eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Sensitivity to Letter Sequences

in word processing and word learning

doctoral dissertation by

María Borragan Salcines

In candidacy for the Degree of Doctor of Philosophy by
the University of the Basque Country

supervised by

Prof. Jon Andoni Duñabeitia and Dr. Angela de Bruin

Donostia-San Sebastian, 2020

r rr rr rr r rr r r rr r r rr r r rr rr rr r rr r r rr rr rr r

Sensitivity to Letter Sequences

IN THE WORD PROCESSING AND WORD LEARNING

dxtb dxtbd xtbd xtbd xtbd xtbd xtbdx tbdx tbdx tbdx tbdxt bdxt

bqhlw bqhlwb qhlwb

María Borragan

hlwbq hlwbqh lwbqh

lwpz lwp zlwp zlw pzlw pzl wpz lwpz lwp zlwp zlw pzlw pzl wpzl

wpz wp zwp zwp zwp zw pz wp zwp zw pz wp zw pzw pzw pzw pz wpz

fs f sf s fs fs f sf s fs fs f sf sf s fs fs f sf sf s fs f sf

s s ss s ss s ss s ss s ss ss s ss ss ss ss s ss s ss s ss s s

kv kv kvk vk vk vkv kv kv kv kvk vk vkv kv kvk vk vk vkv kv kv

gcn gcng cngc ngc ngc ngcn gcng cngc ngc ngc ngc ngcn gcng cng

cnrm cnrmc nrmc nrmc nrmc nrmc nrmc nrmcn rmcn rmcn rmcnr mcnr

mjdxt mjdxtm jdxtm jdxtm jdxtmj dxtmj dxtmj dxtmj dxtmjd xtmjd

xtbq xtb qxtb qxt bqxt bqx tbq xtbq xtb qxtb qxt bqxt bqx tbqx

tbq tb qtb qtb qtb qt bq tb qtb qt bq tb qt bqt bqt bqt bq tbq

2020

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Sensitivity to Letter Sequences

in word processing and word learning

doctoral dissertation by

María Borragan Salcines

In candidacy for the Degree of Doctor of Philosophy
by the University of the Basque Country

supervised by

Prof. Jon Andoni Duñabeitia and Dr. Angela de Bruin

BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

Universidad del País Vasco Euskal Herriko Unibertsitatea

"la Caixa" Foundation

The research presented in this thesis was supported by a fellowship from the "La Caixa" Foundation (*ID 100010434*), fellowship code LCF/BQ/ES16/11570003.

A doctoral dissertation by *Maria Borragan Salcines*
Cover image and internal insertions are adapted works from *Jose Luis Castillejo.* In the cover two works from el libro de las líneas, 1972. In the insertions five works from el libro del trío (*Versión II trios de dos letras*), 1972. Jose Luis is an artist and experimental writer. His work streches the limits of writing using tipography and letters as sound systems able to make the graphisms reso¬nate in the pages through our capacity of reading.
Conteptual desing of the book: Alfonso Borragán.
Desing of the book: Cris Cendoya.
English editing: dr. Magda Altman.
Printed in:

BCBL Basque Center on Cognition, Brain and Language
Paseo Miketegi, 69
Donostia-San Sebastián, Spain
March 2020

Donostia-San Sebastián, 2020

Para Pedro
que me trajiste hasta aquí ...

Acknowledgments

Después de tres años y medio de un largo recorrido puedo decir que estoy ya al final. Este doctorado ha significado mucho para mi vida, en el que he aprendido, he caído y he disfrutado. Pero sobre todo me quedo con la gente que he añadido a mi vida, y este es el espacio que les dedico a todos ellos, por esta ahí, apoyarme, mimarme y sobre todo hacerme feliz. Porque al final de todo este camino no va solo de tener un titulo sino de haber hecho un doctorado en la vida.

First of all, I want to thank my advisors, Prof. Jon Andoni Duñabeitia and Dr. Angela de Bruin, for supervising my PhD these three years. They gave me the stones of the road that helped me to overcome myself, discover, investigate, learn, but above all to learn to handle situations in which I never thought of being before. Because as I said I have not only completed a Phd in Psycholinguistics, I've also earned a PhD in life. These three years have been a roller coaster with many ups and downs. In the process, I learned how to deal with many situations and people in order to finish and reach my goal, finishing this PhD.

But also, I have to thank to some people for their kindly help and their warm support because without them this thesis would not be finished. First, to Dr. Magda Altman who corrected my awful English and made the thesis coherent and well reading. And to Candice Frances, Jose Aguasvivas and Ioanna Taouki who were always free for my doubts and willing to help me with any statistical or procedural difficulty. But above all, we could rely on each other in bad times to motivate ourselves.

Pero esta tesis no podría haber sido sin el apoyo incondicional Jacobo Pacin. Tendré que agradecerte siempre por estar ahí cada día en los días buenos y en los malos, dándome ánimos para seguir y no abandonar. Porque contigo he aprendido la capacidad de aguantar y resistir un huracán. Porque me motivaste a seguir y me levantaste cuando más lo necesite. Porque el camino se anda mejor cuando vas acompañado de alguien que te hace sonreír y te motiva a seguir cuando te caes. Porque él ha hecho posible esta tesis, cuidándome y mimándome. Y puedo decir que ha sido parte de ella, que ambos hemos realizado un doctorado en la vida, juntos.

También tengo que agradecer esta tesis mi compañera de fatigas, Ionna Taouki (ευχαριστώ από καρδιάς μαλάκα). No solo ha sido un gran apoyo en la parte procedimental de la tesis, sino que ha sido otro pilar emocional importante e incondicional. Siempre dispuesta a escucharme y aconsejarme. Siempre con una sonrisa en la boca. Siempre trasmitiendo felicidad. Porque sé, que si algo voy a echar de menos de esta etapa de mi vida es no tenerte cerca día a día. Sé que me llevo una gran amiga para siempre.

Y como no, no podía faltar el apoyo de mis chicas purple, Pavlina Heinzova y Dana Scarici. Con todos los buenos momentos, las tardes de terrazas, o en la playa en Donostia. Ni sin mis chicas Sara Estelles, Lucia Arenillas, Noemi Lopez y Capilla Mata. Porque sin esos buenos momentos no hubiera podido acabar ni aguantar la presión de estos años. Porque todas vosotras habéis sido un apoyo y una parte importante. Porque no era solo un salir de casa, sino salir con las mejores compañeras y amigas. Y a Cristina Cendoya, mi familia vasca. El destino me dio uno de los regalos más bonitos que fue crecer como persona contigo en un piso de Urbieta. Porque gracias a ti, veo la vida diferente, con más alegría. Con esa sonrisa y ese positivismo que hace que todo parezca fácil y bonito en esta vida. Y claro está, ha hecho que esta tesis esté más bonita. Eskerrik asko. Tengo que de agradecer a La Caixa por la beca que me dio. Tengo que reconocer que la chica que hizo esa entrevista hace cuatro años, no es la misma que esta hoy acaba un doctorado. Supongo que la motivación fue desapareciendo, y que los mazazos a lo largo de estos años y ver que la ciencia no es lo que imagina han ido haciendo mella en mi. Pero puedo decir, que gracias a ellos, he podido realizar un doctorado y que no solo he ganado un título, sino un aprendizaje en el recorrido. Y que me llevo a 17 grandes personas conmigo (Dana, Ander, Víctor, Ana, Lidia, Alberto, Andrea, Maria Gracia, Maria, Jorge, Juanma, Luis, Patxi, Simón, Laura, Jesús, Pablo), mis compañeros de beca. Y a Elisabeth Linero, que sin su cuidado y mimo junto con su magnífica organización, esos trainings no hubieran sido lo mismo. Gracias a La Caixa he podido compartir 3 grandes semanas con ellos, en los que risas no faltaron. Y ver que gente que está realizando doctorados en campos tan diversos, están pasando por los mismo estados emocionales y dificultades que tú ayuda a pensar que no estás loco.

Y claro no puedo no dar las gracias a mis compañeros de "apartamento" del BCBL que me han aguantado este tiempo, mis charlas con Candice o mis risas con Ioanna. Porque toda la comunidad de predocs del BCBL mola mucho, y mola el compañerismo que hay. También dar las gracias a todos los que hacen posible que se pasen experimentos en el BCBL y que me han ayudado en mis pasitos como experimentadora, a Itziar Basterra, Amets Esnal y Mamen González. Gracias a ellas pude realizar mis experimentos sin grandes dificultades y sobre todo pase muy buenos momentos con ellas. Y gracias a Eider Juaristi que me ha ayudado en este último tramo, con todos los papeles y mis mil dudas del procedimiento de depositar.

No podría irme sin agradecer a Dr. Aina Casaponsa todo lo que ha hecho. Me dio la motivación necesaria cuando ya la había perdido. Me enseño muchísimo sobre ciencia y metodología pero sobre todo y lo que más agradezco me dio su amistad. Porque no podría haber tenido una estancia en el extranjero mejor, ni haber aprendida más. Porque la admiro como científica y como persona, por lo que consigue día a día. Ya sabes que lo bueno es que no tenemos a pocos kilómetros y te añado a la lista de personas que me llevo en esta etapa para siempre.

Y claro, no podría entenderse este camino de rosas con espinas sin entender de donde viene el origen. Desde bien pequeña observando el cuidado, la ternura que mi padre daba a sus pacientes, dije que yo quería seguir sus pasos. Y cuando tuve oportunidad de escoger que hacer con mi vida escogí ser logopeda. Porque la logopedia no es trabajo sin más, es una forma de ser y de vivir y sobre todo la sensación de ayudar a los demás es muy gratificante. And one day my curiosity took me to the other side of the ocean, United States, to learn about other speech pathologist and other ways of work, thinking and learning. There, I met amazing people Dr.Verdolini, Dr. Orbelo, Dr.Rubin, Dr.Coca, Dr.Knorr, Elit Kenedy) that touch me many things, but above all, they increased my desire to learn and investigate more. Thus, I though a Phd was my perfect feet to complete my training in speech pathologist. My path wasn't in US, because gracias a mi pequeño sobrino Pedro volví a pisar suelo español, donde estoy ahora.

Debo agradecer a mi familia su apoyo incondicional, por estar ahí, por quererme tanto y por enseñarme que los sueños se consiguen con persistencia. Mi padre, Dr. Alfonso Borragan, por ser mi maestro y mi guía tanto en lo profesional como en lo personal. Porque tú nos enseñaste el arte de aprender más y que descubrir es fascinante. A mi madre, Maria Salcines, por ser mi pepito grillo y darme toda la paz necesaria. Porque nadie me entiende como tú y sigues siendo ese lugar donde te sientes bien. A mi hermano, Alfonso, que nos demuestras que todo es fácil y divertido en esta vida. Nada me puede hacer más feliz que hagas parte de tu magia en esta tesis. A mi hermana, Carmen, por ser ese lucero con mimos siempre. Porque nos enseñas que con esfuerzo todo es posible. Porque mi familia me dan ese chute de energía cuando los necesito. Y claro está mi pequeña Alicia. Que nací y crecí contigo de la mano, y aunque la vida quiere que estemos en países diferentes, estas cerca de mí siempre dándome tu alegría y compartiendo tu forma de vida.

Gracias a todos por ser parte de esta tesis, pero sobre todo por ser parte de mi vida.

Abstract

The final goal of reading is comprehending a text. But accessing the meaning of each word (the visual word recognition process) is an important first step in reading. In order to access the meaning of each word, decoding letters-to-sounds (grapheme-to-phoneme) is required. As a reader becomes more skillful, words are recognized by sight without the need for letter-to-sound decoding (reading automatization; Lyon, 1998; Samuels & Flor, 1997). Some models have attempted to describe the process of reading familiar and unfamiliar words (see Chapter 1, section I.i. Models of visual word recognition; the dual route cascade model by Coltheart and colleagues, 2001 and the triangle model by Seidenberg & McClelland, 1989). Recognizing whole words by sight speeds up the reading process, producing more fluency in visual word recognition. It has been shown that fluency in visual word recognition is a strong predictor of good reading comprehension (Adams, 1994; Juel, 1988; Vellutino, Tunmer, Jaccard, & Chen, 2007). However, considering that many children have reading disabilities, the process of letter-to sound decoding, which is then generalized to global word recognition, might not be that easy (Bishop & Snowling, 2004; Goswami, 2011; Snowling & Hulme, 2012).

To this end, reading research initially focused mainly on phonological processing (Anthony & Francis, 2005; Torgesen & Hudson, 2006). But this type of processing alone is not sufficient to explain all variance in word recognition skills. So, recent lines of research have focused on the contribution of orthographic processing (Cunningham, Perry, & Stanovich, 2001). It has now clear those minimal sub-lexical units such as letters play a fundamental role in visual word recognition. Processing letters and letter sequences occurs in the early stages of visual word recognition and these processing units provide considerable information to the reader. Early models showed that this orthographic processing speeds up lexical access (see Chapter 1, section I.ii. Orthographic processing in visual word recognition; interactive model by McClelland and Rumelhart, 1981 and the dual route model by Grainger and Ziegler, 2011).

It has been shown that after little exposure to written words, individuals acquire expectations about letters and letter sequences, which have different average usage frequencies and/or tend to appear in different word positions. Thus, readers become sensitive to the distribution of the letter sequences in their language(s) (on sensitivity to orthographic regularities see Chetail, 2015; Chetail & Content, 2017; Samara & Caravolas, 2014). Even pre-readers can detect distinctive letter sequences (orthographic markedness) in their languages (Ehri, 1995, 2005), such as the word look in English, or violations of typical letter sequences, such as the non-word ffoge (Levy, Gong, Hessels, Evans, & Jared, 2006; Ouellette & Senechal, 2008a, 2008b).

This sensitivity to orthographic regularities and especially to distinctive and salient orthographic regularities (orthographic markedness) is employed by readers during visual word recognition to identify letters and letters strings and to reduce lexical competitors (see Chapter 1, section II.i. The role of orthographic regularities in visual word recognition). However, when two languages coexist in the bilingual mind, orthographic markedness is also used for language attribution (see Chapter 1, section II.ii. Sensitivity to orthographic markedness in bilinguals).

The majority of studies that have investigated how bilingual readers recognize words with the help of orthographic markedness use language decision tasks, in which words are either marked or unmarked. The critical manipulation for these conditions is the frequency of use of the constituent bigrams. Bigrams are two consecutive letters and they are considered an important sub-lexical unit because they include the maximum number of letters that an individual can handle at once and the minimal familiar unit of pronunciation (see Chapter 1, section II. Orthographic regularities; the Cambridge text by Velan and Frost, 2007 and transposed letter task by Perea & Carreiras, 2008). Thus, marked words are the words that include at least one bigram that does not exist in the other language (frequency of use in the other language is 0; e.g., the word txakurra, the Basque word for dog, is a marked word in Basque because the bigram tx does not exist in Spanish). In unmarked words all bigrams exist in both a bilingual's languages (e.g., the word ardi, the Basque word for sheep, and the word cerdo, the Spanish word for pig, both share the bigram rd that has a frequency of use in Spanish of 0.24 and in Basque of 0.30; note that the rest of the bigrams exist in both languages, too).

Researchers have shown that bilinguals use orthographic markedness to speed up conscious language attribution (Oganian, Conrad, Aryani, Heekeren, & Spalek, 2016; Vaid & Frenck-Mestre, 2002; Van Kesteren, Dijkstra, & de Smedt, 2012). This is demonstrated by more rapid responses in the correct attribution of language when words were marked. Bilinguals used these salient letter sequences in order to attribute the language of the word. The salient letter sequence helps to activate only the language with that orthographic markedness and inhibits unnecessary cross-language lexical competitors. The target word only competes with words within the language that have similar letter sequences, and this accelerates the decision on language attribution. This demonstrates that the orthographic (sub-lexical) language node is accessed before the lexical language node, to which it is directly linked. This is in line with the BIA+ extended model updated by Casaponsa, Carreiras and Duñabeitia (2014; see Chapter 1, section I.iii. Model of bilingual visual word recognition), which allows for direct feedforward and feedback connections between orthographic (sub-lexical) language nodes and the lexicon. Note that previous version of this model considers the native language node to be stable across time and assumes that it is the second language that its modulated by the native language (Dijkstra & van Heuven, 2002).

It is worth mentioning that while orthographic markedness helps bilinguals with visual word recognition, this is not the case when they learn new words. Previous studies on bilinguals learning new words have demonstrated that bilinguals are better at learning new words when they exhibit letter sequences common to the language that they already know (Ellis, 2002; Ellis & Beaton, 1993). In other words, these studies show that bilinguals prefer to learn new words when the letter sequences are already in their orthographic (sub-lexical) lexicon (see Chapter 1, section II.iii. The role of orthographic regularities in novel words learning in bilinguals).

Up to this point, we can consider that orthographic markedness is very important for reducing lexical competition, speeding up language attribution in bilinguals, and for learning new words. With this in mind, the present thesis aims to better understand the function of orthographic regularities in the developing word processing system. Specifically, we would like to better understand the function of sensitivity to orthographic markedness in word recognition and in learning to process novel words across development. Knowing more about the role of sensitivity to letter sequences would help us better understand their impact in visual word recognition, a critical question in the fields of psycholinguistics and bilingualism. To this end, we propose a two-pronged approach to investigating our research question, where we examine both word processing and word learning. First, we examine how sensitivity to orthographic markedness changes with development. Second, we observe how processing orthographic regularities impacts novel word learning at different ages.

Following the first approach, we explore whether sensitivity to orthographic markedness is stable across the lifespan (in the BIA+ model proposed by Dijkstra and van Heuven, 2002, the native language is considered to be stable across time while the second language can be modulated by the native language) or whether it is subject to change (see Chapter 2). Also, of interest is if sensitivity to orthographic regularities can be changed by learning a second language early and/or later in life. To investigate this question, we conducted two experiments to investigate sensitivity to orthographic markedness using a language decision task, with words that could be marked (at least one bigram does not exist in the other language) or unmarked (all bigrams exist in both languages). Experiment 1 focused on the study of natural bilingual development. Therefore, four balanced bilingual groups of different ages did the task. Experiment 2 investigated changes in sensitivity to orthographic markedness as a consequence of learning a second language. In this experiment, older monolingual adults learned a second language and were tested in a language decision task at three different moments during the learning process in a longitudinal study.

In the case of the second approach, we explored how sensitivity to orthographic markedness interacts with novel word learning in children and adults (see Chapter 3). In this sense, we wanted to observe whether sensitivity to orthographic markedness in the known languages (orthography language node) impacted storage of new letter sequences in the orthography lexicon. To answer this question, we investigated whether novel words that included letter sequences that do not exist in any of the bilingual languages (illegal sequences, e.g., bx is an illegal bigram because do not exist in Basque nor Spanish) are learned, and thus processed, differently than legal orthographic sequences (all bigrams exist in both the bilingual's languages). In this process, we also considered the role of bilingualism, in the sense of whether being bilingual per se has an

Abstract

impact on learning novel words (legal/illegal) or whether this impact is modulated by experience with two sets of orthographic regularities and with handling the differences between the orthographic regularities in the bilingual's two languages. To do so, we conducted two experiments, one with children and the other with adults performing a novel word learning task with legal and illegal novel words. Experiment 3 investigates whether children learn legal and illegal novel words. Three different groups of children participated in the experiment: a monolingual group, a bilingual group whose two languages had similar orthographic regularities and a bilingual group whose two languages' orthographic regularities differed. Then, two groups of adults, a monolingual group and a bilingual group whose two languages had different orthographic rules, performed the same task (Experiment 4). Finally, because results from Spanish-Basque bilingual children suggested that learning novel words was driven by handling different orthographic rules, we performed an extra experiment. Experiment 5 aimed to observe the relationship between sensitivity to orthographic markedness and novel word learning in the critical group.

Results from Chapter 2 showed that sensitivity to orthographic markedness changes with development, both due to natural bilingual development (Experiment 1) and as a consequence of second language learning (Experiment 2). Children at the age of twelve show changes in the way they are sensitive to orthographic markedness in the second language. This is in line with previous research showing that children at this age exhibit changes in their sensitivity to implicit learning (Janacsek, Fiser, & Nemeth, 2012), verbal fluency (Sauzéon, Lestage, Raboutet, N'Kaoua, & Claverie, 2004), and undergo cerebral maturation (Giedd et al., 1999). Also, it is shown that adults change the way in which they are sensitive to orthographic markedness in the native language with age (Experiment 1) and after learning a second language (Experiment 2). This demonstrates that sensitivity to orthographic markedness is not stable throughout life, and it is not only the L2 that changes as a function of the L1, but also the L1 that changes due to learning the L2. These findings are in line with the idea of an adaptive system (Kroll, Bobb, & Hoshino, 2014; Schmid, 2008). They indicate that the BIA+ extended model (Casaponsa et al., 2014) needs to be updated to account for the fact that sensitivity to orthographic markedness in the orthographic language node is not stable across time and that it includes feedback loops.

Results from Chapter 3 revealed that sensitivity to orthographic markedness plays an important role in learning novel words and that the nature of this sensitivity depends on age. Regardless of whether individuals were better at learning legal or illegal novel words, we observed that sensitivity to orthographic markedness was modulated by the impact of the orthographic language node on the storage of new letter sequences. However, children and adults learned differently. Bilingual children whose languages had different orthographic regularities learned novel words with illegal orthographic sequences equally well as legal ones (Experiment 3). In contrast, monolinguals and bilingual children whose languages shared similar orthographic regularities were better at learning legal than illegal novel words. This effect was no longer visible in adults (Experiment 4). This demonstrated that bilinguals around the age of twelve, who have to cope with different orthographic regularities in their languages (language nodes), tend to incorporate strange letter sequences in their sub-lexical lexicon. To observe the relationship between learning legal/illegal novel words and sensitivity to orthographic markedness in their languages, Experiment 5 was conducted. Results showed that the bilingual children group whose languages differed in orthographic regularities showed a negative correlation.

Our general results (Chapter 4) made two new contributions to the field. The first is the finding that sensitivity to orthographic markedness changes across development. Sensitivity to orthographic markedness in the second language changes around the age of twelve, a critical age for other developmental changes (cerebral maturation, Giedd et al., 1999; implicit learning, Janacsek et al., 2012; verbal fluency, Sauzéon et al., 2004). Sensitivity to orthographic markedness in the L1 changes with age in early second language learners but also after learning a second language later in life, indicating that the language system is adaptive (Kroll et al., 2014; Schmid, 2008). The second new contribution was the finding that bilingual children whose languages differed in terms of their orthographic regularities were able to learn both legal and illegal novel words to the same extent while this effect was no longer found in adults. Again, developmental changes in early learners as well as language specific bilingual experience appear to influence this aspect of novel word learning.

