with a tilde (e.g., é, ñ) with their

[4] There was an additional padding character in the vocabulary used internally for processing multiple words in batches. Therefore, in total there were 464 parameters in the Embedding layer, 74,752 parameters in the LSTM layer, and 3,612 parameters in the Output layer.

corresponding roman alphabet counterpart (e.g., e, n) and removed any duplicate words in each dataset. Recent evidence suggests that processing does not differ between the tilde and non-tilde versions of these letters in Spanish speakers (Perea et al., 2020). Moreover, this step reduced the final number of characters to 26 letters common to all languages under study.

The next step involved matching the word length distributions in the datasets. I sampled the minimum number of words from each dataset at every possible word length to accomplish this. The final datasets contained 19,516 words each. I randomly selected 1,640 words for testing from each dataset, leaving 17,876 words for training. The test words were used to calculate the models' performance on unseen words. I constructed a Spanish-English dataset by combining the Spanish and English training and test words and a Spanish-Basque dataset by combining the Spanish and Basque words. Importantly, even though I randomly sampled the words to assign them to the train and test sets, the length was not significantly different between the languages in either of them (all $p > 0.05$).

The monolingual CLOUD version performed the character prediction task mentioned above over the 17,876 Spanish words. The Spanish-English bilingual version (used to construct Figure 14) learned from a subset of 60% of the Spanish words (10,725) and 40% (7,150) words selected at random from the Spanish-English dataset. Finally, the Spanish-Basque bilingual version used the same proportion (60/40) from the Spanish-Basque dataset. I selected these percentages to reflect the reality of bilingual exposure within Spain and approximate participants' total exposure in Experiment 6. Notably, the vocabulary size used in the present modeling work represents an almost ten-fold input size increase from previously reported models, containing around 1,700 words at most (Li & Grant, 2019). My choice of vocabulary aligns with a real-world scenario considering that the average young adult controls more than 25,000 words (Aguasvivas et al., 2020; Brysbaert et al., 2016b).

Adapted Spanish Lexical Decision Task. It is essential to have a measure that adequately accounts for the model's performance. Traditionally, the measure of choice has been the *top k* accuracy in predicting the successor character—i.e., is the correct letter in the model's *top k* predictions?. However, language models typically do not score high on this measure, as a precise prediction of the successor character is unlikely. Suppose a character-level model only learned the words "care" and "cars" with equal proportion. In that case, the probability of predicting the letter "e" after seeing "car" is approximately equal to that of predicting "s". Thus, the maximum theoretical top 1 accuracy for this simplified model would be 50%, while the top 2 accuracy would be close to 100% as there are only two possible choices. Extending this idea to scaled-up scenarios, it is rare to see character-level models perform above 35-40% using top k accuracy metrics. Indeed, performance above the 50% level could indicate severe overfitting to the training data, leading to poor performance on unobserved words.

Prior research has suggested that using a 2AFC adaptation of the Lexical Decision Task (LDT) provides an adequate alternative to measure and compare character-level models' performance (Le Godais et al., 2017). In this adapted LDT, the character-level model calculates the probability of two words matched in length (one target and one foil) by multiplying its output probabilities at each time step (see Figure 14B for a visual depiction with one word). Ideally, the probability for existing words should be higher than that of non-words. This adapted LDT addresses the issue of multiplicative probabilities, whereby multiplying probabilities will yield smaller values for longer words. Setting a strict threshold for determining whether the seen item is a word or not requires non-trivial normalization of the probabilities (Lau et al., 2017). Notably, this adaptation has been previously used in the psycholinguistic literature with human participants (Baddeley et al., 1993), and thus represents a reliable metric of human and model performance. Therefore, I constructed an adapted Spanish LDT dataset for measuring and matching the CLOUD models' performance.

I selected a set of non-words from the SPALEX dataset (Aguasvivas et al., 2018). These non-words were already rated by human participants, yielding information about how well they could categorize them as non-words. This information was essential to selecting and matching the words and non-words. Initially, there were 56,861 non-words in the dataset. I selected those non-words rated by more than 30 participants and with an average percent correct in the range from 72 to 95%. In other words, these non-words were neither extremely easy nor extremely challenging to discriminate. I further restricted the non-words length between 3 and 10 characters, replaced any tilde character, removed duplicates, and matched the length distributions to the Spanish words. The final list contained 19,516 non-words, paired by length to the existing Spanish dataset and separated into training and testing datasets. Notably, I did not enforce any control beyond these measures (e.g., using a similar orthographic structure). An initial analysis with 20 untrained CLOUD models indicated an average performance did not differ significantly from chance level (M_{acc} = 49.7%, SD = 1.0, $t_{(19)}$ = 1.342, p = 0.196), confirming that the dataset was not inherently biased towards words or non-words.

Flavian vocabulary. The Flavian vocabulary was the same as in Experiment 6. For more information about the construction of this vocabulary, see the Methods section in Experiment 6. Appendix B3 contains the vocabulary and additional metrics.

Procedure

Model pre-training. Since the CLOUD architecture was the same for monolingual and bilingual models, it was essential to pre-train the models to approximate the orthographic lexicon of an adult individual. I call this step *pre-training*. It entailed exposing different versions of the CLOUD model to words from Spanish, English, or Basque. In other words, while the architecture was the same, the models only differed in their exposure to the orthography of different words. I trained three different CLOUD model versions: a Spanish monolingual, a Spanish-English bilingual, and a Spanish-Basque bilingual. The monolingual (MONO) version of the CLOUD model learned from 17,876 (100%) Spanish words. In contrast, the bilingual (SP-EN and SP-BQ) versions were trained on 60% (10,726) Spanish and 40% (7,150) L2 words, randomly selected from their respective pre-training datasets. There were ten runs with different random initializations within each version to add variability to the model's output and statistically compare their results. Notably, the words were presented at random to these models, meaning that a bilingual model could see a word in Spanish followed by a word in the L2.

During each pre-training epoch, the models performed the character prediction task mentioned above over their entire respective datasets. This process was repeated for all words in the monolingual or bilingual training sets, shuffled at each epoch's start. I employed three techniques to improve computational performance during pre-training. The first technique was to allow the model to process batches of 82 words instead of processing words one by one. The second technique was to apply a baseline learning rate of 0.002, adapted during pre-training using the Adam optimizer (Kingma & Ba, 2015). I did not explore how different learning rates affected the results presented in this study and instead chose a commonly used value in the deep learning literature (Le Godais et al., 2017; McMahan & Rao, 2019). Finally, I trained the model using the GPU capabilities, which allowed the parallelization of all vector and matrix operations.

Adapted Spanish LDT. The models were trained to criterion using the adapted Spanish LDT. Empirical evidence from a Spanish lexical decision megastudy suggests that monolingual and bilingual participants do not differ in their LDT accuracy (Aguasvivas et al., 2020). Therefore, the different versions and runs of the CLOUD models needed to have comparable performance on this adapted Spanish LDT. Each CLOUD version and run performed an identical adapted Spanish LDT where they calculated the probability of a pair of items matched in length: one target and one foil. The most probable item was selected as the word, and an accuracy measure was calculated based on this response for all words in the LDT datasets. I used the adapted Spanish LDT training set to match each run and version of the CLOUD model by setting an adaptive accuracy threshold that started at 65%—given that the models quickly reached this threshold after 1 or 2 training epochs. Upon reaching this threshold, a state of that model's parameters was automatically saved, and the threshold increased by 5%. I used the states corresponding to the maximum threshold achieved by all models for all analyses and simulations. The pre-training LDT accuracy and loss over 100 epochs for the CLOUD versions are presented in Appendix C3.

Adapted Flavian vocabulary learning task. After pre-training and matching the models' LDT performance, the monolingual and bilingual CLOUD versions performed an adapted version of the Flavian vocabulary learning tasks. However, since the CLOUD models only learned to produce the successor character distribution when prompted by a context, the models performed adapted versions of the *recognition* and *production* tasks used in Experiment 6. For the *recognition* task, I randomly assigned a numeric label from 0 to 47 to each Flavian word. I added a recognition module on top of the CLOUD architecture that predicted the Flavian word's numeric label. This "read-out" module performed a multinomial logistic regression, taking Flavian words' 128-dimensional hidden representation averaged over all time steps as input and predicting the probability of the correct label, written as

$$label = W_l \cdot h_w + b_l \ ; \ W_l \in \mathbb{R}^{48\,x\,128};\ b_l \in \mathbb{R}^{48} \tag{3}$$

where *label* is a 48-dimensional vector containing each label's logits from a Flavian word's hidden representation h_w, W_l and b_l are the weight and bias learnable parameters initialized as in Eq. 1. Importantly, this module is loosely analogous to the task subsystem employed by bilingual models such as the BIA+ (Dijkstra & van Heuven, 2002). In other words, it takes the orthographic representations of a word and executes an action, which in this case entails selecting its appropriate label.

Since participants in Experiment 6 performed a 4AFC task, I adjusted the recognition scores by selecting the logits from the target label and three randomized labels and then calculating the softmax function—i.e., the exponent of the logits divided by the exponents' sum. I then extracted the model's probability of selecting the correct label for the seen Flavian word from this adjusted vector. This process is analogous to performing a transformation based on the Luce choice axiom (Jessie & Saari, 2016). The production task needed minimal adaptation, as the models could already produce the succeeding character from a prior context. Instead of computing the produced word's accuracy as in Experiment 6, I calculated the model's probability of producing this word by multiplying the probability obtained for each word's characters during learning. I further boosted the probability by multiplying it by the length of each Flavian word (5), thus shifting the probabilities upwards. All analyses and visualizations were performed on the logarithm of the probabilities.

The models learned the Flavian words by backpropagating the loss during the recognition and production tasks using the same hyperparameters as in the pre-training task. Each model performed a familiarization task, followed by five alternating recognition and production learning blocks, as participants did in Experiment 6. However, since there is no information decay in the CLOUD model (i.e., the model does not forget over time), I did not attempt to replicate the test phase in Experiment 6.

Data analysis

The analysis plan was divided into three stages. The first step involved matching the different CLOUD models' versions and runs based on their LDT accuracy. It was essential to verify that the different versions of the models performed comparably on this task. However, while the adaptive threshold was used to select and match the models' performance on the training LDT dataset, I also computed each model's performance on the unseen test LDT words. The accuracy scores were then analyzed using ANOVAs for each model version (SP-EN, SP-BQ, MONO) and set (train or test) separately. The version factor was coded using a Helmert contrast to compare the monolingual and bilingual versions first and then the two bilingual versions. Finally, I compared the achieved accuracy of all models to an empirical chance level obtained by averaging the scores of 20 untrained CLOUD models on the same LDT datasets.

The second step was to evaluate the language organization within the different versions of the CLOUD models. I incrementally obtained the representations for every character in every word in the Spanish-English and Spanish-Basque validation and test sets. Then, a series of logistic regression classifiers were employed to categorize the representations at each time step into one of two classes: Spanish (L1) or English/Basque (L2). The logistic regression classifiers used the L2 (ridge) regularization, with a regularization strength of 1. They were trained using the "lbfgs" solver as implemented by the scikit-learn Python module (Pedregosa et al., 2011), with 10^9 maximum iterations and a fixed seed for replicability (404). The classifiers were only trained on test set words—unseen during training by the models—to minimize computation time. Half of the test sets' representations (820 words) were used to train the models. The other half (820 words) was used to calculate the probability of the logistic regression model selecting the correct language. This procedure was repeated at every possible character position, from 0 (the space symbol) to 10 (the character before the space symbol). The resulting values indicate the probability of the logistic regression model of selecting the appropriate language of a word based on the

CLOUD representations at a specific character. Welch t-tests with Bonferroni corrections for multiple comparisons (0.05/11) were used to contrast the probability scores across versions.

The last step was to simulate the results from the Flavian vocabulary learning experiment. The resulting data from the five learning blocks—measured in log(probabilities)—were modeled using linear growth curve analyses (GCA) for the recognition and production tasks, respectively. I included second-order orthogonal polynomials to reflect linear and quadratic changes across the blocks, using the version (i.e., MONO, SP-EN, SP-BQ) and condition (ES+ versus ES-) factors as fixed effects on all time-terms. The condition factor was deviation coded as -1 (ES-) and 1 (ES+). The version factor was reverse Helmert coded first to contrast the SP-EN and SP-BQ versions (coded as -1 and 1), then contrasting the bilingual (BIL) versions against the MONO version (coded as -2 and 2, respectively). I included the main effects and interactions of the time, condition, and version factors into the fixed effects but excluded any three-way interactions with the time-terms from the GCA models. The final GCA models only converged with the by-run intercepts. Estimating the Bayes Factor for these models was not informative due to the small number of runs in each version.

4.2.4. Results

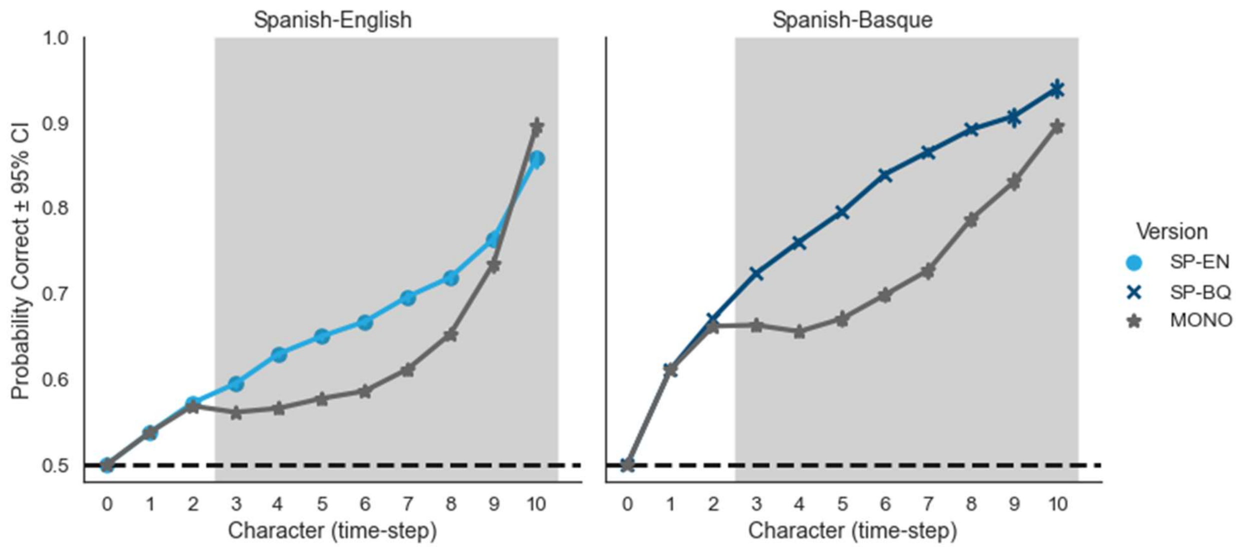
CLOUD Spanish LDT performance. An ANOVA comparing the LDT training dataset accuracy by version revealed no differences between the three versions ($F_{(2, 27)} = 1.716$, $p = 0.199$, $BF_{10} = 0.626$). Similarly, the Helmert contrast indicated no differences between the monolingual and bilingual versions ($p = 0.150$) or between the two bilingual versions ($p = 0.275$). The average accuracy achieved by the monolingual version was 85.3% (SD = 0.2), the Spanish-Basque version averaged 85.3% (SD = 0.2), and the Spanish-English 85.2% (SD = 0.2). I calculated the mean and standard deviation achieved by 20 randomly initialized but untrained CLOUD models on the same LDT training set as an empirical chance level. These randomly initialized models averaged 49.7% (SD = 1.0) on the training set. I then contrasted each version's achieved accuracy against this empirical chance-level using a Wilcoxon test. The results indicated that all versions performed above the empirical chance level (all $p < 0.001$, $BF_{10} > 100$). Appendix C4 shows more details and word clouds for the monolingual and bilingual models after pre-training.

An ANOVA comparing the LDT test dataset accuracy by version indicated no differences between the versions ($F_{(2, 27)} = 0.612$, $p = 0.550$, $BF_{10} = 0.315$). Additionally, the Helmert contrasts between the monolingual and bilingual versions and both bilingual versions did not reach significance ($p = 0.585$, p 0.347, respectively). The empirical chance-level performance for this task was 49.5% (1.5%). All versions performed significantly above this level (all $p < 0.001$, $BF_{10} > 100$). The monolingual versions averaged

77.3% (SD = 0.6), the Spanish-Basque 77.1% (SD = 0.7), and the Spanish-English versions 77.3% (SD = 0.6). These results were surprising, considering these words were unseen by the models during training. As a point of contrast, a prior study has also shown an accuracy score of around 85% for a more complex one-layer LSTM character-level model trained on monolingual input (Le Godais et al., 2017). Also, monolingual and bilingual human participants scored around 70% on average in a standard LDT that used the same items as here (Aguasvivas et al., 2020).

Language organization in the CLOUD. The two main features of the CLOUD architecture are its unified orthographic lexicon and its distributed word representations. As the second step, I evaluated whether the hidden representations of the pre-trained models contained sufficient information about the language category despite their unified structure. To test this idea, I obtained the hidden representations of every word in the test set while processing one character at a time. Half of the words served as input to a logistic regression model trained to decode each word's language category from the hidden representations at every possible character position. I then tested the logistic regression model's performance on the remaining half of the representations at each time step. I repeated this procedure to compare the MONO and SP-EN versions on the Spanish-English dataset and the MONO and SP-BQ versions on the Spanish-Basque dataset. Figure 15 depicts the logistic regression's probability of selecting the correct language at every character position.

Figure 15. *Probability of selecting the correct language at every character position.*



Note. Shading indicates significance at the Bonferroni corrected level. Time-step 0 is the space symbol's representation, which is identical for all words in all the models. MONO = monolingual; SP-EN = Spanish-English bilingual; SP-BQ = Spanish-Basque bilingual; CI = confidence interval.

Surprisingly, the logistic regression models performed significantly above chance level (0.5) after just the first character in all versions of the CLOUD model and datasets ($p < 0.001$, $BF_{10} > 100$). This does not imply that performance was good at this time step, as the probability of selecting the correct language was about 0.55. In the Spanish-English dataset, the logistic regression had a higher probability of correctly selecting the language category from the SP-EN versions than the MONO versions after the third character ($p_{bonf} < 0.001$, $BF_{10} > 100$), and until the ninth character ($p_{bonf} < 0.001$, $BF_{10} > 100$). Interestingly, decoding was better from the MONO version than the SP-EN version at the last time-step ($p_{bonf} < 0.001$, $BF_{10} > 100$). At the tenth character, the MONO versions had an average probability of selecting the correct language of 0.895, whereas the SP-EN versions averaged 0.858.

In the Spanish-Basque dataset, the decoding was more accurate from the SP-BQ version than the MONO version from the third ($p_{bonf} < 0.001$, $BF_{10} > 100$) to the tenth character ($p_{bonf} < 0.001$, $BF_{10} > 100$). At the tenth character, the MONO versions had an average probability of selecting the correct language of 0.896, whereas the SP-BQ versions averaged 0.940. Regardless of the differences between the bilingual and monolingual versions, these results suggest that, even in the MONO version, the internal representations of an unseen foreign word may be sufficiently different to distinguish whether it belongs to a known language or not—as opposed to identifying the exact language of the word.

Adapted Flavian vocabulary learning. Having matched the CLOUD models' performance after pre-training and evaluated the models' ability to categorize the languages despite having a unified lexicon, I focus on this experiment's central question. That is, how does exposure to words from two versus one language facilitate foreign vocabulary learning with varying degrees of orthographic similarity? Here is where the *dynamic* constraint becomes relevant. The pre-trained models learned the Flavian vocabulary, adjusting their weights to incorporate these novel words. The results for the recognition task are shown in Table 12. The GCA indicated a significant effect of the Linear and Quadratic time-terms (both $p < 0.001$). The condition factor was also significant ($p < 0.001$), but the effect was the opposite as in Experiment 6. In other words, words in the ES- condition had an overall higher score than those in the ES+ condition. As in Experiment 6, there were no differences between the bilingual versions ($p = 0.938$), but both bilingual versions outperformed the monolingual version ($p < 0.001$). The bilingual versions also had a significantly different Linear slope than the monolingual version ($p = 0.036$). Finally, there was a significant interaction between the Condition and the monolingual versus bilingual contrast ($p = 0.007$). Contrary to Experiment 6, the difference between the monolingual and bilingual versions was more considerable for the ES- condition. There were no further two-way interactions (all $p > 0.05$). The CLOUD simulation results for the recognition task are depicted in Figure 16A.