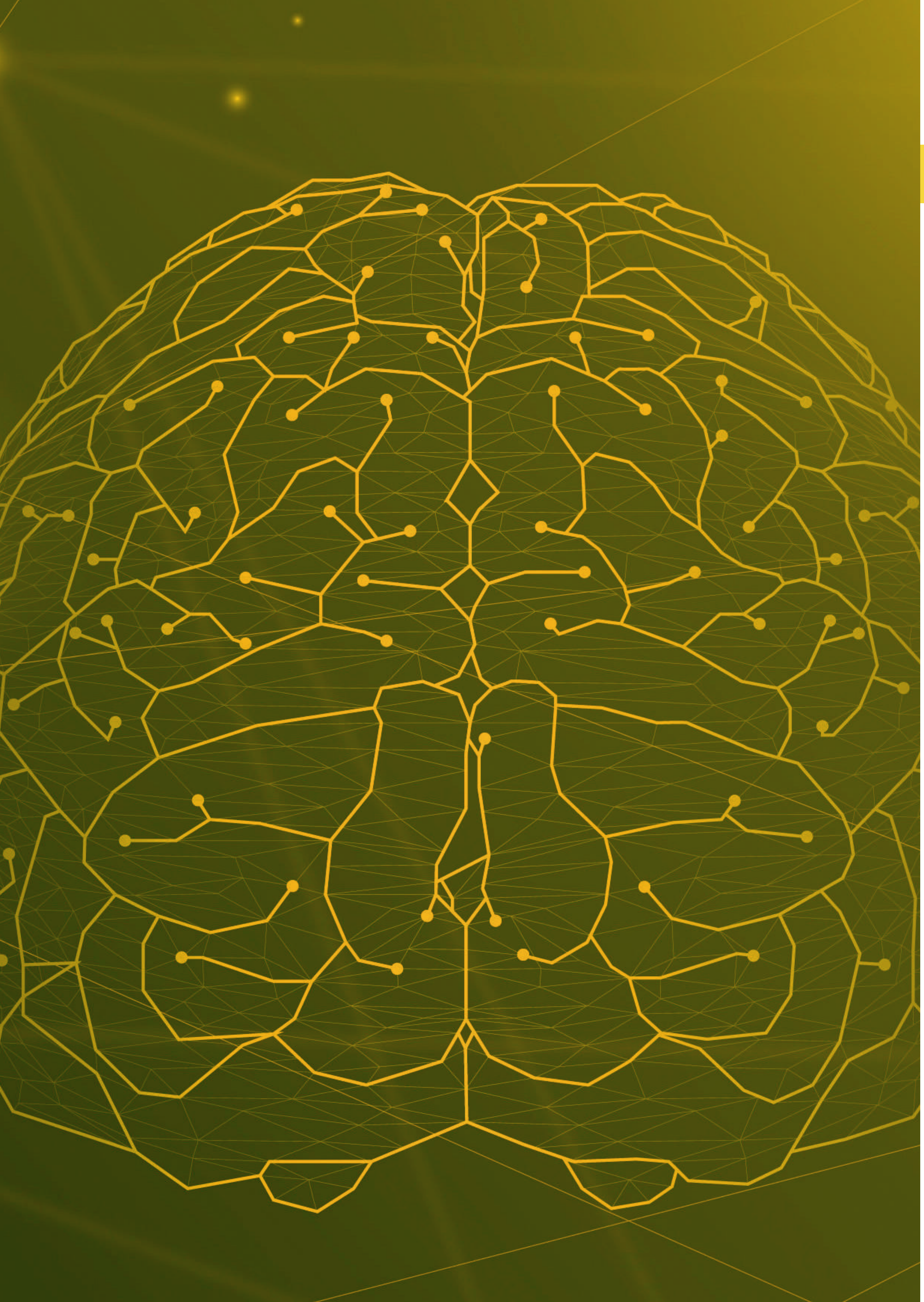
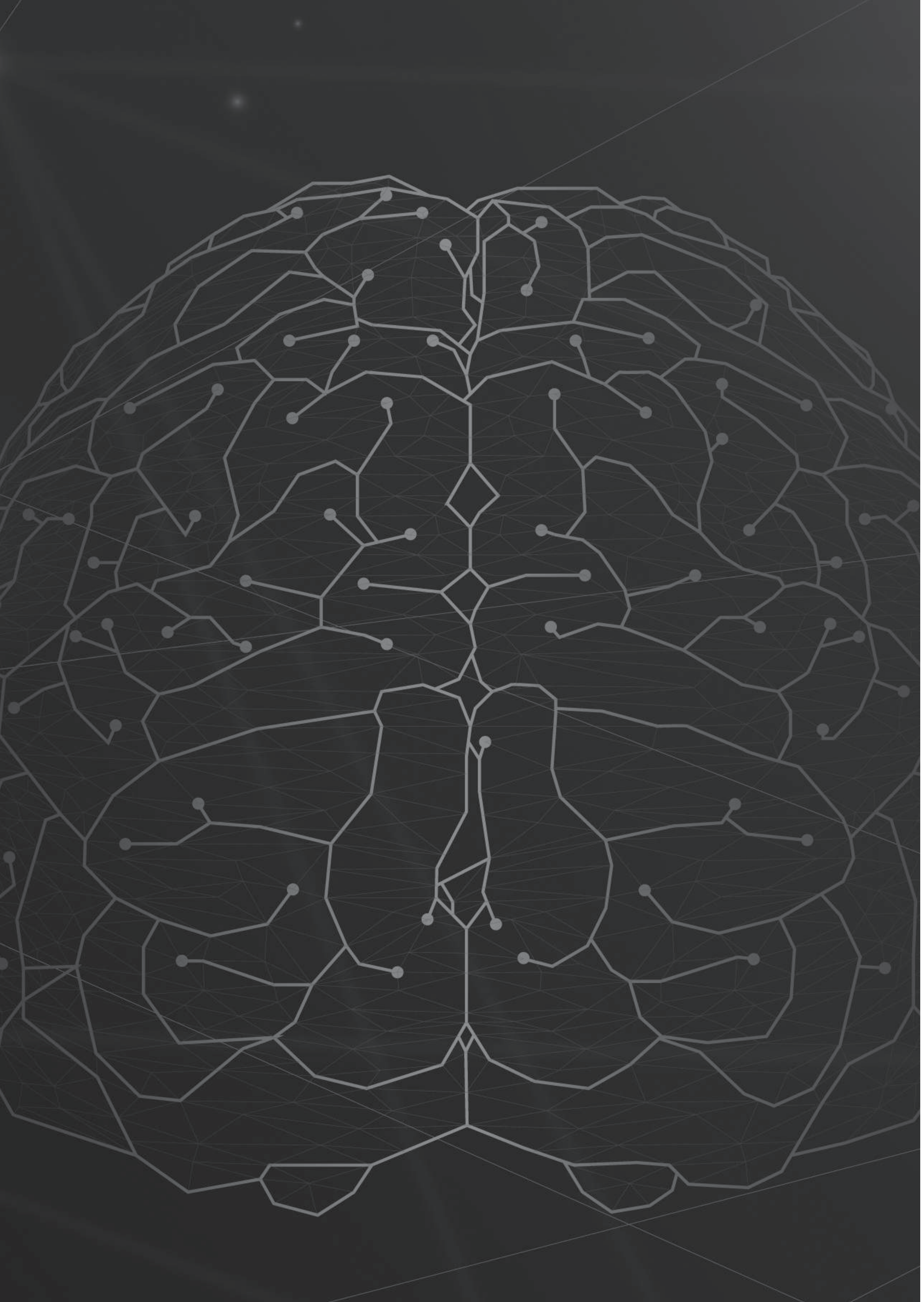
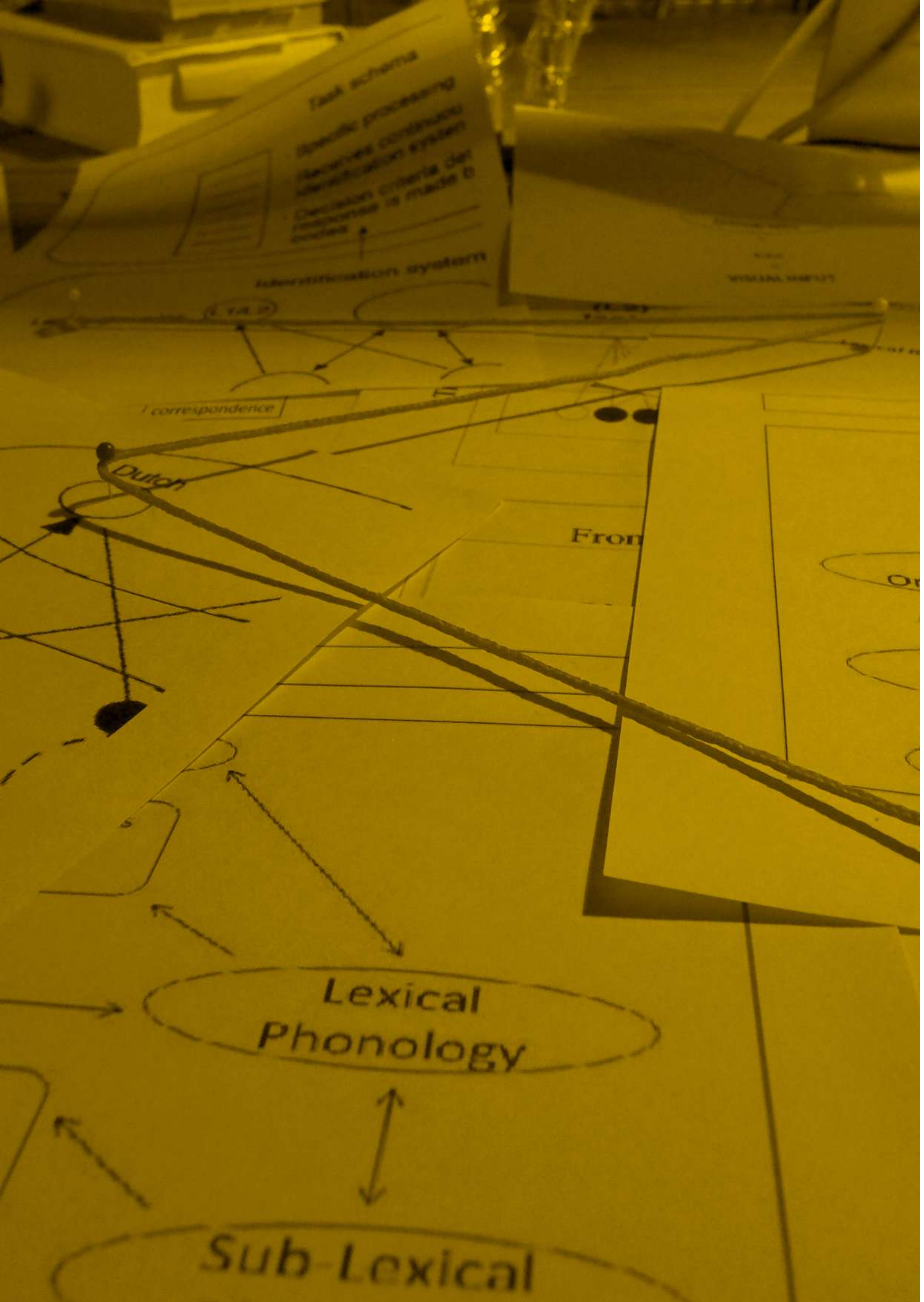


Table of contents





Chapter 1
General introduction

Long ago, humans realized the value of recording information in writing since spoken language is so ephemeral. Writing also permits a more optimal form for learning, sharing knowledge and communicating with others. Thus, humans developed ways to write down their oral languages. Written language can be defined as an artifact of culture. They represent spoken language using a standardized orthographic system, a set of rules that permits the visual representation of spoken words (*e.g., spelling, capitalization, emphasis, punctuation*). These representations make use of different types of symbol systems. For instance, logographic systems map one word to one symbol (*e.g.,* 犬 *is the symbol for dog in Japanese*). Another way to represent spoken words is to represent each individual sound in the spoken language (*sounds/phonemes*) by one visual symbol (*letters/graphemes*). These are alphabetic systems. The majority of written languages in Europe (*Spanish, English, Italian, etc.*) are alphabetic.

Reading is the act of extracting information from written words. The ultimate goal of reading is to understand a whole text. But in order to do this, readers first have to access the meanings of the words in the text. Readers get access to the meaning of written words by decoding the letters (*orthography*) into sounds (*phonology*), then activating the representation of the words. Learning to read is typically accompanied by formal instruction in how to decode written words. Children generally acquire reading around the age of six, but it takes years for them to become skillful readers.

At the beginning of reading instruction, individuals start to map letters (*graphemes*) onto sounds (*phonemes*). Decoding skills require an ability to connect each individual letter (*grapheme*) with its corresponding sound (*phoneme*). However, as readers become more skillful, they process words as a unit instead of decoding each letter-to-sound. Skilled readers manage to process frequently occurring sequences of letters very efficiently because they have already seen them and converted them into phonological forms so many times. Then, individuals gain the ability to process larger orthographic strings of unfamiliar words, although they are not aware of this automatic chunking and rapid processing.

However, considering that many children experience difficulties with reading, it is clear that decoding mechanisms are not that straightforward. As a result, research has focused on understanding the causes of individual differences in fluent word recognition. This is crucial for improving intervention programs that support students with reading disabilities. Research has demonstrated the importance of phonological processing in reading skills. But, phonological processes are not sufficient to explain all the variance in word recognition skills (*Anthony & Francis, 2005; Torgesen & Hudson, 2006*). Thus, recent attention has focused on the orthographic domain as an additional source of variance (*Cunningham et al., 2001*).

One aspect of learning in the orthographic domain is that readers develop sensitivity to the specific pattern of letter sequences in their language (*sensitivity to orthographic regularities*). Before receiving formal reading instruction, individuals are less aware of letter-to-sound (*grapheme-to-phoneme*) decoding (*Adams, 1994*). However, individuals are able to recognize words as belonging to their languages even before knowing how to decode letters-to-sounds. Ehri (*1995, 2005*) showed that pre-reading children could identify words only on the basis of visual cues such as the specific letter sequences that occur in their languages (*orthographic regularities; e.g., the word look in English might be familiar due to the typically English double vowel*). Pre-readers can detect that such words belong to their language. This is probably because they acquire sensitivity to orthographic regularities which they pick up via implicit learning even after limited exposure to written words. So, it should be clear that implicit learning of orthographic regularities is strongly related to solid acquisition of the written language. This strongly suggests that specific letter sequences (*orthographic regularities*) play an important role in decoding and recognizing written words.

A clear example of the importance of the specific letter sequences in word recognition is when words with different meanings are visually similar (*orthographic ambiguity; e.g., rock* and *lock* or *rack*). Those words have different meanings but the only visual difference between them is a single letter. In such cases, individuals may have some trouble identifying the correct word in order to quickly access its meaning, because other visually similar words are also activated in the process of visual word recognition. These similar looking words compete both visually and semantically in the individual's mind with the target word. Thus, it appears that sensitivity to orthographic regularities not only helps a reader identify written words (*letter processing, identification and encoding*) but also reduces lexical competition during lexical access.

Bilingual readers who have to deal with two languages may additionally face ambiguous language contexts. If languages have different scripts (*e.g. Spanish and Chinese*), this provides strong cues that help determine the language of a word. But, when languages share the same script, such as Spanish and English, bilinguals have to deal with written words that look very similar but may have different meanings in the two languages (*e.g., library in English and libreria in Spanish, which means bookstore*). In such cases, during visual word recognition, a word may compete with other words in an individual's lexicon (*word storage*), not only within a language but also between languages (*Dijkstra & van Heuven, 2002; Van Kesteren, Dijkstra, & de Smedt, 2012*). Thus, bilingual readers need to find cues that help with language identification to ensure they access the correct meaning of a particular word in the target language. In this case, one salient cue would be the unique letters in one the language, such as the letter ñ that exists in Spanish but not in Basque or English. However, when there are no simple letter-based cues, another intrinsic property of languages such as the specific letter sequences (*orthographic regularities*) that are allowed, can help to differentiate languages.

Indeed, previous research has found that bilingual readers rely on the orthographic regularities specific to each of the languages they know to identify the language of words (*Casaponsa et al., 2014; Casaponsa & Duñabeitia, 2016; Chetail, 2015; Dijkstra & van Heuven, 2002b; Oganian et al., 2016; Van Kesteren et al., 2012*). Orthographic regularities that are specific to a language are considered orthographically marked. For instance, in the case of an English-Spanish bilingual reader, the word *writing* is a marked word in English because the letter sequence *wr* does not exist in Spanish. These salient letter sequences help bilinguals identify the language of a word more quickly and easily.

It is worth mentioning that letters and letter sequences play an important role in reading. Through implicit learning an individual acquires sensitivity to orthographic regularities (*the specific patterns of letter sequences*) and this helps in visual word recognition and language detection. However, little is known about the function of orthographic regularities in the development of the word processing system, specifically in word recognition and in learning to process novel words. The present thesis aims to contribute to our understanding of how specific letter sequences help readers process and acquire written words. This first chapter reviews previous findings. It first provides an overview of visual word recognition theories and the importance of orthographic processes in reading. Then, it offers a deeper review of the research on orthographic regularities and sensitivity to letter sequences and explores how letter sequences are learned.

I. Visual word recognition

The ultimate goal of reading is the comprehension of written language. Readers receive information from written words. Thus, visual word identification is an important initial stage in reading. The goal of visual word identification is to access the correct meanings of words. Automaticity in word recognition (when readers are not aware of the automatic and rapid processing they use to recognize a word, Snow & Kang, 2006) facilitates comprehension of a text. Hence, it is an important achievement that has a great impact on schooling. It has been shown that the more easily a child can identify words, the better their comprehension of texts will be (*Cunningham & Stanovich, 1997*). In fact, fluent visual word recognition is the strongest predictor of reading comprehension from first to at least third grade (*Adams, 1994; Juel, 1988; Vellutino et al., 2007*). Juel (*1988*) also provided longitudinal evidence reporting that students with slow word recognition at the end of first grade are very likely to have difficulties with reading comprehension by the end of fourth grade.

Therefore, research in psycholinguistics has focused on how individuals are able to read a single word (*Grainger & Holcomb, 2010*), from recognizing written words to rejecting meaningless linguistic stimuli. Additionally, many authors have focused on the description of reading processing models. Since the 70s and 80s, several models of visual word recognition have been developed, such as the Logogen (*Morton, 1969*) and Cohort (*Marslen-Wilson, 1989*) models, among others. These models attempt to demonstrate how readers read familiar words by sight and how a reader's ability to decode unfamiliar words is based on the use of letters. In this section, we first describe models of visual word recognition, and then examine visual word recognition in bilinguals and, finally, models that focus on orthographic processes.

i. Models of visual word recognition

As mentioned above, there is a strong relationship between fluent word recognition and reading comprehension. Thus, models attempt to explain theoretically and empirically how readers recognize words. These theoretical frameworks differ in their explanations of visual word recognition. They assume that word recognition moves from a (*highly demanding*) intentional process that requires difficult letter-sound translation to a (*less demanding*) direct process which incorporates the automatic recognition of letters and instantaneous identification of specific words (*Perfetti, 2007*).

The dual route cascade (*DRC*) model (*Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001*) is an example of a dual-route model (*see Figure 1*). This model assumes that two different but parallel routes are used to access the meaning of a word. The first is the lexical route, which processes familiar words, and the other is the sub-lexical route, which processes unfamiliar words. The sub-lexical route decodes words by applying letter-to-sound (*grapheme-to-phoneme*) conversion rules. In this route, readers have to pay careful attention to individual letters. They have to associate sounds with each of these letters in order to construct units that activate the lexicon (*total stock of words*). When they have decoded the letters into sounds correctly, they can access the meaning through both the orthographic and phonological lexicon. After it has been read many times, a word is stored in the orthographic lexicon, and readers can access it using the lexical route. Each word the reader has learned in its printed form has an individual entry in the lexicon. Then, the lexical route recognizes written words without any letter-to-sound (*grapheme-to-phoneme*) mediation. This route is usually faster than the sub-lexical route because it activates single unit representations (*just the orthographic lexicon*) before it activates the semantics. During the learning process, readers rely more on the sub-lexical route, but when a word is already known, the lexical route is predominant (*Coltheart, 2001*).

This model describes how new words are read, but also, how skillful readers are able to read quickly and accurately. The DRC model assumes that letter-to-sound (*grapheme-to-phoneme*) mappings are predefined. However, this model fails to explain how readers cope with non-words or words with different letter sequences: if letter-to-sound (*grapheme-to-phoneme*) routes were predefined, how would it be possible to read letter sequences that do not exist in the orthographic lexicon (*e.g., th does not exist in Spanish*) and therefore do not have lexical representations? Another problem with this model is that there is little transfer of these skills to the reading of texts.

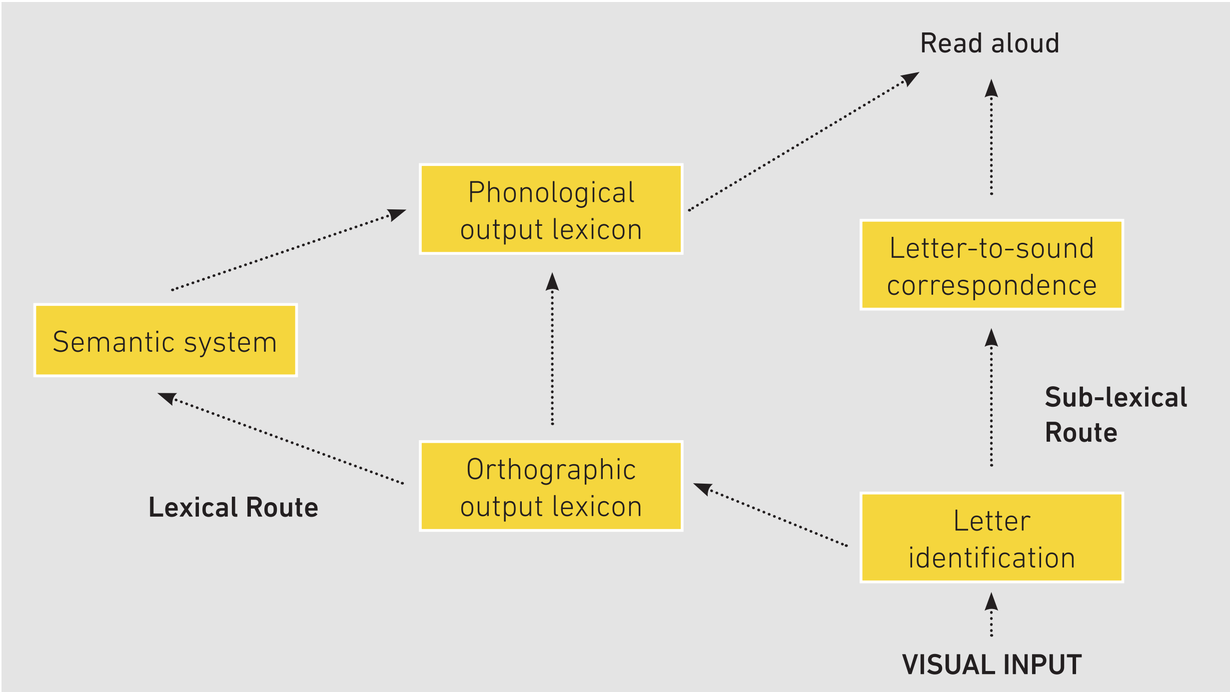


Figure 1. *The DRC model of visual word recognition: different stages of visual word recognition in the two different routes, lexical and sub-lexical (adapted from Coltheart, 2006).*

To solve this issue, the triangle or parallel distributed processing model (*PDP*) model emerged as a main competitor to the DRC model (*Harm & Seidenberg, 1999; Seidenberg & McClelland, 1989*). PDP is a connectionist model that assumes the same mechanism is used to read both familiar and unfamiliar words (*see Figure 2*). This model does not require letter-to-sound (*grapheme-to-phoneme*) conversion rules, because phonological representations are a property of the system. They consist of three layers that are interconnected: orthography, phonology and semantics. Information about words is stored in a distributed manner across these three different layers. When a reader reads a word, the activation spreads in parallel across the layers resulting in an accurate pronunciation of the word. Thus, all information about a word is linked through a neural network, and activation in any part of the network produces activation in related parts. Thus, this model assumes access to meaning across all layers. Unlike the schemas in the DRC model, PDP makes use of representations distributed across a number of units. Orthographic units function as parts of the representations for many different words. For example, ab is a part of the representations for *table, able,* and *baby*. In this way, orthographic representations are assigned phonological representations in a single route. This depends on connections that allow for interactions between a large number of simple codes.

Importantly, this model does not require a predefined set of letter-to-sound (*grapheme-to-phoneme*) mapping rules. Readers learn though exposure to written words, using their knowledge of how spoken words sound and contextual clues that help to make predictions and inferences. With repeated exposure to printed words, stronger connections are formed between different units in the network. In this way, the model can capture statistical learning regularities between orthography and phonology at the sub-lexical level (*Harm & Seidenberg, 1999*). This also allows the system to process unfamiliar words. Unlike dual route models that propose separate paths for familiar and unfamiliar words and the sequential activation of different processors, connectionist models describe a dynamical system that activates the meaning of the word over several steps. Nevertheless, both models assume that letters play an important role in processing written words. The following section will focus on models that attempt to describe orthographic processing as an important cue in visual word recognition.

Chapter 1 General introduction

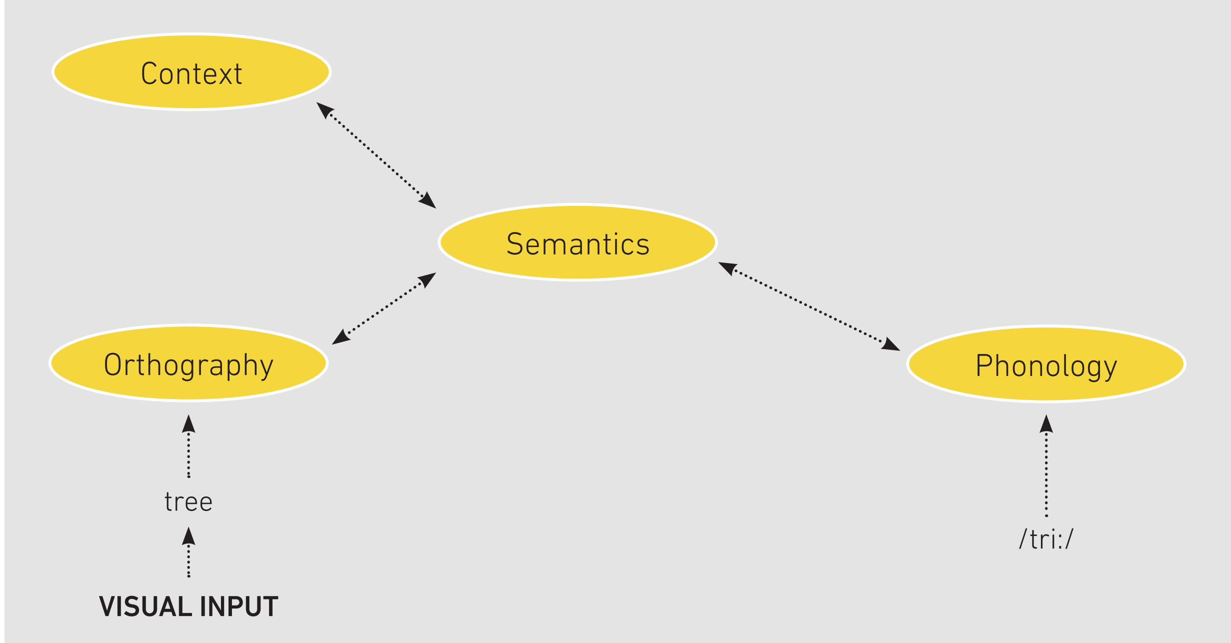


Figure 2. *The PDP or "triangle" model of visual word recognition (adapted from Seidenberg & McClelland, 1989)*

ii. Orthographic processing in visual word recognition

Learning to map letters onto sounds is an important stage of visual word recognition, not only for beginner readers but also for skillful readers who need to read new words (*new vocabulary, pseudowords, other languages*). Thus, there is a heavy reliance on orthographic processing in the development of reading fluency and visual word recognition. Letters need to be identified in order to be decoded into sounds that help the reader recognize a word and access its meaning.

In visual word recognition, the eyes are fixed on the target word and they send visual information to the brain. The brain detects the visual features of the target word. After this, the brain starts to identify the letters. In this step, the complications for recognition begin: the same letter can be written in different manners, such as in different fonts, cases or sizes (*e.g.,* ***M***, m, m, m, *etc.*). However, letters maintain common features that help the reader to identify them. Thus, identifying individual letters is an important and critical step in visual word recognition (*Grainger & Ziegler, 2011; Polk et al., 2009*).

Additionally, readers have to distinguish between words that are visually similar but where just one letter can completely change the meaning (*e.g., rock and lock or rack*). In such cases, it is clear that not only letter identification but also encoding the locations of letters in a word are crucial first steps in visual word recognition. This is especially important in alphabetic languages because of the large number of letter combinations they allow for. Estes and colleagues (*1976*) found that the percentage of bigrams (*two consecutive letters*) reported in a letter report task was higher for low-frequency bigrams (*see also, Frankish & Barnes, 2008; Perea & Lupker, 2003*). In line with this findings, other researchers have studied letter position coding though the transposed letter effect. This effect tests how a word is processed when the positions of two letters in a word are switched to create a new non-word. Perea and Lupker (*2003*) suggested that letter positional coding may have a phonological component. However, years later, Perea and Carreiras (*2006*) tested the effect of transposed letter legality in a primed lexical decision task. The authors found that an illegally transposed letter (*e.g., comsos, does not exist in Spanish, thus, is an illegal bigram*) facilitated word processing (*e.g., cosmos*), resulting in faster reaction times compared to legally transposed letters (*e.g., vebral and verbal*). They concluded that these effects were orthographic and failed to find any clear evidence supporting a role for phonology. Current evidence supports word recognition models that take letter analysis into account. Thus, models of visual word identification should try to explain the sub-processes for letter identification and letter location, which many older models did not account for (*see Section i. Models of visual word recognition*).

An important example is the Interactive model (*IA*) model developed by McClelland and Rumelhart (*1981*). These authors established different levels involved in letter recognition: feature detectors, letter detectors and word detectors (*see Figure 3*). They suggested that position is coded as if the words were strings, with letter identification and position processed in parallel. For instance, if a reader reads the word tear, the t is in position 1, e in position 2, and so on. Each position is associated with a single identification of the letter. This model is used to explain the word superiority effect, which refers to the phenomenon that people recognize letters more easily if they are presented within words than when they are presented in isolation or within non-words (*Cattell, 1886; Reicher, 1969; Wheeler, 1970*).

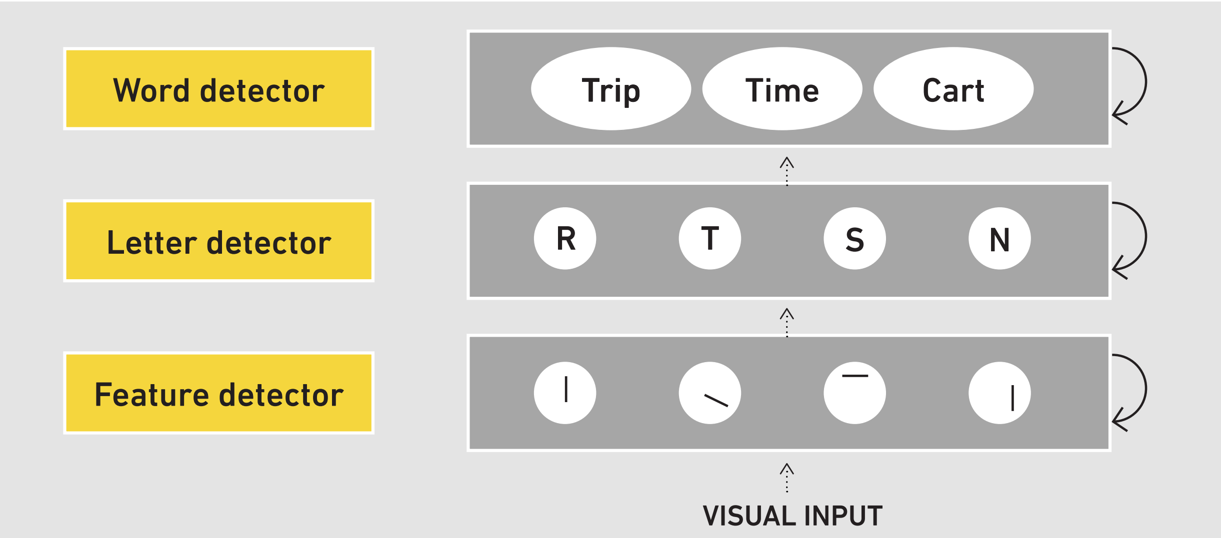


Figure 3. *The interactive model (adapter from McClelland & Rumelhart, 1981). The picture represents the identification of t in the first position in a word and model interconnections.*

The interactive model is useful for the recognition of anagrams (*i.e., words created by rearranging the letters of the original word to make a new word or phrase; e.g., elbow-below*) but it does not explain readers' tolerance for letter strings with transposed characters. It is known that readers can read texts with altered letter order (*e.g., Cambridge text "Aoccrding to a rscheerch at Cmabrige Uinervisty..."; see Massol, Duñabeitia, Carreiras, & Grainger, 2013 for a review of the transposed letter effect*). Researchers have tried to account for letter positions in their visual word identification models. For instance, Grainger and Van Heuven (*2004*) assume that readers encode the relative position of letters by activating pairs of letters. They propose an open bigram model (*see Figure 4*), in which the position of the letters is determined by the anagrams activated. They further describe two different types of positional bigram activations: contiguous and noncontiguous (*e.g., the word tear has the contiguous bigrams te, ea, ar and the non-contiguous bigrams, ta, tr, er*). Activation of both contiguous and noncontiguous bigrams would explain the tolerance and lack of difficulty skilled readers experience when reading words with transposed letters.

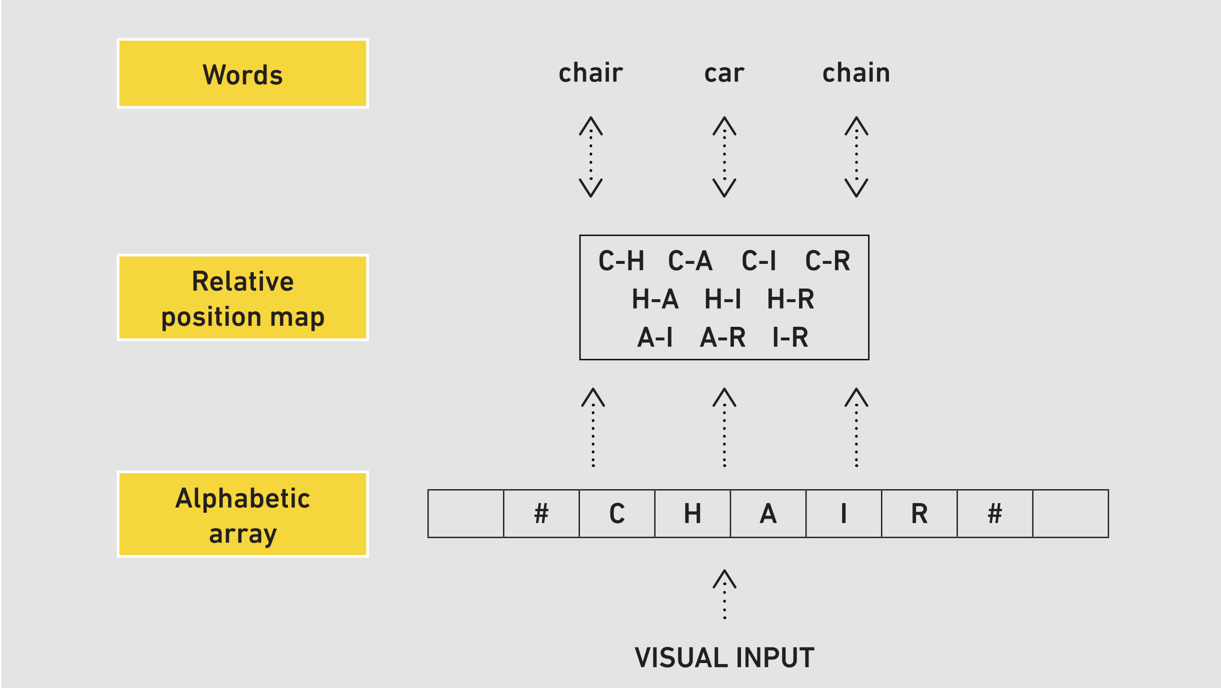


Figure 4. *Open bigrams model (adapted from Grainger & Van Heuven, 2004).*

In order to address some weak aspects of the previous model, Grainger and Ziegler tried to represent the visual word recogniton model in a more optimal way. They suggested a dual route model for orthographic processing (*Grainger & Ziegler, 2011; see Figure 5*). This model is a hybrid model that implements the strengths of the two main types of word recognition models (*see Section i. Models of visual word recognition*): dual-route models (*e.g., DRC model*) and connectionist neural models (*e.g., PDP model*).

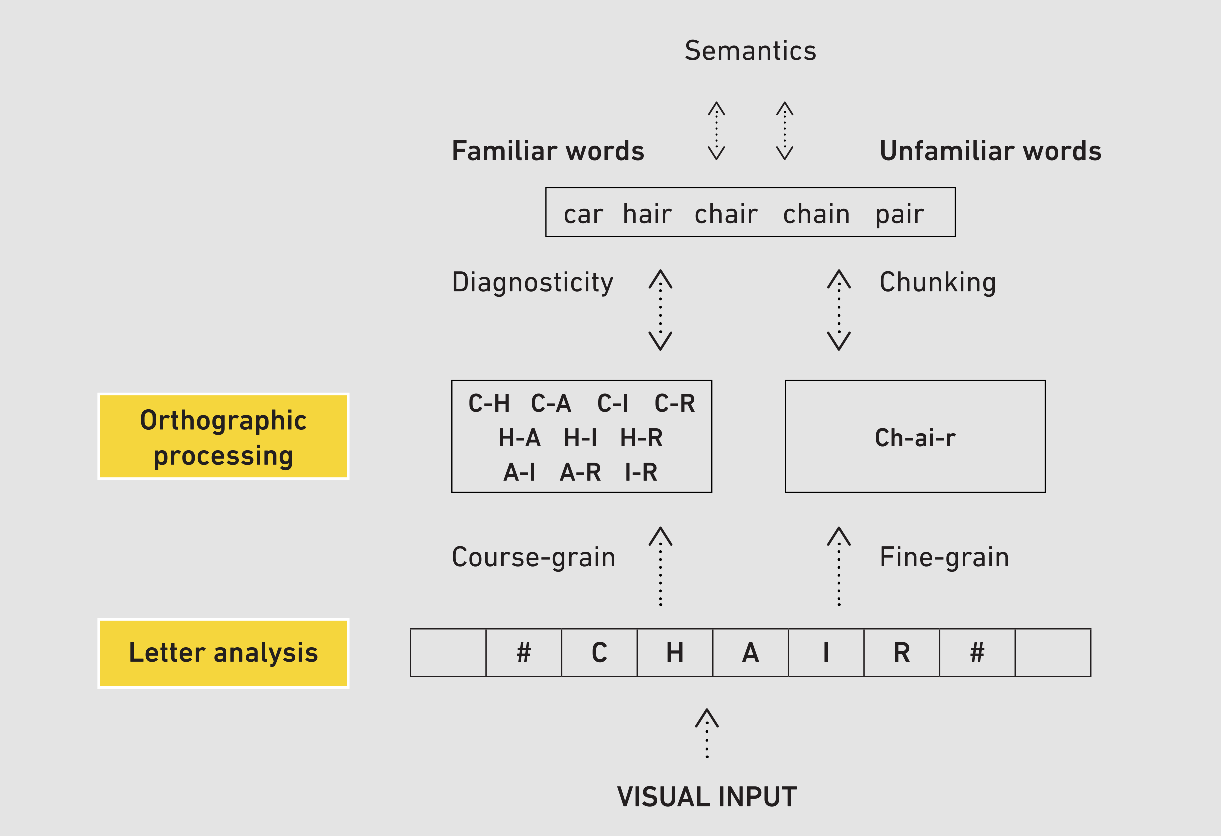


Figure 5. *Dual route model (adapted from Grainger & Ziegler, 2011)*

The authors implemented separate routes for reading familiar and unfamiliar words (*as in the DRC model*) but described routes based on connectionist neural networks (*as in the PDP model*). The model describes two types of orthographic processing that take place after the visual feature analysis of letters: coarse-grained and fine-grained. These two types of orthographic processing differ in terms of how they process letters (*identification and recognition of letter position*). Familiar word recognition is based on coarse-grain orthographic processing, which is defined by the diagnosticity constraint. Diagnosticity optimizes the processing of orthographic input by aiming to process letter identity and the approximate letter positions of the minimal number of letter(*s*) that allow for identification of the word (*i.e., the activation of the correct orthographic representation*). On the other hand, the fine-grained route processes unfamiliar words, as defined by the chunking constraint. The fine-grained orthographic processing assumes a more detailed analysis of letter processing. Thanks to this fine-grained orthographic processing, readers may access word meaning though phonological representations. To do this, it is necessary to encode different letter identities in their precise order. Thus, fine-grained processing differs from coarse-grained processing in that it requires both precise letter identity and letter position coding. In these models written words are processed based on visual feature detectors that lead to position-specific letter activation, providing more concrete implementations of visual aspects of orthographic processing.

At the level of orthographic processing, a combination of these two models is optimal. It helps adapt orthography and visual demands for lexical reading. This allows reading to be maximally efficient because different orthographic constraints can be placed on the processing of familiar and unfamiliar words (*Grainger & Ziegler, 2011*). This model is in line with a previous study of transposed-letter primes (*Acha & Perea, 2010*) which showed that altering the order of letters in pseudo-homophones disturbs fine-grained orthographic processing because misplaced letters do not prompt activation of phonological representations. In this experiment, the authors found that both transposed-letter primes based on real words and pseudo-homophone primes without letter transpositions generated priming effects, but there was no priming effect with transposed letter primes based on pseudo-homophones.

iii. Models of bilingual visual word recognition

Previously presented models of visual word recognition only described how a reader recognizes a word when it competes with other words within their language. But what happens in the bilingual mind, where two languages with different words (*maybe using different sounds or letters*) compete for activation? Bilingual models attempt to disentangle how a bilingual reader recognizes a word. As mentioned above, when any reader reads a word, it competes with other words within their language. Previous models have shown that the target word competes with its orthographic neighbors. But, in the bilingual mind, the target word also competes across languages. For instance, the target word table will compete with words within one language such as *tank* or *taste* as well as with words across languages such as *tabla* or *tambor* (*the Spanish words for board and drum*). In this section, some visual word recognition models for bilinguals are reviewed.

Models of visual word recognition in monolinguals have described the recognition of a word as the retrieval of orthographic representations from the mental lexicon corresponding to the input letter string Grainger & Dijkstra, 1996). Thus, some authors who have attempted to present a model in the bilingual domain took the basic architecture and parameter settings from the monolingual Interactive Activation model (*McClelland & Rumelhart, 1981*). This model is the Bilingual Interactive Activation model (*BIA+ model; Dijkstra, Van Heuven, & Grainger, 1998; see Figure 6*), which implements top-down language-to word inhibition. Lexical access is considered to be language non-selective and to take place within a single integrated bilingual lexicon. Authors have described that the visual input, formed by a string of letters, affects particular features at each letter position. Subsequently, this activates letters that contain these features and at the same time inhibits letters in which those features are absent. Then, the activated letters activate words in both languages in which the activated letter occurs at the same position. The language nodes collect activation from words in the language they represent and inhibit active words in the other language. The activation of the language nodes reflects the amount of activity in each language. This model was based on the empirical evidence available at the time. However, this model fails to explain the representation of the processing of orthographic, phonologic and semantic codes, of cognates, of language node confounds, linguistic and non-linguistic context effects and the relationship between word identification and task demands.

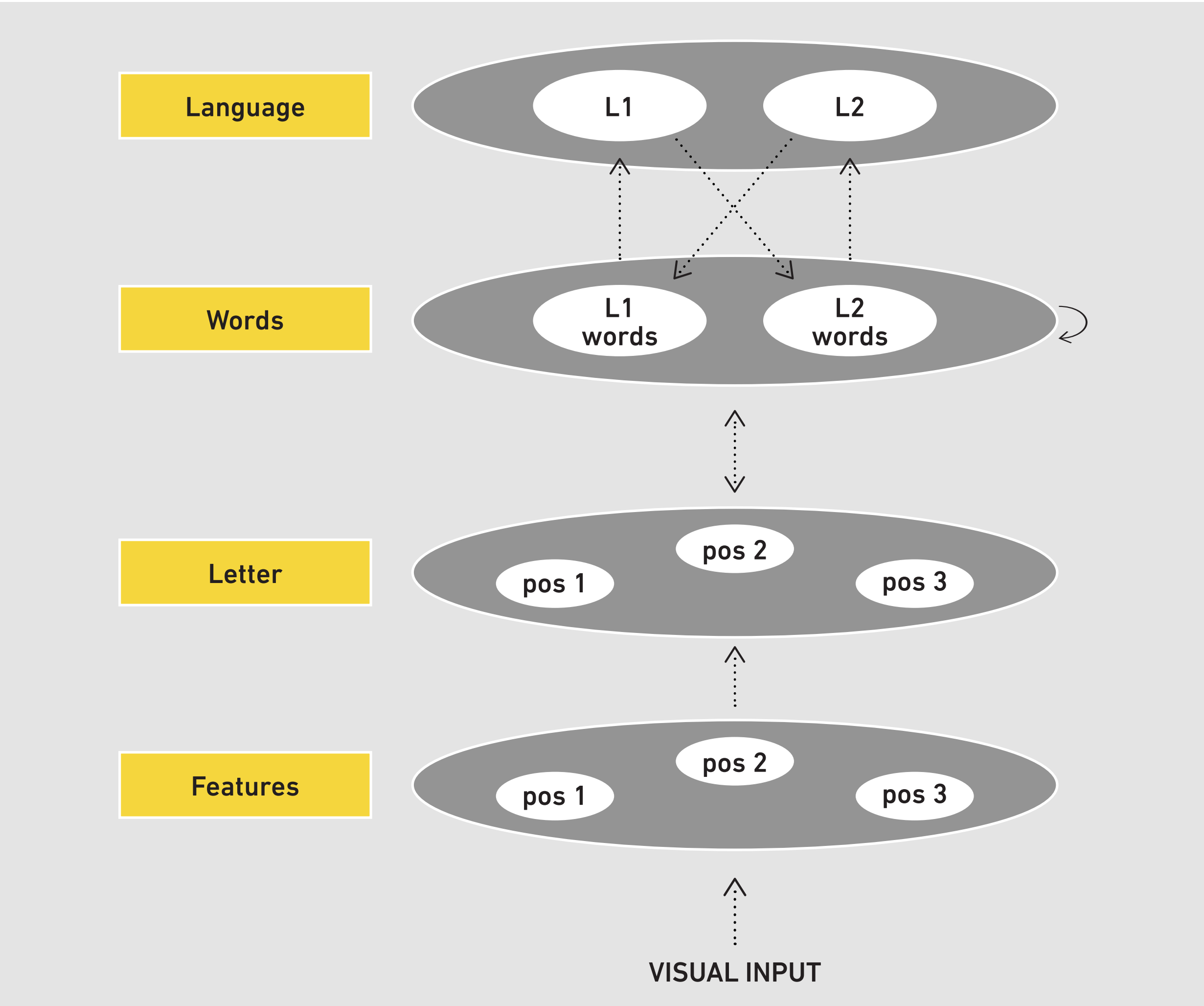


Figure 6. *The Bilingual Interactive Activation (BIA) model (adapted from Dijkstra et al., 1998*

To address these issues, Dijkstra and van Heuven (*2002*) proposed an updated model, in the Bilingual Interactive Activation Plus model (*BIA+ model; Dijkstra & van Heuven, 2002*). These researchers tried to solve previous problems with the BIA model by introducing three major changes: implementing the distinction between a word recognition system and a task/decision system, modifying the function of the language node and implementing orthographic, phonologic and semantic processing (*see Figure 7*). The first change reflects Green's (*1998*) ideas on task schemes and task control because he said that the inhibitory control model exercises control over processing in the lexico-semantic system as specified by task condition.

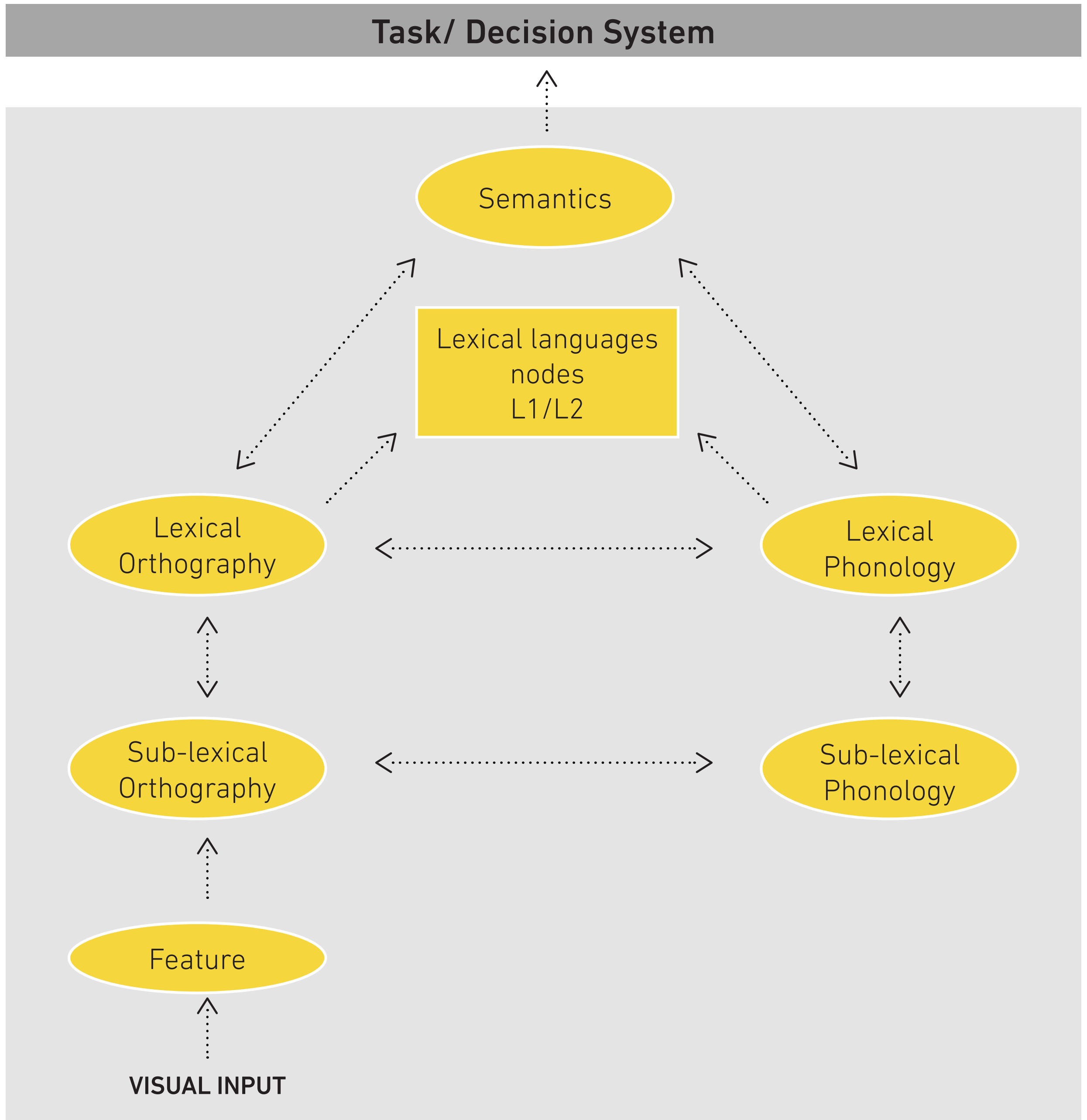


Figure 7. *The Bilingual Interactive Activation Plus model (adapted from Dijkstra & Heuven, 2002)*

Second, this model includes language nodes as key factors in language comprehension in bilinguals. The model makes two assumptions. First, lexical access is language non-selective by nature. That is, there is co-activation of information in both linguistic systems (*De Groot, 2011*). When a bilingual encounters a written word, activation takes place across both linguistic subsystems. And second, only one integrated lexicon exists. The integrated lexicon assumes interactivity between the visual representation of word orthography, and the phonological, and semantic representations of the word. According to this model, when a bilingual reads a word, the initial bottom-up information activates lexical representations in both languages. Then, language selection takes place in the language nodes, where information from the input is integrated with contextual information about the probability of the current input being from a particular language. The non-target language representations are then inhibited in a top-down manner proceeding from the language nodes to the lexical level representations. Note that the authors describe the native language as impermeable but consider that the second language can be influenced by native language processing.

Then, when a visual input is presented, the first stages of word recognition proceed in the same way as in the BIA model. Lexical orthographic candidates are activated in parallel depending on their similarity to the input string. Next, activated orthographic word candidates activate their corresponding phonological, semantic representations. According to the model it is the similarity of the input word to the internal lexical representations that determines their activation and not the word's language membership. The larger the overlap between the input string and a representation in the mental lexicon, the more the internal representation is activated. As a consequence, in the case of two languages with alphabetical writing systems, the number of activated orthographic candidates is determined by factors such as the neighborhood density and frequency of the target word and its within-and between-language neighbors.

However, this model fails to account for the lack of masked translation effects in the L2-L1 direction found in several studies masked translation paradigms exhibit the automatic co-activation of two languages by using target words preceded either by their translation or by unrelated words in the other language and ; e.g., ####-*CASA*-*HOUSE* vs. ####-*VASO*[glass]-*HOUSE*; (*Dimitropoulou, Duñabeitia, & Carreiras, 2011; Duñabeitia, Dimitropoulou, Uribe-Etxebarria, Laka, & Carreiras, 2010*), or the early language switching effects found in ERP components typically related to the mapping of sub-lexical units onto lexical representations (*Duñabeitia et al., 2010*). Thus, Van Kesteren and collegues (*2012*) proposed an extension of the BIA+ model, the BIA+ extended model *(see Figure 8*). They demonstrated a language decision advantage for words that contain language-specific orthography in one of the languages. They proposed a direct link between sub-lexical information and language membership, adding a sub-lexical node (*see Figure 8*).

Furthermore, some authors attempted to update the model based on experimental results. For instance, Oganian and collegues (*2016*) suggested the strength of the connections between sub-lexical and lexical units in the language nodes. They said that word recognition is modulated by both sub-lexical information (*bigram frequency*) and lexical information (*neighborhood size*).

Casaponsa and collegues (*2014*) also corroborate the presence of two routes to access language membership information, the sub-lexical and lexical routes. These two interconnected routes follow interactive activation principles and appear to be operating in parallel during reading. The presence

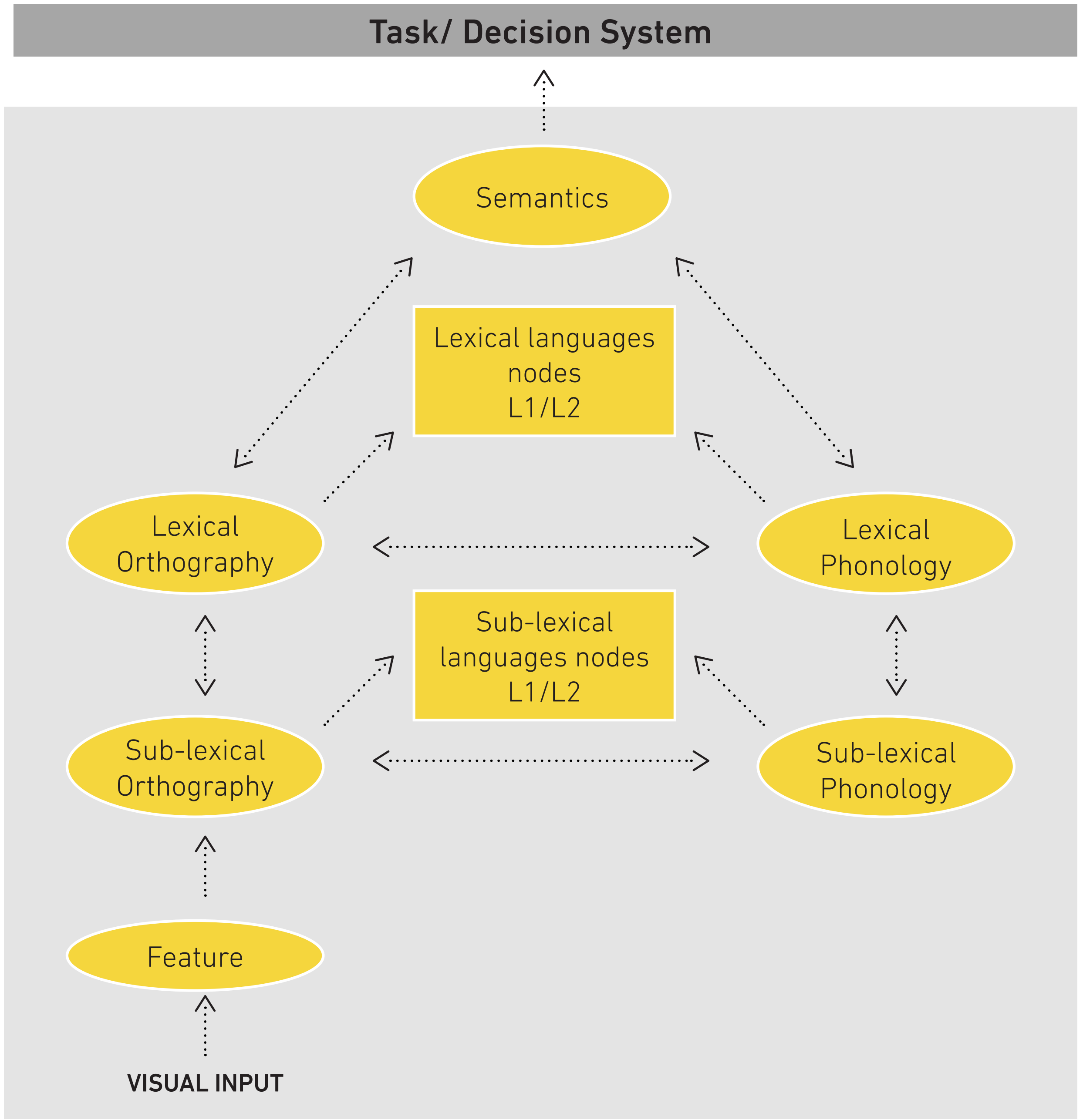


Figure 8. *The Bilingual Interactive Activation Plus model extended (adapted from Van Kesteren et al., 2012). A sub-lexical language node has been added to the BIA+ model.*

of language specific orthographic information mediates lexical activation and ultimately visual word recognition. Studies have shown that bilingual readers recognize the language of a given word faster when the word has letter sequences that only exist in the target language (*e.g., the letter sequence **ck** in English, as in the word duck, does not exist in Spanish*) (*Casaponsa et al., 2014; Van Kesteren et al., 2012*). Integrating these findings with the BIA + extended model (*Van Kesteren et al., 2012*) suggests that sub-lexical orthography supports conscious language attribution by restricting unnecessary cross-language lexical activation (*Casaponsa et al., 2014; Oganian et al., 2016; Van Kesteren et al., 2012*). But Casaponsa and collegues also demonstrated the automaticity of sub-lexical language detection mechanisms. Casaponsa, Carreiras, and Duñabeitia (*2015*) conducted an experiment examining event-related potentials (*ERPs*). Participants completed a masked language switching priming paradigm in a single language, in which half of the masked words were marked with language specific orthography. Their results demonstrate that the bilingual mind unconsciously detects language changes only when sub-lexical language (*language specific orthography*) information is available and it can do this as early as 200 ms.