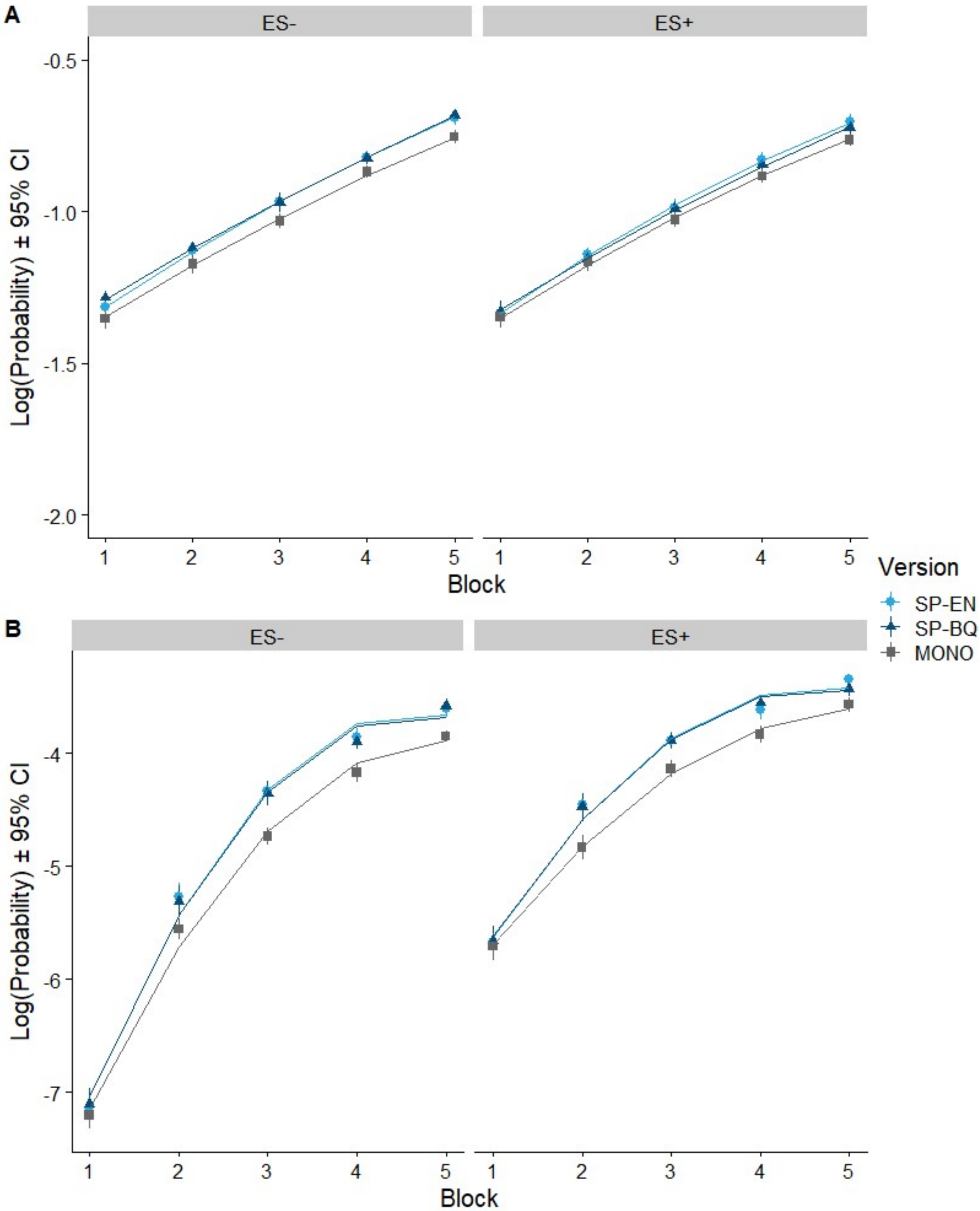
Table 12. *Linear GCA of the simulated recognition learning task.*

Fixed Effects		Estimate	SE	df	t	*p*
(Intercept)		**-24.175**	**0.073**	**27.0**	**-330.747**	**< 0.001**
Linear		**11.522**	**0.104**	**259.0**	**110.333**	**< 0.001**
Quadratic		**-0.708**	**0.104**	**259.0**	**-6.777**	**< 0.001**
Condition		**-0.196**	**0.047**	**259.0**	**-4.197**	**< 0.001**
SPBQ-SPEN		-0.007	0.090	27.0	-0.079	0.938
MONO-BIL		**-0.357**	**0.052**	**27.0**	**-6.899**	**< 0.001**
Linear x Condition		-0.014	0.104	259.0	-0.133	0.894
Quadratic x Condition		-0.076	0.104	259.0	-0.724	0.470
Linear x SPBQ-SPEN		-0.216	0.128	259.0	-1.685	0.093
Linear x MONO-BIL		**-0.156**	**0.074**	**259.0**	**-2.109**	**0.036**
Quadratic x SPBQ-SPEN		0.150	0.128	259.0	1.175	0.241
Quadratic x MONO-BIL		-0.008	0.074	259.0	-0.105	0.916
Condition x SPBQ-SPEN		-0.103	0.057	259.0	-1.794	0.074
Condition x MONO-BIL		**0.090**	**0.033**	**259.0**	**2.728**	**0.007**
Random Effects	*Group*	*Variance*	*SD*	*Correlation*		
Run	(Intercept)	0.095	0.308			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

The production task results are shown in Table 13 and depicted in Figure 16B. In this case, the analysis indicated significant Linear and Quadratic time-terms (both $p < 0.001$). Contrary to the recognition task, the Condition effect was in the expected direction, with ES+ words having a higher probability than ES- words throughout the learning blocks ($p < 0.001$). There were no differences between the bilingual versions ($p = 0.723$), but both bilingual versions outperformed the monolingual version ($p < 0.001$). As in Experiment 6, there were Linear by Condition and Quadratic by Condition interactions (both $p < 0.001$), indicating that performance increased and reached an inflection point more rapidly in the ES+ condition than the ES- condition. Furthermore, there were Linear and Quadratic interactions with the bilingual versus monolingual contrast ($p = 0.047$, $p < 0.001$, respectively). These interactions suggested a faster acquisition rate and a faster inflection point for the bilingual than the monolingual CLOUD versions. The rest of the two-way interactions were not significant (all $p > 0.05$).

Figure 16. *Experiment 6 simulation results.*



Note. Log-probabilities (symbols, vertical lines indicate ± 95% confidence intervals) by version and condition for the recognition (A) and production (B) simulation tasks. The solid lines depict the average GCA model predicted values. The conditions are plotted separately to avoid cluttering. SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

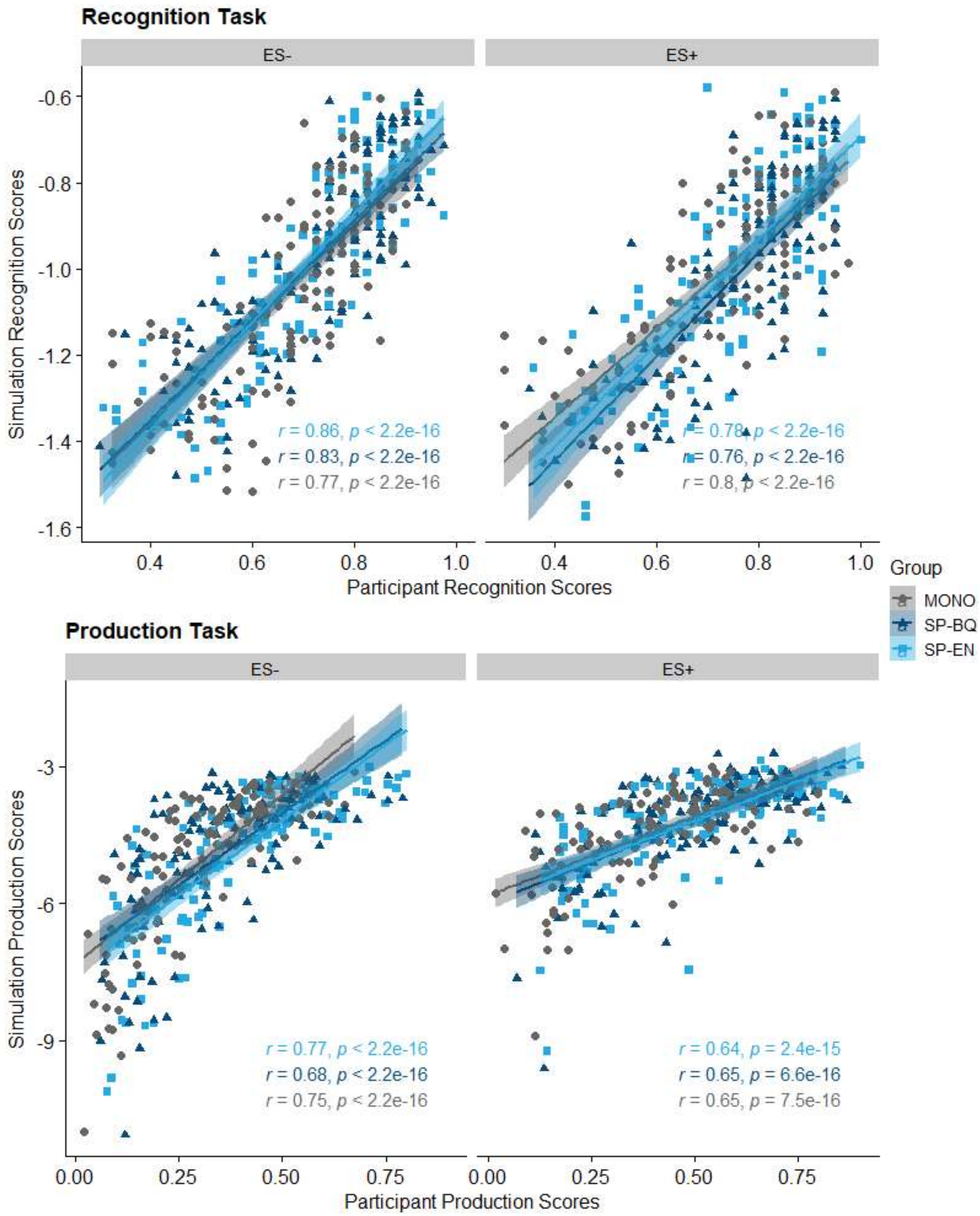
Table 13. *Linear GCA of the simulated production learning task.*

Fixed Effects		Estimate	SE	df	t	*p*
(Intercept)		**-110.461**	**0.251**	**27.0**	**-440.440**	**< 0.001**
Linear		**52.149**	**0.468**	**259.0**	**111.398**	**< 0.001**
Quadratic		**-17.372**	**0.468**	**259.0**	**-37.110**	**< 0.001**
Condition		**7.928**	**0.209**	**259.0**	**37.870**	**< 0.001**
SPBQ-SPEN		-0.110	0.307	27.0	-0.358	0.723
MONO-BIL		**-1.906**	**0.177**	**27.0**	**-10.750**	**< 0.001**
Linear x Condition		**-11.244**	**0.468**	**259.0**	**-24.019**	**< 0.001**
Quadratic x Condition		**4.157**	**0.468**	**259.0**	**8.879**	**< 0.001**
Linear x SPBQ-SPEN		-0.280	0.573	259.0	-0.488	0.626
Linear x MONO-BIL		**-0.662**	**0.331**	**259.0**	**-1.999**	**0.047**
Quadratic x SPBQ-SPEN		0.072	0.573	259.0	0.125	0.901
Quadratic x MONO-BIL		**1.420**	**0.331**	**259.0**	**4.290**	**< 0.001**
Condition x SPBQ-SPEN		0.037	0.256	259.0	0.145	0.885
Condition x MONO-BIL		0.165	0.148	259.0	1.113	0.267
Random Effects	*Group*	*Variance*	*SD*	*Correlation*		
Run	(Intercept)	0.572	0.756			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

Correlation with Experiment 6 results. As a final exploratory step, I calculated the mean log probabilities achieved by all CLOUD versions on each word by learning block and condition, separately for the recognition and production tasks. I also computed the participants' average scores per word using the same procedure. These values were subjected to a Pearson correlation to verify that the model produced similar results as those observed in the participants for the same words and learning blocks. The correlations were performed for the recognition and production tasks and each group separately. The results are depicted in Figure 17. The results indicated that the recognition and production scores were significantly and positively correlated (all $p < 0.001$), ranging from 0.77 to 0.86 in the recognition task and from 0.64 to 0.77 in the production task. These results were expected considering that performance increased during the learning phase for the participants and the models. Interestingly, overall, higher participant scores for a specific word at a learning block corresponded to higher scores for that word in the simulations.

Figure 17. *By-item correlations between the simulation and participant scores.*



Note. Shading indicates the 95% confidence interval. Each shape is an individual observation. **(top)** Recognition task correlations. **(bottom)** Production task correlations. SP-EN = Spanish-English bilinguals; SP-BQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals.

4.2.5. Summary

In Experiment 7, I introduced a model (CLOUD) to account for the findings from Experiment 6. The CLOUD model is instantiated as a recurrent neural network with three constraints: unified, distributed, and dynamic orthographic lexicon. I have shown that this simple model can accomplish several feats. First, it can successfully perform an adapted Spanish Lexical Decision Task without explicit instruction on word meaning. The three versions of the CLOUD model (i.e., Spanish monolingual, Spanish-English bilingual, and Spanish-Basque bilingual) were matched and significantly above chance in their adapted LDT performance. Critically, this suggests that the model could perform the adapted LDT without ever accessing any meaning for the seen words and purely based on orthotactic patterns. Prior work has shown that non-human primates (baboons) can use orthotactics to successfully distinguish words from pseudowords in an LDT, with both trained and unseen items (Grainger et al., 2012; Rajalingham et al., 2020).

Second, it can learn distributed representations of the orthotactic patterns in words from different languages, and these representations contain information about a word's language category. Neuroimaging studies have identified an area in the neocortex that gets activated with orthotactic information and has been called "Visual Word Form Area" (Dehaene et al., 2002). Its activation seems to be modulated by the orthotactics of natural languages, even more so than word frequency (Woolnough et al., 2021). While I do not claim that the CLOUD model is a one-to-one mapping to this area, the similarities are striking. Indeed, future work could employ the CLOUD architecture to replicate the activation patterns observed in this area and extend the findings to bilingual contexts.

Third, the CLOUD model can (partially) replicate the results from Experiment 6, showing an advantage of bilingual exposure and differences arising from orthographic similarity. Curiously, the recognition results revealed an effect of orthographic similarity in the opposite direction as the behavioral experiment. If the ES+ words were also more similar to one another, the CLOUD model might have had a more challenging time disentangling their representations to predict the appropriate label. Regardless, the recognition and production simulation results corroborated that experience with the orthography of two languages facilitated learning orthographically similar and dissimilar words over experience with a single language. Critically, for learning to occur, the models had to adapt their existing representations learned during the pre-training (see Appendix C5). Future studies could elaborate on exactly how and when the mental lexicon reorganizes its representations.

Discussion

The present experiment explored how orthographic similarity and bilingual experience influence foreign vocabulary learning through a computational approach. In the simulations presented above, I have shown a computational model of the orthographic lexicon for written vocabulary learning. The model could reproduce the bilingual advantage for vocabulary learning and orthographic similarity effects as evidenced in the behavioral results of Experiment 6. Here, I discuss how the CLOUD model's unique constraints—namely its unified, distributed, and dynamic nature—account for the behavioral data.

All orthographic word forms are stored in a unified lexicon in the CLOUD model, regardless of the language. While I am not the first to propose the idea of an integrated mental lexicon (Baxter et al., 2021; French & Jacquet, 2004), the findings presented here suggest that this constraint is essential for understanding native and foreign vocabulary learning. A direct implication stemming from a unified lexicon is the ability to re-use information during vocabulary learning. For instance, it is unnecessary (and frankly inefficient) to construct completely different representations for the words "model" and "modelo", as information from the former could help acquire the latter. It may seem like a trivial realization, but the most prominent computational models of the mental lexicon have largely overlooked vocabulary learning, opting to implement whole-word or language-specific tags (Dijkstra et al., 2019b; Dijkstra & van Heuven, 2002). Instead, words share their orthographic representations in the CLOUD architecture so long as they share their script.

I presented evidence that a unified orthographic lexicon facilitates learning orthographically similar words. As shown in Experiment 6, participants' performance was better for the orthographically similar words in the recognition and production tasks. The model could reproduce the orthographic similarity effects in the production task. However, there were some discrepancies in the recognition task. Aside from the different learning trajectories arising from the distinct nature of the behavioral and simulation tasks, perhaps the most relevant difference is the inverse effect of orthographic similarity in the recognition task. A possible explanation is that, while orthographically similar words are more straightforward to produce, their representations are more difficult to disentangle when assigning a label (see also Appendix C5). Previous research has suggested that, for language learners, similarity can also produce interference rather than facilitation in the absence of cues to direct attention to the differences between words (Baxter et al., 2021; P. Nation, 2000). These discrepant findings provide testable predictions regarding the interplay of attention and word form similarity during receptive vocabulary learning.

The second constraint that characterizes the CLOUD model is its use of distributed representations. Distributed representations allow the model to learn from the orthographic patterns composing words from different languages. Despite the unified lexicon, the results show that this bottom-up model can store information about the words' language category in its distributed representations without requiring any top-down language tags. While theory suggests that top-down signals can modulate language activation in the bilingual mental lexicon (Dijkstra & van Heuven, 2002; Grainger et al., 2008; Van Heuven et al., 1998), word form representations are sufficient to distinguish the language category. In other words, individuals do not need to constantly know the language of every word they read, as it can potentially be inferred from the word itself if necessary.

Prior modeling efforts have shown that distributed representations in the bilingual lexicon can cluster according to their respective language (French, 1998; Li & Farkas, 2002). This experiment extends these findings in several ways. First, I showed that language discrimination could occur at the sub-lexical orthographic level. Some specific orthographic patterns are unique to words in different languages (e.g., "th" in English; "tx" in Basque), and the model seems to capture those patterns within its distributed representations. These results are consistent with a growing literature investigating written language identification in bilingual contexts (e.g., Casaponsa et al., 2014, 2015). In brief, these studies suggest that bilinguals can identify the language of a word using the orthotactics of their known languages. Thus, like the CLOUD model, these studies indicate that sub-lexical orthographic information is enough for languages to organize in the mental lexicon (Casaponsa et al., 2014, 2015). Second, I show that language separation is not a unique feature of the bilingual lexicon but rather a by-product of distributed representations. Even in the monolingual version, the representations were enough to distinguish words that belong to a known language from words that do not—without ever learning words in other languages. Third, the results suggest that as the word's length increases, so does its individuality within the mental lexicon, with longer words having more language-specific representations than shorter words. Finally, there was also a bilingual advantage for language classification. The bilingual versions were better than the monolingual counterpart in classifying the language from the distributed representations, particularly with shorter words.

This experiment's central finding is the bilingual advantage for foreign written vocabulary learning, evidenced in both the model and human performance, receptive and productive vocabulary tasks, and orthographically similar and dissimilar words. Studies have shown that bilinguals outperform monolinguals when learning dissimilar words (Kaushanskaya & Marian, 2009a, 2009b), but the results seem to indicate that this is also the case for similar words. In the CLOUD model, experience with words from two languages shapes the distributed representations, making the orthographic lexicon more flexible to incorporate similar and dissimilar patterns. In other words, the representations become less tuned to specific orthographic patterns to encode the input's variable statistics more efficiently. Interestingly, the effects were evident for both bilingual versions of the model, regardless of the language combination, pointing to bilingual exposure as the source for this advantage. Future studies could manipulate word form similarity to multiple languages to dissociate the specific effects of similarity to one language from the overall effects of bilingual experience.