Casaponsa and collegues proposed two separate sub-lexical language routes, orthographical and phonological, in order to allow for differential language specific effects to emerge from orthographic

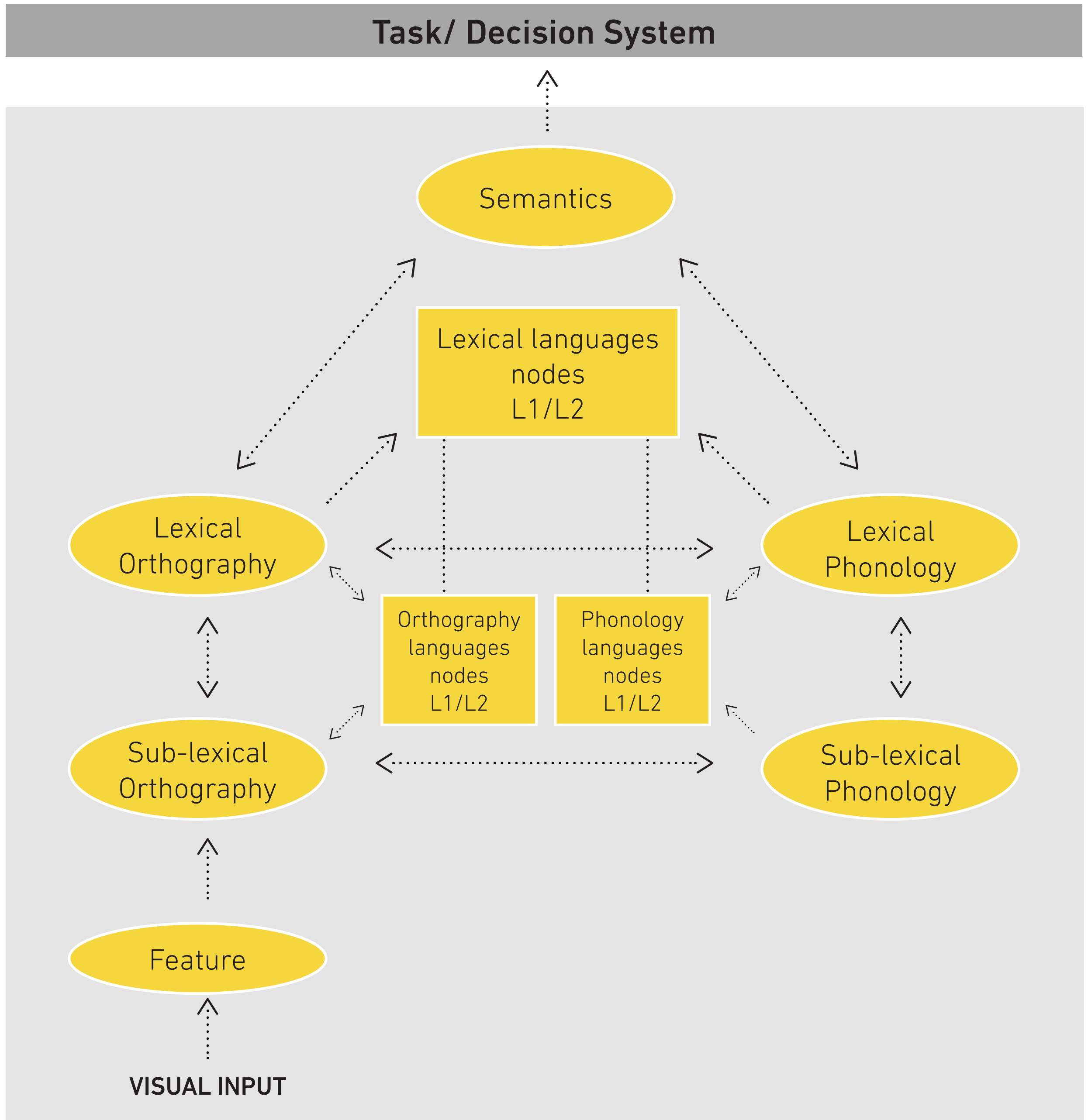


Figure 9. *The BIA + extended model (adapted from Casaponsa et al., 2014).*

regularities and phonological regularities. The interconnections of these two routes are expected to be mediated by the intrinsic characteristics of the languages. In addition, they suggested that a) sub-lexical and lexical language nodes are interconnected with sub-lexical and lexical representations permitting top-down inhibitory/excitatory regulations of activation and b) sub-lexical language nodes are interconnected with the lexicon, permitting language-specific lexical access (*see Figure 9*).

II. Orthographic regularities

Orthographic regularities refer to the distribution of specific letter sequences (*orthographic sub-lexical units of words*), without direct reference to high-order levels of processing of, for instance, phonological and morphological units (*Henderson, 1980; Massaro, Jastrzembski, & Lucas, 1981; Seidenberg, 1987*). Orthographic regularity connotes an understanding of the letter sequences that govern an individual's own language (*Apel, 2011*). These letter sequences include knowledge of how a letter(*s*) may represent speech sounds, positional and contextual restrictions on the use of letters, and how letters can and cannot be combined (*e.g., tx is not a legal combination in Spanish*). People learn orthographic regularities by extracting regularities from visual stimulus inputs (*Conway, Bauernschmidt, Huang, & Pisoni, 2010; Krogh, Vlach, & Johnson, 2013*). After minimal exposure to written words, individuals acquire implicitly expectations for letters and letter sequences which occur at different frequencies and/or in different word positions. Thus, readers become sensitive to the distribution of the letter sequences in their language (*s*) (*Chetail, 2015; Chetail & Content, 2017; Samara & Caravolas, 2014*).

It has been shown that even before formal reading acquisition, individuals are sensitive to orthographic regularities. Five-year-old children, who have almost no experience with reading, are already sensitive to orthographic regularities (*Apel, 2010; Wolter & Apel, 2010*) and they effectively identify violations to orthographic regularities, that is, letter sequences that do not occur in a given language (*Levy, Gong, Hessels, Evans, & Jared, 2006; Ouellette & Senechal, 2008a, 2008b*). Cassar and Treiman (*1997*) conducted an experiment in which they presented five-year-old English-speaking children with pseudowords that contained orthographic violations, namely initial consonant doublets (*e.g., ffoge*). They found that children detected this violation easily, showing that children with almost no reading experience showed sensitivity to orthographic regularities. Starting in kindergarten, children acquire implicit knowledge of specific letter(*s*) (*e.g. one sound corresponds to a consonant doublet*) and accept or reject different letter sequences as belonging to their languages (*Cassar & Treiman, 1997; Treiman & Kessler, 2006*). These findings were replicated with six-year-old French speaking children (*Pacton & Fayol, 2000*) and seven-year-old Finnish speaking children (*Lehtonen & Bryant, 2005*), showing that once acquired, sensitivity to orthographic regularities persists over time. Orthographic regularities are developed early on and continue to develop over the elementary school years. Along this trajectory, Treiman and Kessler (*2006*) described developmental changes in sensitivity to orthographic regularities: English speaking children in first grade were sensitive to initial consonants but did not become sensitive to final consonants until fourth grade.

The idea that sensitivity to orthographic regularities is used during visual word recognition and that readers employ their implicit knowledge of orthographic regularities to process letter strings emerged decades ago because it was noted that this knowledge influences the way readers perceive written words (*Estes, 1975; Henderson & Chard, 1980; Miller, Bruner, & Postman, 1954*). A body of evidence suggests that when reading, people make use of orthographic (*sub-lexical*) units (*e.g., grain size*) that are more complex than single letters to facilitate word perception. The most frequently examined sub-lexical clusters or grain sizes are bigrams (*a cluster of two consecutives letters*), (*Andrews, 1992; Conrad, Carreiras, Tamm, & Jacobs, 2009; McClelland & Johnston, 1977*) and trigrams (*a cluster of three letters*) (*Massaro, Venezky, & Taylor, 1979; Pacton, Fayol, & Perruchet, 2005*). However, bigrams are considered to be a more critical sub-lexical or grain size unit than trigrams (*Shattuck-Hufnagel, 1983*), because bigrams contain the maximum number of letters that an individual can handle at once and also contain the minimal number of letters that can have meaning.

First, bigrams are considered a sub-lexical or grain size unit because bigrams are the maximum number of letters that an individual can handle at once. One clear example is the Cambridge text (*"Aoccrding to a rscheerch at Cmabrige Uinervisty…"; see Velan & Frost, 2007*). Despite the fact that the letter order of many words was distorted (*e.g., judge-jugde*), people were able to read the text without problems. This demonstrated that readers are tolerant to some order changes of the letters in words (*Rayner, White, Johnson, & Liversedge, 2006; Velan & Frost, 2007*) even if they know that the distorted word isn't an exact match (*Norris, Kinoshita, & van Casteren, 2010*). But even more precisely, it demonstrated that consecutive letters can be misplaced and still read (*Kinoshita & Norris, 2013*). The special way in which the human brain encodes the position of letters in texts is also in line with the transposed letter effect (*see Section ii. Orthographic processing in visual word recognition*). Experiments done using the transposed letter task demonstrate that primes formed by transposing two adjacent letters in a target word can facilitate word recognition compared with control primes (*e.g., judge-jugde*; Perea & Carreiras, 2008; Perea & Lupker, 2003; Schoonbaert & Grainger, 2004).

However, recognizing a "sub-lexical or grain size unit" is also facilitated by the perception of the internal structure of words. Cottrell and Mason (*2009*) affirmed that the structure of the language determines the nature of the "sub-lexical unit". First, the internal structure of the word should be determined by the characteristic mapping between orthography and phonology in a given language. Opaque languages have more letter-to-sound (*grapheme-to-phoneme*) correspondences at the level of bigrams than trigrams. In English, it is more common to map from bigrams to phonemes (*e.g., the bigram ea as in head to /e/*) than trigrams to phonemes (*e.g., the trigram eou as in gorgeous to /ə/*). Berndt and collegues (*1987*) showed that in English there were 98 bigrams mapping onto one phoneme compared with 13 trigrams (*see also Peereman & Content, 1999 for more examples*). However, this letter-to-sound (*grapheme-to-phoneme*) attribution does not happen in transparent languages (*note that we are referring to bigrams/trigrams formed with consonant(s) and vowel(s) that correspond to one sound*). In transparent languages (*e.g., Spanish, Catalan, Basque*), each letter corresponds to a sound. Thus, in these languages, the internal structure of the word is determined by the minimal familiar unit of pronunciation that can take on one vowel sound. The fact that bigrams more frequently correspond to syllables than trigrams may support the idea that the bigram is the "sub-lexical or grain size unit" (*e.g., 333.303 vs. 244.287 occurrences per million of bigrams and trigrams in French, respectively, based on Chetail & Mathey, 2010*).

It is worth mentioning that measures of orthographic regularities have mostly been based on the frequency of letter co-occurrences rather than transitional probabilities. Most of the literature on orthographic regularities has manipulated bigram frequency. For example, it is more frequent in Spanish to see the letter sequence *ra* than *xa* (*2.02 vs.0.02 frequency by the appearances per percentage respectively*). Thus, *ra* is a high frequency bigram compared with *xa*, a low frequency bigram. Orthographic redundancy is the general term used to refer to high-frequency and low-frequency bigrams, which are used as manipulations (*Andrews, 1992; Conrad et al., 2009; Seidenberg, 1987*). But going beyond frequencies, researchers have also manipulated the co-occurrence of letters in letter sequences that violate the orthographic regularities of a language, using so-called illegal bigrams (*e.g., jc does not exist in Spanish*). Decades of research has accounted for these slightly different types of orthographic regularities and used illegal bigrams as an experimental manipulation in order to observe sensitivity to the orthographic regularities in a given language (*Andrews, 1992; McClelland & Johnston, 1977; Pacton et al., 2005; Seidenberg, 1987*).

i. The role of orthographic regularities in visual word recognition

Orthographic regularities are part of our implicit language knowledge and they influence the way we perceive written words. Research has studied the influence of orthographic regularities on word recognition and its impact on letter string processing. Thus, orthographic processing (*letter processing*) receives a position in current models of visual word processing due to the fact that orthographic regularities play different roles during visual word recognition, from the analysis of stimulus input to the access to word meaning. We can define four different roles orthographic regularities play during visual word recognition.

One important role is that orthographic regularities facilitate the identification of letters in the initial steps of word perception. It has been shown that letter recognition is easier when letters are embedded within words than when they are presented in isolation or within non-words (*word superiority effect; Cattell, 1886; McClelland & Johnston, 1977; Reicher, 1969; Wheeler, 1970*). But, it is also known that letter perception increases with single letter frequency. A study conducted by New and Grainger (*2011*) reported that the time required to decide whether a symbol is a letter or not decreases with single letter frequency, meaning that high-frequency letters are easier to detect in any given word.

Another role is that orthographic regularities help readers to encode the order of letters in strings. Sometime the perception of a letter can be noisy during visual word recognition because depending on the sequence of the letters they could form different words (*e.g., cat, act*). Using a forced-choice task, it has been shown that good readers (*adults and children*) detect letters better in their most frequent positions, especially in the first and last positions in the word, (*i.e., participants had to choose the most frequent position at which a given letter occurs; Katz, 1977; Pitchford, Ledgeway, & Masterson, 2008*). Thus, these models have implemented flexible encoding of letter position (*see Section ii. Orthographic processing in visual word recognition models*).

Another role of orthographic regularities is that they help to reduce lexical competitors during lexical access. Words with low-frequency bigrams are processed more efficiently than those with high-frequency bigrams (*Broadbent & Gregory, 1968; Owsowitz, 1963; Rice & Robinson, 1975; Westbury & Buchanan, 2002*) because low-frequency bigrams are remarkable and not many words share this bigram (*thus the words are also low-frequency*). In line with these findings, Rice and Robinson (*1975*) conducted a lexical decision task. They found that words with high-frequency bigrams were recognized more slowly than words with low-frequency bigrams. Westbury and Buchanan (*2002*) replicated this result years later. It should be noted, however, that although high-frequency letters are easier to identify (*New & Grainger, 2011*), this does not necessarily imply that words composed of high-frequency bigrams will also be easier to identify. The problem is that high-frequency bigrams are less informative with respect to word identity than low-frequency bigrams, because high-frequency bigrams are embedded in more words and thus more lexical competitors are activated during visual word recognition. This is why sensitivity to orthographic regularities increases with low-frequency bigrams leading the reader to recognize these words more quickly (*Lupker, Perea, & Davis, 2008*).

In line with low-frequency bigrams, research has found that sensitivity to orthographic regularities increases when the letter sequences belong to an individual's own language (*Miller et al., 1954*).

In this case, the authors created non-words that could either follow the orthographic rules of the language (*legal*) or violate the orthographic rules (*illegal*). Miller and colleagues (*1954*) showed that English language participants were more likely to correctly report the number of letters when the letter sequences of the non-words were more similar to the letter sequences of English words (*e.g. vernalist vs. ozhgpmtj*). The same findings were reported by Chambers and Forster (*1975*) with a different task. These researchers demonstrated that readers responded faster to "same answer" in a same-different task when they had to decide on non-words with legal bigrams (*e.g., foon-foon*) than illegal bigrams (*e.g., ftre-ftre*). Baron (*1975*) replicated this same finding. And again, results consistently showed that non-words with orthographic regularities close to those of real words were harder to reject (*Henderson & Chard, 1980*) in a language decision task.

The fourth role of orthographic regularities is that they play an important part in bilingual language detection. When two languages share the same script, orthographic regularities should help to identify language membership. This role will be discussed in more depth in the next section (*ii. Sensitivity to orthographic markedness in bilinguals*).

ii. Sensitivity to orthographic markedness in bilinguals

As seen above, orthographic regularities help readers to process letter strings, to encode the order of letters and to reduce lexical competitors within a language. But when readers have to deal with more than one language, they can also make use of the statistical orthographic regularities of these languages to disambiguate the language that words belong to. In a bilingual environment in which languages have different alphabets (*e.g, Hebrew and Spanish*), the dissimilar scripts themselves provide enough information to easily attribute a word to one of the languages. However, this is not the case with languages that share the same script (*e.g. Spanish and Basque*). Then, readers might have difficulty determining the language of each individual word. Several studies have suggested that in an ambiguous language context, language information is accessed through the lexical representations of words (*Chauncey, Grainger, & Holcomb, 2011; Chauncey, Holcomb, & Grainger, 2008; Dijkstra & van Heuven, 2002b*). However, recent research on visual word recognition in bilinguals using two languages with the same script has identified the sub-lexical characteristics of words as a cue to language selection and recognition. These letter combinations or sub-lexical characteristics of words help to speed up conscious language attribution (*Oganian et al., 2016; Vaid & Frenck-Mestre, 2002; Van Kesteren et al., 2012*) and word identification (*Lemhöfer & Radach, 2009*).

Previous studies have suggested that the frequency of letters and their combinations within a language may play an important role in bilingual language detection (*Grainger & Beauvillain, 1987; Lemhöfer, Koester, & Schreuder, 2011; Lemhöfer & Radach, 2009; Oganian et al., 2016; Thomas & Allport, 2000; Vaid & Frenck-Mestre, 2002; Van Kesteren et al., 2012*). The legality of bigrams in the two orthographies is a reliable marker of language identity. Thus, research has focused on the role of bigram frequency or orthographic markedness. Bigrams (*or words*) can be marked or unmarked. Marked bigrams (*or words*) are those that only exist in one of the languages and have a null frequency of use in the other language (*'tx' as in txakurra, the Basque word for dog, does not exist in Spanish*). Unmarked bigrams (*or words*) are those that contain bigrams that exist in both languages, and normally, in experimental contexts, they are controlled to have the same frequency of use in both languages (*e.g., igel in Basque, rana in Spanish*). And again, bilinguals become sensitive to orthographic markedness after little exposure to written words in both languages.

With this in mind, previous research that has explored language detection mechanisms by manipulating orthographic markedness has found that adults recognize sub-lexical differences in words very quickly (*Lemhöfer et al., 2011; Vaid & Frenck-Mestre, 2002; Van Kesteren et al., 2012*). For instance, Vais and Frenck-Mestre (*2002*) conducted an experiment with high proficient French-English bilinguals doing a language decision task using words with marked bigrams (*e.g., oeuf* (*egg*) in French or kick in English, which contain bigrams that do not exist in the other language) and unmarked bigrams in both languages (*e.g., pont the word bridge in French or drop in English, which contain bigrams that exist in both languages*). They found that participants responded more quickly to marked than unmarked words in their second language (*English*). These results showed that language detection is mediated by the sub-lexical characteristics of the words at early stages of visual word recognition. This means, that when participants saw the word kick they didn't rely on the meaning of the word but rather on a sub-lexical strategy to decide on the language of the word. They appear to have rejected the possibility that this word was French because of its orthographic regularities. Van Kesteren and collegues (*2012*) tested Norwegian-English bilinguals in a language decision task. They demonstrated a language decision advantage for marked words, reinforcing the link between sub-lexical information and language membership. They suggested that language membership could be accessed via lexical information but could also be directly accessed via the sub-lexical information in words.

In line with these findings, some studies have shown that this sensitivity to markedness occurs regardless of the level of proficiency in an L2. Casaponsa and collegues (*2014*) conducted a lexical decision task with Spanish-Basque bilingual adults and Spanish monolinguals with no previous experience with the target language. Participants had to decide whether the target word (*marked or unmarked*) belonged to the first language (*Spanish*) or second language (*Basque*). Both groups

were faster at detecting marked words in Basque, suggesting that even monolinguals are very sensitive to orthographic markedness that violates their orthographic knowledge. These results are observed regardless of language proficiency levels, and even include monolinguals that do not know second languages. In other words, even when adults do not know a language, they develop a certain degree of sensitivity to letter sequences that do not conform to the orthographic rules of their first language and are still able to easily detect words from other languages. This again suggests that orthographic regularity processing occurs at an early stage. Consequently, orthographic markedness recognition is said to take place before access to meaning. Language detection may be guided by sub-lexical processes associated with the detection of non-native bigram combinations (*see section iii. Models of bilinguals visual word recognition, the BIA+ extended model adapted from Van Kesteren, Dijkstra, & de Smedt, 2012*).

So far it has been demonstrated that when bilinguals read words in both of their languages, they use sub-lexical and lexical routes to access language membership information (*Casaponsa et al., 2014*) and also that sub-lexical characteristics reduce unnecessary cross-language activation at the lexical level (*Duñabeitia, Ivaz, & Casaponsa et al., 2014; Oganian et al., 2016*). In addition, Casaponsa and colleagues (*2015*) found that sub-lexical orthographic cue effects on ERP patterns are related to automatic and unconscious processing of language switches. This demonstrates that sub-lexical language information plays a critical role in bilingual word processing and in conscious and unconscious language detection. And it corroborates the idea that intrinsic characteristics of languages such as orthographic regularities are used by bilinguals to disentangle the language of words prior to the activation of the lexical entries in the lexicon.

In a similar vein, Oganian and collegues (*2016*) conducted a related experiment to observe language membership attribution and whether it is helped by sub-lexical statistics and/or lexical statistics. To do so, they conducted a language decision task and created pseudowords where they measured sub-lexical statistics in terms of bigram frequency and lexical statistics in terms of orthographic neighborhood size. Bigram frequency was manipulated as marked and unmarked and orthographic neighbors were manipulated by the size of the orthographic neighborhood in both languages. They found that bilinguals were sensitive to orthographic markedness in both languages. However, the pattern for the languages was different: in the native language, bilinguals were slower and less accurate when sub-lexical and lexical levels provided conflicting language membership information, that is, when the marked pseudoword in the native language had more L2 neighbors. This suggests that language membership information from both levels is integrated in order to make a language decision. This effect was absent in the L2, probably because participants could discard these pseudowords as non-L1 based solely on orthographic markedness.

iii. The role of orthographic regularities in novel word learning in bilinguals

When readers are exposed to a language (*or to several languages*), they pick up statistical orthographic regularities in an unconscious manner. These letter sequences seem to be automatically extracted and guide language processing. Learning orthographic regularities means learning letter order and letter identity. Thus, this implicit learning is very important in acquiring good reading strategies to process words. For instance, a Spanish-English bilingual can easily determine that the word *txakurra* (*the Basque word for dog*) does not belong to Spanish or English. The reader can make this decision simply on the basis of the statistical orthographic regularities of its letter sequences because the bigram *tx* does not exist in any of their known languages.

When learning a new language it is important to acquire new orthographic regularities and thereby a new set of words. In general terms, it has been shown that bilinguals may be better at word learning than monolinguals (*see Hirosh & Degani, 2018 for review*). These researchers suggest that this is because bilinguals have more experience with language learning. This has been shown with both children and adults (*Kaushanskay & Marian, 2009; Kaushanskaya, Gross, & Buac, 2014; Margarita Kaushanskaya, 2012*). However, this set of experiments focused on bilingual languages that did not share the same script (*e.g., English-Mandarin*), and only tested the effect of being bilingual per se without testing the effect of specific orthographic regularities on learning.

More specifically, authors have studied the effects of orthographic regularities in acquiring new words (*Pitts, White, & Krashen, 1989; Sagari, Nation, & Meister, 1978*). Pitts and colleagues (*1989*) replicated a previous finding by Sagari (*1979*) with second learners of English. Participants read a text taken from the movie 'Clockwork Orange', in which some words were atypical vocabulary. As in the film, the characters spoke in English but mixed in Russian slang words (*e.g., using the word bitva for battle and glazz for eye*). Results showed that second language learners acquired these atypical words through reading exposure. This suggests that second language learners could learn new words though exposure even when they were not familiar with the second language's orthographic regularities.

New words may catch the attention of the reader, and this may be why participants accurately learned these atypical words. In line with these findings, Lutjeharms (*1994*) illustrated that bilingual readers are not attracted to words that resemble their first and second languages, because they do not pay attention to words with high frequency letter sequences. However, words that have low

probability in both languages are commonly rated as salient and are more likely to be acquired. They suggested that rare letter sequences are easily acquired because they are more easily noticed and more likely to trigger new representations.

On the other hand, Ellis and Beaton (*1993*) went further and conducted a study in which second language readers learned words in their second language. They showed a different trend for learning new words than Lutjeharms (*1994*) had suggested. Words in the second language that conformed to first language orthographic regularities were memorized more easily. They argued that letter sequences that conform to learned expectations are easier to learn (*Ellis, 2002; Rast & Dommergues, 2003*).