The final—and possibly the more controversial—constraint in the CLOUD model is the dynamic lexicon. Namely, learning new words modifies the existing representations of words in the lexicon (see Figure 16 for an example). It is safe to assume that the human brain never turns off its learning mechanisms, as evidenced by implicit learning studies (Christiansen, 2019; Perruchet & Pacton, 2006). Thus, it would make sense that the language context and environmental demands continually modify the existent representations. However, this poses additional questions regarding the mental lexicon's stability, particularly in adulthood—for example, do the representations for certain words ever stabilize? How much exposure is necessary to produce a change? Research suggests that, even in adult monolinguals, the foreign language learning experience can reshape their native language knowledge (Bice & Kroll, 2015). Furthermore, adult monolinguals exposed to heterogeneous linguistic contexts show a vocabulary learning advantage over those from a more homogeneous linguistic context (Bice & Kroll, 2019). Together, these findings suggest that the mental lexicon is not static, and mere exposure to different words forms can affect language processing and foreign vocabulary learning. The extent to which the mental lexicon is dynamic remains to be studied.

There are, of course, limitations when comparing such a simple model to the nuances of human behavior. In my simulations, I did not model differences arising from early versus late second language learning. Instead, I approximated an adult monolingual and bilingual lexicon using exposure to random words in one or more languages as a proxy. However, in practice, bilingualism involves many more factors than exposure, ranging from individual learning experiences to cultural and identity aspects. Additionally, most mental lexicon theories argue for the idea of word resting level activation—established due to the frequency of exposure to different words—and how this crucial factor intervenes in processes such as competition during word recognition (Baxter et al., 2021). In pre-training the CLOUD models, however, I assumed that all words were equally frequent. While the influence of second language age of acquisition and word frequency could be easily put to the test by modifying the CLOUD model's pre-training datasets (e.g., using complete texts rather than single words), these aspects remain beyond this study's scope.

One critical observation is that I explicitly avoided meaning representations, opting to identify the minimal structure that could account for written vocabulary learning. A limitation of this decision is that there is no distinction between cognates (e.g., "pera"-"pear") and false friends (e.g., "carta"-"letter" and not "card") in the CLOUD model. The nature of meaning representation has been consistently problematic to implement. Some authors reduce it to manually coded distributed representations, and others treat these as localist (single-valued) labels (French & Jacquet, 2004; Laszlo & Plaut, 2012). Recent advances in artificial intelligence suggest that meaning representations are also dynamic and highly context-dependent, like the orthographic representations in the present work (Devlin et al., 2019). The meaning representations pose an additional challenge when dealing with multilingual data—for instance, is the meaning of two words precisely the same across languages? Is it even the same for two individuals? While it will be helpful for future work to explore the possibility of extending the model by adding meaning representations, my aim here was to report the efficacy of a minimal model for learning orthographic representations in multiple languages.

Finally, I do not claim by any means that the learning mechanisms employed by artificial neural networks are humans-like or biologically plausible. Other models employing distinct learning mechanisms or architectures to produce unified, distributed, and dynamic representations could match or even surpass the CLOUD model's capacity to explain the behavioral data. Still, this model provides a first large-scale attempt towards understanding the role of prior experience on foreign written vocabulary learning. Future work could employ more biologically plausible architectures to corroborate the ideas presented here, also including phonological and meaning representations.

Chapter 4 Conclusion

Foreign vocabulary learning is a multi-faceted and complicated process influenced by multiple variables. Amongst them, past research has highlighted similarity and bilingual experience as catalysts in foreign vocabulary learning. The present study is first in unifying these seemingly disparate findings under a common computational framework, whereby distributed representations of word forms are stored in a unified space and dynamically modified by learning experiences. To some extent, the model can simulate human behavior, corroborating the influence of orthographic similarity and showing a bilingual advantage for receptive and productive vocabulary. This conceptualization has implications about how consistent experience with specific words in different linguistic contexts can influence foreign vocabulary learning.

Chapter 5: General Discussion

Language learning is a complex human activity comprised of multiple interconnected analytic levels (e.g., sub-lexical, lexical, morphological, semantics). Having experience with two languages could provide additional implicit or explicit knowledge to exploit during language learning at one or multiple analytic levels. This idea is backed by many experimental and theoretical studies, highlighting that experience with more languages facilitates subsequent language learning (for recent reviews, see Festman, 2021; Hirosh & Degani, 2018; Montanari, 2019). In this work, I was interested in whether monolingual and bilingual—not multilingual—participants, matched in a series of variables, differed in learning contexts that are novel and as equal as possible for all of them. I sought to address a single question from multiple angles: *Are adults who already know two languages (bilinguals) better at learning a foreign language than those who only know one (monolinguals)?* I started by decomposing this question into three specific research questions regarding "where" (RQ1), "what" (RQ2), and "how" (RQ3) are bilinguals and monolinguals different in foreign language learning.

I limited the research context by considering initial artificial language learning as a proxy for foreign language learning ability. Moreover, I recruited participants from Spain with three linguistic profiles: Spanish monolinguals, Spanish-English bilinguals, and Spanish-Basque bilinguals. The bilingual participants were highly proficient in both languages, had minimal exposure to a third language, and started learning their second language on average before the age of six, making them relatively early but not necessarily balanced bilinguals —as per the definitions in Chapter 1. Throughout the experimental part of this work, participants in each group were exposed to artificial linguistic materials over one learning session and at most two test phases separated by one day. Consequently, any resulting claims arising from this work are limited to the context of initial exposure and learning in an artificial language of individuals with the language profile described above.

In what follows, I will briefly summarize the results from all experiments in this work. I will then present my arguments regarding the three specific research questions. Next, I will contextualize all these findings into the overall literature, highlighting their theoretical and practical implications, and strengths and limitations. I will outline some outstanding questions for future work regarding this research line. Finally, I will conclude this thesis by answering the overarching research question as originally posed.

5.1. Summary of findings

Chapter 2 targeted implicit foreign language learning at the sub-lexical phonology, morphology, syntactic, and lexical-semantics level using four well-established statistical language learning experiments. In Experiment 1, I compared the ability of bilinguals and monolinguals to segment words from three auditory artificial speech streams that varied in their sub-lexical phonological patterns (i.e., phonotactics). The results suggested that, while the complexity of the phonotactic patterns hindered participants' performance in the task, there were no differences between monolinguals and bilinguals in their overall performance.

Experiment 2 extended the findings from the previous experiment by testing participants' ability to generalize the knowledge acquired through statistical language learning to novel words. Specifically, I designed the artificial speech streams to mimic affixal morphology (e.g., **un**break**able**, **un**touch**able**), and participants had to generalize this knowledge to new items (e.g., **un**fill**able**). Again, the bilingual and monolingual participants did not differ in this task, and thus, there was no bilingual advantage.

Experiment 3 targeted the syntax level. In this experiment, I tested participants' ability to segment words from an ambiguous speech stream that could be segmented based on a syntactic property of their known languages: word order. The results indicated that Spanish-Basque bilinguals outperformed the other two groups in this task, better segmenting the words congruent with the predominant word order of Basque (SOV). Regardless, as in the previous two experiments, there was no overall bilingual advantage. This experiment was vital to disentangle the effects of bilingual experience from those stemming from specific language knowledge.

Finally, in Experiment 4, I targeted the lexical-semantics level using a statistical word-referent learning task. In this audio-visual task, participants had to discover the names of non-existing visual referents implicitly through a series of scenes (Cross-Situational Statistical Learning). Some objects only had one name—i.e., they were exclusive mappings—, but other objects could have two distinct names (synonyms), or two distinct objects could have the same name (homonyms). In experiencing two languages, participants learn multiple types of word-referent mappings, so the natural prediction was that bilinguals would outperform monolinguals in learning these three types of mappings. However, the results indicated a bilingual advantage only for the one-to-one (exclusive) mappings and no differences between monolinguals and bilinguals on the multiple mappings.

Chapter 3 addressed explicit foreign language learning using two additional experiments that targeted morphology and lexical-semantics learning. Experiment 5 compared monolinguals and bilinguals when learning new suffixes for existing Spanish stems (e.g., laboral**suti**). Since participants already knew the definitions for these stems, the experiment addressed only the group differences in learning the artificial suffixes' orthographic form. The results revealed that, while participants could learn to discriminate the combined stem and suffixes from other types of foils, the groups did not differ in their performance. In other words, as in Experiment 2, there were no differences between bilinguals and monolinguals in learning morphological information. Moreover, an exploratory correlation analysis only indicated a weak correlation between Experiment 2's scores and one of the conditions in Experiment 3.

Experiment 6 combined the sub-lexical orthography and the lexical-semantics level. In this experiment, participants were tasked with learning an artificial written vocabulary named Flavian. The critical manipulation was that each words' sub-lexical orthography was either more similar (ES+) or more dissimilar (ES-) to Spanish—which is the common language that all participants speak. There were three important results from this experiment. First, orthographic similarity facilitated learning the written words throughout the learning blocks, consistent with prior findings in the literature (Hayakawa et al., 2020; Marecka et al., 2021). Second, bilinguals were better than monolinguals at learning the novel written words, regardless of their similarity to Spanish. This finding extends prior experimental work suggesting a bilingual advantage using only dissimilar and auditory novel words (e.g., Antoniou et al., 2015; Kaushanskaya & Marian, 2009b). Third, the differences between bilinguals and monolinguals were more prominent for the ES+ words than the ES- words during the recognition learning phase, but the differences between the groups and the interaction disappeared at the test. Conversely, orthographic similarity and bilingual status did not interact during the learning or test phases for the production task, with bilinguals being better across the board in producing the Flavian words.

Chapter 4 focuses on understanding how the findings from Experiment 6 could emerge from exposure to bilingual input. Despite several computational models addressing the monolingual and bilingual mental lexicon, none of these models adequately accounts for the results mentioned above. Therefore, in Experiment 7, I proposed and developed a new computational model specifically focused on the orthographic lexicon, named CLOUD. I employed this model to examine how orthographic similarity and bilingual experience could influence foreign vocabulary learning. Among other findings, the highlight of Experiment 7 is that the computational models could simulate—with some differences—the effects of orthographic similarity and bilingual status observed in Experiment 6. This experiment unifies the seemingly disparate findings of orthographic similarity and bilingual experience under a common computational framework, whereby distributed orthographic representations reside in a unified lexicon and are modified by learning experiences.

I can combine the findings from this thesis to answer the three specific research questions and the main question as outlined in Chapter 1.

RQ1. At which analytic level, if any, do bilinguals and monolinguals differ? "Where"

Overall, the findings from this thesis indicated that bilinguals and monolinguals did not differ in learning at the sub-lexical, morphology (implicitly or explicitly), and syntax levels. They did, however, differ in implicitly and explicitly learning word-referent pairs at the lexical-semantics level, wherein both bilingual groups outperformed monolinguals in learning novel vocabulary but did not differ from each other. Therefore, based on these results, my answer to **RQ1** is that bilinguals and monolinguals mainly seem to differ at the vocabulary (lexical-semantics) level. I will henceforth refer to these differences as "bilingual vocabulary learning advantage."

Notably, throughout this work, I have shown more null than positive findings. A critical aphorism to consider here is that the *absence of evidence is not evidence of the absence*. In other words, the fact that my experimental designs, participants, and stimuli selection did not lead to observed differences between the groups is not proof that these do not exist. Indeed, different experiments and manipulations targeting other types of bilinguals or language combinations could—and have—shown there could be potential benefits arising from experience with more than one language (e.g., Antoniou et al., 2015; Poepsel & Weiss, 2016; Wang & Saffran, 2014). My experiments target a fraction of the multiverse of possible designs and manipulations, therefore, I cannot firmly conclude that there are absolutetly no differences between monolinguals and bilinguals on the analytic levels mentioned above. Nevertheless, as a counterargument, I believe it is highly likely that studies that do not find differences are also more susceptible to file-drawer effects and publication bias. There is evidence of these phenomena in the analogous literature regarding the so-called bilingual cognitive advantage (de Bruin et al., 2015, 2021). Open science practices such as replications and pre-registration could undoubtedly address these issues, providing a reliable answer to whether bilinguals and monolinguals indeed differ on aspects beyond vocabulary learning. Note, however, that Bayesian statistics showed more evidence for the null hypothesis when no differences were reported.

Compared to the rest of the experiments, Experiments 4 and 6 also included visual referents during learning. In Experiment 4, these were color depictions of non-existing objects, and in Experiment 6, black and white drawings of existing objects. The combined results of these experiments argue in favor of bilingual vocabulary learning advantage. Vocabulary learning studies have employed translation pairs (e.g., dog – perro) or depictions of existing objects in Experiments 4 and 6. The former approach is perhaps more similar to the classroom learning environment but restricts the language context in which the translation pairs appear (e.g., Spanish-Foreign Language). In this regard, while some studies have shown a bilingual vocabulary learning advantage using translation pairs (e.g., Kaushanskaya & Marian, 2009b, 2009a), a study comparing English monolinguals to Spanish-English bilinguals found no differences between the groups when learning Swahili-English translation pairs (Bakker-Marshall et al., 2021), possibly due to the Spanish-English group performing the task in their second language.

In contrast, using object depictions as referents during learning could allow individuals more robustly activate the meaning representations. The main challenge is that the vocabulary in these studies is limited to the objects that can potentially be depicted (i.e., concrete objects). Research suggests that the differences between monolinguals and bilinguals are more prominent for novel vocabulary associated with concrete, rather than abstract, referents (Kaushanskaya & Rechtzigel, 2012). It is possible that the observed differences between monolinguals and bilinguals primarily rely on participants' ability to integrate these word-referent mappings into the mental lexicon rather than any other factor. Future studies could add these referents to the sub-lexical, morphological, and syntactic levels and test whether these give rise to differences between monolinguals and bilinguals.

RQ2. What are the differences? "What"

The findings are consistent in showing a bilingual vocabulary learning advantage in one-to-one mappings but not in multiple mappings. Experiment 4 tested both one-to-one and multiple mappings, revealing a bilingual advantage only for the former type. If I consider Experiment 5 as a paired-associates task, it is easy to see how the stems and novel suffixes represented one-to-many mappings—i.e., each stem was paired with eight different suffixes. Therefore, the lack of differences in this experiment corroborates the findings from the multiple mapping condition in Experiment 4. Conversely, Experiment 6 presented participants with one-to-one word-referent mappings in an explicit learning task that combined recognition and production. Again, the results indicated a bilingual vocabulary learning advantage, as evidenced in Experiment 4.

Experiment 6 also tested the influence of sub-lexical orthographic similarity during foreign vocabulary learning. Words that are more similar in orthographic form to others in a known language were recognized and produced better during learning and testing. The analyses suggested that the differences between monolinguals and bilinguals were more considerable for the more similar than dissimilar words in the recognition, but not the production learning tasks. Nevertheless, I observed a bilingual vocabulary learning advantage for both similar and dissimilar words. Experiments 4 and 6 were the only ones to test participants at multiple points throughout the learning phase, suggesting that the bilingual vocabulary learning advantage seemed to emerge throughout the course of learning. In Experiment 4, the differences between monolinguals and bilinguals in recognizing the words became more prominent as learning progressed. In contrast, Experiment 6 showed a bilingual vocabulary learning advantage for the recognition learning task, but this advantage disappeared during the testing phase, regardless of test delay. Still, bilinguals were better than monolinguals at producing the Flavian words throughout learning and testing.

The answer to **RQ2** is that bilinguals seem to be better than monolinguals at learning one-to-one word-referent mappings regardless of the words' orthographic similarity to a known language. Also, the bilingual vocabulary learning advantage is more evident for productive rather than receptive vocabulary, possibly due to productive vocabulary being typically harder than its receptive counterpart (De Groot & Keijzer, 2000; Loucky, 2006). Notably, this advantage seems to develop throughout learning but does not necessarily emerge during testing, which could be why I did not find any differences in Experiments 1-3 and 5.

RQ3. How could these differences emerge from the bilingual experience? "How"

Rather than settle with ad hoc explanations (e.g., metalinguistic awareness, phonological working memory), I directly tested a hypothesis for how the bilingual vocabulary learning advantage might emerge. Concretely, I hypothesized that the observed facilitatory effects of orthographic similarity and bilingual experience were rooted in exposure to the orthographic patterns—orthotactics—of one versus two natural languages.

These ideas led to Experiment 7, where I attempted to replicate the findings from Experiment 6 employing the CLOUD model. Critically, the model met several criteria to help understand how bilingual experience and orthographic similarity influence vocabulary learning. First, a single unified, distributed, and dynamic architecture learned orthotactics in a monolingual or bilingual setting. Second, the resulting monolingual and bilingual versions only differed in their input (i.e., monolingual or bilingual), but not in the number of words that they learned. Third, like the rest of the experiments in this thesis, I verified that any effects were consistent for two bilingual contexts (Spanish-Basque, Spanish-English) and not a specific language combination. Fourth, consistent with prior findings (Aguasvivas et al., 2020), the monolingual and two bilingual versions of the model were comparable in performing an adapted Spanish lexical decision task. Lastly, the models performed an adapted version of Experiment 6's learning phase, having the same exposure to the Flavian words as the participants. Consequently, I can argue that the results were primarily due to the model's exposure to the orthotactics of one or two languages rather than the model's architecture or performance differences.

The simulations from Experiment 7 replicated the findings from Experiment 6. There was an effect of orthographic similarity and an effect of exposure to bilingual input. The orthographic similarity effect operated in the same direction as in Experiment 6 for the production but not the recognition task. In other words, the models' performance was worse for the ES+ words in the adapted recognition task than the ES- Flavian words, possibly due to their representations being more similar to one another. Nevertheless, both bilingual versions outperformed the monolingual counterparts, regardless of orthographic similarity, and the model's outputs for the artificial vocabulary were highly correlated with participants' scores. Hence, my response to **RQ3** is that the bilingual vocabulary learning advantage might emerge from mere exposure to bilingual input. In other words, through active or passive exposure, bilinguals might develop sub-lexical representations that are more flexible to integrate novel vocabulary than monolinguals. It seems that experiencing words in two languages could act as a regularization, allowing the model to develop flexible representations to incorporate similar and dissimilar novel words. I will further discuss this idea in the following section.

5.2. Theoretical implications

The theoretical framework I employed to select and plan the experimental part of this thesis is that of direct and indirect effects (Hirosh & Degani, 2018). This framework proposes that experience with multiple languages can *directly* influence subsequent language learning due to the transfer of linguistic information from any known language towards the target language to learn. In essence, learners try to exploit similarities during foreign language learning to acquire constructions more rapidly. A large body of literature from experimental linguistics backs this idea (e.g., Rothman, 2015; Westergaard et al., 2017). Studies have also shown that learners can draw from perceived or objective similarities to kickstart their vocabulary knowledge (e.g., Bartolotti & Marian, 2017; Hayakawa et al., 2020). Besides the orthographic similarity effect from Experiment 6, these direct effects do not explain the bilingual vocabulary learning advantage considering the artificial linguistic materials they had to learn in my experiments. In other words, there was minimal margin for cross-linguistic transfer in the experimental part of this thesis.

Multilingual experiences might also *indirectly* enhance non-linguistic or linguistic cognitive abilities that, in turn, improve language learning success (Hirosh & Degani, 2018). In the experimental part of this work, I have addressed both types of abilities. In Chapter 2, I contrasted monolingual and bilingual participants across four statistical/implicit learning experiments—a non-linguistic ability according to the direct/indirect framework. Previous studies have suggested that bilingual experience might enhance this ability (e.g., Onnis et al., 2018; Poepsel & Weiss, 2016; Wang & Saffran, 2014). However, the overall pattern of results from Chapter 2 contradicts this idea. I found that specific linguistic knowledge (e.g., SOV word order) can modulate learning in these tasks in Experiment 3.

Moreover, the results from Experiment 4 (in combination with Experiments 6 and 7) suggested that the differences between monolinguals and bilinguals emerge at the vocabulary level and are not due to an enhanced statistical/implicit learning ability in bilinguals. The current literature on statistical learning moves towards which individual differences can explain the outcome in these implicit tasks (e.g., Assaneo et al., 2019; Siegelman, Bogaerts, Christiansen, et al., 2017; Siegelman & Frost, 2015). Therefore, based on my results, I argue that bilingual experience may not be an individual difference that affects the statistical learning process but instead how learners filter the information in these tasks. In particular, studies should focus on the level of exposure to input in one or more languages as a source of variability in statistical learning experiments.

In terms of the indirect linguistic abilities, the theoretical framework posits that multilingual experience might strengthen or enhance the orthographic, phonological, or lexical-semantic networks (Hirosh & Degani, 2018). The results from Experiment 6 could very well fit the idea of enhanced orthographic or lexical-semantic networks. However, I argue that these ad hoc explanations are not sufficient, in part because there is no theory explaining how this strengthening/enhancement occurs or how neural networks implement it. As an alternative, in Experiment 7, I demonstrated that a simple computational model that learns the orthotactics from words in one or two languages was enough to replicate the findings from Experiment 6. These findings, of course, do not imply that my model is correct but certainly highlight the role of sub-lexical orthotactics in monolingual and bilingual vocabulary learning.

Therefore, my argument is that experience with two languages does not enhance or strengthens these networks but instead acts as a form of *regularization*. Regularization is a common concept in cognitive and computational learning theories (e.g., Ferdinand et al., 2019; Girosi et al., 1995; Hudson Kam & Newport, 2005). In computational terms, regularization implies imposing sparsity or strength constraints—through penalty terms or data variability—on a model's representations to reduce generalization errors (Goodfellow et al., 2016). In humans, this process typically refers to how learners adapt their production (or representations) to reduce the entropy in the information they process (Hudson Kam & Newport, 2005). Critically, I have demonstrated that this regularization process occurs even at the sub-lexical level when the information about orthographic patterns is unpredictable—as it comes from two distinct languages. In the context of my experiment, the orthotactic representations became less tuned to a specific language—due to the more variable and unpredictable statistics—and were generally more flexible to accommodate the novel artificial vocabulary.

Since the differences between monolinguals and bilinguals were mainly at the vocabulary level, other points of contrast to my work are the theoretical and computational models of bilingual word processing and learning. On the one hand, localist models, like the Bilingual Interactive Activation (Dijkstra & van Heuven, 1998, 2002; Grainger et al., 2008) and the Revised Hierarchical Model (RHM) (Kroll et al., 2010), have been highly regarded for their ability to account for multiple phenomena in bilingual word recognition and language representation, such as inter-language competition, and language switching effects. These models focus primarily on word processing rather than word learning and provide only verbal accounts of foreign vocabulary acquisition. For example, the RHM proposes that, during second language learning, individuals initially establish a link between form representations in the second language (L2) and their corresponding form in the native language (L1), without a direct link to meaning representations. With increased exposure to the L2, the connection between L2 words and meaning strengthens, while the original link between L1 and L2 word forms weakens. It is unclear how these models could account for the facilitatory effect of orthographic similarity and bilingualism during foreign language learning, as the representations of word forms are localist in nature and unique for each language.

On the other hand, three distributed models have suggested relevant ideas regarding bilingual mental lexicon development. The Bilingual Simple Recurrent Network—trained with simple sentences in English and French (e.g., *man lift boy; garcon voit ballon*)—was one of the first models to reveal that the lexical representations clustered according to each language without the need for explicit language tags (French, 1998). Following these results, the Dev-Lex-II model simulated the developmental differences caused by early versus late second language learning, showing similar language separation for phonological word forms (Li & Farkas, 2002; X. Zhao & Li, 2010). Finally, the Conceptual Feature Mapping model (Lee et al.,

1999) provides a verbal explanation for the facilitatory effects of word form similarity on foreign vocabulary learning. According to this model, learning a foreign word requires modifying the connections from the word form layer to an intermediate lemma layer or from this intermediate layer to the meaning representations, depending on whether the target word's form or meaning overlaps with its native language counterpart. Still, it is hard to extend this idea to more than two languages, as these lemma nodes are localist, and it would be necessary to add a new node for every "new" language.

Localist models are highly interpretable and can be compared to human behavior, but they require extensive manual engineering of connections and activations, which complicates explaining their learning mechanisms. Conversely, distributed models can learn representations using simple learning mechanisms, but these models have yet to be compared to human behavior directly (Holman & Spivey, 2016). Importantly, all of these models only present small-scale simulations using carefully selected subsets of words (Li & Grant, 2019), which hinders their generalizability. Finally, none of them have been extended beyond second language learning to explain the differences between monolingual and bilingual vocabulary learning.

In this regard, the CLOUD model from Experiment 7 provides a first large-scale attempt to understand how the bilingual vocabulary learning advantage emerges. Although I developed this model to understand how the pattern of results from Experiment 6 emerged, the ideas contained in this work extend beyond the bilingual vocabulary learning advantage. First, the model contains a unified mental lexicon, where character and orthotactic representations lie in a common space for multiple languages—so long as they share scripts. Adding a unified phonological layer could establish the correspondence between different scripts (e.g., Russian "Г" and English "G" partially map to the same phoneme /g/). Second, the model exploits the expressivity of distributed representations, making this architecture flexible to words within and outside its vocabulary and ideal for probing questions about native and foreign vocabulary learning. Finally, the idea of the lexicon being dynamic and continually modified by experiences poses novel challenges and exciting ideas to investigate learning and processing in the mental lexicon. Studies suggest that learning a foreign language inadvertently reshapes native language knowledge (Bice & Kroll, 2015; Borragan et al., 2020). A recent study has shown that even English monolinguals exposed to a less predictable orthographic context (e.g., living in California versus living in Pennsylvania) show a vocabulary learning advantage over monolinguals exposed to a more predictable context (Bice & Kroll, 2019).

The CLOUD architecture implemented in this work is a simplified model of the orthographic lexicon. In this regard, it requires vertical and horizontal expansions to become a fully-fledged model of the (multilingual) mental lexicon. For instance, the distributed letter representations could be computed based on the letters' visual features rather than using an Embedding layer. Similarly, how orthographic patterns interact with phonology and semantics is not captured by this model in its current version. It is also well-known that individuals process more than one character and sometimes entire phrases in one glance. Implementing this sort of window approach—e.g., using windowed attentional mechanisms (Vaswani et al., 2017)—combined with phonology and semantics into the CLOUD architecture could further boost its validity. A natural next step for the current CLOUD model is to examine how the model responds to the lexical properties of words (e.g., frequency, orthographic neighborhood, length) to validate it as a multilingual visual word recognition model.

5.3. Practical implications

Although my purpose was not to inform language educational practices, the results of this work offer some practical applications. Recent work has elaborated on how the organization of the multilingual mental lexicon facilitates vocabulary learning (for a review, see Baxter et al., 2021). Because of this, I will only briefly outline some of how this work could inform foreign language instruction:

It is essential to consider the learner's linguistic background. The particularities of known languages, and similarities and differences with the target foreign language, can implicitly or explicitly be exploited to maximize language learning achievement.

Experience with two languages facilitates vocabulary learning. This statement does not mean to discourage monolingual learners. Instead, the essential aspect to consider is that learning a foreign language can facilitate further vocabulary acquisition. Surprisingly, even passive exposure to a non-native language seems to confer vocabulary learning advantages in monolinguals (Bice & Kroll, 2019).

Objective and perceived similarity can facilitate and hinder language learning. For instance, as shown in Experiment 7, teaching lists of very similar words might hinder rather than facilitate their recognition. In this regard, and as suggested by other work (Baxter et al., 2021; Marian et al., 2021), it is vital to balance similar and dissimilar vocabulary instruction.

Acquiring multiple mappings is harder than one-to-one mappings. To avert the difficulties of acquiring synonym and homonyms, it is perhaps best to space their presentation or present them in variable contexts to facilitate their learning (Benitez et al., 2016).

Consistent practice yields the best results. The participants in my experiments could learn constructions with only a few repetitions, but it is unlikely they could retain them. In Experiments 4 and 6, participants learned to recognize most of the words and produce over half of them within a 40 minutes session of repeated implicit exposure or explicit practice.

5.4. Strengths and Limitations

The fundamental strength of my approach to the research question and hypothesis in this work is that I probed it from multiple angles. Admittedly, all these experiments are only a pixel in the broader picture of language learning and learning in general. Furthermore, future endeavors could part from the basis of these experiments to further probe the idea of monolingual and bilingual language learning. Since the limitations of each experiment are highlighted within their respective chapters, here I offer some points that limit the generalizability of my work:

Recognition versus lexicalization. I have mainly used recognition (and sometimes production) tasks to measure participants' learning outcomes throughout this work. These tasks have several methodological and theoretical limitations. For instance, performance in these tasks is bounded by the number of test items, and longer recognition tests do not address the issues present in these simple decision tasks. Moreover, these tasks do not address the lexicalization of the items presented. In other words, I cannot disentangle whether participants integrated the constructions in their lexicons or simply memorized them. The difference is that when an item is lexicalized, it begins competing for selection with other items within the lexicon (Baxter et al., 2021). Consequently, while I can certainly talk about differences in performance on these tasks, I cannot claim that participants truly learned the constructions.

Initial versus long-term learning. My experiments only targeted the initial exposure to artificial constructions. It is certainly theoretically and practically important to understand whether bilinguals and monolinguals differ on initial exposure. However, my findings are temporally bounded to the experimental context, and I cannot argue anything about long-term language learning differences. Language learning is a life-long process wherein individuals consistently acquire and discover new constructions. In this regard, future studies could certainly target learning over multiple sessions or even controlled classroom environments.

Artificial versus natural languages. The use of artificial linguistic materials in my experiments reduced the influence of nuisance factors such as prior experience with the stimuli or direct transfer effects. Compared to the nuances of natural languages, artificially constructed materials are limited in their scope and usage. It is possible that, despite telling participants they would be learning a new language, they were primed to employ memory processes rather than contextualize the artificial materials using the structure and nuances of a natural language. This issue can be certainly addressed using a carefully constructed artificial language. However, even an unknown natural language could suffice to study the language learning process using more ecologically valid settings.

Manipulating proficiency. A strength of my work is that I carefully controlled for participants' linguistic profiles to make the results comparable across experiments. At the same time, this disallowed me from testing bilingualism as a continuum. In other words, it is possible that, since participants vary in their proficiency and experience with the language, there might be differences even within the bilingual groups. Operationalizing and measuring bilingual experiences as a unitary construct (i.e., a bilingual quotient) is an ongoing question (Marian & Hayakawa, 2021). Nevertheless, future studies could benefit from treating bilingualism as a continuum rather than a dichotomy.

5.5. Outstanding questions

There are many ways in which this work opens new interesting questions about foreign language learning, language experience, and even learning as a cognitive and neural mechanism:

Foreign language learning. When exactly is a foreign language construction learned? What are the neural and computational bases of foreign language learning? What are the language learning strategies employed by participants with different linguistic profiles?

Language experience. Do monolingual and bilingual learning trajejctories differ through becoming proficient in a foreign language? Do they differ in learning other types of materials? At which stage of second language learning do individuals become proficient enough to show a vocabulary learning advantage?

Learning. What does it mean to "learn"? How does the brain incorporate novel knowledge into pre-existent networks? Is there a domain-general learning mechanism, or is it domain-specific? How can the principles of efficient encoding and retrieval be implemented into computational models? What is the exact code employed by neuros to learn?

Conclusion

Are adults who already know two languages (bilinguals) better at learning a foreign language than those who only know one (monolinguals)? If I consider foreign language learning as a whole, the answer to my central question is a resounding no. Instead, what I can conclude is that (1) bilinguals and monolinguals do not differ in general foreign language learning, but rather in specific analytic levels; (2) experience with specific properties of one or more languages (i.e., word order) can benefit foreign language learning; (3) similarities can be exploited to facilitate language learning; (4) bilinguals seem to be consistently be better at learning vocabulary, more so than other aspects of a foreign language; (5) this advantage could be rooted in how information is learned and organize within the monolingual and bilingual mental lexicon.

My results primarily suggest that bilinguals might be better at learning vocabulary during the initial stages of foreign language learning. Critically, vocabulary is an essential aspect of a foreign language and a prerequisite for other more abstract and complex analytic levels. Likewise, general proficiency tests that measure writing, speaking, reading, and listening depend on vocabulary knowledge. Therefore, it would not be surprising if individuals that are quicker to acquire vocabulary can develop their proficiency faster in a foreign language. However, there is a significant leap from this thesis' findings to overall achievement in a foreign language. Language learning requires substantial practice and exposure over extended periods. While I have shown that previous linguistic experiences can facilitate this process, learning achievement—in language or otherwise—ultimately depends on individual factors such as motivation, age, learning strategies, and time devoted to learning.

Lastly, I believe it is essential to consider that monolinguals are a dying race. Knowing at least two languages is now the norm rather than the exception in the current globalized world. Comprehending how language learning and processing differ due to distinct language experiences can be theoretically and practically useful. Yet, it is more important to focus not on advantages or differences but on the outstanding capacity of the human brain to efficiently and flexibly accommodate and operate in multiple languages.

List of publications derived from this thesis

Aguasvivas, J.A. & Carreiras, M. (Under Review). Does bilingual experience influence statistical language learning? *Cognition.*

Aguasvivas, J.A., Zorzi, M., Testolin, A., & Carreiras, M. (Submitted). Words in the CLOUD: How orthographic similarity and bilingual experience facilitate foreign vocabulary learning. *Psychological Review.*

Aguasvivas, J.A., Carreiras, M. (Under Review). Bilingualism, foreign language learning, and cognition: Insights for education. In *Cognitive Sciences and Education in Non-WEIRD Populations - A Latin American Perspective.* Springer Nature.

Resumen en Castellano

Casi todos hemos estado en esta situación, sentados en un aula, tratando de aprender palabras en japonés, francés, inglés, o cualquier otro idioma. Los individuos tienen que aprender un vocabulario de varios miles de palabras, cómo pronunciarlas, escribirlas—a veces en un sistema totalmente distinto--, y combinarlas para formar oraciones y comunicarse de manera correcta. Aprender un nuevo idioma es, sin duda, una experiencia desafiante pero gratificante. Los seres humanos somos únicos en nuestra habilidad para aprender no sólo uno, sino múltiples idiomas a lo largod e nuestras vidas. Sin embargo, es posible que aprender nuevos idiomas se haga más fácil mientras más idiomas sabemos.

Esta observación me llevó a la pregunta que motiva esta tesis: *Son los adultos que ya hablan dos idiomas (bilingües) mejores al aprender un nuevo idioma que aquellos que sólo hablan uno (monolingües)?* Intuitivamente, conocer dos idiomas provee a los individuos con un conocimiento más extenso que pueden utilizar a la hora de aprender un tercer idioma, en comparación con hablar un solo idioma. No obstante, esta tesis se enfoca en habilidades que van más allá del conocimiento o similitudes que los aprendices pueden utilizar a la hora de estudiar un nuevo idioma. En otras palabras, si todas las condiciones son lo más parecidas posible, son los bilingües inherentemente mejores que los monolingües para aprender un nuevo idioma?

Como sucede con cualquier pregunta de investigación, existe un largo trecho desde la concepción hasta las metodologías que mejor acaparan dicha pregunta. Por esto, en el Capítulo 1 de esta tesis, abarco lo que significa aprender un nuevo idioma. En resumen, los idiomas tienen distintos niveles de análisis que van desde aspectos sub-léxicos (ortografía y fonología) hasta otros más abstractos como pragmática. Aprender un nuevo idioma involucre adquirir elementos de uno o varios niveles. Este aprendizaje puede ser implícito, si el individuo no está consciente de que está adquiriendo los elementos, o explícito, si existe un esfuerzo por parte del aprendiz en captar y memorizar el contenido. Entonces, para responder si son mejores los bilingües que los monolingües al aprender un nuevo idioma, debemos saber en qué nivel lingüistico son mejores (**“dónde”**), cuáles son las diferencias (**“qué”**), y cómo pueden surgir estas diferencias de la experiencia con dos idiomas (**“cómo”**).

Asimismo, es necesario tomar en consideración lo que significa ser bilingüe. Las personas bilingües adquiren un segundo idioma en edades tempranas o tardías, consiguiendo un nivel de dominio de ambos idiomas que puede ser similar (balanceado), o distinto (desbalanceado). Para esta tesis, tomé en consideración a personas con un alto nivel en ambos idiomas, hablantes castellano-inglés, y castellano-vasco. Ambos grupos fueron comparados entre sí y contra un grupo de hablantes monolingües castellanoparlantes.

En el Capítulo 1 también reviso la evidencia a favor de que los bilingües pueden aprender idiomas mejor que los monolingües. En síntesis, algunos estudios indican que los bilingües son mejores a la hora de aprender construcciones en distintos niveles. Sin embargo, el nivel más estudiado ha sido el léxico-semántico (vocabulario). Por ello, en la parte experimental de esta tesis, desarrollé experimentos para comparar los tres grupos de participantes mencionados anteriormente utilizando materiales lingüísticos artificiales y experimentos de aprendizaje implícito (Capítulo 2) y explícito (Capítulo 3).

El Capítulo 2 trabaja el aprendizaje implícito de idiomas a través de cuatro experimentos conocidos de *aprendizaje estadístico*. Estos experimentos estaban dirigidos al aprendizaje de elementos sub-léxicos fonológicos, morfológicos, sintácticos, y léxico-semánticos, respectivamente. En el Experimento 1, comparé la habilidad de bilingües y monolingües al segmentar palabras de diferentes streams de voz artificiales que variaban en función de los patrones sub-léxicos fonológicos (i.e., fonotáctica). Los resultados de este experimento indicaron que, a pesar de que una fonotáctica más compleja reduce el rendimiento en este tipo de tareas, no existen diferencias entre los monolingües y bilingües en su desdempeño en esta tarea.

El Experimento 2 extiende los hallazgos del experimento previo al medir la habilidad de los participantes de generalizar el contenido aprendido a través del aprendizaje estadístico a palabras que no habían visto. Específicamente, diseñé el stream de voz artificial para simular la morfología de afijos (ej., **i**rrompi**ble**, **in**toca**ble**), y los participantes tenían que generalizar el conocimiento a nuevos ítemes (ej., **in**venci**ble**). Al igual que en el Experimento 1, aunque los participantes desempeñaron por encima del nivel de azar, no evidencié ninguna diferencia entre los grupos, y por lo tanto no había una ventaja bilingüe.

El Experimento 3 trabaja el nivle sintáctico. En este experimento, medía la capacidad de los participantes al segmentar palabras de un stream artificial ambiguo que podía ser segmentado basado una propiedad sintácticas de idiomas conocidos: el orden de las palabras. Los resultados indicaron que los hablantes de castellano-vasco desempeñaron esta tarea mejor que los otros dos grupos (castellano y castellano-inglés) en esta tarea. En otras palabras, ellos podían segmentar las palabras de una manera congruente con el orden de palabras predominante del vasco (sujeto-objeto-verbo). No obstante, como en los dos experimentos anteriores, no encontré diferencias entre bilingües y monolingües más allá de las que se explican por las propiedades de los idiomas que conocen. Este experimento fue vital para separar los efectos de la experiencia bilingüe de aquellos que surgen debido a las propiedades de los idiomas conocidos.

En el Experimento 4, trabajé el nivel léxico-semántico utilizando una tarea de aprendizaje estadístico en la que los participantes aprendieron los nombres de objetos desconocidos. Esta tarea es conocida como cross-situational statistical learning. La manipulación en este experimento fue la relación entre los nombres y objetos. Algunos objetos solo tenían un nombre, y por lo tanto eran exclusivos. Otros objetos podían tener dos nombres (sinónimos), y algunos nombres podrían referirse a dos objetos (homónimos). Al enfrentarse a dos idiomas, los individuos aprendern diferentes tipos de relaciones entre nombres y objetos, por lo que mi predicción era que los bilingües desempeñarían mejor que los monolingües en los tres tipos de relaciones. Sin embargo, los resultados indicaron que los bilingües sólo desempeñaron mejor que los monolingües en aprender las relaciones exclusivas entre nombres y objetos, y no había diferencias entre los grupos al aprender otros tipos de relaciones múltiples.