Following this line of findings, Bordag and collegues (*2017*) said that the lack of consistency in these results was due to the fact that they tested two different types of new word learning: incidental and intentional. On one hand, the authors found that bilinguals showed an advantage for learning low frequency letter sequences through incidental learning (*new words learned through reading*). This is in line with Lutjeharms (*1994*) results. On the other hand, when bilinguals learned new words intentionally, there was an advantage for high frequency letter sequences as Ellis and Beaton (*1993*) showed. This suggests that orthographic probability influences not only the extent to which a new word is identified as familiar/unfamiliar, but also the degree of attention directed at the word and its role in a given textual context. Thus, in incidental learning, low frequency sequences are easier to remember. They are more salient in the text because detection is easier due to their significant mismatch with the other words in the text. In intentional learning, bilinguals prefer to memorize high frequency sequences. This is because they have similar vocabulary items in their memory storage, and it is easier to remember words with similar sequences. Orthographic neighbors converge to facilitate the triggering of lexical acquisition, speeding up the detection of new inputs that are similar to existing representations. It is likely that low frequency sequences make greater demands on short term memory and the retention of new items is more difficult (*S. Gathercole, Willis, & Baddeley, 1991*) since greater effort is needed to encode the word.

III. The current thesis

To sum up, when the two languages of a bilingual reader share different written language systems (*e.g. Spanish and Greek*), the dissimilarity of the scripts provides enough information to recognize the language of the word. However, this is not the case with many language pairs. In countries such as Spain, there are regions where two or more languages officially coexist together, as is the case with Spanish and Basque in the Basque Country. While Spanish comes from Latin, Basque comes from a different pre-Indo-European root. Nevertheless, these two languages share the same script. The influence of Spanish on Basque means both languages share most phonemes and graphemes. Given this situation, Spanish-Basque bilingual readers might have difficulties determining the language of each individual word. Therefore, research on visual word recognition with same script language combinations has focused on how bilinguals identify the characteristics of the words that facilitate bilingual language selection and recognition. One of the intrinsic differences between the two scripts are the unique letters in one of the language (*e.g., the letters ñ,c,q,v,w are unique in Spanish because they do not exist in Basque*) that facilitate word identification (*Van Kesteren et al., 2012*). However, these characteristic letters are not the only letters present in written words, so bilinguals must rely on other intrinsic cues. In this case, language specific letter sequences (*orthographic regularities*) help word processing.

Orthographic regularities are learned implicitly by exposure to written words. After only a little exposure, readers become sensitive to these orthographic regularities (*Chetail, 2015*). They help the reader recognize written words and access their meaning, which is the ultimate goal of reading text comprehension. During visual word recognition, orthographic regularities help in the processing of letter strings (*New & Grainger, 2011*), in encoding the order of letters (*Estes et al., 1976; Perea & Carreiras, 2008)* and in reducing lexical competitors within the language (*Chetail, 2015; Owsowitz, 1963; Rice & Robinson, 1975*). But also, in the case of bilinguals whose languages share the same script, orthographic regularities facilitate the identification of the language, making visual word recognition easier.

Bilinguals develop sensitivity to orthographic markedness (*the use of language-specific letter sequences*) in order to identify the language membership of words. Marked words, the ones that have bigrams that do not exist in the other language, are processed faster than unmarked ones (*Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002*). And, it has been shown that orthographic markedness helps with identifying language membership before meaning is accessed (*Casaponsa & Duñabeitia, 2015; Oganian et al., 2016*). Also, it has been shown that unmarked words, which have bigrams common to all the languages that the person knows, are easier to learn as new words (*Ellis, 2002; Ellis & Beaton, 1993*). However, little is known about the function of orthographic regularities across the development of the word processing system, in word recognition and in learning to process new words.

Recent models of bilingual visual word recognition have tried to introduce the importance of letters and/or letter sequences (*sub-lexical*) information (*Casaponsa et al., 2014; Van Kesteren et al., 2012*). Thus, the BIA+ extended model (*Van Kesteren et al., 2012*) proposed two separate language nodes, one for lexical and the other for sub-lexical information. This was based on their finding that bilinguals could detect the language of the word more easily and faster when they could rely on language-specific orthographic regularities. Recently, Casaponsa and collegues (*2014*) suggested a further modification to the BIA+ extended model based on her findings. She suggested a critical role for sub-lexical language cues for bilinguals in both conscious and unconscious language detection mechanisms. Bilinguals access language membership by using the sub-lexical route activated via orthographic markers and the lexical route for unmarked letter sequences. In this case, the sub-lexical route modulates speed by restricting unnecessary cross-language lexical activation. In addition, Casaponsa and collegues proposed two separate sub-lexical language routes, orthographical and phonological, which are expected to be mediated by the intrinsic characteristics of the languages.

Despite the fact that previous versions of this updated model (*BIA+ model; Dijkstra & van Heuven, 2002*) assumed the stability of the system through development, many challenges for the bilingual language system are dynamic (*Kroll et al., 2014*). In this context, we should consider that little is known about the development of the underlying mechanisms for using of orthographic regularities in word processing and learning to process new words. And we can ask how are the integrated-yet-differentiated representations for two languages established in the orthographic (*sub-lexical*) language node during development? To what extent does age play a role in the acquisition of L2 orthographic regularities? To what extent are there transfer effects between first and second languages in terms of orthographic regularities? How is an L2/L3 best acquired? By forming initial associations with an existing L1/L2? By taking advantage of orthographic markedness, we can investigate the mechanisms involved in processing orthographic regularities.

i. Purpose of this thesis

A better understanding of orthographic sub-lexical information development is critical for the field of psycholinguistics and bilingualism. Understanding the function of specific letter sequences will help us evaluate their impacts on visual word processing and in acquiring new words. We propose that orthographic markedness is the key to observing how sensitivity and learning is mediated by sub-lexical information. For this reason, we will investigate sensitivity to orthographic markedness in word processing and learning. This will also allow us to better understand whether the function of orthographic regularities in the visual word system is stable or dynamic (*the BIA+ extended model; Casaponsa et al., 2014; Van Kesteren et al., 2012*). This is important since although models of bilingual visual word recognition have started to account for the specific role of orthographic regularities in the bilingual mind, there are many other aspects of bilingual word recognition that should be considered such as how the model develops over time and during learning.

Therefore, the aim of the present thesis is twofold: we will explore the function of orthographic regularities in the system across the lifespan and when learning new words. First, we will observe whether sensitivity to orthographic markedness is stable across the lifespan or whether it is subject to change. It is possible that second language learning early in life impacts sensitivity to orthographic regularities in the system during childhood but second language learning later in life, when the native language is fully established, might also impact such sensitivity. As we saw previously, orthographic regularities are learned by implicit exposure to written words of the individual's language(s) but acquiring a new language also entails learning a new set of orthographic rules.

On the other hand, we will investigate how orthographic markedness interacts with new word learning in children and adults. We aim to explore whether orthographically illegal novel words are learned differently from those with legal combinations, and in this process, we will also consider the role of bilingualism. The question is whether bilingualism per se has an impact on learning novel legal and illegal words, or whether this potential impact depends on the languages that bilinguals handle and the extent to which they are used to dealing with different orthographic regularities.

The BIA+ model (*Dijkstra & van Heuven, 2002*) assumes that the native language is stable across time, and that the second language is the one that is modulated by the native language, but recent authors have proposed a more dynamic language system (*Kroll et al., 2014*). In both scenarios, word processing and word learning, we can attempt to observe the stability or dynamism of the system.

ii. Overview of the experiments

To examine this, the present thesis studies how orthographic markedness is processed in visual word categorization and in learning. Following our research question, the thesis is divided into two experimental chapters that aim to study word processing (*Chapter 2*) and learning (*Chapter 3*) (*see Table 1 for a description of the experimental chapters*).

Chapter 2 aims to investigate whether sensitivity to orthographic markedness during word processing is stable across the lifespan or whether it is subject to change in bilinguals. More specifically, we aimed to determine whether changes occur during childhood development in early second language learners and/or as a consequence of learning a new language later in life. To address these two possible alternatives, two experiments were conducted. Experiment 1 sought to ascertain whether sensitivity to orthographic markedness changed in bilinguals throughout childhood and early adulthood. To this end, four groups of balanced bilinguals (*Spanish-Basque*) of different ages performed a language decision task with words and pseudowords. We designed two conditions to observe sensitivity to orthographic markedness, using marked and unmarked words. In marked words, one of the bigrams in the word does not exist in the other language (*e.g., txakurra, the word for dog in Basque, is a marked word in Basque because tx does not exist in Spanish*). By contrast, unmarked words are composed of bigrams that both exist and have the same frequency of use in both languages. In addition to words, pseudowords were included in the task. Pseudowords do not have any meaning, so when readers have to process them, in this case identify their language, they have to trust sublexical information in order to do the task. Thus, pseudowords can provide information about the mechanisms that underpin bilinguals' use of orthographic regularities.

On the other hand, Experiment 2 examined whether sensitivity to orthographic markedness changed after learning a second language in later life. To do so, an older monolingual group (*Spanish*) learned a second language (*Basque*) for a year and had to perform a language decision task with marked/unmarked pseudowords before starting to learn the second language, at the end of the year of learning, and after one year after finishing the year of learning. In this case, only pseudowords were included in the language decision task because they give us more information about the underpinning mechanisms for detecting orthographic regularities.

Chapter 3 examines how orthographic markedness interacts with new word learning in children and adults. For this propose, we proposed two questions. First, whether orthographically illegal novel words are learned differently from those with legal combinations. As we saw previously, illegal novel words refer to words in which some bigrams do not exist in the languages that a person knows (*e.g., ubxijla is an illegal novel word because bx and jl do not exist in either of the languages of a Spanish-Basque bilingual*). Second, whether learning these legal/illegal novel words is modulated by bilingualism per se or whether any potential impact depends on the specific languages that a bilingual knows and therefore, the orthographic regularities that differ between the languages they know. Experiment 3 attempted to address these questions in children. Three different groups of children, one monolingual group (*Spanish*) and two bilingual groups whose languages either shared (*Spanish-Catalan*) or differed (*Spanish-Basque*) in terms of orthographic regularities, had to learn legal and illegal novel words. An additional Experiment (*3B*) was included in order to replicate the critical findings from Experiment 3A. Then, Experiment 4 aimed to observe whether legal/illegal word learning in adults was modulated by bilingualism. One group of monolinguals (*Spanish*) and another group of bilinguals

(*Spanish-Basque*) learned the same legal/illegal words as the children. Finally, Experiment 5 was designed to test for any relationship between learning novel words and sensitivity to orthographic markedness in the languages bilinguals already knew. In this experiment, we used a linear regression to compare performance in learning legal/illegal novel words and sensitivity to orthographically marked words and marked pseudowords.

Table 1. Summary of the experiments corresponding each experimental chapter

Experimental chapters	Experiment	Paradigm	Lexicality	Markedness	Language operation	Participants age	Participants language
Chapter 2. Changes in sensitivity to orthographic markedness	1	Language decision task	Words Pseudowords	Marked Unmarked	Spanish Basque	Young children Older children Teenager Young adults	Spanish-Basque bilinguals
	2	Language decision task	Pseudowords	Marked Unmarked	Spanish Basque	Older adults	Spanish monolinguals learning Basque
Chapter 3. Learning novel words with illegal orthographic regularities	3A	Novel word learning task	Novel words	Legal Illegal	Novel	Older children	Spanish monolingual Spanish-Catalan bilingual Spanish-Basque bilingual
	3B	Novel word learning task	Novel words	Legal Illegal	Novel	Older children	High proficient Spanish-Basque Low proficient Spanish-Basque
	4	Novel word learning task	Novel words	Legal Illegal	Novel	Yung adults	Spanish monolingual Spanish-Basque bilingual
	5	Language decision task	Words Pseudowords	Marked Unmarked	Spanish Basque	Older children	Spanish-Basque bilinguals
		Novel word learning task	Novel words	Legal Illegal	Novel		

Chapter 2
Changes in sensitivity to orthographic markedness

I. Overview

The current chapter aims to better understand the function of orthographic regularities in the system across the lifespan. More specifically whether sensitivity to orthographic markedness during visual word processing is stable across the lifespan or whether it is subject to change in bilinguals. In the BIA+ model, Dijkstra and van Heuven (*2002*) assumed that the native language is stable across time and it is the second language that is modulated by the native language. Thus, we aimed to observe whether the change in sensitivity to orthographic markedness in both languages, the native and the second language, arose across normal individual bilingual development in early second language learners and/or was a consequence of learning a new language late in life.

Bilingual adults are very sensitive to orthographic markedness and they rely on orthographic regularities as a strategy to differentiate between languages with the same alphabets. A study conducted by Casaponsa and colleagues (*2014*) investigated sensitivity to markedness in Spanish-Basque bilinguals. Participants performed a speeded language decision task. They had to decide, indicating with a key press, whether the words presented on the screen belonged to their first language (*L1, Spanish*) or second language (*L2, Basque*). The critical manipulation was the frequency of use of the bigrams. In this sense, some words included bigrams that only existed in one of the two languages ('*marked words*') and other words included bigrams that existed in both languages (*'unmarked words' see Figure 10*). Also, all words were controlled to have a low number of orthographic neighbors because this measure correlates with language specificity (*Thomas & Allport, 2000*). Finally, all words had the same word length and high frequency of use. Results showed that young balanced bilingual adults were faster to recognize marked than unmarked words as belonging to a different language. This implies that bilinguals are able to identify languages very quickly based on sub-lexical cues in words.

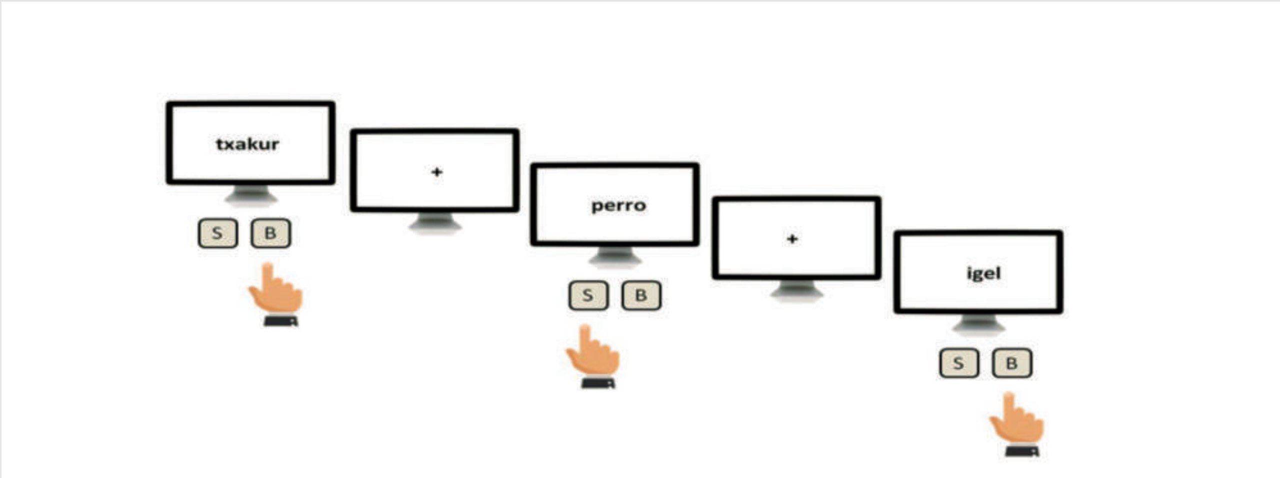


Figure 10. *Speeded language decision task presented with marked (e.g., txakur) and unmarked (e.g., perro, igel) words. 'S' key board is the key answer for Spanish and 'B' for Basque.*

These results were observed regardless of language proficiency level. A group of unbalanced Spanish-Basque bilinguals as well as a group of Spanish monolinguals with no previous experience with Basque performed the same speeded language decision task (*Casaponsa et al., 2014*). Both groups were faster at detecting letter strings that violated Spanish orthographic regularities, showing a recognition advantage for Basque words with marked bigrams. These results showed that even monolinguals who are not familiar with one of the languages could easily detect letter sequences that did not align with their previous orthographic knowledge. Importantly, this suggests that people develop a certain degree of sensitivity to letter sequences that do not conform to their known orthographic rules, even if they do not know the language of these novel words.

As seen above, monolinguals are able to easily detect words from another language that do not follow the orthographic rules of their language. This suggests that orthographic processing occurs at early stages of visual word recognition (*see previous section, i. Models of bilingual visual word recognition, the BIA+ extended model adapted from Van Kesteren, Dijkstra, & de Smedt, 2012 as well as Casaponsa et al., 2014 thesis contribution*). This idea was corroborated by Casaponsa and Duñabeitia (*2016*). These authors used the masked prime language switching paradigm together with event-related potentials (*ERPs*) to observe the impact of the automatic and unconscious language detection mechanism. They suggested that sub-lexical orthographic cues are related to automatic and unconscious processing. Thus, orthographic markedness provides enough information to determine language through a fast-acting detection mechanism. Casaponsa and collegues (*2014*) incorporated the orthographic language node to the BIA+ extended model to account for sub-lexical language-related effects in bilinguals.

However, little is known about the specific role of this orthographic language node. We know that children are very sensitive to orthographic regularities (*Chetail, 2015*) even before they become skillful readers, but is this sensitivity to orthographic markedness stable across the lifespan? Are there differences in sensitivity to markedness between the native language and the second language? If so, do they change with age? Or do they change depending on when the second language was learned early or late in life?

As discussed above, adults are very sensitive to markedness (*Casaponsa et al. 2014*), but it is not clear whether sensitivity to markedness is developed throughout childhood or after childhood. Previous research on implicit learning has shown differences throughout the lifespan (*Howard & Howard, 2013; Janacsek et al., 2012*). Implicit learning refers to people becoming more sensitive to regularities in the environment without being aware of this sensitivity (*Reber, 1957*). Janacsek et al. (*2012*) showed that around the age of 12, older children showed a change in implicit learning. They become more accurate at learning high probability events, those that occur more frequently, than low probability events, the less frequent events. This showed that at this age, older children are faster when learning new information and more sensitive to probabilistic events in the environment.

However, it is possible that another factor may affect system change, namely, learning a second language early or later in life. It has been shown that second language learning exhibits differences depending on the age it is learned and the type of exposure to the new language (*e.g., differences in accent, grammar; Flege, 1987; Hu, 2016; Martin et al., 2013*). Thus, we hypothesize that acquiring the new language early in life might also influence the native language differently than learning a new language later in life when the native system has already been completely established for a long time. Learning a new language not only involves acquiring new grammar, vocabulary and phonology rules but also acquiring the implicit statistical probabilities of orthographic regularities. As we saw previously, one does not need to know the language or understand the meaning of a target word to decide that it does not belong to the native language. As an individual learns a second language its orthography maps onto the native language. So, another possible cue in studying sensitivity to orthographic markedness and its stability in the system is whether changes in sensitivity to orthographic markedness are affected by learning a second language early or later in life.

To sum up, the experiments in this chapter were designed to explore the role of orthographic markedness in word processing and more specifically, whether sensitivity to orthographic markedness is stable across the lifespan. Accordingly, since it is possible that age of acquisition may influence sensitivity to orthographic markedness differently in early or late second language learners, we conducted two experiments that attempted to address these two possible scenarios. Experiment 1 focused on the bilingual's sensitivity to markedness throughout childhood and early adulthood (*with four groups of Spanish-Basque bilinguals of different ages*) who had acquired a second language early in life. Experiment 2 used a longitudinal study to determine whether adults may change their sensitivity to markedness after learning a second language later in life (*one group of Spanish monolinguals learning Basque as a second language*). Note that this longitudinal study would test sensitivity to orthographic markedness both before and after these adults started acquiring their second language, so it provided a more controlled study of the actual effects of L2 learning.

To observe possible changes in sensitivity to orthographic markedness, participants performed a language decision task with words and pseudowords (*Experiment 1*) and only pseudowords (*Experiment 2*). Pseudowords do not have meaning, so when readers have to identify the language that they could belong, they have to trust sub-lexical information in order to be able to perform the task. Thus, pseudowords give us valuable information about the mechanisms underpinning the orthographic regularities used by bilinguals. Participants have to decide on the language of the word (*L1 or L2*) and words were selected as being marked or unmarked in term of their corresponding bigram frequency. Marked words were defined as having at least one bigram that did not exist in the other language (*as measured by frequency of use per percentage*) and unmarked words with bigrams that exist in both languages. Therefore, the difference between marked and unmarked words was the presence of bigrams that did not exist in the other language, rather than the probability of the entire word. Also, word length, orthographic neighbors and frequency of use were controlled.

II. Experiment 1. Changes in sensitivity to orthographic markedness during childhood and early adulthood

i. Rationale

This first experiment aimed to examine the function of orthographic regularities in the system across bilingual development. We will examine how sensitive bilinguals who acquired their second language early in life are to markedness throughout childhood and early adulthood. The purpose was to observe the development of the recognition of marked and unmarked words from the languages the bilinguals knew across different age groups. The final goal was to see whether this sensitivity to markedness changed or remained stable throughout life. To do so, we used a speeded language decision task in line with the task used by Casaponsa et al., (*2014*) , and we also aimed to replicate their findings with different age groups. If results varied between different age groups, we could infer that children and adults differ in their sensitivity for recognizing marked words. Therefore, the results would indicate whether development during childhood changes the way language distinctiveness is detected, as shown in previous experiments on implicit learning (*Janacsek et al., 2012*). On the other hand, if the results did not show such changes in sensitivity, we could infer that sensitivity is not related to maturational stages.

ii. Methods

Participants

One hundred twenty Spanish-Basque bilinguals from the Basque Country participated in this experiment. They were grouped into four age groups of thirty participants each: Younger Children (*17 females; mean age=8.67 years, SD=0.47*), Older Children (*18 females; mean age=12.40 years, SD=0.62*), Teenagers (*22 females; mean age=16.97 years, SD=0.31*), and Adults (*20 females; mean age=23.01 years, SD=2.74*). All participants were right handed and none had been diagnosed with language disorders, learning disabilities, or auditory impairments.

Younger children, older children and teenagers were recruited from a bilingual school in Vitoria. Adults were recruited from the University of the Basque Country in San Sebastian. Adults, children, and children´s families were appropriately informed. Adult participants signed consent forms prior to the experiment. In the case of underage participants, their parents or legal guardians received an information letter with a short questionnaire, in which parents had to rate their children's language proficiency and socioeconomic status. Parents had to return the signed written informed consent before testing. The protocol was carried out according to the guidelines approved by the BCBL Ethics and Scientific Committees. Adults were economically compensated and children were rewarded with a present.

All participants were assessed in terms of their language proficiency, socioeconomic status, and IQ (*see Table 2 for results*). Three measures were used to evaluate language proficiency: one subjective and two objective measures. In the subjective measure, adult participants or parents in the case of students had to rate their language competence compared with their peers on a scale from 0 to 10. The first objective measure was a lexical decision task (*LexTale*). Participants completed the task in the three languages that they knew, Spanish (*Izura, Cuetos, & Brysbaert, 2014*), Basque (*de Bruin, Carreiras, & Duñabeitia, 2017*), and English (*Lemhöfer & Broersma, 2012*). The other objective measure was a naming task, adopted from a picture naming task (*de Bruin et al., 2017*). Participants had to name twenty common objects. In addition, we measured English proficiency, which is not a relevant language for the task, but was included to make sure that the participants' English level was relatively low and would not interfere with performance on the main experimental task (*see Table 2*). Next, socioeconomic status was measured with a short questionnaire in which participants or parents rated on a scale from 1 to 10 how they perceived their economic situation as compared to other members of their community (*Adler & Stewart, 2007*). Finally, IQ was measured with a 6-minute abridged version of the K-BIT (*Kaufman, 2004*) in which participants had to complete as many matrices as they could in the allotted time.

Participant groups were matched in terms of their percentage of exposure to the three languages (*Spanish, Basque, and English*), subjective language competence in these three languages, Spanish picture naming results, and socioeconomic status (*see Table 2*). The age groups could not be matched on the lexical decision task (*LexTale*), IQ, and picture naming due to differences related to development. For instance, vocabulary size increases with age thanks to exposure to new vocabulary (*Hamilton, Plunkett, & Schafer, 2000*) and IQ also increases with age (*Ramsden et al., 2013*).

Table 2. Descriptive statistics of demographic and language variables

	Younger children	Older children	Teenagers	Adults	ANOVAs	
					F (df)	p
Age	8.67 (0.47)	12.40 (0.62)	16.97 (0.31)	23.01 (2.74)	F(3,116)=55.98	<.001
Age of Basque acquisition	3.40 (1.99)	3.50 (1.45)	3.30 (1.36)	2.97 (1.84)	F(3,116)=0.56	0.639
Spanish exposure	62.67 (10.2)	63.00 (12.9)	62.67 (13.9)	60.83 (11.3)	F(3,116)=0.19	0.899
Basque exposure	24.67 (7.64)	22.00 (7.83)	22.17 (8.97)	26.00 (11.7)	F(3,116)=1.36	0.259
English exposure	12.67 (5.68)	15.00 (7.65)	15.17 (7.59)	13.17 (5.64)	F(3,116)=1.07	0.364
Spanish self-rated proficiency	9.33 (0.75)	9.36 (0.71)	9.43 (0.72)	9.46 (0.73)	F(3,116)=0.21	0.892
Basque self-rated proficiency	5.73 (1.61)	5.93 (0.98)	6.20 (1.44)	6.50 (1.67)	F(3,116)=1.56	0.202
English self-rated proficiency	4.57 (1.63)	4.70 (1.51)	4.96 (1.21)	5.16 (1.48)	F(3,116)=1.01	0.395
Spanish LexTale	69.36 (11.1)	87.69 (6.77)	92.97 (3.67)	93.05 (3.45)	F(3,116)=77.01	<.001
Basque LexTale	51.16 (11.9)	68.86 (10.2)	75.63 (13.2)	76.70 (12.3)	F(3,116)=29.11	<.002
English LexTale	52.08 (5.13)	53.87 (6.64)	57.29 (7.34)	56.45 (15.2)	F(3,116)=1.92	0.13
Spanish picture naming	19.90 (0.30)	19.96 (0.18)	20.00 (0)	20.00 (0)	F(3,116)=2.11	0.103
Basque picture naming	11.13 (2.83)	14.06 (3.62)	14.30 (4.47)	14.53 (4.81)	F(3,116)=4.74	0.004
English picture naming	7.13 (3.97)	12.90 (3.38)	13.23 (4.31)	13.76 (3.80)	F(3,116)=19.17	<.001
Economic status	6.30 (1.29)	6.43 (1.67)	6.33 (1.34)	6.50 (1.10)	F(3,116)=0.13	0.939
IQ	17.30 (2.15)	19.73 (2.39)	20.20 (2.65)	20.50 (2.94)	F(3,116)=9.76	<.001

Note. Values reported are means with standard deviation in parentheses for age (in years), age of acquisition (in years), language exposure (in % exposed), subjective language proficiency (0-10 scale), LexTale (averaged % correct), picture naming(0-20 scale), economics status (1-10 scale), and IQ (correct answerers). The last column shows the results from one-way ANOVAs comparing the four age groups on the different assessments.

Materials

Corpus of bigrams

A corpus of bigrams was compiled from the Spanish (*B-PAL; Davis & Perea, 2005*) and Basque (*E-HITZ; Perea et al., 2006*) databases. First, diacritics and words containing letters that do not exist in one of the languages (*ñ, c, q, v, w*) were removed. All words were broken down into bigram units (*e.g., the Spanish word for house, casa, was deconstructed as ca-as-sa*). All bigram combinations were then averaged based on their appearance rates relative to all bigrams in terms of percentage (*percentage frequency*) in the languages. For example, the bigram *ca* appears in the Spanish language 3482 times. The average number of appearances in the language is 1.57% (*number of times a specific bigram appears*100/total number of bigrams that appear in the language*).

Language decision task

The language decision task was similar to the one that Casaponsa et al. (*2014*) used in their experiment. In total, we selected one-hundred and sixty words. Half of the words were in Basque (*selected from Perea et al., 2006*) and the other half were in Spanish (*taken from Davis & Perea, 2005*). As our manipulation was also focused on markedness, words in both languages were selected to be marked and unmarked words.

Marked words contained one bigram that only existed in the target language and was illegal in the other language (*e.g., txakurra, the Basque word for dog, is a marked word for Basque because tx only exists in Basque*). We defined marked bigrams as those that had 0 percentage frequency of use in the other language and a bigram percentage frequency of use higher than 0.1 percent in the target language. Following this rule, we picked four marked bigrams: two marked bigrams for Basque (*tx and ts; bigram percentage frequency of use in Basque: 0.42 and 0.39, respectively*) and two for Spanish (*mp and mb; bigram percentage frequency of use in Spanish: 0.31 and 0.28, respectively*). On the other hand, unmarked words contained only bigrams that existed in both languages (*e.g., ardi, the Basque word for sheep, and ardilla, the Spanish word for squirrel, are unmarked words because all bigrams exist in both languages*) and that have a high bigram percentage frequency of use (*higher than 0.1*) (*see Appendix 1 to see the words used in the task*).

Words were matched on certain characteristics to control the influence of how words are recognized and read (*see Table 3*). First, we controlled that the average of bigram percentage frequency of use (*per percentage*) in each condition was matched (*see Table 3, Spanish and Basque bigram frequency*). In this case, Spanish marked words had the same average of bigram percentage frequency in Spanish as the Basque marked and unmarked words to control that use of any of the specific marked bigrams chosen was not more salient in one of the languages. Also, bigrams were controlled such that the bigrams had a high frequency or occurrence in each position in the word. Also, we controlled the percentage frequency of use of the words making sure they were high in both languages (*the frequency of use was bounded between 1 and 100 per million*). It has been shown that lexical frequency is a variable that influences word recognition, with high frequency words being recognized faster than low frequency ones (*Rubenstein, Garfield, & Millikan, 1970*). Also, we controlled for word length and we ensured that bigram frequency in the two languages was comparable (*see Table 3*).

Table 3. Descriptive statistics of characteristics of the materials

WORDS	Spanish		Basque	
	Marked	Unmarked	Marked	Unmarked
Word frequency (ZIP)	3.98 (0.67)	4.18 (0.29)	4.06 (0.59)	4.17 (0.58)
Word length	7 (1.43)	7 (1.46)	7 (1.45)	6.95 (1.35)
Spanish bigram frequency	0.71 (0.22)	0.72 (0.21)	0.53 (0.18)	0.69 (0.20)
Basque bigram frequency	0.52 (0.21)	0.69 (0.21)	0.71 (0.16)	0.72 (0.19)
Orthographic neighbors in Spanish	1.07 (1.43)	1.05 (1.31)	0.08 (0.22)	0.16 (1,49)
Orthographic neighbors in Basque	0.13 (1.43)	0.17 (1.08)	0.85 (1.31)	1.07 (1.28)
Corrected LD	0.14 (0.11)	0.12 (0.09)	0.13 (0.11)	0.13 (0.10)

Pseudowords	Spanish Marked	Basque Marked	Unmarked
Word length	7 (1.43)	7 (1.43)	7 (1.43)
Spanish bigram frequency	0.71 (0.16)	0.52 (0.17)	0.71 (0.23)
Basque bigram frequency	0.54 (0.18)	0.72 (0.16)	0.70 (0.26)
Orthographic neighbors in Spanish	0.1 (0.30)	0.02 (0.15)	0.22 (0.61)
Orthographic neighbors in Basque	0.02 (0,15)	0.42 (0.78)	0.12 (0.46)

Note. Values reported are means with standard deviation in parentheses for word frequency (ZIP scale), word length (number of letters), Spanish bigram frequency (percentage per million), Basque bigram frequency (percentage per million), orthographic neighbors (number of words), and corrected LD (scale from 0 to 1).

Furthermore, we ensured that the number of orthographic neighbors in both languages, that is, the number of novel words that could be created by changing one letter of the word (*Coltheart, Davelaar, Jonasson, Besner, & Dornic, 1977*), was low or almost null. It is shown that the more orthographic neighbors a word has, the more difficult it is to identify that word due to lexical competition. We also controlled for translation equivalents in the other language by calculating the corrected orthographic Levenshtein distance. This measure accounts for the number of letters that differ in the equivalent translation of the target word and is established by comparing the target word and its translation in the other language. It ranges from 0 to 1, in which 0 refers to totally different translation equivalents for the target word in the other language and 1 which corresponds to a completely overlapping cognate. For instance, the word *fiction* in English and its translation equivalent in Spanish, *ficción*, has a Levenshteein distance of 0.86. This number shows a high translation equivalence between the two languages. We wanted to avoid widespread overlap, so we picked words that had corrected LDs of 0.4 or lower (*see Table 3*).

Furthermore, one-hundred sixty pseudowords were built. Pseudowords were generated with Wuggy (*Keuleers & Brysbaert, 2010*) from the words described in the previous section. As previously mentioned, pseudowords were added to the experiment because when participants have to process them, they would have to base their answer on sub-lexical cues because there is no possible access to (*lexical*) meaning. Pseudowords were divided into Spanish-marked, Basque-marked and unmarked pseudowords. Marked bigrams in pseudowords were the same as those used in the word task *(tx and for Basque and mp and mb for Spanish*). The rest of the bigrams included in the marked pseudoword were unmarked bigrams that existed in both languages (*e.g., alitxo, see Appendix 2 for more examples*). Unmarked pseudowords included only bigrams that existed in both languages.

Procedure

Participants were tested individually. Students did the experiment during school hours in a room alone. Adults were tested during lab hours in a soundproof cabin. The whole experiment lasted approximately thirty minutes, including the language decision task and the language assessment. First, participants performed a language decision task. The experiment was presented on a 13-inch MacBook® running Experiment Builder®. In the middle of the screen a fixation cross appeared for 500 ms. Next a word appeared until a response was given or up to 5000 ms. In the experiment by Casaponsa et al. (*2014*), this interval lasted 1500 ms but this time was not sufficiently long for children to identify the language.

Participants were asked to respond as fast as they could while indicating to which language (*Basque or Spanish*) each word/pseudoword belonged. Responses were given by two keys on the keyboard, 'C' key if the word belonged to Spanish or 'B' if it belonged to Basque. Words and pseudowords were presented together in random order. So, participants were informed that they would see a mix of pseudowords and words and had to decide which language words/pseudowords could belong to. After the language decision task, participants completed the language assessment and the IQ test with an online survey.

Data analysis

The dependent variables of interest collected in this experiment were Accuracy and Reaction time (see Table 4). The statistical environment R (R core team, 2013) was used to remove outliers. Responses below 200 ms, considered to be chance responses (0.89% outliers), were excluded from the analyses as were responses above 5000 ms (timeout responses; 1.04% outliers). Moreover, responses above the 0.75 and below the 0.25 interquartiles from the participant-based (word outliers: 2.25%; pseudowords outliers: 1.15%) and item-based (words outliers: 2.16%; pseudowords outliers: 0.90%) means in each condition were also removed for the analysis (in total, words outliers: 3.31%; pseudowords outliers: 1.88%). After outlier removals, the percentage of errors and reaction times removing erroneous responses were analyzed using Jamovi 0.9.6.7 using repeated.

In this experiment, four independent variables of interest were manipulated: Lexicality (Words|Pseudowords), Language (Basque|Spanish), Markedness (Marked| Unmarked), as within-subjects and Group (younger-children|older-children|teenagers|adults) as between- subjects. However, words and pseudowords were analyzed separately because there is no correct answer in the unmarked pseudowords condition. We first analyzed performance on the lexical decision task for words by using a three-way ANOVA[1] with the within-subject predictors Language (Basque|Spanish) and Markedness (Marked| Unmarked) and the between-subject predictor Group (younger-children|older-children|teenagers|adults) on percentage of errors and reaction times removing erroneous responses.

Then, pseudowords were analyzed separately because the unmarked pseudowords were equally likely to be Basque-like or Spanish-like. We therefore analyzed performance on the unmarked pseudowords as the probability of a participant deciding 'Basque'. The probability of Basque responses

Note 1: that this data was also analyzed using Mixed-effects models, accuracy with logistic mixed-effects models and reaction times with linear mixed-effects models (Baayen, Davidson & Bates, 2008; Jaeger, 2008; Barr, 2013), using the lme4 package in R (Bates, Maechler, Bolker & Walker, 2014), including subjects and items as random factors. These analyses showed the same results as repeated measures ANOVAs. In the present thesis, we report ANOVAs to be consistent with the analyses reported in other chapters.

was a recoded variable coded as '1' if participants responded Basque and '0' if they responded Spanish. Thus, unmarked pseudowords were analyzed using one-way ANOVA by Group on the probability of 'Basque' responses. In the analysis of reaction times for unmarked pseudowords, trials were considered to be Spanish if the participant responded 'Spanish' and Basque if the participant responded 'Basque', so two-way ANOVAs were carried out by Group and Response Language on the reaction times. Marked pseudowords were also analyzed in two-way ANOVAs with the effects of Language (*Basque|Spanish*) as a within-subject factor and Group as a between-subjects factor for the percentage of errors and reaction times without erroneous responses.

Table 4. Descriptive statistics for the Language decision task.

	Words				Pseudowords			
	Basque		Spanish		Marked		Unmarked	
%errors	Marked	Unmarked	Marked	Unmarked	Basque	Spanish	Probability of 'Basque' response	Probability of 'Spanish' response
Younger children	14.12 (17.9)	23.98 (20.07)	18.2 (19.3)	19.0 (18.61)	15.4 (18.5)	53.03 (18.39)	0.65 (0.17)	0.35 (0.17)
Older children	3.14 (4.17)	6.58 (6.72)	7.74 (9.28)	7.89 (8.68)	5.04 (7.94)	56.31 (21.84)	0.74 (0.15)	0.26 (0.15)
Teenagers	3.35 (4.18)	7.11 (6.58)	4.45 (4.54)	4.56 (4.31)	4.23 (4.26)	45.63 (24.6)	0.68 (0.15)	0.32 (0.15)
Adults	1.84 (2.3)	3.44 (3.65)	3.51 (3.56)	3.86 (2.9)	6.22 (5.47)	25.65 (16.42)	0.56 (0.11)	0.44 (0.11)
RT	Marked	Unmarked	Marked	Unmarked	Basque	Spanish	Probability of 'Basque' response	Probability of 'Spanish' response
Younger children	1736 (521)	1955 (591)	1831 (597)	1912 (590)	1882 (590)	2184 (683)	2205 (711)	2160 (593)
Older children	1063 (249)	1199 (316)	1006 (206)	1061 (226)	1238 (340)	1663 (431)	1822 (567)	1557 (479)
Teenagers	810 (119)	935 (151)	782 (111)	821 (131)	959 (247)	1435 (414)	1592 (509)	1298 (389)
Adults	755 (115)	828 (129)	815 (106)	850 (124)	1000 (207)	1329 (360)	1500 (373)	1305 (254)

Note. Values reported are means with standard deviations in parentheses for accuracy (%errors), and reaction times (milliseconds)

iii. Results

Words

Analysis of percentage of errors:

First, the analysis of percentage of errors was carried out. Results showed that there was no effect of Language (*F1(1,116)=0.91, p=.343,* η_p^2*=.001; F2(1,79)=0.13, p=.720,* η_p^2*=.000*), suggesting that Basque and Spanish words overall were recognized equally well. However, there was a significant main effect of Markedness (*F1(1,116)=37.39, p<.001,* η_p^2*=.010; F2(1,79)=3.70, p=.026,* η_p^2*=.004*), showing that marked words were recognized better than unmarked words, and there was a significant interaction between Markedness and Language (*F1(1,116)=22.35, p<.001,* η_p^2*=.008; F2(1,79)=4.035, p=.028,* η_p^2*=.003*), showing that there was a markedness effect in Basque (*Tukey Post-Hoc test: t1(230)=7.61, p<.001; t2(158)=3.61, p=.009*) but not in Spanish (*Tukey Post-Hoc test: t1(230)=0.61, p=.929; t2(158)=0.17, p=.998*). Also, the variable Group was significant (*F1(3,116)=18.1, p<.001,* η_p^2 *=.252; F2(3,237)=222.99, p<.001,* η_p^2*=.154*) suggesting that people of different ages performed differently on the language decision task. The variable Group also interacted significantly with Markedness and Language (*F1(3,118)=3.28, p=.023,* η_p^2*=.003; F2(3,237)=4.72, p=.023.,* η_p^2*=.004).*

To follow up on this three-way interaction, we performed a two-way measure ANOVA with the within-subjects level Markedness (*Marked|Unmarked*) and Group as between-subjects on Basque and Spanish words separately to observe if sensitivity to markedness was different in the L1 and L2. Results showed that on the Spanish words, there was a significant main effect of Group (*F1(3,116)=13.0, p<.001,* η_p^2*=.238; F2(3,237)=106.67, p<.001,* η_p^2*=.157*), suggesting that young children performed worse than the other groups (*see Table 5*). However, there was no main effect of Markedness, (*F1(1,116)=0.539, p=.464,* η_p^2*=.000; F2(1,79)=0.03, p=.864,* η_p^2*=.000*) nor any interaction between Markedness and Group (*F1(3,116)=0.123, p=.946,* η_p^2*=.000; F2(3,237)=0.04, p=.988,* η_p^2*=.000*). Results of sensitivity to Spanish (*L1*) suggested that people are not sensitive to markedness in L1, performing similarly in response to marked and unmarked words (*see Figure 11*). On the other hand, sensitivity to Basque (*L2*) showed different results. There was a significant main effect of Markedness (*F1(1,116)=44.15, p<.001,* η_p^2*=.035; F2(1,79)=6.66, p=.012.,* η_p^2*=.030*), of Group (*F1(3,116)=18.0 p<.001,* η_p^2*=.272; F2(3,237)=138.39, p<.001,* η_p^2*=.068*) and an interaction between Markedness and Group (*F1(3,116)=6.52, p<.001,* η_p^2*=.016; F2(3,237)=3.14, p=.025,* η_p^2*=.004; see Figure 11*). This showed that people are sensitive to markedness in Basque, and the significant interaction suggests that this sensitivity to markedness in Basque changes with age.

Specifically, our research question was to identify when sensitivity to markedness changes with development. Therefore to follow up on the significant interaction in Basque, we carried out one-way $ANOVA_1$ on the Basque Markedness effect by groups as a within-subject level. The Basque markedness effect is calculated as the difference between the unmarked and marked percentage of errors. Results showed a significant effect of age group (*F1(3,63)=4.06, p=.011,* η_p^2*=.144; F2(3,237)=3.18, p=.019,* η_p^2*=.016*). The Tukey follow-up test showed that younger children had different sensitivity to markedness in Basque than the other groups (*see Table 5*), while the other three groups had similar sensitivity to markedness between them (*see Table 5*). These results suggested that children show larger differences in detecting marked and unmarked words in their L2. But during later childhood/ early adulthood, bilinguals show smaller differences between marked and unmarked words in their L2 (*see Figure 11*).

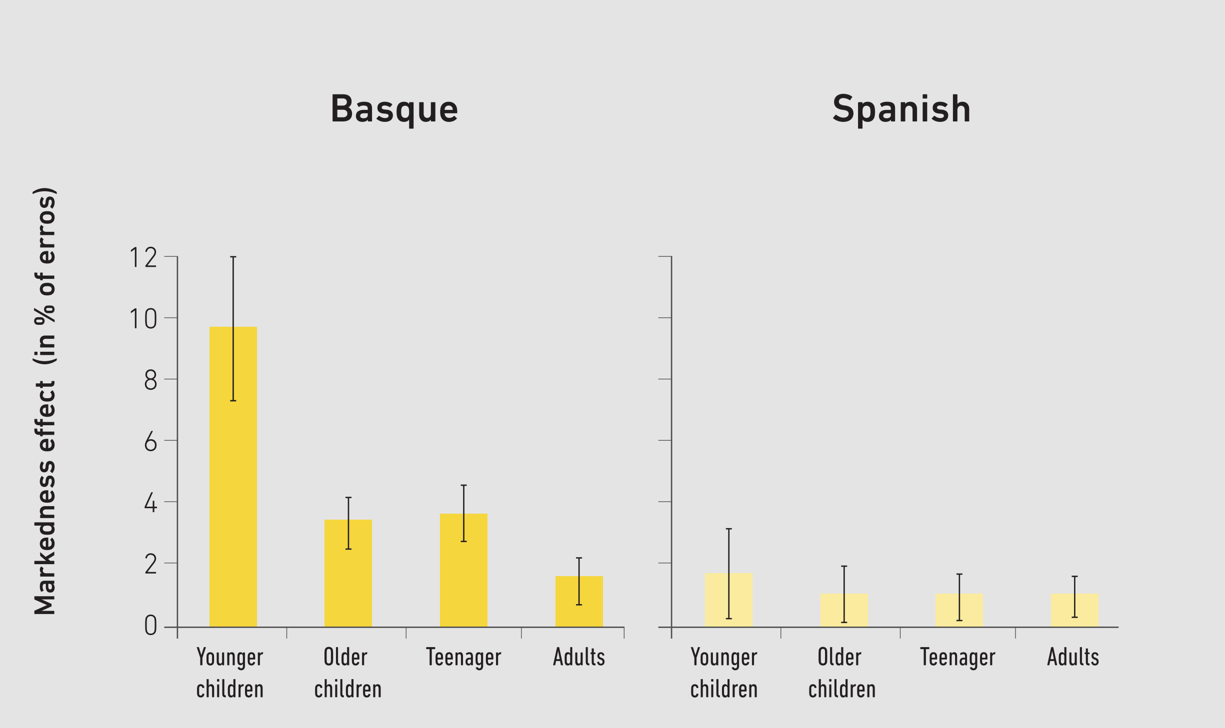


Figure 11. *Bar plots for the effect of markedness effect in groups for Basque and Spanish words. The markedness effect is the difference between unmarked marked percent of errors. Error bars refer to the standard error (SE) of each mean*

Table 5. Tukey Post-Hoc test values for Spanish words by group and Basque markedness effect by group.

t test of percentage of errors

Group comparisons		Spanish words t1	p	t2	p	Basque words t1	p	t2	p
Younger children	Older children	t1(116)=3.98	<.001	t2 (237)= 11.72	<.001	t1(63)=6.41	.009	t2(237)=3.15	.019
Younger children	Teenagers	t1(116)=35.2	<.001	t2 (237)= 14.94	<.001	t1(63)=6.10	.014	t2(237)=3.01	.022
Younger children	Adults	t1(116)=5.52	<.001	t2(237)=15.84	<.001	t1(63)=8.25	<.001	t2(237)=3.45	.015
Older children	Teenagers	t1(116)=1.22	.613	t2(237)=1.98	.368	t1(63)=0.31	.999	t2(237)=1.23	.606
Older children	Adults	t1(116)=1.52	.235	t2(237)=1.23	.696	t1(63)=1.84	.791	t2(237)=0.66	.911
Teenagers	Adults	t1(116)=0.31	.990	t2(237)=0.89	.806	t1(63)=1.58	.712	t2(237)=1.89	.232

t test of reaction times

Group comparisons		Spanish words t1	p	t2	p	t test of percentage of errors t1	p	t2	p
Younger children	Older children	t1(116)=10.64	<.001	t2(237)=42.98	<.001	t1(87)= 3.94	.007	t2(237)=2.98	.012
Younger children	Teenagers	t1(116)=13.91	<.001	t2(237)=54.46	<.001	t1(87)= 3.46	.009	t2(237)=2.87	.013
Younger children	Adults	t1 (116)= 13.33	<.001	t2(237)=54.13	<.001	t1(87)= 3.12	.013	t2(237)=2.46	.019
Older children	Teenagers	t1 (116)= 2.06	.208	t2(237)=1.87	.239	t1(87)= 0.23	.996	t2(237)=0.45	.964
Older children	Adults	t1 (116)= 1.68	.641	t2(237)=1.68	.267	t1(87)= 1.34	.538	t2(237)=0.42	.973
Teenagers	Adults	t1 (116)= 0.57	.939	t2(237)=2.02	.198	t1(87)= 1.13	.682	t2(237)=0.35	.985

Note. Values reported are values of t tests with degrees of freedom in parentheses and p values.

Analysis of reaction times

The analysis of reaction times showed similar results as the analysis of percentage of errors. The main effect of language was not significant ($F1(1,116)=0.83$, $p=.243$, $\eta_p^2=.001$; $F2(1,79)=0.144$, $p=.706$, $\eta_p^2=.001$), but the main effect of markedness was significant ($F1(1,116)=48.34$, $p<.001$, $\eta_p^2=.061$; $F2(1,79)=9.04$, $p=.004$, $\eta_p^2=.001$), as well as the effect of age group ($F1(3,116)=80.6$, $p<.001$, $\eta_p^2=.636$; $F2(3,237)=198.46$, $p<.001$, $\eta_p^2=.096$). The three-way interaction was significant too ($F1(3,116)=3.38$, $p=.021$, $\eta_p^2=.001$; $F2(3,237)=9.94$, $p<.001$, $\eta_p^2=.010$).

Then, Basque and Spanish words were split for the analysis and two-way ANOVAs with Markedness (Marked|Unmarked) as within-subjects and Group as between-subject were carried out to follow up on the triple interaction (*see Figure 12*). Spanish words showed a main effect of Group ($F1(3,116)=83.8$, $p<.001$, $\eta_p^2=.666$; $F2(3,237)=56.79$, $p<.001$, $\eta_p^2=.098$), showing that younger children needed more time to respond to Spanish words than the rest of the groups (*see Table 5*). However, the main effect of Markedness ($F1(1,116)=2.56$, $p=.185$, $\eta_p^2=.001$; $F2(1,79)=1.20$, $p=.991$, $\eta_p^2=.001$) and the interaction between Group and Markedness ($F1(3,116)=0.65$, $p=.581$, $\eta_p^2=.001$; $F2(3,237)=0.05$, $p=.983$, $\eta_p^2=.001$) were not significant. On the other hand, reaction times to Basque word showed a main effect of Markedness ($F1(1,116)=69.33$, $p<.001$, $\eta_p^2=.017$; $F2(1,79)=3.42$, $p=.019$, $\eta_p^2=.003$), Group ($F1(3,116)=70.7$, $p<.001$, $\eta_p^2=.615$; $F2(3,237)=41.93$, $p<.001$, $\eta_p^2=.036$) and an interaction between these variables ($F1(3, 116)=3.32$, $p=.020$, $\eta_p^2=.002$; $F2(3,237)=104.91$, $p<.001$, $\eta_p^2=.059$).

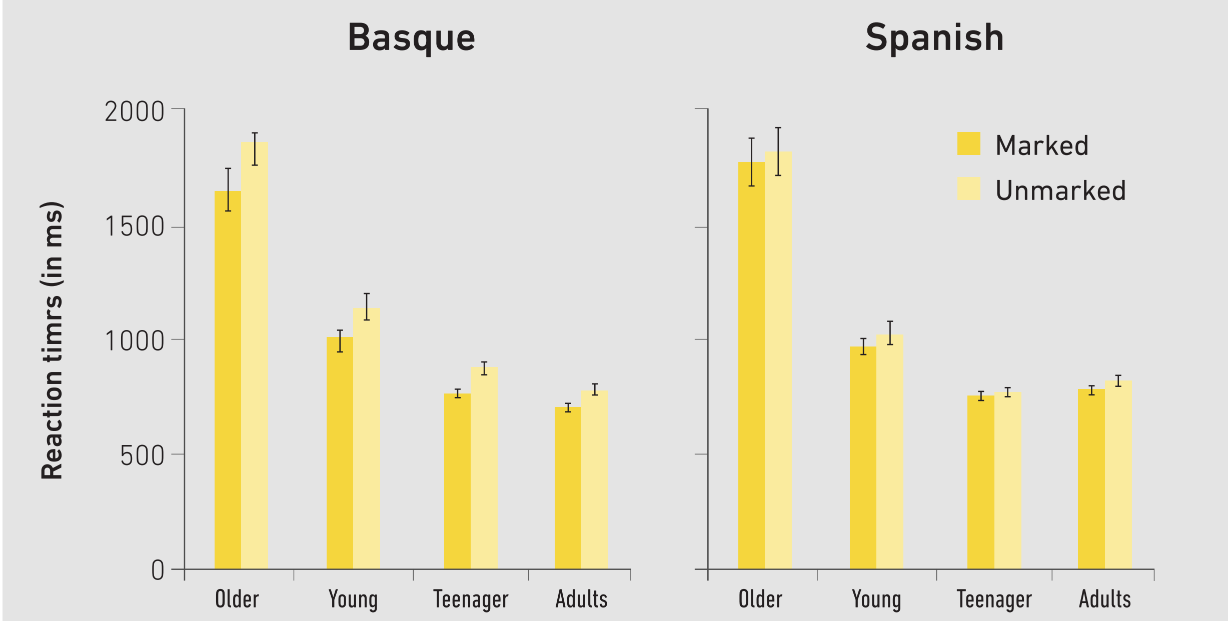


Figure 12. *Bar plots for reaction times in groups for Basque and Spanish words in milliseconds. Error bars refer to the standard error (SE) of each mean*

As a follow up on this interaction, the effect of markedness in Basque was calculated to observe the sensitivity to markedness of the different groups. Groups showed a significant effect of Markedness (*F1(3,87)=3.33, p=.019, η_p^2=.079; F2(3,237)=2.97, p=.021, η_p^2=.003*). In line with the percentage of errors results, reaction times also showed that children performed differently than the other groups (*see Table 5*). The other three groups did not differ in response times (*see Table 5*).

Pseudowords

Unmarked pseudowords

Analysis of probability of Basque response

First, we analyzed unmarked pseudowords examining participants' responses. We carried out one-way ANOVA by Group on the probability of Basque responses. There was a main effect on Group (*F1(3,87)=7.41, p<.001, η_p^2=.160; F2(3,237)=28.91, p<.001, η_p^2=.079*), showing that younger children, older children and teenagers tended to classify unmarked pseudowords as Basque (*see Table 6*). However, adults tended to attribute unmarked pseudowords to either language interchangeably (*see Table 6 and Figure 13*).