El Capítulo 3 se enfoca en el aprendizaje explícito de nuevos idiomas y utiliza dos experimentos adicionales que trabajan el aprendizaje morfológico y léxico-semántico. El Experimento 5 comparaba a los monolingües y bilingües al aprender nuevos sufijos para raíces conocidas del castellano (ej., laboral**suti**). Como los participantes ya conocían la definición de dichas raíces, el experimento se enfocaba fundamentalmente en el aprendizaje de la forma ortográfica de los sufijos. Los resultados indicaron que, pese a que los participantes aprendieron a discriminar los sufijos nuevos de otras palabras distractoras, los grupos no difirieron en su rendimiento en esta tarea. En otras palabras, así como en el Experimento 2, no hallé diferencias entre monolingües y bilingües al aprender información morfológica. Además, un análisis correlacional indicó correlaciones débiles entre los puntatjes de los participantes en los experimentos 2 y 5.

El Experimento 6 combinaba la ortografía sub-léxica y el nivel léxico-semántico. En este experimento, los participantets tenína que aprender un vocabulario artificial llamado Flaviano a través de cinco bloques de reconocimiento y producción. La manipulación de este experimento consistió en que el vocabulario contenía palabras que eran ortográficamente más similares al castellano (ES+) o más disímiles del castellano (ES-)—siendo este idioma el común para todos los participantes. Este experimento sugirió tres resultados importantes. Primero, la similitude ortográfica facilitó el aprendizaje de vocabulario escrito a través de los distintos bloques de aprendizaje, consistente con otros hallazgos de la literatura. Segundo, los bilingües desempeñaron mejor que los monolingües al aprender el vocabulario, independientemente de su similitud con el castellano. Este hallazgo extiende el trabajo experimental previo sugiriendo que los bilingües aprenden vocabulario similar y disímil mejor que los monolingües. Tercero, las diferencias entre monolingües y bilingües eran más pronunciadas para las palabras ES+ que las ES- durante los bloques de reconocimiento, pero las diferencias desaparecieron al probar su conocimiento luego de un tiempo. Por el contrario, la similitud ortográfica y la experiencia bilingüe no interactuaron durante la tarea de producción, siendo los bilingües mejores en ambos tipos de palabras.

Finalmente, el Capítulo 4 se enfoca en comprender cómo las diferencias en el Experimento 6 surgen a base del input bilingüe. Pese a varios modelos computacionales enfocados en el léxico mental de monolingües y bilingües, ninguno de estos modelos puede explicar los resultados de este último experimento. Por esto, en el Experimento 7, propuse y desarrollé un modelo computacional enfocado en el léxico ortográfico, llamado CLOUD. Este modelo fue entrenado con input monolingüe o bilingüe (castellano-inglés o castellano-vasco). Utilicé este modelo para examinar cómo la similitud ortográfica y la experiencia con input bilingüe pueden influenciar el aprendizaje de vocabulario. Entre otros hallazgos, el Experimento 7 indicó que este modelo puede simular los efectos de similitud ortográfica y experiencia bilingüe observados en el Experimento 6. Este experimento unifica los hallazgos dispares de similitud ortográfica y experiencia bilingüe bajo un mismo marco computacional, en el que representaciones distribuídas de las palabras residen en un léxico unificado y son modificadas por las experiencias de aprendizaje.

Los hallazgos de todos estos experimentos permiten responder a las tres preguntas especificadas al inicio (dónde, qué, y cómo). En primer lugar, los hallazgos de esta tesis indicaron que los monolingües y bilingües no difieren en aprender elementos de los niveles sub-léxico, morfológico (implícito o explícito), y sintáctico. Sin embargo, sí difieren en el aprendizaje implícito y explícito en el nivel léxico-semántico, donde ambos grupos de bilingües desempeñaron mejor que el grupo de monolingües al aprender vocabulario, pero no difirieron el entre ellos. Basado en estos resultados, mi respuesta a la pregunta de **"dónde"** es que los monolingües y bilingües parecen diferir primordialmente en el aprendizaje de vocabulario (nivel léxico-semántico).

En segundo lugar, las manipulaciones experimentales de los Experimentos 4 y 6 dan respuesta a la pregunta de **"qué"**, sugiriendo que los bilingües son mejores al aprender relaciones exclusivas entre nuevas palabras y los objetos a los que se refieren. Además, esta ventaja de los bilingües al aprender vocabulario parece ser más evidente en tareas de producción que en tareas de reconocimiento y es independiente de la similitud de las palabras con idiomas conocidos. Esta ventaja también parece desarrollarse a través del curso del aprendizaje, pero no se manifiesta necesariamente durante los tests realizados luego de un tiempo.

Las simulaciones del Experimento 7 dan respuesta a la pregunta de **"cómo"**. En concreto, el modelo CLOUD es susceptible a la similitud ortográfica, y, al entrenarse con input bilingüe, desempeña mejor que la contraparte monolingüe, independientemente de la similitud ortográfica. En otras palabras, la ventaja de los bilingües al aprender vocabulario puede surgir como consecuencia de su exposición (pasiva o activa) a palabras en dos idiomas. Es possible que los bilingües desarrollen representaciones ortográficas de las palabras que son más flexibles para integrar nuevo vocabulario que las de los monolingües. Parece ser que la experiencia con palabras en dos idiomas actúa como una regularización. La regularización es un concepto relativamente común en teorías cognitivas y computacionales del aprendizaje. En términos computacionales, la regularización implica constreñir las representaciones a través de términos de penalidad o variabilidad en los datos para reducir los errores de generalización en el aprendizaje automático. En humanos, este proceso se refiere a cómo los aprendizes adaptan su producción (y representacions) para reducir la entropía en la información que procesan. En este sentido, mis resultados demuestran que este proceso puede ocurrir en niveles sub-léxicos cuando los patrones ortográficos son menos predecibles—al provenir de dos idiomas distintos.

En fin, mi respuesta a la pregunta original de si los adultos que ya hablan dos idiomas (bilingües) son mejores al aprender un nuevo idioma que aquellos que sólo hablan uno (monolingües) es un rotundo no si considero el aprendizaje de idiomas como un todo. En cambio, lo que puedo concluir es que (1) los bilingües y monolingües no difieren en el aprendizaje de idiomas en general, sino en niveles específicos; (2) la experiencia con propiedades específicas de los idiomas (ej., orden de palaras) puede beneficiar el aprendizaje de idiomas; (3) las similitudes también pueden ser utilizadas para facilitar el aprendizaje de vocabulario; (4) los bilingües parecen ser consistentemente mejores al aprender vocabulario, más que en cualquier otro aspecto del aprendizaje de idiomas; (5) esta ventaja parece estar relacionada con cómo la información es aprendida y organiziada en el léxico mental de los bilingües en comparación con los monolingües.

REFERENCES

Acha, J., & Carreiras, M. (2014). Exploring the mental lexicon: A methodological approach to understanding how printed words are represented in our minds. *The Mental Lexicon*. https://doi.org/10.1075/ml.9.2.03ach

Acha, J., Laka, I., Landa, J., & Salaburu, P. (2014). EHME: A new word database for research in basque language. *Spanish Journal of Psychology*. https://doi.org/10.1017/sjp.2014.79

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. In *Review of Educational Research*. https://doi.org/10.3102/0034654310368803

Aguasvivas, J. A., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E., & Duñabeitia, J. A. (2018). SPALEX: A Spanish Lexical decision database from a Massive Online Data Collection. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.02156

Aguasvivas, J. A., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E., & Duñabeitia, J. A. (2020). How do Spanish speakers read words? Insights from a crowdsourced lexical decision megastudy. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01357-9

Aitchison, J. (2012). Words in the Mind: An Introduction to the Mental Lexicon. In *Language*. John Wiley & Sons. https://doi.org/10.2307/416662

Alonso, R. A. (2016). Cross-linguistic influence in second language acquisition. In *Multilingual Matters*. https://doi.org/10.1002/9781405198431.wbeal0292

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. C. M. (2015). The bilingual advantage in phonetic learning. *Bilingualism*, *18*(4), 683. https://doi.org/10.1017/S1366728914000777

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, *22*(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55–65. https://doi.org/10.1111/j.2044-8260.1993.tb01027.x

Bakker-Marshall, I., Takashima, A., Fernandez, C. B., Janzen, G., McQueen, J. M., & Van Hell, J. G. (2021).

Overlapping and distinct neural networks supporting novel word learning in bilinguals and monolinguals. *Bilingualism*, *24*(3), 524–536. https://doi.org/10.1017/S1366728920000589

Bardel, C., & Falk, Y. (2012). Behind the L2 status factor : A neurolinguistic framework for L3 research. In *Third Language Acquisition in Adulthood* (pp. 61–78).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartolotti, J., & Marian, V. (2017). Bilinguals' Existing Languages Benefit Vocabulary Learning in a Third Language. *Language Learning*. https://doi.org/10.1111/lang.12200

Bartolotti, J., Marian, V., Schroeder, S. R., & Shook, A. (2011). Bilingualism and inhibitory control influence statistical learning of novel word forms. *Frontiers in Psychology*, *2*, 324. https://doi.org/10.3389/fpsyg.2011.00324

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

Batterink, L. J., Westerberg, C. E., & Paller, K. A. (2017). Vocabulary learning benefits from REM after slow-wave sleep. *Neurobiology of Learning and Memory*, *144*, 102–113. https://doi.org/10.1016/j.nlm.2017.07.001

Baxter, P., Bekkering, H., Dijkstra, T., Droop, M., van den Hurk, M., & Léoné, F. (2021). Grounding Second Language Vocabulary Instruction in Cognitive Science. *Mind, Brain, and Education*, 1–11. https://doi.org/10.1111/mbe.12278

Bengtson, J. (2011). The Basque Language: History and Origin. *International Journal of Modern Anthropology*, *1*(4), 43–59. https://doi.org/10.4314/ijma.v1i4.3

Benitez, V. L., Yurovsky, D., & Smith, L. B. (2016). Competition between multiple words for a referent in cross-situational word learning. *Journal of Memory and Language*, *90*, 31–48. https://doi.org/10.1016/j.jml.2016.03.004

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive Control and Lexical Access in Younger and Older Bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics*, *35*, 177–191. https://doi.org/10.1017/S0142716412000288

Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport*. https://doi.org/10.1097/WNR.0000000000000453

Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*. https://doi.org/10.1016/j.bandl.2019.104644

Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. *Journal of Memory and Language*, *44*, 235–249. https://doi.org/10.1006/jmla.2000.2750

Blanco-Elorrieta, E., & Pylkkänen, L. (2018). Ecological Validity in Bilingualism Research and the Bilingual Advantage. *Trends in Cognitive Sciences*, *22*(12), 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001

Borragan, M., Casaponsa, A., Antón, E., & Duñabeitia, J. A. (2020). Incidental changes in orthographic processing in the native language as a function of learning a new language late in life. *Language, Cognition and Neuroscience*, 1–10. https://doi.org/10.1080/23273798.2020.1784446

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016a). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01116

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016b). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2016.01116

Cabrelli Amaro, J., Flynn, S., & Rothman, J. (2012). *Third language acquisition in adulthood*. https://doi.org/10.1075/sibil.46.02ama

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.11.005

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: The impact of orthographic markedness. *Frontiers in Psychology*, *5*, 424. https://doi.org/10.3389/fpsyg.2014.00424

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2015). How do bilinguals identify the language of the words they read? *Brain Research*, *1624*, 153–166. https://doi.org/10.1016/j.brainres.2015.07.035

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. In *International Journal of Bilingualism*. https://doi.org/10.1177/13670069030070010501

Cenoz, J. (2013). The influence of bilingualism on third language acquisition: Focus on multilingualism. *Language Teaching*. https://doi.org/10.1017/S0261444811000218

Chan, J., & Monaghan, P. (2019). Simulating Bilingual Word Learning: Monolingual and Bilingual Adults' Use of Cross-Situational Statistics. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Chen, T., & Hartshorne, J. K. (2021). More evidence from over 1.1 million subjects that the critical period for syntax closes in late adolescence. *Cognition*. https://doi.org/10.1016/j.cognition.2021.104706

Christiansen, M. H. (2019). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, *5*(2), 82–88. https://doi.org/10.1016/S1364-6613(00)01600-4

Chuang, Y. Y., Bell, M. J., Banke, I., & Harald Baayen, R. (2021). Bilingual and Multilingual Mental Lexicon: A Modeling Study With Linear Discriminative Learning. *Language Learning*, *71*(S1), 219–292. https://doi.org/10.1111/lang.12435

Church, K. W. (2017). Word2Vec. *Natural Language Engineering*, *23*(1), 155–162. https://doi.org/10.1017/s1351324916000334

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Clyne, M., Hunt, C. R., & Isaakidis, T. (2004). Learning a community language as a third language. *International Journal of Multilingualism*, *1*(1), 33–52. https://doi.org/10.1080/14790710408668177

Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition*, *44*(3), 197–226. https://doi.org/10.1016/0010-0277(92)90001-X

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? In *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn3709

Council of Europe. (2001). The Common European Framework of Reference for Languages. *Europe*.

Cox, J. G. (2017). Explicit instruction, bilingualism, and the older adult learner. *Studies in Second Language Acquisition*, *39*(1), 29–58. https://doi.org/10.1017/S0272263115000364

de Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology*, *8*, 522. https://doi.org/10.3389/fpsyg.2017.00522

de Bruin, A., Dick, A. S., & Carreiras, M. (2021). Clear Theories are Needed to Interpret Differences: Perspectives on the Bilingual Advantage Debate. *Neurobiology of Language*, 1–46. https://doi.org/10.1162/nol_a_00038

de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive Advantage in Bilingualism: An Example of Publication Bias? *Psychological Science*. https://doi.org/10.1177/0956797614557866

de Groot, A. M. B. (1992). Bilingual Lexical Representation: A Closer Look at Conceptual Representations. *Advances in Psychology*. https://doi.org/10.1016/S0166-4115(08)62805-8

De Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. https://doi.org/10.1080/713755891

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*. https://doi.org/10.1111/0023-8333.00110

Dehaene, S., Le Clec'H, G., Poline, J. B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport*. https://doi.org/10.1097/00001756-200203040-00015

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*.

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No bilingual advantage for executive function: Evidence from a large sample of children in the Adolescent Brain and Cognitive Development (ABCD) Study. *Nature Human Behaviour*, *3*(7), 692–701.

Dijkstra, T. (2012). The multilingual lexicon. In *The Oxford Handbook of Psycholinguistics*. https://doi.org/10.1093/oxfordhb/9780198568971.013.0015

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In *Localist connectionist approaches to human cognition*.

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/s1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019a). Modelling bilingual lexical processing: A research agenda and desiderabilia. In *Bilingualism*. https://doi.org/10.1017/S1366728918000986

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019b). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism*. https://doi.org/10.1017/S1366728918000287

Donnelly, S., Brooks, P. J., & Homer, B. D. (2015). Examining the Bilingual Advantage on Conflict Resolution Tasks: A Meta-Analysis. *CogSci*.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*. https://doi.org/10.3758/s13428-013-0326-1

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly*

Journal of Experimental Psychology (2006). https://doi.org/10.1080/17470218.2017.1310261

Dutoit, T., Pagel, V., Pierret, N., Bataille, E., & van der Vrecken, O. (1996). MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. *Proceeding of Fourth International Conference on Spoken Language Processing, ICSLP*, 1393–1396. https://doi.org/10.1109/icslp.1996.607874

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism*. https://doi.org/10.1017/S1366728913000722

Ellis, N. C. (1993). Rules and Instances in Foreign Language Learning: Interactions of Explicit and Implicit Knowledge. *European Journal of Cognitive Psychology*, *5*(3), 289–318. https://doi.org/10.1080/09541449308520120

Ellis, N. C. (1998). Emergentism, connectionism and language learning. In *Language Learning*. https://doi.org/10.1111/0023-8333.00063

Ellis, N. C. (2002). Frequency effects in language processing: A Review with Implications for implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Ellis, N. C. (2015). *Implicit AND explicit language learning*. https://doi.org/10.1075/sibil.48.01ell

Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, And Grammatical Structure. *Machine Learning*. https://doi.org/10.1023/A:1022699029236

Endress, A. D., & Bonatti, L. L. (2016). Words, rules, and mechanisms of language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(1), 19–35. https://doi.org/10.1002/wcs.1376

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Escudero, P., Mulak, K. E., Fu, C. S. L., & Singh, L. (2016). More limitations to monolingualism: Bilinguals outperform monolinguals in implicit word learning. *Frontiers in Psychology*, *7*, 1218. https://doi.org/10.3389/fpsyg.2016.01218

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. M. (2016). The Relationship Between Artificial and Second Language Learning. *Cognitive Science*, *40*(4), 822–847. https://doi.org/10.1111/cogs.12257

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Festman, J. (2021). Learning and Processing Multiple Languages: The More the Easier? *Language Learning*, *71*(S1), 121–162. https://doi.org/10.1111/lang.12437

Filippi, R., D'Souza, D., & Bright, P. (2019). A developmental approach to bilingual research: The effects of multi-language experience from early infancy to old age. *International Journal of Bilingualism*, *23*(5), 1195–1207. https://doi.org/10.1177/1367006917749061

Finkbeiner, M., & Coltheart, M. (2009). Letter recognition: From perception to representation. *Cognitive Neuropsychology*, *26*(1), 1–6. https://doi.org/10.1080/02643290902905294

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499. https://doi.org/10.1016/j.cognition.2008.04.002

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *International Journal of Multilingualism*, *1*(1), 3–16. https://doi.org/10.1080/14790710408668175

Foote, R. (2009). Transfer in L3 acquisition: The role of typology. In *Third Language Acquisition and Universal Grammar*.

French, R. M. (1998). A simple recurrent network model of bilingual memory. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*.

French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Sciences*, *8*(2), 87–93. https://doi.org/10.1016/j.tics.2003.12.011

Frost, Ram, Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Frost, Rebecca, & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74.

https://doi.org/10.1016/j.cognition.2015.11.010

García Mayo, M. D. P., & Villarreal Olaizola, I. (2011). The development of suppletive and affixal tense and agreement morphemes in the L3 english of basque-spanish bilinguals. *Second Language Research*, *27*(1), 129–149. https://doi.org/10.1177/0267658310386523

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0806530105

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese-Italian cross-linguistic study. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2007.12.001

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, *9*, 249–256. https://doi.org/10.1.1.207.2059

Gollan, T. H., & Acenas, L. A. R. (2004). What Is a TOT? Cognate and Translation Effects on Tip-of-the-Tongue States in Spanish-English and Tagalog-English Bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.30.1.246

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism*, *15*, 594–615. https://doi.org/10.1017/S1366728911000332

González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*. https://doi.org/10.1093/applin/amy057

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://doi.org/10.1016/B978-0-12-391420-0.09987-X

Grainger, J., & Dijkstra, T. (1992). On the Representation and Use of Language Information in Bilinguals. *Advances in Psychology*, *83*, 207–220. https://doi.org/10.1016/S0166-4115(08)61496-X

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., & Fagot, J. (2012). Orthographic processing in baboons (Papio papio). *Science*, *336*(6078), 245–248. https://doi.org/10.1126/science.1218152

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1996.0033

Grainger, J., & Jacobs, A. M. (1996). Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model. *Psychological Review*. https://doi.org/10.1037/0033-295X.103.3.518

Grainger, J., Manjaly, K., & Holcomb, P. J. (2008). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In *Langauge Acquisition across Linguistic and Cognitive Systems*.