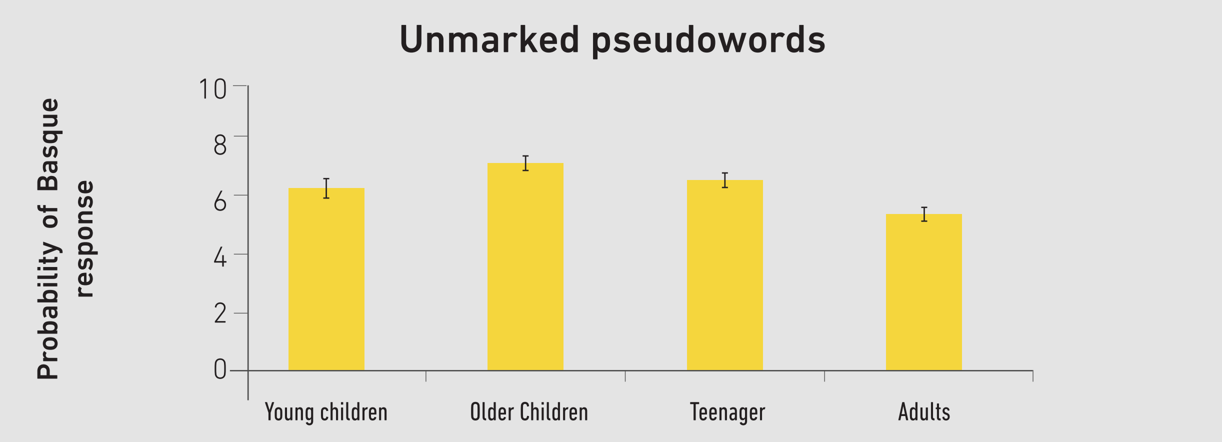


Figure 13. *Bar plots for unmarked pseudowords represented by the probability of Basque response (coded as:'1' Basque response and '0' Spanish response). Error bars refer to the standard error (SE) of each mean.*

Analysis of reaction times

On the other hand, the analysis of reaction times in a two-way ANOVA for the unmarked pseudowords revealed a main effect of Response language (*F1(1,116)=64.45, p<.001, η_p^2=.028; F2(1,21)=29.3, p<.001, η_p^2=.098*), showing faster responses when participants decided that an unmarked pseudoword could belong to Basque language (*Tukey Post-Hoc test: t1(116)=8.03, p<.001; t2(21)=13.4, p<.001), a main effect of Group (F1(3,116)=16.4, p<.001, η_p^2=.272; F2(3,63)=19.2, p<.001, η_p^2=.080*) and in the interaction between the variables (*F1(3,116)=4.99, p=.003, η_p^2=.007; F2(3,63)=8.93, p<.001, η_p^2 =.074*). This interaction showed that young children didi not differ in response time across languages (*Tukey Post-Hoc test: t1(116)=0.91, p=.985; t2(41)=1.21, p=.783*), but the other groups were faster when responding Basque. This shows that young children needed time to decide the language of the unmarked pseudowords compared with the other groups (*see Table 6*), while the other groups did not differ between them (*see Table 6 and Figure 14*).

Table 6. Tukey Post-Hoc test values for Basque marked and Spanish marked percentage of errors and reaction times, and for unmarked pseudowords by the probability to Basque response and reaction times based on the Response Language

Analysis of percentage of errors

Group comparisons		Unmarked pseudowords				Basque marked pseudowords				Spanish marked pseudowords			
		t1(87)	p	t2(237	p	t1(87)	p	t2(120	p	t1(87)	p	t2(120	p
Younger children	Older children	2.25	.118	1.92	.288	3.71	.002	8.28	<.001	0.62	.926	1.45	.471
Younger children	Teenagers	0.76	.872	1.53	.420	3.99	<.001	8.84	<.001	1.39	.507	1.83	.128
Younger children	Adults	2.98	.047	4.54	<.001	3.29	.008	7.12	<.001	5.15	<.001	11.6	<.001
Older children	Teenagers	1.49	.447	1.78	.338	0.29	.991	0.56	.945	2.01	.192	1.98	.101
Older children	Adults	4.61	<.001	9.13	<.001	0.42	.975	1.17	.646	5.76	<.001	13.1	<.001
Teenagers	Adults	3.12	.013	6.08	<.001	0.71	.892	1.73	.314	3.76	.002	8.82	<.00

t test of reaction times

Group comparisons		Unmarked pseudowords				Basque marked pseudowords				Spanish marked pseudowords			
		t1(116	p	t2(63)	p	t1(87)	p	t2(120	p	t1(87)	p	t2(120	p
Younger children	Older children	3.94	<.001	7.43	<.001	6.63	<.001	18.49	<.001	4.15	<.001	5.22	<.001
Younger children	Teenagers	5.89	<.001	6.34	<.001	9.51	<.001	26.06	<.001	5.97	<.001	13.15	<.001
Younger children	Adults	6.24	<.001	7.34	<.001	9.09	<.001	24.93	<.001	6.82	<.001	14.91	<.001
Older children	Teenagers	1.95	.212	1.92	.288	2.45	.075	2.32	.087	1.81	.273	2.75	.061
Older children	Adults	2.30	.104	1.53	.420	2.21	.098	1.98	.115	2.67	.064	2.19	.075
Teenagers	Adults	0.34	.986	0.82	.823	0.42	.975	1.13	.673	0.85	.829	1.77	.294

Note. Values reported are values of t tests with degrees of freedom in parentheses and p values.

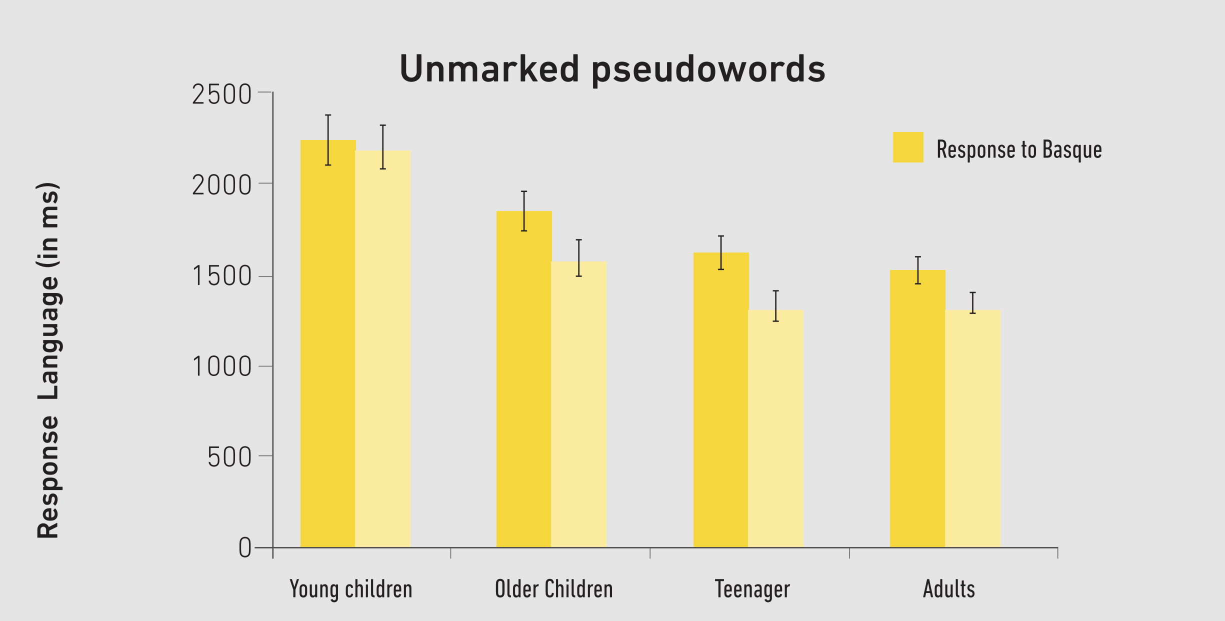


Figure 14. *Bar plots of unmarked pseudowords represented by reaction time (in milliseconds) for the language identified in the response. Error bars refer to the standard error (SE) of each mean.*

Marked pseudowords

Analysis of percentage of errors

On the marked pseudowords, we conducted two-way ANOVAs on Language by groups on percentage of errors. Results showed a main effect of Language (*F1(3,116)=276.24, p<.001,* η_p^2*=.509; F2(1,22)=263.8, p<.001,* η_p^2*=.621*), reflecting that participants responded more accurately to Basque-like than Spanish-like pseudowords, and of Group (*F1(3,116)=16.80, p<.001,* η_p^2*=.069; F2(3,66)=44.6, p<.001,* η_p^2*=.090*). Also the interaction between the two variables was significant (*F1(3,116)=8.73, p<.001,* η_p^2*=.048; F2(3,66)=21.2, p<.001,* η_p^2*=.047*).

To follow up on this interaction, Basque-marked and Spanish-marked pseudowords were analyzed separately using one-way ANOVA by groups on percentage of errors. The Basque-marked pseudowords analysis showed a significant effect (*F1(3,87)=6.88, p<.001,* η_p^2*=.160, F2(3,120)=33.7, p<.001,* η_p^2*=.256*),with younger children making more errors than the other groups (*see Table 6*). The other three groups showed similar performances (*see Table 6 and Figure 15*).

However, Spanish-marked pseudowords showed a different pattern in performance by group (*F1(3,87)=13.4, p<.001,* η_p^2*=.258; F2(3,120)=68.9, p<.001,* η_p^2*=.360*). Adults had fewer errors than the rest of the groups (*see Table 6*), suggesting that adults are the only group that rely on Spanish-marked orthographic regularities to categorized Spanish-marked pseudowords (*see Figure 15*).

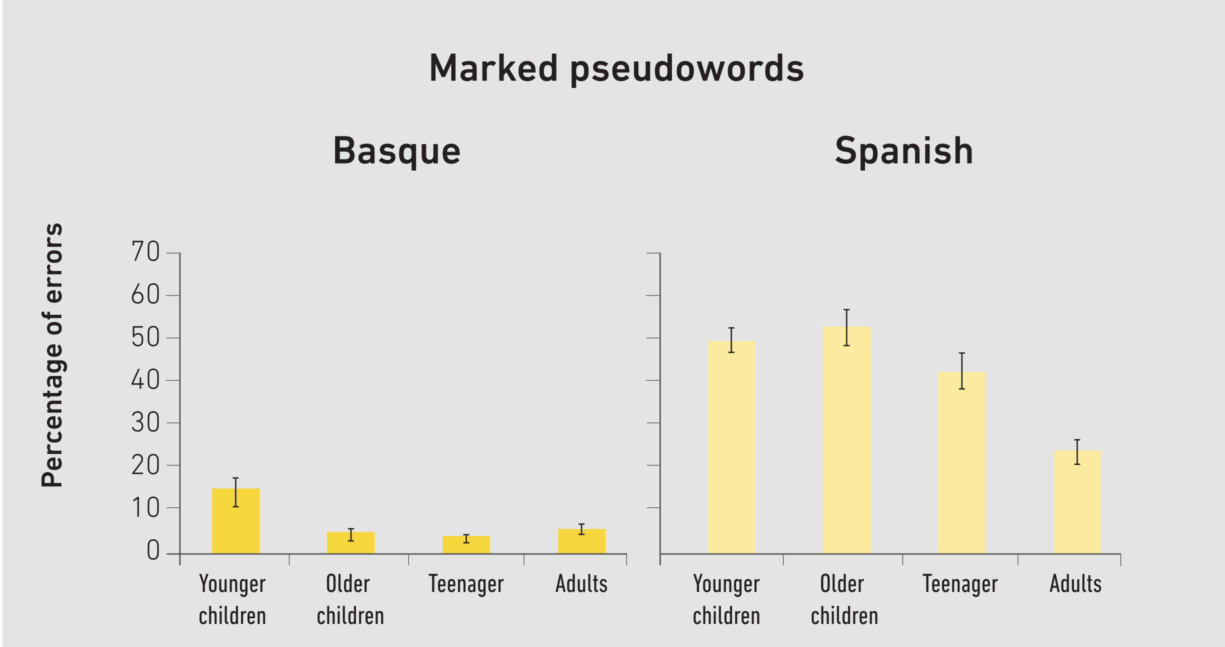


Figure 15. *Bar plots of marked pseudowords represented by the percentage of errors. Error bars refer to the standard error (SE) of each mean..*

Analysis of reaction times

The analysis of marked pseudowords on reaction times was carried out. The main effect of Language (*F1(1,116)=128.87, p<.001,* η_p^2*=.107; F2(1,22)=24.97, p<.001,* η_p^2*=.128*) was significant, showing that participants needed more timed to respond to Spanish-marked pseudowords. Also, Group (*F1(3,116)=31.1, p<.001,* η_p^2*=.353; F2(3,66)=171.26, p<.001,* η_p^2*=.534*), and the interaction between the variables (*F1(3,116)=5.32, p=.026,* η_p^2*=.004; F2(3,66)=9.64, p<.001,* η_p^2*=.025*) was significant.

To follow up on this interaction, Basque-marked pseudowords and Spanish-marked pseudowords were analyzed separately. Basque-marked pseudowords showed a group effect (*F1(3,87)=38.6, p<.001,* η_p^2*=.499; F2(3,120)=290, p<.001,* η_p^2*=.689*), reflecting that younger children performed worse than the rest of the groups as they needed more time (*see Table 6*). The other groups performed similarly in this condition (*see Table 6*). These results are in line with the percentage of error results (*see Figure 16*).

However, reaction times in Spanish-marked pseudowords showed different results than the percentage of errors. Although the effect of group was significant (*F1(3,87)=18.4, p<.001,* η_p^2*=.320; F2(3,120)=97, p<.001,* η_p^2*=.549*) as we saw in the other analysis, younger children performed worse than the other groups (*see Table 6*), while the other groups needed a similar amount of time to decide the language of the Spanish-marked pseudowords (*see Table 6 and Figure 16*).

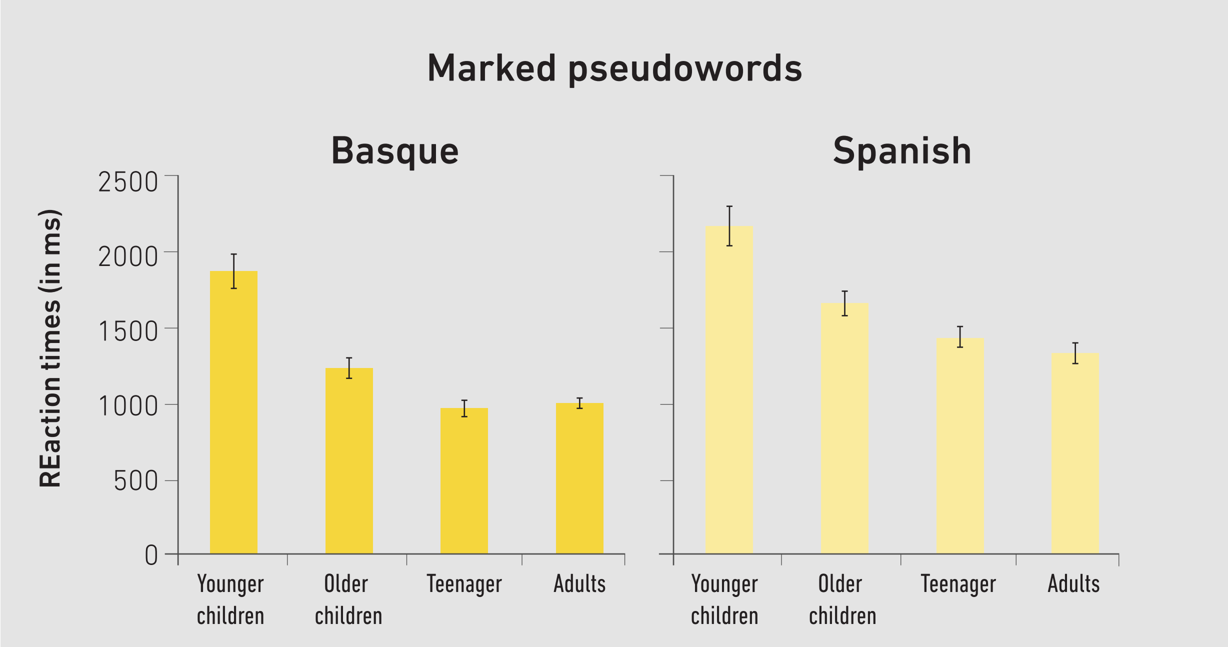


Figure 16. *Bar plots of marked pseudowords represented by reaction times in milliseconds. Error bars refer to the standard error (SE) of each mean.*

iv. Summary of results

The aim of this study was to investigate whether sensitivity to orthographic markedness changes across childhood and early adulthood in bilinguals whose languages share the same alphabet but differ in terms of orthographic regularities, as is the case with Spanish and Basque. We compared four age groups of Spanish-Basque bilinguals that acquired their second language early in life (*younger children, older children, teenagers, and adults*) in a language decision task with words and pseudowords. The results gave us a better understanding of developmental stages, showing that sensitivity to markedness changes with words and pseudowords.

In other words, results showed that people have different sensitivity to markedness depending on whether they have to categorize words in the mother tongue (*Spanish*) or the second language (*Basque*). In Basque, people detected words more easily when they contained marked bigrams (*e.g., tx is a marked bigram in Basque*) compared to words with bigrams shared by the two languages. This suggests that people are highly sensitive to markedness in the second language and they trust marked bigrams when attributing words to a language, consistent with prior research (*Casaponsa et al., 2014; Casaponsa & Duñabeitia, 2016; Chetail, 2015; Lemhöfer & Dijkstra, 2004; Van Kesteren et al., 2012*). However, this sensitivity to markedness in the second language seemed to change with age. Although younger children performed worse overall in the task than the rest of the group, they benefited more from Basque-marked words and that is why they showed a larger markedness effect. This trend seemed to change at the age of twelve. Older children began to rely less on marked bigrams to decide the language of second language words, much like the older age groups. Although it seems that as bilinguals grew up they depended less on markedness, what probably happened is that by the age of twelve, bilinguals needed fewer language cues to determine language due to changes in their implicit knowledge (*Janacsek et al., 2012*). Even though in Spanish, younger children performed worse than the other groups, results showed that participants were not very sensitive to markedness, performing equally well when presented with marked and unmarked words. These results with Spanish words suggested that people might already be very good at detecting words in their native language and may therefore not have paid attention to orthographic cues.

On the other hand and surprisingly, when we used pseudowords, which are visually similar to words (*orthographic regularities*) but don't include word meaning, the development of sensitivity to markedness changed in L1 with age. Basque-marked pseudowords showed the same results as words, with younger children performing worse than the other groups. However, Spanish-marked pseudowords showed different results. The adults performed differently from the other groups. It seems that adults benefited more from using orthographic cues to determine whether marked pseudo-words were Spanish-like. This could be due to their language proficiency and experience.

III. Experiment 2. Changes in sensitivity to orthographic markedness and second language learning

i. Rationale

This second experiment aimed to examine the function of orthographic regularities in the system after learning a second language later in life. Thus, the present longitudinal study aims to investigate whether adults change their sensitivity to orthographic markedness after learning a second language. Specifically, here we tested whether language learning late in life and progressive improvement in L2 skills could modulate learners' sensitivity not only to L2 orthographic regularities but also to the orthographic regularities of the L1. Note that sensitivity to orthographic markedness was tested before and after learning the second language to better understand the impact of second language learning on sensitivity to orthographic regularities.

It seems plausible that orthographic cues may play an important role in detecting new information in the absence of other language-based cues, for instance, when learning a second language. While learning a language a process for extracting statistical orthographic regularities takes place in learners' minds (*Bordag et al., 2017; Comesaña, Soares, Sánchez-Casas, & Lima, 2012*). Consequently, orthographic regularities should have direct implications for second language learning to correctly allow for language categorization. It seems reasonable that as learners become more proficient in a second language, the statistics of orthographic regularities would be better interiorized. If so, sensitivity to orthographic regularities should also vary across learning. Given that new regularities have entered into the system, general sensitivity to the orthographic regularities of both first and second language might change.

As learning requires so much time, we recruited older adults for this experiment. Older adults would have more time and might have more motivation to learn a new language. To this end, we used a project run in conjunction with the Basque government and the BCBL (*Basque Center on Cognition, Brain and Language*) which promoted older adults learning Basque as a second language (*the Garuna project*). In this project, older adults were native Spanish speakers immersed in a Basque language-learning course for one consecutive academic year but learners agreed to be tested for two consecutive years in order to observe possible changes while learning and after learning the second language. Participants performed a language decision task at three critical moments alongside regular testing of the learning experience. This data was collected in 2014-2015. The data were included in this thesis because they were of great relevance for the topic of this thesis. We carried out the data pre-processing and analysis.

ii. Methods

Participants

Thirty retired Spanish monolingual adults took part in this longitudinal experiment. However, only twenty participants remained throughout the two years (*8 females; mean age = 66.57; SD = 5.56*). All participants were living in the Basque Country, where the Spanish and Basque languages coexist. None of the participants had prior formal knowledge of Basque and they could not understand or produce linguistic structures in any other language than Spanish. However, they were exposed to basic Basque in everyday life such as street signs or spoken language in the street. All participants reported having normal or corrected-to-normal vision, and none of them had any history of chronic neuropsychological disorders.

The Garuna project was coordinated by the Department of Education, Linguistic Policy and Culture of the Basque Government. Thus, this project was carried out at The Center of Continuing Education for Adults in Bilbao (*Basque Country*). Participants were recruited by advertisements. The Center of Continuing Education for Adults advertised free Basque lessons for older adults with the condition that they had no prior knowledge of formal Basque. The benefit of going to free Basque classes was in exchange for participating in experimental tests. Every participant signed a written consent form approved by the Ethics and Research Committees of the Basque Center on Cognition, Brain, and Language (*BCBL*) before the start of research or educational actions.

Participants undertook Basque lessons for one whole academic year at the Center of Continuing Education for Adults, starting in October and finishing in June. The maximum number of participants per class was 10. In total, participants attended 5.30 hours of classes per week, distributed across three sessions held on working days. Classes were taught by native Basque-Spanish bilingual professional language trainers who were specialized in adult teaching.

Participants were tested at three different moments by a research assistant in the Center of Continuing Education for Adults, at the beginning of the academic year (*pre-test*), at the end of that same academic year (*post-test 1*), and one year after the second test (*post-test 2*). However, the assessments for cognitive and language proficiency tasks were carried out just at the beginning of the first academic year (*pre-test*) and at the end of the academic year (*post-test 1*). At post-test 1, a general measure of intelligence (*IQ*) was obtained based on participants composite scores in the verbal and non-verbal tasks of the Kaufman Brief Intelligence test (*K-BIT; see in Kaufman, 2004; see Table 7*). Also, age-related cognitive functioning was assessed using the Spanish version of the Mini-Mental State Examination (*MMSE; see Lobo, Ezquerra, Gómez, Sala, & Seva, 1979*) to ensure that participants had no cognitive decline that might prevent adequate Basque learning (*Harris et al., 2009; Ware et al., 2017*). After the cognitive task, the language proficiency tasks were carried out. Participants had to characterize their linguist profile in a subjective manner. They had to score their proficiency in Spanish and Basque via self-report, rating their knowledge of these languages on a scale from 1 to 10 (*see Table 7*). To ensure that participants had no prior knowledge of Basque, teachers also provided a subjective proficiency rate. They had to report their perception of participants' Basque proficiency before the lessons started. Subjective measures of Basque proficiency were carried out at the end of the academic year to ascertain whether participants had learned Basque. In addition to these subjective measures, extra objective measures of Basque knowledge were tested. First, participants had to name sixty-five common names in Basque in an adapted version of the picture naming test (*de Bruin et al., 2017*) (*see Table 7*). Also, a beginner standardized language test was performed (*A1 level*) of the Common European Framework for Reference (*CEFR, Council of Europe, 2011*). The maximum score on this test is 20.

Table 7. Descriptive statistics of assessments

	Pre-test	Post-test 1
Age	65.2 (3.81)	-
Cognitive function (MMSE)	28.8 (1.24)	-
Self-perceived Spanish competence	8.1 (0.55)	-
Self-perceived Basque competence	0	5.75 (1.45)
Teacher-perceived Basque competence	0	6.15 (1.09)
A1 level score	-	19.7 (4.28)
Picture naming	-	27.85 (10.26)
IQ	82.3 (28.34)	-

Note. Values reported are means and standard deviation in parenthesis on age (in years), scholar age (in years), self perceived language competence in Spanish/Basque (0-10 scale), teacher perceived Basque (0-10 scale), A1 level (direct punctuation), Picture naming (% correct), IQ (verbal and non-verbal task), Cognitive function (direct punctuation in which 30 is the maximum score indicating a "normal" range)

Materials

Corpus of bigrams

A similar corpus of bigrams to that used in Experiment 1 was used for this experiment. But in this database all letters were included, even ones that did not exist in the other language such as c or v which exist in Spanish but not in Basque.

Language decision task

A hundred and thirty-five pseudowords were generated using Wuggy (*Keuleers & Brysbaert, 2010*). Forty-five of these pseudowords were Spanish-marked pseudowords, forty-five were Basque-marked pseudowords, and forty-five were unmarked pseudowords (*see Appendix 3*). As in Experiment 1, marked pseudowords were created using at least one bigram that does not exist in the other language. Marked bigrams were controlled to have 0 frequency of use in the other language (*e.g., Basque-marked bigrams: ko, kz, ok, ts, tx, uj, zk, uk, np and Spanish-marked bigrams: ci,cl, ev, iv, lc, ña, ni, ño, ps, rc, sv, uc, vu*). The other bigrams that were used to create the marked pseudowords were legal bigrams that exist in both languages (*see Table 8 and to see all pseudowords created in Appendix 1*). In addition, unmarked pseudowords were created using bigrams that exist in both languages (*see Table 8*). Furthermore, the number of orthographic neighbors in Spanish and Basque were controlled to be equal in unmarked pseudowords and in marked pseudowords (*see Table 8*).

Table 8. Descriptive statistics of characteristics of the materials

	Spanish-marked	Basque-marked	Unmarked
Word length	6.11 (1.35)	6.11 (1.54)	5.89 (1.47)
Spanish bigram frequency	0.75 (0.26)	0.31 (0.19)	0.75 (0.28)
Basque bigram frequency	0.28 (0.15)	0.78 (0.15)	0.78 (0.15)
Orthographic neighbors in Spanish	1.1 (0.69)	1.64 (2.67)	1.6 (2,15)
Orthographic neighbors in Basque	1.71 (2.19)	0.98 (1.19)	1.76 (2.5)

Note. Values reported are means and standard deviation in parenthesis on word length (number of letters), bigram frequency (percentage per million), orthographic neighbors (number of words).