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In *Language Acquisition across Linguistic and Cognitive Systems* (pp. 267–283). https://doi.org/10.1075/lald.52.18gra

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology*, *22*(6), 197–206. https://doi.org/10.1016/j.cub.2012.02.012

Grey, S., Sanz, C., Morgan-Short, K., & Ullman, M. T. (2018). Bilingual and monolingual adults learning an additional language: ERPs reveal differences in syntactic processing. *Bilingualism*, *21*(5), 970–994. https://doi.org/10.1017/S1366728917000426

Harald Baayen, R., Chuang, Y. Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning. *Complexity*. https://doi.org/10.1155/2019/4895891

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Hawkins, J., Lewis, M., Klukas, M., Purdy, S., & Ahmad, S. (2019). A framework for intelligence and cortical function based on grid cells in the neocortex. *Frontiers in Neural Circuits*. https://doi.org/10.3389/fncir.2018.00121

Hayakawa, S., Bartolotti, J., & Marian, V. (2020). Native Language Similarity during Foreign Language Learning: Effects of Cognitive Strategies and Affective States. *Applied Linguistics*.

Hayakawa, S., Ning, S., & Marian, V. (2019). From Klingon to Colbertian: Using Artificial Languages to Study Word Learning. *Bilingualism*, *23*(1), 74. https://doi.org/10.1017/S1366728919000592

Hermas, A. (2014). Restrictive relatives in L3 english: L1 Transfer and ultimate attainment convergence. *Australian Journal of Linguistics*, *34*(3), 361–387. https://doi.org/10.1080/07268602.2014.898227

Hickey, R. (2005). Legacies of colonial English: Studies in transported dialects. In *Legacies of Colonial English: Studies in Transported Dialects*. https://doi.org/10.1017/CBO9780511486920

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin and Review*, *25*(3), 892–916. https://doi.org/10.3758/s13423-017-1315-7

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*. https://doi.org/10.1162/neco.1997.9.8.1735

Holman, D., & Spivey, M. J. (2016). Connectionist models of bilingual word reading. In *Methods in Bilingual Reading Comprehension Research*. https://doi.org/10.1007/978-1-4939-2993-1_9

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 442–451. https://doi.org/10.1007/978-3-540-39432-7_47

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the spanish vocabulary size. *Psicológica*.

Jacquet, M., & French, R. M. (2002). The BIA++: Extending the BIA+ to a dynamical distributed connectionist framework. *Bilingualism: Language and Cognition*, *5*(3), 202–205. https://doi.org/10.1017/s1366728902223019

Jarvis, S., & Pavlenko, A. (2007). Cross-linguistic influence in language and cognition. In *Cross-linguistic Influence in Language and Cognition*. Routledge. https://doi.org/10.4324/9780203935927

Jessie, D. T., & Saari, D. G. (2016). From the Luce Choice Axiom to the Quantal Response Equilibrium. *Journal of Mathematical Psychology*, *75*, 3–9. https://doi.org/10.1016/j.jmp.2015.10.001

Jurafsky, D., & Martin, J. (2014). Speech and Language Processing. In *Speech and Language Processing.*

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kahn-Horwitz, J., Schwartz, M., & Share, D. (2011). Acquiring the complex English orthography: A triliteracy advantage? *Journal of Research in Reading*, *34*(1), 136–156. https://doi.org/10.1111/j.1467-9817.2010.01485.x

Kan, P. F., & Sadagopan, N. (2014). Novel word retention in bilingual and monolingual speakers. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.01024

Kaufman, A. S., & Kaufman, N. L. (2014). Kaufman Brief Intelligence Test, Second Edition. In *Encyclopedia of Special Education*. https://doi.org/10.1002/9781118660584.ese1325

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism*. https://doi.org/10.1017/S1366728911000472

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism*, *23*(5), 945–950. https://doi.org/10.1017/S1366728919000038

Kaushanskaya, M., & Marian, V. (2009a). Bilingualism Reduces Native-Language Interference During Novel-Word Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/a0015275

Kaushanskaya, M., & Marian, V. (2009b). The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review*, *16*(4), 705–710. https://doi.org/10.3758/PBR.16.4.705

Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-012-0271-5

Keuleers, E. (2013). vwr: Useful functions for visual word recognition reserach. *R Package*.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), 35–42. https://doi.org/10.1016/S0010-0277(02)00004-5

Klein, E. C. (1995). Second versus Third Language Acquisition: Is There a Difference? *Language Learning*, *45*(3), 419–466. https://doi.org/10.1111/j.1467-1770.1995.tb00448.x

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, *78*(9), 1464–1480. https://doi.org/10.1016/S0925-2312(98)00030-7

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*. https://doi.org/10.1080/20445911.2013.799170

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. In *Bilingualism*. https://doi.org/10.1017/S136672891000009X

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2010.05.001

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software*. https://doi.org/10.18637/jss.v082.i13

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible Parallel Distributed Processing model of Event-Related Potential word reading data. *Brain and Language*. https://doi.org/10.1016/j.bandl.2011.09.001

Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, Acceptability, and Probability: A Probabilistic View of Linguistic Knowledge. *Cognitive Science*, *41*(5), 1202–1241. https://doi.org/10.1111/cogs.12414

Le Godais, G., Linzen, T., & Dupoux, E. (2017). Comparing character-level neural language models using a lexical decision task. *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 125–130. https://doi.org/10.18653/v1/e17-2020

Lee, M.-W., de Groot, A. M. B., & Kroll, J. F. (1999). Tutorials in Bilingualism: Psycholinguistic Perspectives. *Language*. https://doi.org/10.2307/417531

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. *Advances in Psychology*. https://doi.org/10.1016/S0166-4115(02)80006-1

Li, P., & Grant, A. (2019). Scaling up: How computational models can propel bilingualism research forward. In *Bilingualism*. https://doi.org/10.1017/S1366728918000755

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *34*(3), 408–415. https://doi.org/10.1080/15326900701399905

Loucky, J. P. (2006). Maximizing vocabulary development by systematically using a depth of lexical processing taxonomy, CALL resources, and effective strategies. *CALICO Journal*. https://doi.org/10.1558/cj.v23i2.362-399

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(1), 1–17. https://doi.org/10.18637/jss.v088.i02

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60). https://doi.org/10.21105/joss.03139

Mady, C. (2014). Learning French as a second official language in Canada: Comparing monolingual and bilingual students at Grade 6. *International Journal of Bilingual Education and Bilingualism*, *17*(3), 330–344. https://doi.org/10.1080/13670050.2013.767778

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(8), 4398–4403. https://doi.org/10.1073/pnas.070039597

Makowski, D. (2018). The psycho Package: an Efficient and Publishing-Oriented Workflow for Psychological Science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01272-8

Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Foryś-Nogala, M., Kutyłowska, K., & Wodniecka, Z. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition*, *206*(September 2020), 104477. https://doi.org/10.1016/j.cognition.2020.104477

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND. *PLoS ONE*.

Marian, V., Bartolotti, J., van der Berg, A., & Hayakawa, S. (2021). Costs and Benefits of Native Language Similarity for Non-native Word Learning. *Frontiers in Psychologyr*, *12*, 1906. https://doi.org/https://doi.org/10.3389/fpsyg.2021.651506

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

Marian, V., & Hayakawa, S. (2021). Measuring bilingualism: The quest for a bilingualism quotient. *Applied Psycholinguistics*, *42*(2), 527–548. https://doi.org/10.1017/S0142716420000533

Masters, R. S. W., van Duijn, T., & Uiga, L. (2020). Advances in implicit motor learning. In *Skill Acquisition in Sport* (p. 20). https://doi.org/10.4324/9781351189750-5

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111. https://doi.org/10.1016/S0010-0277(01)00157-3

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

McMahan, B., & Rao, D. (2019). Natural Language Processing with PyTorch - Build Intelligent Language Applications Using Deep Learning. In *O'Reilly Media*.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378. https://doi.org/10.1111/j.1467-7687.2009.00822.x

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology*, *64*(6), 1200–1220. https://doi.org/10.1080/17470218.2010.538211

Meyer, L. L., & Schmitt, N. (2002). Vocabulary in Language Teaching. *TESOL Quarterly*. https://doi.org/10.2307/3588334

Mihi, P. (2016). Third language acquisition among early bilinguals. In *PQDT - Global*.

Mikolov, T., Karafiát, M., Burget, L., Jan, C., & Khudanpur, S. (2010). Recurrent neural network based language model. *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*.

Mirkovic, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLoS ONE*, *11*(4), e0152489. https://doi.org/10.1371/journal.pone.0152489

Mirman, D. (2017). *Growth Curve Analysis and Visualization Using R*. CRC press. https://doi.org/10.1201/9781315373218

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Misyak, J., & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *29*(29).

Misyak, J., & Christiansen, M. H. (2012). Statistical Learning and Language: An Individual Differences Study. *Language Learning*, *62*(1), 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Modirkhamene, S. (2006). The reading achievement of third language versus second language learners of english in relation to the interdependence hypothesis. *International Journal of Multilingualism*, *3*(4), 280–295. https://doi.org/10.2167/ijm043.0

Montanari, S. (2019). Facilitated Language Learning in Multilinguals. In *Multidisciplinary Perspectives on Multilingualism* (pp. 302–324). https://doi.org/10.1515/9781501507984-014

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism*, *17*(1), 56–72. https://doi.org/10.1017/S1366728912000715

Moschkovich, J. (2007). Using two languages when learning mathematics. *Educational Studies in Mathematics*, *64*, 121–144. https://doi.org/10.1007/s10649-005-9005-1

Muñoz, C. (2008). Age-related differences in foreign language learning. Revisiting the empirical evidence. *IRAL - International Review of Applied Linguistics in Language Teaching*, *46*(3), 197–220. https://doi.org/10.1515/IRAL.2008.009

Nair, V. K. K., Biedermann, B., & Nickels, L. (2016). Consequences of late bilingualism for novel word learning: Evidence from Tamil–English bilingual speakers. *International Journal of Bilingualism*, *20*(4), 473–487. https://doi.org/10.1177/1367006914567005

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). Package "performance." *Journal of the Royal Society Interface*.

Nation, P. (2000). Learning Vocabulary in Lexical Sets: Dangers and Guidelines. *TESOL Journal*. https://doi.org/10.1002/j.1949-3533.2000.tb00239.x

Nation, R., & Mclaughlin, B. (1986). Novices and experts: An information processing approach to the "good language learner" problem. *Applied Psycholinguistics*, *7*(1), 41–55. https://doi.org/10.1017/S0142716400007177

Nikolov, M., & Djigunović, J. M. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, *26*, 234–260. https://doi.org/10.1017/S0267190506000122

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2013.08.003

Onnis, L., Chun, W. E., & Lou-Magnuson, M. (2018). Improved statistical learning abilities in adult bilinguals. *Bilingualism*, *21*(2), 427. https://doi.org/10.1017/S1366728917000529

Onnis, L., Frank, S. L., Yun, H., & Lou-Magnuson, M. (2016). Statistical learning bias predicts second-language reading efficiency. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society (CogSci 2016)*.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268–284. https://doi.org/10.1016/j.cognition.2012.10.008

Otwinowska, A., Foryś-Nogala, M., Kobosko, W., & Szewczyk, J. (2020). Learning Orthographic Cognates and Non-Cognates in the Classroom: Does Awareness of Cross-Linguistic Similarity Matter? *Language Learning*. https://doi.org/10.1111/lang.12390

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in PyTorch. *NIPS 2017 Autodiff Workshop: The Future of Gradient-Based Machine Learning Software and Techniques*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607. https://doi.org/10.1126/science.1072901

Perea, M., Fernández-López, M., & Marcet, A. (2020). What is the letter é? *Scientific Studies of Reading*, *24*(5), 434–443. https://doi.org/10.1080/10888438.2019.1689570

Perea, M., Urkia, M., Davis, C. J., Agirre, A., Laseka, E., & Carreiras, M. (2006). E-Hitz: A word frequency list and a program for deriving psycholinguistic statistics in an agglutinative language (Basque). *Behavior Research Methods*. https://doi.org/10.3758/BF03193893

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory and Cognition*. https://doi.org/10.3758/MC.36.7.1299

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

https://doi.org/10.1016/j.tics.2006.03.006

Pineño, O., & Miller, R. R. (2005). Primacy and recency effects in extinction and latent inhibition: A selective review with implications for models of learning. *Behavioural Processes*, *69*(2), 223–235. https://doi.org/10.1016/j.beproc.2005.02.006

Place, S., & Hoff, E. (2011). Properties of Dual Language Exposure That Influence 2-Year-Olds' Bilingual Proficiency. *Child Development*, *82*(6), 1834–1849. https://doi.org/10.1111/j.1467-8624.2011.01660.x

Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition*, *152*, 9–19. https://doi.org/10.1016/j.cognition.2016.03.001

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d /–/ð / perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190–2201. https://doi.org/10.1121/1.1362689

Ponsford, D., Wiggins, G., & Mellish, C. (1999). Statistical learning of harmonic movement. *International Journal of Phytoremediation*, *28*(2), 150–177. https://doi.org/10.1076/jnmr.28.2.150.3115

Poorebrahim, F., Afsharrad, M., & Ghonsooly, B. (2020). Bilingualism, monoliteracy, and third language writing: A case from Turkish-Persian context in Iran. *IIUM Medical Journal Malaysia*, *10*(2), 369–381. https://doi.org/10.17509/ijal.v10i2.28608

Portocarrero, J. S., Burright, R. G., & Donovick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*. https://doi.org/10.1016/j.acn.2007.01.015

Rafat, Y. (2016). Orthography-induced transfer in the production of English-speaking learners of Spanish. *Language Learning Journal*, *44*(2), 197–213. https://doi.org/10.1080/09571736.2013.784346

Rajalingham, R., Kar, K., Sanghavi, S., Dehaene, S., & DiCarlo, J. J. (2020). The inferior temporal cortex is a potential cortical precursor of orthographic processing in untrained monkeys. *Nature Communications*, *11*(1), 1–13. https://doi.org/10.1038/s41467-020-17714-3

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija*, *46*(4), 377–396. https://doi.org/10.2298/PSI1304377R

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*. https://doi.org/10.1038/4580

Rapp, B., & Caramazza, A. (1997). From Graphemes to Abstract Letter Shapes: Levels of Representation in Written Spelling. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 1130. https://doi.org/10.1037/0096-1523.23.4.1130

Rastle, K. (2015). Visual Word Recognition. In *Neurobiology of Language*. https://doi.org/10.1016/B978-0-12-407794-2.00021-3

Rastle, K., & Davis, M. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7–8), 942–971. https://doi.org/10.1080/01690960802069730

Rauch, D. P., Naumann, J., & Jude, N. (2012). Metalinguistic awareness mediates effects of full biliteracy on third-language reading proficiency in Turkish-German bilinguals. *International Journal of Bilingualism*, *16*(4), 402–418. https://doi.org/10.1177/1367006911425819

Ringbom, H. (2006). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters LTD. https://doi.org/10.1080/14790718.2011.552237

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Rosenshine, B. V. (1986). Synthesis of research on explicit teaching. *Educational Leadership*, *43*(7), 60–69.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*(1), 107–127. https://doi.org/10.1177/0267658310386439

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism*. https://doi.org/10.1017/S136672891300059X

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In *Backpropagation: Theory, architectures, and applications.* (pp. 1–34).

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*. https://doi.org/10.1017/S0261444819000053

Schroeder, S. R., & Marian, V. (2017). Cognitive consequences of trilingualism. *International Journal of Bilingualism*. https://doi.org/10.1177/1367006916637288

Schunk, D. H. (2012). Learning theories: An educational perspective. In *Reading*. Pearson.

Schwartz, M., Kahn-Horwitz, J., & Share, D. L. (2014). Orthographic learning and self-teaching in a bilingual and biliterate context. *Journal of Experimental Child Psychology*, *117*, 45–58. https://doi.org/10.1016/j.jecp.2013.08.008

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223. https://doi.org/10.1017/s0267190500200135

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*. https://doi.org/10.1111/j.0963-7214.2005.00372.x

Serratrice, L., Sorace, A., Filiaci, F., & Baldo, M. (2009). Bilingual children's sensitivity to specificity and genericity: Evidence from metalinguistic awareness. *Bilingualism*, *12*(2), 239–257. https://doi.org/10.1017/S1366728909004027

Sheng, L., Lu, Y., & Gollan, T. H. (2014). Assessing language dominance in Mandarin-english bilinguals: Convergence and divergence between subjective and objective measures. *Bilingualism*, *17*(2), 364. https://doi.org/10.1017/S1366728913000424

Shopen, T. (2007). Language typology and syntactic description: Second edition Volume III: Grammatical categories and the Lexicon. In *Language Typology and Syntactic Description*. https://doi.org/10.1017/CBO9780511618437

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160059. https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, *49*(2), 418–432. https://doi.org/10.3758/s13428-016-0719-z

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120. https://doi.org/10.1016/j.jml.2015.02.001

Singer, J. D., & Willett, J. B. (2009). A Framework for Investigating Change over Time. In *Applied Longitudinal Data Analysis* (pp. 115–139). https://doi.org/10.1093/acprof:oso/9780195152968.003.0001

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U., & Baltes, P. B. (2003). The fate of cognition in very old age: six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging*.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1271–1278. https://doi.org/10.1038/nature04286

Stoianov, I. (1998). *Tree-based analysis of simple recurrent network learning*. https://doi.org/10.3115/980691.980820

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2006/085)

Struys, E., Mohades, G., Bosch, P., & van Den Noort, M. (2015). Cognitive control in bilingual children: Disentangling the effects of second-language proficiency and onset age of acquisition. *Swiss Journal of Psychology*, *74*(2), 65–73. https://doi.org/10.1024/1421-0185/a000152

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-{A figure is presented}/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–388. https://doi.org/10.1016/j.cognition.2005.04.007

Sundermeyer, M., Oparin, I., Gauvain, J. L., Freiberg, B., Schluter, R., & Ney, H. (2013). Comparison of feedforward and recurrent neural network language models. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. https://doi.org/10.1109/ICASSP.2013.6639310

T., K. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Testolin, A., Stoianov, I., Sperduti, A., & Zorzi, M. (2016). Learning Orthographic Structure With Sequential Generative Neural Networks. *Cognitive Science*. https://doi.org/10.1111/cogs.12258

Thiessen, E. D., & Saffran, J. R. (2003). When Cues Collide: Use of Stress and Statistical Cues to Word Boundaries by 7- to 9-Month-Old Infants. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.39.4.706

Thomas, M. S. C., & McClelland, J. L. (2012). Connectionist Models of Cognition. In *The Cambridge Handbook of Computational Psychology*. https://doi.org/10.1017/cbo9780511816772.005

Touretzky, D. S., & Hinton, G. E. (1988). A distributed connectionist production system. *Cognitive Science*, *12*(3), 423–466. https://doi.org/10.1016/0364-0213(88)90029-8

Trapman, M., van Gelderen, A., van Steensel, R., van Schooten, E., & Hulstijn, J. (2014). Linguistic knowledge, fluency and meta-cognitive knowledge as components of reading comprehension in adolescent low achievers: Differences between monolinguals and bilinguals. *Journal of Research in Reading*, *37*(S1), S3–S21. https://doi.org/10.1111/j.1467-9817.2012.01539.x

Tremblay, M.-C., & Sabourin, L. (2012). Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals. *The Journal of the Acoustical Society of America*, *132*(5), 3465–3474. https://doi.org/10.1121/1.4756955

Tuncer, U. (2009). How do monolingual and bilingual language learners differ in use of learning strategies while learning a foreign language? Evidences from Mersin University. *Procedia - Social and Behavioral Sciences*, *1*(1), 852–856. https://doi.org/10.1016/j.sbspro.2009.01.153

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Van Der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. J. (2020). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin and Review*, 1–14. https://doi.org/10.3758/s13423-020-01798-5

Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2003). Roles of linguistic knowledge, metacognitive knowledge and processing speed in L3, L2 and L1 reading comprehension: A structural equation modeling approach. *International Journal of Bilingualism*, *7*(1), 7–25. https://doi.org/10.1177/13670069030070010201

Van Hell, J. G., & Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*. https://doi.org/10.1111/0023-8333.00018

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language*, *29*(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

VanPatten, B., & Benati, A. G. (2015). *Key terms in second language acquisition*. Bloomsbury Publishing.

Varnosfadrani, A. D., & Basturkmen, H. (2009). The effectiveness of implicit and explicit error correction on learners' performance. *System*, *37*(1), 82–98. https://doi.org/10.1016/j.system.2008.04.004

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, *44*(1), 121–133. https://doi.org/10.1016/j.neuron.2004.08.031

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in Psychology*, *5*, 953. https://doi.org/10.3389/fpsyg.2014.00953

Weiss, D. J. (2020). Introduction: The use of artificial languages in bilingualism research. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728919000750

Werbos, P. J. (1990). Backpropagation Through Time: What It Does and How to Do It. *Proceedings of the IEEE*. https://doi.org/10.1109/5.58337

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Cross-linguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, *21*(6), 666–682. https://doi.org/10.1177/1367006916648859

Williams, W., Prasad, N., Mrva, D., Ash, T., & Robinson, T. (2015). Scaling recurrent neural network language models. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. https://doi.org/10.1109/ICASSP.2015.7179001

Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2021). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, *5*, 389–398. https://doi.org/10.1038/s41562-020-00982-w

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2014.09.004

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology*, *2*, 210. https://doi.org/10.3389/fpsyg.2011.00210

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, *137*(33), 1452–1463. https://doi.org/10.1121/1.4913457

Zhao, X., & Li, P. (2007). Bilingual Lexical Representation in a Self-Organizing Neural Network Model. *English*.

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2010.488284

Zhao, X., & Li, P. (2013). Simulating cross-language priming with a dynamic computational model of the lexicon. *Bilingualism*. https://doi.org/10.1017/S1366728912000624

Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. https://doi.org/10.1080/713755906

Appendix

APPENDIX A: Chapter 2

Appendix A1. Linguistic profile and contrasts per group in Experiment 1.

	Spanish Monolinguals		Spanish-Basque Bilinguals		Spanish-English Bilinguals		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 114)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.8	2.6	21.2	1.9	20.9	2.3	1.610	0.204	0.303	0.088	0.586
Non-verbal IQ	101.5	8.3	100.7	8.6	99.8	6.6	0.481	0.619	0.120	0.406	0.592
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	-	-
Total Exposure L1 (%)	91.4	8.1	60.3	11.8	64.5	9.9	111.685	**< 0.001**	**> 100**	**< 0.001**	0.070
Self-rated proficiency L1 (1-10)	9.8	0.4	9.8	0.2	9.9	0.3	0.876	0.419	0.166	0.411	0.294
BEST L1 (0-65)	64.6	0.6	64.6	0.7	64.8	0.5	1.425	0.245	0.260	0.217	0.243
Age of Acquisition L2	5.3	1.8	3.0	1.6	3.5	1.7	20.508	**< 0.001**	**> 100**	**< 0.001**	0.237
Total Exposure L2 (%)	7.7	7.2	32.5	13.1	33.5	9.4	80.766	**< 0.001**	**> 100**	**< 0.001**	0.664
Self-rated proficiency L2 (1-10)	4.3	1.8	8.6	0.7	8.1	0.8	149.854	**< 0.001**	**> 100**	**< 0.001**	0.074
BEST L2 (0-65)	25.9	8.2	55.9	6.9	55.9	4.2	264.291	**< 0.001**	**> 100**	**< 0.001**	0.991
LexTALE L1 (%)	-	-	92.9	5.6	93.8	6.3	-	-	-	-	0.506[b]
LexTALE L2 (%)	-	-	90.7	6.1	88.8	6.0	-	-	-	-	0.176[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

Appendix A2. Linguistic profile and contrasts per group in Experiments 2 and 5.

	Spanish Monolinguals		Spanish-Basque Bilinguals		Spanish-English Bilinguals		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 117)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.7	2.4	21.8	2.2	21.0	2.4	1.405	0.250	0.252	0.478	0.132
Non-verbal IQ	99.8	6.6	101.1	8.8	100.1	6.6	0.320	0.727	0.103	0.578	0.567
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	-	-
Total Exposure L1 (%)	91.4	8.6	61.3	8.1	64.8	9.9	136.974	**< 0.001**	**> 100**	**< 0.001**	0.081
Self-rated proficiency L1 (1-10)	9.9	0.3	9.8	0.4	9.9	0.3	1.084	0.342	0.194	0.925	0.144
BEST L1 (0-65)	64.7	0.6	64.6	0.7	64.8	0.4	1.784	0.172	0.344	0.829	0.063
Age of Acquisition L2	5.6	1.5	3.8	1.0	3.5	1.7	25.839	**< 0.001**	**> 100**	**< 0.001**	0.435
Total Exposure L2 (%)	8.6	8.6	34.0	9.6	33.4	9.3	99.097	**< 0.001**	**> 100**	**< 0.001**	0.762
Self-rated proficiency L2 (1-10)	4.2	1.7	8.5	1.0	8.1	0.8	146.243	**< 0.001**	**> 100**	**< 0.001**	0.145
BEST L2 (0-65)	25.6	9.6	56.3	6.5	55.9	4.3	243.908	**< 0.001**	**> 100**	**< 0.001**	0.802
LexTALE L1 (%)	-	-	93.4	5.6	93.3	6.8	-	-	-	-	0.986[b]
LexTALE L2 (%)	-	-	87.1	7.9	88.9	6.0	-	-	-	-	0.262[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

Appendix A3. Linguistic profile and contrasts per group in Experiment 3.

	Spanish Monolinguals		Spanish-Basque Bilinguals		Spanish-English Bilinguals		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 117)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.7	2.5	21.9	1.9	21.0	2.4	1.772	0.175	0.341	0.696	0.068
Non-verbal IQ	102.2	8.0	101.5	6.2	100.1	6.6	0.916	0.403	0.169	0.312	0.372
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	-	-
Total Exposure L1 (%)	91.9	6.3	62.5	10.8	64.8	9.9	126.895	**< 0.001**	**> 100**	**< 0.001**	0.276
Self-rated proficiency L1 (1-10)	9.9	0.3	9.8	0.3	9.9	0.3	1.012	0.367	0.182	0.817	0.163
BEST L1 (0-65)	64.6	6.0	64.7	0.5	64.8	0.4	2.026	0.136	0.420	0.074	0.374
Age of Acquisition L2	5.5	2.0	2.9	1.0	3.5	1.7	27.259	**< 0.001**	**> 100**	**< 0.001**	0.079
Total Exposure L2 (%)	7.5	5.8	30.5	9.3	33.4	9.3	116.518	**< 0.001**	**> 100**	**< 0.001**	0.125
Self-rated proficiency L2 (1-10)	3.9	1.6	8.5	0.9	8.1	0.8	257.160	**< 0.001**	**> 100**	**< 0.001**	0.069
BEST L2 (0-65)	26.9	9.2	56.2	7.3	55.9	4.3	217.372	**< 0.001**	**> 100**	**< 0.001**	0.817
LexTALE L1 (%)	-	-	91.1	7.6	93.3	6.8	-	-	-	-	0.170[b]
LexTALE L2 (%)	-	-	88.7	9.0	88.9	6.0	-	-	-	-	0.919[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

Appendix A4. Linguistic profile and contrasts per group in Experiment 4.

	Spanish Monolinguals		Spanish-Basque Bilinguals		Spanish-English Bilinguals		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 114)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	21.7	2.5	21.9	1.9	21.0	2.4	1.712	0.185	0.328	0.696	0.072
Non-verbal IQ	102.2	8.0	101.5	6.2	100.6	6.5	0.477	0.622	0.120	0.313	0.583
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	-	-
Total Exposure L1 (%)	90.7	7.5	62.5	10.8	65.0	9.4	111.191	**< 0.001**	**> 100**	**< 0.001**	0.242
Self-rated proficiency L1 (1-10)	9.9	0.3	9.8	0.3	9.9	0.2	1.54	0.219	0.286	0.962	0.082
BEST L1 (0-65)	64.7	0.5	64.7	0.5	64.8	0.4	1.618	0.203	0.304	0.145	0.284
Age of Acquisition L2	5.5	2.0	2.9	1.0	3.6	1.7	26.798	**< 0.001**	**> 100**	**< 0.001**	0.067
Total Exposure L2 (%)	8.9	7.2	30.5	9.3	33.1	8.9	95.487	**< 0.001**	**> 100**	**< 0.001**	0.183
Self-rated proficiency L2 (1-10)	3.9	1.6	8.5	0.9	8.1	0.9	180.184	**< 0.001**	**> 100**	**< 0.001**	0.082
BEST L2 (0-65)	26.9	9.2	56.2	7.3	55.7	4.3	210.094	**< 0.001**	**> 100**	**< 0.001**	0.766
LexTALE L1 (%)	-	-	91.1	7.6	93.9	5.9	-	-	-	-	0.077[b]
LexTALE L2 (%)	-	-	88.7	8.9	88.9	6.1	-	-	-	-	0.900[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

APPENDIX B: Chapter 2

Appendix B1. Learning stimuli in Experiment 5.

List A		List B	
actualune	generalutet	actualbire	generalodi
alcoholibe	genialpol	alcoholisos	genialule
anormalboru	horrorpol	anormalodi	horrorodi
anteriorune	hospitalbur	anteriorbire	hospitalatut
anualutet	hostilsuti	anualbire	hostilatut
autoromed	humorsuti	autorsotu	humorodi
bestialomed	idealsuti	bestialter	idealbire
brutalbur	infantilibe	brutalpel	infantilsotu
canalomed	jovialbur	canalbire	jovialsotu
capilaribe	laborboru	capilarule	laborpel
capitalune	legalutet	capitalisos	legalatut
casualibe	liberalbur	casualsotu	liberalsotu
centralpol	literalsuti	centralpel	literalule
civilbur	manualsuti	civilter	manualodi
colorpol	materialsuti	colorbire	materialule
cordialpol	mayorutet	cordialisos	mayorodi
corporalsuti	mensualpol	corporalisos	mensualter
criminalomed	mentalboru	criminalule	mentalpel
cristalomed	militarune	cristalter	militarpel
dualomed	mineralutet	dualisos	mineralpel
ejemplarboru	moralibe	ejemplarsotu	moralter
escolarpol	mortalibe	escolarsotu	mortalter
especialune	motorbur	especialodi	motorbire
eventualboru	musicalsuti	eventualatut	musicalodi
exteriorune	mutualutet	exteriorisos	mutualbire
familiaribe	nacionalbur	familiarisos	nacionalatut
fatalomed	natalboru	fatalsotu	natalatut
federalboru	naturalbur	federalatut	naturalule
finalune	neutralboru	finalule	neutralule
florutet	normalpol	florter	normalisos
formalune	oficialibe	formalpel	oficialter
futbolomed	originalutet	futbolatut	originalpel

Appendix B2. Linguistic profile and contrasts per group in Experiment 6.

	Spanish Monolinguals		Spanish-Basque Bilinguals		Spanish-English Bilinguals		ANOVA			Helmert contrasts p-value	
	M	*SD*	*M*	*SD*	*M*	*SD*	$F_{(2, 117)}$	*p*	BF_{10}	MONO-BIL	SPBQ-SPEN
Age	20.4	2.1	21.5	2.6	21.0	2.5	2.242	0.111	0.501	0.064	0.326
Non-verbal IQ	102.9	5.2	103.2	5.7	102.9	4.2	0.036	0.965	0.082	0.909	0.809
Age of Acquisition L1 [a]	0.0	0.0	0.0	0.0	0.0	0.0	-	-	-	-	-
Total Exposure L1 (%)	91.2	8.1	60.9	10.1	64.8	9.9	123.029	**< 0.001**	**> 100**	**< 0.001**	0.068
Self-rated proficiency L1 (1-10)	9.8	0.5	9.8	0.2	9.9	0.3	2.425	0.093	0.368	0.059	0.273
BEST L1 (0-65)	64.7	0.6	64.5	0.9	64.8	0.5	1.892	0.155	0.376	0.714	0.059
Age of Acquisition L2	5.3	1.6	3.0	1.8	3.5	1.7	19.809	**< 0.001**	**> 100**	**< 0.001**	0.196
Total Exposure L2 (%)	8.4	7.9	34.4	11.1	33.4	9.3	94.713	**< 0.001**	**> 100**	**< 0.001**	0.641
Self-rated proficiency L2 (1-10)	4.3	1.8	8.4	1.2	8.1	0.8	121.22	**< 0.001**	**> 100**	**< 0.001**	0.227
BEST L2 (0-65)	23.1	13.7	56.2	7.4	55.9	4.3	166.745	**< 0.001**	**> 100**	**< 0.001**	0.867
LexTALE L1 (%)	-	-	92.3	5.1	93.3	6.8	-	-	-	-	0.442[b]
LexTALE L2 (%)	-	-	89.7	5.5	88.8	5.9	-	-	-	-	0.510[b]

Note. Significant contrast terms are highlighted in bold. SD = standard deviation; MONO = Spanish monolinguals; BIL = bilinguals; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals.

[a] Statistic and *p*-value undefined due to zero variance.

[b] Difference calculated using Welch t-tests.

Appendix B3. Flavian vocabulary in Experiment 6.

Word	Syllables	Condition	Composite_SP	Composite_BQ	Composite_EN	Final Composite
rerta	rer-ta	ES+	1.330	0.923	1.131	1.762
perto	per-to	ES+	0.705	0.789	0.551	1.661
ronta	ron-ta	ES+	1.304	0.178	0.824	1.658
lesta	les-ta	ES+	1.227	1.178	0.774	1.547
terta	ter-ta	ES+	1.074	0.948	0.429	1.523
ranta	ran-ta	ES+	0.868	0.408	0.825	1.430
sasta	sas-ta	ES+	1.118	0.958	0.736	1.391
respa	res-pa	ES+	0.966	-0.112	0.446	1.326
relta	rel-ta	ES+	1.053	0.182	0.978	1.248
donte	don-te	ES+	0.815	0.769	1.273	1.248
larta	lar-ta	ES+	1.036	1.392	0.887	1.247
lenda	len-da	ES+	0.671	0.692	0.445	1.197
derdo	der-do	ES+	0.831	0.594	0.488	1.185
sonto	son-to	ES+	1.017	0.244	0.787	1.166
londa	lon-da	ES+	1.042	0.453	0.620	1.016
malga	mal-ga	ES+	0.834	0.363	0.376	0.991
denra	den-ra	ES+	0.622	0.346	0.485	0.962
tunta	tun-ta	ES+	0.648	0.351	0.316	0.961
fosta	fos-ta	ES+	0.545	0.684	0.460	0.945
penra	pen-ra	ES+	0.425	0.122	0.181	0.945
pusta	pus-ta	ES+	0.708	0.750	0.492	0.934
talta	tal-ta	ES+	0.639	0.478	0.219	0.920
tista	tis-ta	ES+	0.667	0.629	0.740	0.875
polta	pol-ta	ES+	0.504	0.363	0.313	0.812
mimbo	mim-bo	ES-	-0.367	-0.899	-0.287	-0.357
bulmo	bul-mo	ES-	-0.432	-0.276	-0.341	-0.518
nulpa	nul-pa	ES-	-0.393	-0.890	-1.162	-0.565
rimpo	rim-po	ES-	-0.466	-1.148	-0.099	-0.651
mulmo	mul-mo	ES-	-0.529	-0.636	-1.250	-0.655
rulgo	rul-go	ES-	-0.512	-1.324	-1.335	-0.712
mupto	mup-to	ES-	-0.565	-0.451	-0.898	-0.748
dagmo	dag-mo	ES-	-0.698	-0.462	-0.592	-0.837
sulbe	sul-be	ES-	-1.093	-1.008	-0.171	-0.892
gurfo	gur-fo	ES-	-0.663	-0.017	-0.627	-0.923
susbe	sus-be	ES-	-1.188	-0.672	-0.391	-0.950
rirbo	rir-bo	ES-	-0.647	-0.403	-1.540	-0.991
sutno	sut-no	ES-	-0.477	0.068	-0.678	-1.020
ripso	rip-so	ES-	-0.634	-1.270	-0.334	-1.029
gunfo	gun-fo	ES-	-0.659	-1.056	-0.492	-1.070
bipto	bip-to	ES-	-0.695	-0.139	-0.913	-1.163
sumne	sum-ne	ES-	-1.399	-0.749	0.061	-1.181
rutno	rut-no	ES-	-0.692	-0.576	-0.384	-1.213

bompu	bom-pu	ES-	-1.322	-1.282	-0.380	-1.214
dutno	dut-no	ES-	-0.744	-0.074	-0.778	-1.219
fimbe	fim-be	ES-	-1.136	-1.465	-0.208	-1.228
dumle	dum-le	ES-	-1.164	-0.784	-0.121	-1.261
gulbe	gul-be	ES-	-1.624	-1.112	-0.438	-1.390
tadmi	tad-mi	ES-	-1.736	-1.130	-1.030	-1.584

Note. SP = Spanish; BQ = Basque; EN = English.

Appendix B4. Recognition test GLMM results from Experiment 6.

Fixed Effects		Estimate	SE	z	*p*	BF_{10}
(Intercept)		**2.296**	**0.109**	**21.007**	**< 0.001**	**-**
Day		-0.022	0.043	-0.516	0.606	0.048
Condition		**0.063**	**0.029**	**2.144**	**0.032**	**0.439**
SPBQ-SPEN		-0.027	0.126	-0.215	0.830	0.046
MONO-BIL		-0.142	0.073	-1.950	0.051	0.301
Day x Condition		0.003	0.029	0.113	0.910	0.001
Day x SPBQ-SPEN		-0.048	0.047	-1.025	0.306	0.003
Day x MONO-BIL		0.012	0.026	0.466	0.641	0.015
Condition x SPBQ-SPEN		0.043	0.037	1.158	0.247	0.004
Condition x MONO-BIL		-0.003	0.020	-0.151	0.880	0.014
Covariates						
Age		**-0.108**	**0.044**	**-2.440**	**0.015**	**0.712**
Non-verbal IQ		0.006	0.009	0.650	0.516	0.263
Gender		0.106	0.290	0.365	0.715	0.047
OSPAN score		0.012	0.007	1.703	0.089	0.954
Random Effects	*Group*	*Variance*	*SD*	*Correlation*		
Participant	(Intercept)	0.999	0.999			
	Day	0.047	0.218	0.09		

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

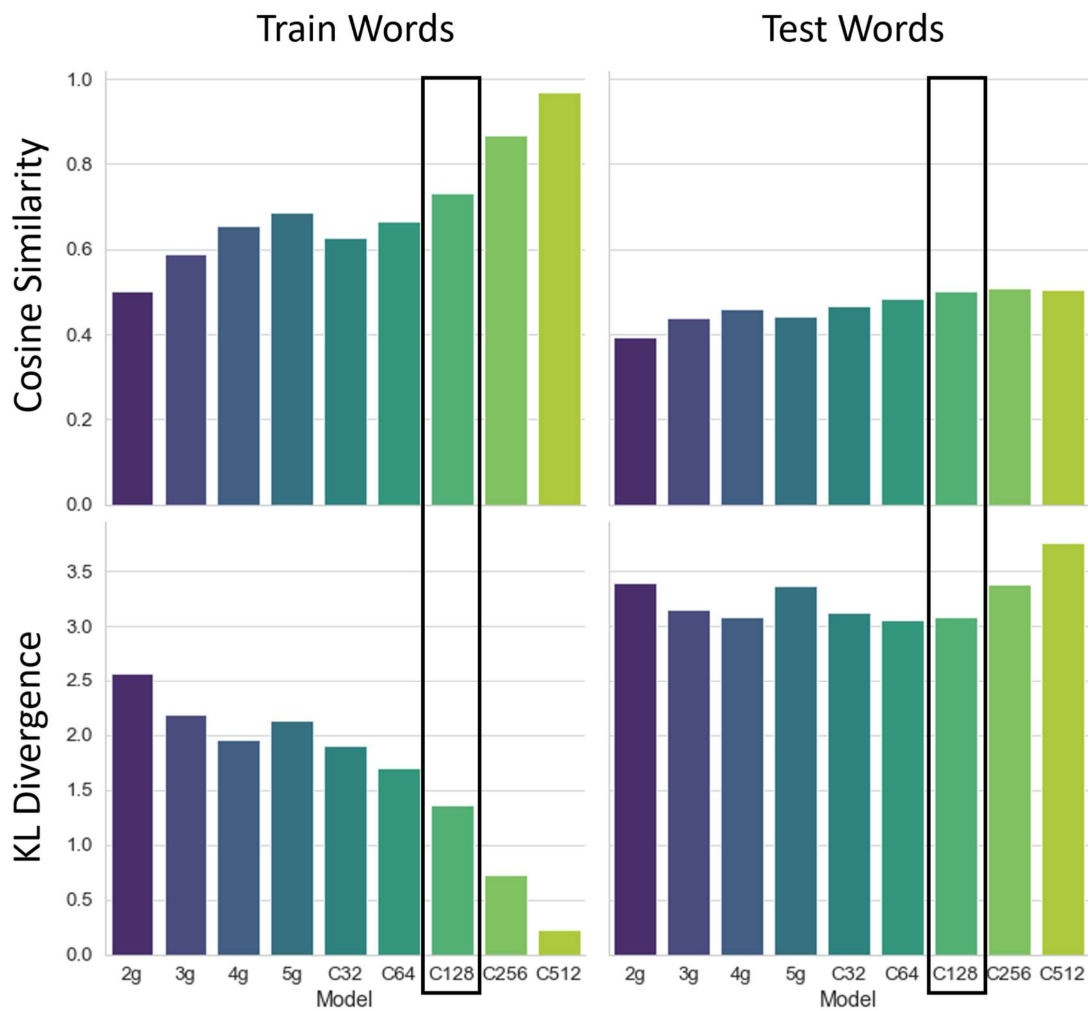
Appendix B5. Production test LMM results from Experiment 6.

Fixed Effects		Estimate	SE	df	t	*p*	BF_{10}
(Intercept)		**13.552**	**0.434**	**111.2**	**31.205**	**< 2e-16**	**-**
Day		**-0.428**	**0.057**	**115.0**	**-7.514**	**0.000**	**> 100**
Condition		**1.486**	**0.116**	**115.0**	**12.782**	**< 2e-16**	**> 100**
SPBQ-SPEN		-0.594	0.502	111.0	-1.183	0.239	0.075
MONO-BIL		**-0.968**	**0.293**	**111.1**	**-3.303**	**0.001**	**3.033**
Day x Condition		0.019	0.046	117.0	0.424	0.672	0.008
Day x SPBQ-SPEN		-0.089	0.070	115.0	-1.271	0.206	0.003
Day x MONO-BIL		-0.037	0.040	115.0	-0.910	0.365	0.048
Condition x SPBQ-SPEN		0.201	0.143	115.0	1.403	0.163	0.005
Condition x MONO-BIL		0.124	0.082	115.0	1.513	0.133	0.158
Covariates							
Age		**-0.366**	**0.176**	**111.0**	**-2.080**	**0.040**	**0.050**
Non-verbal IQ		-0.031	0.035	111.0	-0.904	0.368	0.004
Gender		1.225	1.154	111.0	1.062	0.291	0.216
OSPAN score		0.043	0.029	111.0	1.474	0.143	0.016
Random Effects	*Group*	*Variance*	*SD*	*Correlation*			
Participant by Condition	(Intercept)	2.690	1.640				
Participant	(Intercept)	16.978	4.121				
	Day	0.134	0.366	0.15			

Note. Significant fixed effects terms are highlighted in bold. SE = standard error; SD = standard deviation; SPEN = Spanish-English bilinguals; SPBQ = Spanish-Basque bilinguals; MONO = Spanish monolinguals; BIL = bilinguals.

APPENDIX C: Chapter 3

Appendix C1. Selecting the number of LSTM nodes.



Note. I calculated the successor character probability at each time-step for the Spanish train and test (validation + test) words. The probability was approximated in three ways: (1) I computed different n-gram models (2g to 5g) by counting the occurrence of n-grams within the train words and normalizing them using different Laplace constants ranging from 0 to 1 in increments of 0.2 (Jurafsky & Martin, 2014). (2) I trained monolingual CLOUD models on the train words with different numbers of LSTM nodes (C32 to C512) for 50 epochs and five runs each. (3) I calculated the empirical probabilities separately for the train and test words using a tree-based (Trie) analysis (Stoianov, 1998; Testolin et al., 2016). The plots show the results of comparing each model's (2g to C512) predicted successor character probability at each time-step with the Trie analysis's empirical probability. I compared the probability vectors using the cosine similarity (top) and the Kullback-Leibler (KL) Divergence (bottom). The KL Divergence measures how one probability distribution differs from another by adding the probability of each possible value multiplied by the differences between the two probabilities' logarithmic values. The plots show how much the different models can approximate the empirical probability of the Spanish words. For the cosine similarity, higher values are better, while for the KL Divergence, lower values are better. The plots suggest that values between 64 and 128 LSTM nodes are sufficient to outperform the n-gram models while still avoiding overfitting—i.e., minimizing the difference of performance on the seen train words and the unseen test words. For instance, the C512 model performs outstandingly in the train words, with cosine similarity close to 1 and KL Divergence close to 0. However, the values change significantly for the test words, a clear sign of overfitting to the training data. In contrast, the C128 model could outperform the 5-gram model on all train and test words metrics. Thus, I selected 128 nodes as the hyperparameter for the distributed word representations.

Appendix C2. LSTM forward equations.

The LSTM nodes operate using the following forward equations

$$f_t = \sigma\left(W_f \cdot [x_t, h_{t-1}] + b_f\right) \; ; \quad W_f \in \mathbb{R}^{128\, x\, 160} ; \; b_f \in \mathbb{R}^{128} \tag{1}$$

$$i_t = \sigma(W_i \cdot [x_t, h_{t-1}] + b_i) \quad ; \quad W_i \in \mathbb{R}^{128\, x\, 160} ; \; b_i \in \mathbb{R}^{128} \tag{2}$$

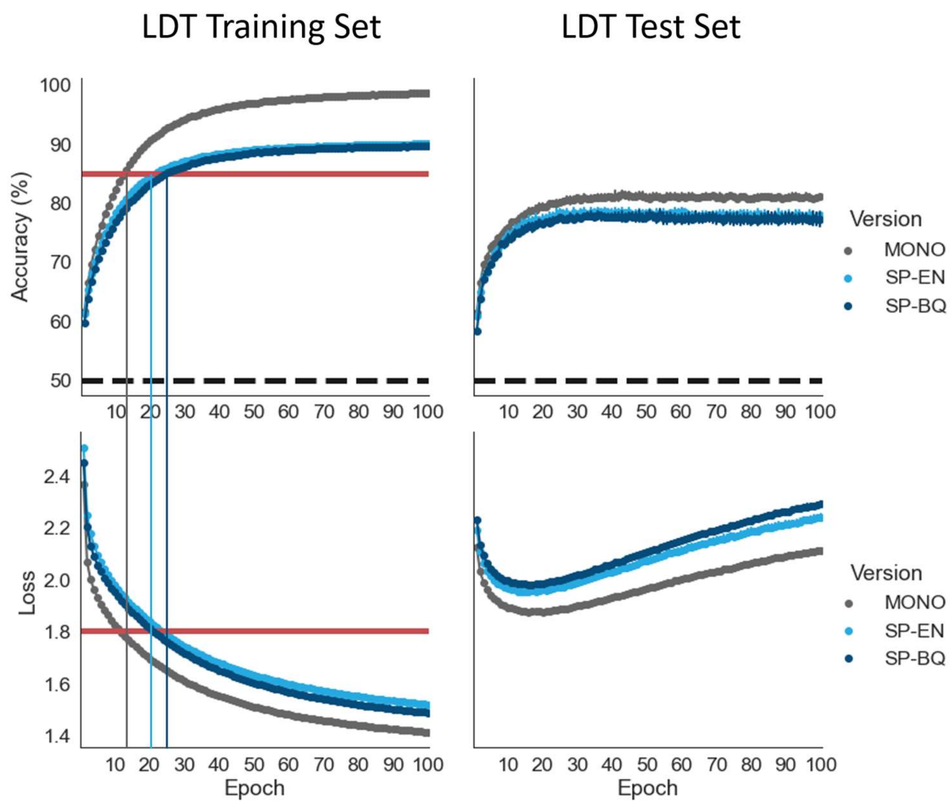
$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_c \cdot [x_t, h_{t-1}] + b_c) \; ; \; W_c \in \mathbb{R}^{128\, x\, 160} ; \; b_c \in \mathbb{R}^{128} \tag{3}$$

$$o_t = \sigma(W_o \cdot [x_t, h_{t-1}] + b_o) \; ; \; W_o \in \mathbb{R}^{128\, x\, 160} ; \; b_o \in \mathbb{R}^{128} \tag{4}$$

$$h_t = o_t \circ \tanh(C_t) \quad ; \quad o_t \in \mathbb{R}^{128} ; \; C_t \in \mathbb{R}^{128} \tag{5}$$

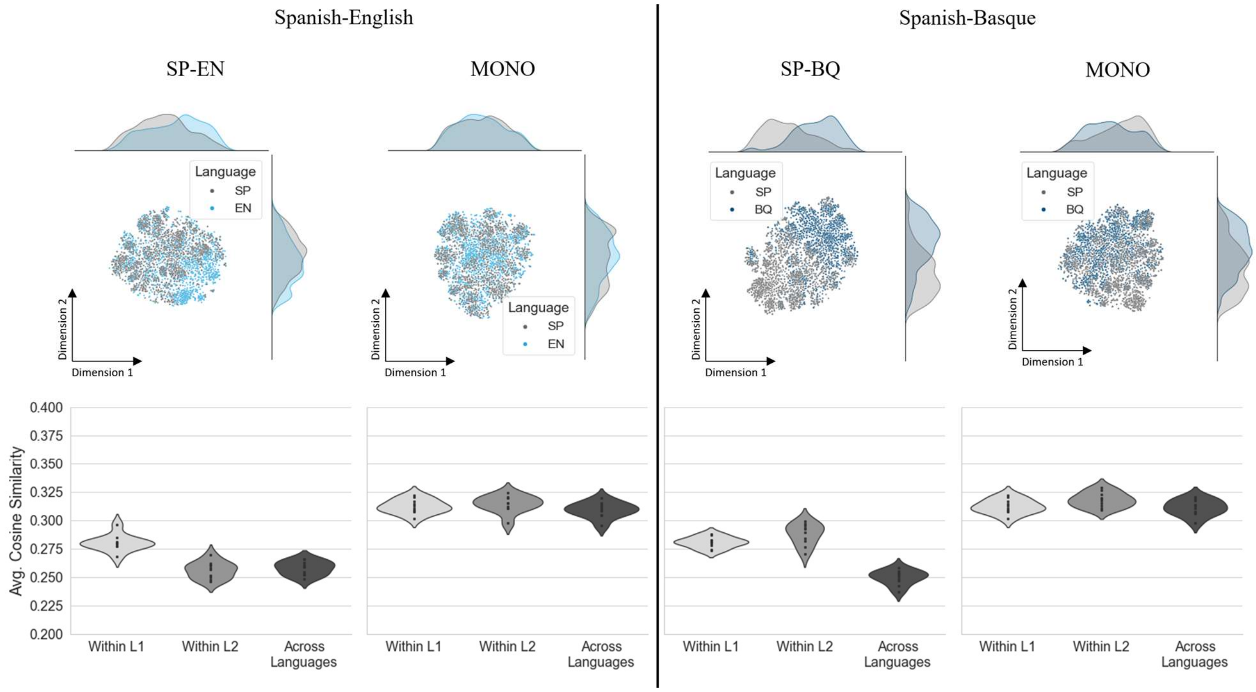
where f_t i_t, o_t are the *forget, input,* and *output gates* at time *t*, σ(·) is the sigmoid activation function, tanh(·) is the hyperbolic tangent activation function, [] indicates vector or matrix concatenation, **·** is the dot product, ° is the element-wise multiplication, C_t is the cell state, and h_t is the hidden state and output of the LSTM layer at time *t*. All *W* are learnable weight matrices, and all *b* are learnable bias vectors. Following common deep learning practice (Glorot & Bengio, 2010), these parameters are initialized from a uniform distribution defined using the square root of the inverse number of nodes, ranging from -$(1/128)^{0.5}$ to $(1/128)^{0.5}$. At time-step zero, the cell and hidden states are initialized as zero vectors.

Appendix C3. Average pre-training LDT accuracy and loss of each CLOUD model version.



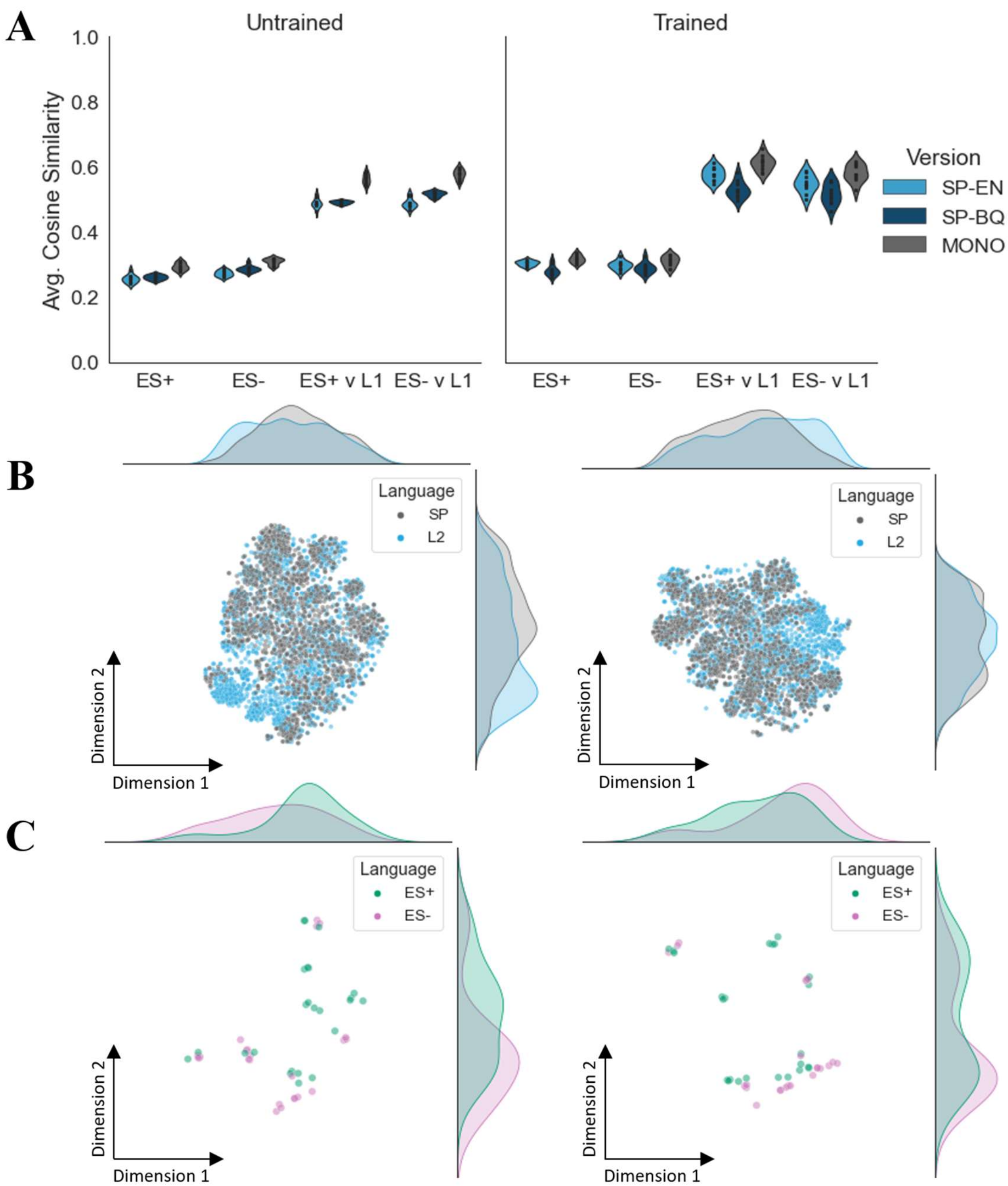
Note. I trained ten runs for each CLOUD version for 100 epochs on their respective training sets. I stored a state of each model's parameters for the simulations by setting an adaptive accuracy threshold on the Spanish LDT training set (left). The average accuracy at every epoch is plotted on top, and the average loss at the bottom. The maximum accuracy value achieved by all model runs was 85% (red horizontal line). Matching the models implied that they did not train for the entire 100 epochs (as indicated by the vertical color lines). Naturally, the MONO version that only trained on the Spanish words reached the maximum threshold before the bilingual versions. Regardless, the MONO version still achieved a slightly lower loss at that epoch than the bilingual versions in later epochs. The average accuracy and loss for the LDT test set are presented on the right. Notably, the models' loss started to increase after epoch 25, a clear sign of overfitting. Similarly, the accuracy on the test set started to decrease slightly after epoch 30. Thus, the selected models are not only comparable in their performance but also minimize overfitting.

Appendix C4. Word clouds and average word similarity for the CLOUD model's versions.



Note. The figures were created using the Spanish-English (left) and Spanish-Basque (right) test set words. **(top)** I computed the representation of each word as the average of the LSTM output at each time step. Each 128-dimensional representation was projected onto a 2-dimensional space using the t-SNE algorithm (Van Der Maaten & Hinton, 2008). I repeated this procedure for each CLOUD version's first run (SP-EN, SP-BQ, and MONO). The same MONO version was used in both the Spanish-English and Spanish-Basque word clouds. As evidenced in the sample word clouds, the bilingual CLOUD versions showed a higher separation for the languages than the MONO version, which clustered most of the representations. **(bottom)**. I calculated the cosine similarity of each word to other words within the same language and across languages. In matrix form, the cosine similarity is defined as the dot product between two L2-normalized matrices. This formulation yields a symmetric matrix where each cell corresponds to the pair-wise similarities between words, with ones over the main diagonal—similar to a correlation matrix. I only computed the average of each matrix's lower triangle to avoid inflating the results by including the similarity of each word to itself (always 1). Higher values indicate more similarity between the word representations. The MONO version showed more similarity than the bilingual versions for all representations, within and across languages, as it only learned from the Spanish orthographic patterns. The differences observed in the SP-EN and the SP-BQ version can be explained by the three languages' different roots. Spanish and English share a large portion of their vocabulary due to their Latin and Greek roots, but since English also possesses words from other sources (e.g., Germanic roots), the English word representations are more spread. This makes the similarity within English words slightly lower than across English and Spanish. Conversely, Spanish and Basque do not share origins. Instead, due to geographical proximity, Basque utilizes many Spanish loan words. These loan words are generally adapted to the Basque orthography—for instance, by replacing the "v" with a "b" in vela/bela (a sail). These replacements, combined with Basque-specific orthographic patterns (e.g., "ts", "tx", "tz") push the representations of Basque words closer together and distances them more from Spanish. ES = Spanish, EN = English, EU = Basque.

Appendix C5. Word similarity and word clouds before and after learning the Flavian vocabulary.



Note. **(A)** Average cosine similarity of Flavian word representations to themselves (ES+ and ES-) and to the Spanish words (ES+ v L1, ES- v L1) for each CLOUD model's version and run. Before training (left), the similarity scores are lower for the bilingual than the monolingual versions. After training (right), the ES+ and ES- words show a slight increase in their similarity to each other, and an overall increase in their similarity to Spanish. **(B)** Word cloud from the Spanish-English validation and test sets. The word clouds were constructed as in Figure 1. Due to the CLOUD model's dynamic constraint, all representations are modified while learning the Flavian words. **(C)** Word cloud of the Flavian words within the Spanish-English word cloud as shown in (B).