Procedure

Participants were tested alone in a quiet room. The procedure for the three tests (*pre-test, post-test 1, and post-test 3*) was the same. A fixation cross appeared in the middle of the screen for 500 ms before each trial. Immediately after the cross fixation, the target word appeared in the middle of the screen for 3000 ms or until the participant's response. At the beginning of the task, participants performed some trials for practice. They were informed that they would see pseudowords and they had to decide whether the string of letters could belong to Spanish or Basque (*i.e., forced-choice*). Participants had to respond as fast as possible by pressing a keyboard, the key 'S' if they thought string could be Spanish and the key 'B' if it could be Basque.

Data analysis

Accuracy and reaction times were collected in this experiment as the dependent variables of interest. Before data analysis, outliers were removed using the statistical environment R (*R core team, 2013*). Responses below 200 ms (*0.01 % outliers*) were considered as change responses and timeouts (*0.04% outliers*) were excluded. Also, responses above 0.75 and below 0.25 interquartile from the participant-based (*1.51% outliers*) and item-based (*2.08% outliers*) mean for all within-factors were excluded from the analyses (*in total, 3.17% of outliers*). After outlier removals, percentage of errors and reaction times removing erroneous responses were analyzed using Jamovi 0.9.6.7 with an $ANOVA_2$ test.

Note 2: Note that this data was also analyzed using Mixed-effects models, accuracy with logistic mixed-effects models and reaction times with linear mixed-effects models (Baayen, Davidson & Bates, 2008; Jaeger, 2008; Barr, 2013), using lme4 package for R (Bates, Maechler, Bolker & Walker, 2014), including subjects and items as random factors. These analyses showed the same results as repeated measures ANOVAs. In the present thesis, we report ANOVAs to be consistent with the analyses reported in other chapters.

Three independent variables of interest were manipulated in this experiment, Language (*Basque|Spanish*), Markedness (*Marked|Unmarked*), Tests (*pre-test|post-test 1|post-test 2*) as within-subjects. Marked and unmarked pseudowords were analyzed separately because decisions made on unmarked pseudowords cannot be characterized as correct or incorrect responses in the absence of language cues. So, unmarked pseudowords were modeled as the probability of Spanish responses to observe the preference that participants had to respond to one language or the other. The probability of Spanish responses was recorded as a variable coded, '1' if participants responded Spanish and '0' if they responded Basque. A one way ANOVA was carried out on the probability of Spanish responses by Tests (*pre-test|post-test 1|post-test 2*) to observe if participants responded differently across time in the tests for unmarked pseudowords. Also, reaction times for unmarked pseudowords were analyzed using two-way repeated measures ANOVAs by Language response (*Spanish|Basque*) and Tests (*pre-test|post-test 1|post-test 2*). On the other hand, marked pseudowords were analyzed taking two factors as variables of interest, Tests (*pre-test|post-test 1|post-test 2*) and Language Markedness (*Basque|Spanish*) with two way repeated measures ANOVAs on percentage of errors and on reaction times without incorrect responses. Averaged reaction times and accuracy rates per condition are presented as well as probability of Spanish responses in Table 9.

Table 9. Descriptive statistics for the language decision task in the three different test moments.

	MARKED		UNMARKED	
%errors	Basque	Spanish	Probability of Basque	Probability of Spanish
Pre-test	92.68 (26.06)	91.82 (27.42)	32.28 (46.78)	67.72 (46.78)
Post-test 1	94.59 (22.63)	94.94 (21.93)	28.2 (45.02)	71.8 (45.02)
Post-test 2	93.5 (24.67)	92.74 (25.96)	28.03 (44.94)	71.97 (44.94)
RT	Basque	Spanish	Probability of Basque	Probability of Spanish
Pre-test	854 (266)	978 (402)	1282 (549)	1042 (443)
Post-test 1	883 (278)	931 (344)	1265 (522)	1011 (439)
Post-test 2	868 (264)	891 (283)	1257 (497)	993 (413)

Note. Values reported are means and standard deviation in parenthesis on accuracy (%errors), and reaction times (milliseconds).

iii. Results

Unmarked pseudowords

Analysis of percentage of errors

The analysis of the probability of Spanish responses across Tests (*pre-test|post-test 1|post-test 2*) was carried out. The analysis revealed no effect on the probability of Spanish responses across Tests (*F1(2,38)=0.28, p=.756,* η_p^2*=.010; F2(2,268)=0.07, p=.933,* η_p^2*=.001*), showing that participants tended to responded to unmarked pseudowords equally across the time points. To observe the preference of response, a paired-sample t-test was made comparing the probability of Spanish response and the probability of Basque response. The test was significant, showing that overall participants preferred to attribute unmarked pseudowords to Spanish (*t1(19)=4.94, p<.001, Cohen's d= 1.10; t2(44)=6.16, p<.001, Cohen's d=0.92*).

Analysis of reaction times

Furthermore, an analysis of reaction times on unmarked pseudowords based on the Language response (*Spanish|Basque*) and Test (*pre-test|post-test 1|post-test 2*) was made, showing that participants were faster at responding to unmarked pseudowords as Spanish than as Basque (*F1(1,18)=12.67, p=.002,* η_p^2*=.041; F2(1,35)=56.89, p p<.001,* η_p^2*=.265*). However, the main effect of the test and the interaction between the variables were not significant (*F1(2,36)=0.56, p=.575,* η_p^2*=.007;*

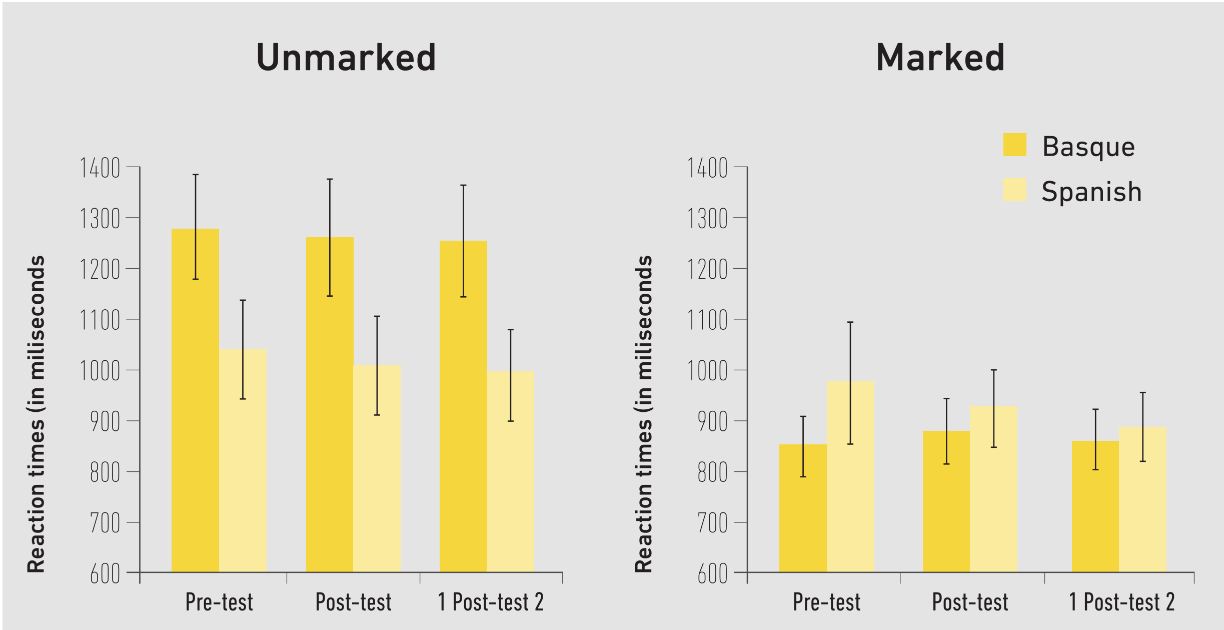


Figure 17. *Bar plots of reaction times (in milliseconds) during the language attribution task for unmarked and marked pseudowords for pre-test, post-test 1 and post-test 2. Error bars refer to the standard error (SE) of each mean.*

F2(2,70)=1.28, p=.284, η_p^2*=.007 and F1(2,36)=0.85, p=.436,* η_p^2*=.003; F2(2,70)=0.19, p=.821,* $^{2}_{p}$ *=.001, respectively*). Putting these results together with the probability of Spanish response, showed that overall participants performed equally with unmarked pseudowords across the time periods, but they preferred to attribute unmarked pseudowords to Spanish, also showing better reaction times when they attributed these pseudowords to Spanish (*see Figure 17*).

Marked pseudowords

Analysis of percentage of errors

The analysis of accuracy in percentage of errors on the presence of language Markedness orthographic regularities (*Spanish-marked, Basque-marked*) across Tests did not reveal any significant main effect or interaction (*all ps > .28*). Overall accuracy ratings were already close to ceiling in both the Spanish-marked and Basque-marked pre-tests. Thus, in this experiment we focus on the reaction time results as the dependent variable of interest.

Analysis of reaction times

Analyses of reaction times on marked pseudowords removing incorrect responses were carried out. Results did not show a significant effect of Test (*F1(2,38)=0.61, p=.553,* η_p^2*=.007; F2(2,88)=5.54, p=.005,* η_p^2*=.017*). However, results revealed a significant language Markedness effect (*F1(1,19)=11.41, p=.003,* η_p^2*=.040; F2(1,44)=11.35, p=.002,* η_p^2*=.068*), showing that overall participants were slower at detecting Spanish-marked than Basque-marked pseudowords. And importantly, an interaction was found between language Markedness and Test (*F1(2,38)=5.19, p=.010,* η_p^2 *=.014; F2(2,88)=12.13, p<.001,* η_p^2 *=.026*).

To follow up this interaction a Tukey Post-Hoc test was carried out. Comparisons revealed that participants in the pre-test were significantly slower at responding to Spanish-marked pseudowords as compared to Basque-marked pseudowords (*t1(46)=4.65, p<.001; t2(76)=5.29, p<.001*). However, this difference was no longer significant after language learning (*post-test 1: t1(46)=1.67, p=.556, t2(76)=2.35, p=.186; post-test 2: t1(46)=1.31, p=.776, t2(76)=1.05, p=.900*). These results suggested that participants reduced the time needed to detect pseudowords in one of the two languages. Thus, multiple paired sample t-test that compared the different Test moments in the Spanish-marked and Basque-marked were carried out. Results showed that the response time between the three tests for Basque-marked pseudowords were the same (*pre-test-post-test 1: t1(19)=1.15, p=.261, Cohen's d=.398; t2(44)=1.54, p=.129, Cohen's d=.230; pre-test-post-test 2: t1(19)*=0.39, *p*=.739, *Cohen's d=.075; t2(44)=0.69, p=.493, Cohen's d=.103*). However, there were differences in reaction times between the tests in the Spanish-marked pseudowords (*pre-test-post-test 1: t1(19)*=1.88, *p=.081, Cohen's d=.397; t2(44)=2.31, p=.026, Cohen's d=.343; pre-test-post-test 2: t1(19)=0.197, p=.063, Cohen's d=.441; t2(44)=4.57, p<.001, Cohen's d=.681*), showing that participants reduced the time to detect Spanish-marked pseudowords after learning Basque (*see Figure 17*).

iv. Summary of results

This longitudinal study investigated changes in sensitivity to orthographic markedness in older adults after learning a second language late in life. Note that sensitivity to orthographic markedness was tested before and after second language learning, which made it possible to investigate the effects of second language learning.

Adults were Spanish native speakers with no prior knowledge of Basque, the target second language. While accuracy in detecting the language of marked pseudowords remained very high and constant across the three tests and no main effects or interactions were found, probably due to ceiling effects, reaction time results showed that before and after learning the second language, the presence of language-specific orthographic cues guided language classification.

When participants classified unmarked pseudowords, without language specific orthographic cues, they tended to classify these as belonging to their native language, Spanish. Participants also had faster reaction times for unmarked pseudowords which they deemed to be Spanish. On the other hand, marked pseudowords results showed sensitivity to markedness. Even though older adults were as accurate at detecting Spanish as they were at detecting Basque pseudowords, they responded more quickly to Basque-marked pseudowords than to Spanish-marked pseudowords across all three tests.

This results confirmed previous findings that showed that adults are highly sensitive to orthographic markedness in a second language even when they do not know the target language (*Casaponsa et al., 2014; Duñabeitia, Ivaz, & Casaponsa, 2016; Oganian, Conrad, Aryani, Heekeren, & Spalek, 2016*). This was shown by faster reaction times in the second language compared with the native language in the pre-test, suggesting that participants could easily establish that Basque-marked pseudowords did not conform to the L1 orthographic regularities. These results persisted during learning. However, post-test results showed that after second language learning, participants also demonstrated increased sensitivity to orthographic markedness in their native language as shown by faster reaction times in the native language at post-test than at pre-test. This strongly suggests that the native orthographic processing skills change due to the accommodation of newly acquired regularities in the second language.

IV. Discussion

Differences between the orthographic rules of two languages that share the same script, such as orthographic markedness, are extremely important for language detection, and ultimately for lexical access in bilingual and monolingual contexts (*Casaponsa et al., 2014; Casaponsa & Duñabeitia, 2015; Oganian et al., 2016; Van Kesteren et al., 2012*). However, to date, it is not clear when sensitivity to orthographic markedness changes across the lifespan and/or if it can occur as a consequence of second language learning early or later in life. Thus, the aim of this chapter is to better understand the function of orthographic regularities in word processing, specifically the development of sensitivity to orthographic markedness in the system across development in the bilingual mind. To do so, we investigated whether sensitivity to orthographic markedness changes with the development of the bilingual mind.

To examine for possible changes in sensitivity to orthographic markedness, we considered two possible factors that might impact orthographic processing: early second language learning in normal concurrent bilingual development (*Experiment 1*) and learning a second language (*L2*) after consolidating the native language (*Experiment 2*). We considered that not only might being bilingual and being raised in a bilingual environment influence the word processing system in the bilingual mind, but also becoming bilingual at a later age due to learning a second language could influence the system (*Cummins, 1993; Genesee, Paradis, & Crago, 2004*). Thus, we proposed a twofold approach to these questions using two experiments to test for possible changes in the system.

In order to test sensitivity to orthographic markedness, we ran a language decision task, in which participants had to decide whether the target word belonged to their first language (*L1*) or the second language (*L2*). Also, words were chosen to belong two different conditions: marked and unmarked. Marked words had bigrams that only exist in one of the languages and unmarked had only bigrams that exist in both languages. In Experiment 1, four groups of participants (*younger children, older children, teenager and adults*) performed the language decision task with words and pseudowords. We included pseudowords, in order to observe whether participants relied on sub-lexical cues to attribute language when they could not rely on semantic contexts. In Experiment 2, older monolingual adults learning a second language performed a language decision task at three different moments along the second language learning trajectory (*pre-test, post-test 1 and post-test 2*). Results showed differences in response to unmarked words/pseudowords and marked words/pseudowords.

On the one hand, results from Experiment 1 demonstrated that marked words resulted in a lower percentage of errors and less time was required to determine what language they belonged to than unmarked words. Overall bilinguals presented an advantage in words that have marked letter sequence (*sensitivity to orthographic markedness*) in their L2, but not in the L1. Vaid and Frenck-Mestre (*2002*) also found that marked words were responded to faster than unmarked words in second language words suggesting that it is driven by a perceptual effect of orthotactic (*sub-lexical*) information rather than complete lexical access. Bilinguals may not need full lexical access to second language representations. These results showed that language detection in the second language is mediated by sub-lexical characteristics of the words at early stages of visual word recognition. Marked orthographic cues facilitate language attribution in the second language but not in the native one. This means, that when a bilingual reader see a marked word in the second language, they don't rely on the meaning but rather on the sub-lexical strategy to attribute the language of the word (*Casaponsa et al., 2014; Casaponsa & Duñabeitia, 2015; Oganian et al., 2016; Vaid & Frenck-Mestre, 2002*).

This is in line with the mechanisms proposed in the BIA+ extended model (*for a review, see Van Kesteren et al., 2012,and Casaponsa et al., 2014 Chapter 1*). The BIA+ extended model stated that when a printed word is presented to a bilingual, both languages will initially be activated. Then, the presence of marked orthographic cues allows for selective lexical access. Orthographic markedness, via the sub-lexical information in letter strings, indicates language membership and thus directly supports the retrieval of lexical information. In short, a decision on language membership can be made based on the direct links established between sub-lexical nodes and language membership, making full lexical access unnecessary. In such cases, marked words in the second language can be detected using only a sub-lexical strategy. Thus, marked words in the second language showed better accuracy and reaction times compared with unmarked words. In contrast, words from the native language required more time to be processed and had a higher percentage of errors than second language words. And results showed that bilinguals did not rely on orthographic markedness to attribute the languages. This could suggest that bilinguals are very good at recognizing their L1 words.

On the question of development, we found that only younger children showed a disadvantage overall in marked and unmarked words, since they sometimes provided wrong answers and spent more time making decisions compared to the older groups in both languages. This may be because children at this age are not very competent readers and require more effort to recognize and decide on the language presented. This is related to the development of reading automatization (*Lyon, 1998; Samuels & Flor, 1997*).

In the initial stages of reading development, learning letter-to-sound decoding and practicing with texts is critical in order to acquire fluency and automaticity. However, considering that readers have only so much capacity for attention and memory, it seems likely that beginner readers read words in a laborious and inefficient manner. They may remember what they have read by relating text content to their background knowledge. Thus, they are less efficient in reading than their older peers (*Samuels & Flor, 1997*). Supporting this idea, Samuels, LaBerge, and Bremer (*1978*) described that the unit of word recognition for beginning readers was the letter, whereas for skilled readers the unit was the word (*which has meaning*). Holding meaningless letters in memory is more difficult than holding meaningful words. Thus, as automaticity develops, the size of the visual unit that must be held in short-term memory increases up to the level of meaningful words.

In this line of research, a study conducted by Garnetta and Fleischner (*1983*) tested younger children (*under 10 years old*) and older children (*over 10 years old*) in a rapid automatic naming task (*RAN*). This task provides a good index of automaticity by measuring the speed of repeated naming-to-stimulus (*Blachman, 1984; Denckla & Rudel, 1976*). They found that younger children were slower that older children, and the RAN was positive correlated with reading comprehension. This suggests that around the age of ten there may be a developmental change towards reading automaticity; the ability to decode and comprehend simultaneously is an important indicator that decoding is automatic. Thus, skillful readers may link the ideas presented in print to their own experiences and develop the necessary vocabulary to make sense of the content being read (*Lyon, 1998*). As our results also suggest, children around the age of twelve may change the way in which they perceive words in both their languages because they become skillful readers as adults. This age-specific transition is in line with some findings showing that around the age of twelve children experience changes in implicit learning (*Janacsek et al., 2012*).

On the other hand, in the absence of orthographic cues, as it is the case of pseudowords, individuals have to rely on their previous knowledge in order to determine the language of the target word. In this case, there is a high degree of activation from lexical candidates from both the L1 and the L2, leading to difficulty in selecting the correct representation from the lexicon. When lexical access is required, it is thought that the sub-lexical route is still in play, but decisions are also mediated by lexical search strategy. Thus, pseudowords give us reliable information about the use of sub-lexical cues. And again, results from pseudowords showed different results in marked and unmarked pseudowords.

Unmarked pseudowords were tricky to respond to because they could belong to either of the two languages since there is no orthographic cue to attribute language membership. From our results it seemed that unmarked pseudowords were attributed inconsistently in the two experiments. In Experiment 1, bilinguals tended to decide that unmarked pseudowords were Basque, until they reached adulthood. Adults, by contrast, preferred to attribute words to both of their languages. Individuals in Experiment 2 were predisposed to attribute unmarked pseudowords to Spanish. One possible explanation for this finding is that bilinguals (*Experiment 1*) may tend to attribute pseudowords to the second language because it is clear that they are not real words in the native language. Thus, in a decision task they may prefer

to categorize them as coming from the other language. On the other hand, monolinguals and bilinguals after second language learning (*Experiment 2*) may consider familiar orthographically regular sequences to be part of their previous knowledge. In line with this assumption, previous research (*Ellis & Beaton, 1993*) has shown that people prefer to learn letter sequences that follow sequences found in their native language, suggesting they have a preference for letter sequences that follow L1 orthographic rules.

However, marked pseudowords have orthographic cues that allow them to be correctly attributed to a specific language. Results showed that individuals seem to be highly sensitive to orthographic marked cues in an L2. Orthographic markedness helped participants to attribute words to to the second language pseudowords at all ages (*Experiment 1*) and even monolinguals (*Experiment 2*) were very good at attributing words with orthographic distinctiveness to a different language quickly and accurately. It seems that people are very sensitive to orthographic statistical regularities in their second language by nature and this helps them to categorize languages even before they access meaning (*Casaponsa et al., 2014; Casaponsa & Duñabeitia, 2015*). And again, these results showed processing markedness through a sub-lexical route, requiring fewer steps to reach the goal of identifying language membership (*BIA+ extended model by Casaponsa et al., 2014*). But, our results also show that the pattern of sensitivity to markedness in the L1 is strikingly different from that for the L2.

Taken together, the present results support the contention that bilinguals with languages that are orthographically different from each other differ in their sensitivity to orthographic markedness in the L1 and L2. Bilinguals from an early age (*as Experiment 1 showed*) and even monolinguals (*as Experiment 2 showed*) are very sensitive to orthographic markedness in the process of visual word recognition in the L2 (*Casaponsa et al., 2014; Casaponsa & Duñabeitia, 2016*), but it seems that the sensitivity to orthographic markedness in the second language may change at the age of twelve (*Janacsek et al., 2012*). Older children rely less on orthographic markedness in the second language to determine the language of the word (*this finding will be discussed in Chapter 4. General discussion*). On the other hand, in the native language where there is access to the meaning of the word, individuals do not need to pay attention to orthographic markedness in order to attribute language membership (*as Experiment 1 showed*). It seems that bilinguals are not sensitive to orthographic markedness of words at any age, but again around the age of twelve older children change the way that they process words. But, in the absence of such lexical familiarity, as is the case with pseudowords, adults start to rely on orthographic markedness in the L1 to identify the language of the pseudoword. Both Experiments 1 and 2 showed that the use of orthographic markedness in L1 seems to change in balanced bilinguals both across normal development and after learning a second language (*this finding will be discussed in Chapter 4. General discussion*). This demonstrates that the native language is dynamic and permeable and changes during second language learning (*Kroll, Bob, & Hoshino, 2014*).

Chapter 3
Novel word learning and orthographic markedness

I. Overview

Up to this point, we have seen that bilinguals who know two languages that share the same script use orthographic markedness to recognize words as well as detect the language of the word (*Casaponsa & Duñabeitia, 2016; Lemhöfer, Koester, & Schreuder, 2011; Oganian, Conrad, Aryani, Heekeren, & Spalek, 2016; Van Kesteren, Dijkstra, & de Smedt, 2012*). In Chapter 2, we saw that sensitivity to orthographic markedness may change with normal development and with second language learning. It seems that children at the age of twelve change the way they process orthographic markedness, in both their native and second languages. And, surprisingly, adults in normal bilingual development and after learning a second language change the way they process orthographic markedness in their native language. These findings suggest that becoming aware of the differences between the orthographic regularities of two languages allows bilinguals to change their sensitivity to markedness in the native language. We noted that this challenges the notion that the native language is stable across time as previously described in the BIA+ model (*Dijkstra & van Heuven, 2002b*).

This increased awareness of differences in orthographic regularities through development suggests that orthographic markedness could also play a role in new word learning. Thus, we propose to investigate how sensitivity to orthographic regularities impacts learning new words. The present chapter focuses on how orthographic markedness interacts with novel word learning in bilingual children and adults. For this purpose, we will focus on whether orthographically illegal novel words (*novel words with illegal bigrams*) are learned differently from those with legal combinations (*novel words with only legal bigrams*). And, whether learning novel words is affected by being a bilingual whose languages share the same script or reflects bilinguals increased experience with the similarities and/or differences between the orthographic regularities of their two languages. For instance, another language that coexists in Spain is Catalan. Spanish and Catalan share the majority of their letters; except that the letter *ñ* only exist in Spanish, while *ç* only exists in Catalan, and there are a few letter sequences distinct to each language (*only ch is a marked bigram in Spanish and ny, ss, tz are marked bigrams in Catalan*). However, Spanish and Basque differ far more both in terms of distinctive letters and distinctive bigrams. Spanish has five unique letters that Basque does not have (*ñ, c, q, v, w*) and the two language differ in terms of the following bigrams: mp, mb are marked bigrams in Spanish; *kr, kt, np, nb, lk, nk, zt, rk, sk, zk, ts, tx, tz* are marked bigrams in Basque.

We included a monolingual group to compare with the bilingual groups. Previous research has suggested that both bilingual adults and children are better at word learning than monolinguals due to their broader experience with language learning (*see Hirosh & Degani, 2018 for a review*). This may be because bilinguals know that objects can have different names in their languages and can therefore link words from another new language more easily to a known concept than monolinguals (*Au & Glusman, 1990; Kaufman, 2004*). Along these lines, it was shown that, from a very early age, bilingual children (*English-Chinese, English-French, English-Spanish, English-Russian, English-Urdu and English-Vietnamese*) learned novel words better than their monolingual peers in a task where they had to associate novel words with a corresponding referent (*Yoshida & collegues, 2011*). These benefits for word learning have been observed both for young bilingual children who learned their languages in a classroom environment (*Kaushanskaya et al., 2014; Mady, 2014*), as well as bilingual children who acquired both languages from birth (*Kahn-Horwitz, Kuash, Ibrahim, & Schwartz, 2014; Yoshida et al., 2011*). In general, these findings have beeen found in bilinguals speaking two languages with distinct orthographic systems, as is the case with English and Chinese.

In line with these findings, Kaushanskaya and Marian (*2009a,2009b*) ran a word-learning task (*see Figure 18*) with bilingual and monolingual adults. The task included novel words which would be unfamiliar to participants. Participants had to learn the new words as translation equivalents of existing words from their native language. Results showed that bilinguals outperformed monolinguals in their learning performance (*see Kan & Sadagopan, 2014, for a review*). However, the bilingual learners in this study had very different language combinations: English-Spanish share the same orthographic code, while English-Mandarin use highly distinct orthographic codes. It is possible that bilinguals who speak more orthographically distinct languages use a more flexible orthographic system that allows them to map orthography to phonology more easily. This in turn might allow these bilinguals to more readily accept alternative letter sequences.

These experiments suggested that the experience of managing two languages, in general, may enhance learning and may change how novel words are acquired. However, these findings do not necessarily imply that all types of bilinguals will learn novel words better than monolinguals, because the above-mentioned studied tested bilinguals with two languages that clearly have different orthographic regularities and phonotactic structures. Werker and Byers-Heinlein (*2008*) underscore the importance of the specific language pairs in the bilingual language system and their interaction. Thus, it could be tentatively hypothesized that a difference in language scripts is the underlying factor that makes it easier for bilinguals to learn new items. What happens if the languages that are mastered share the same script?

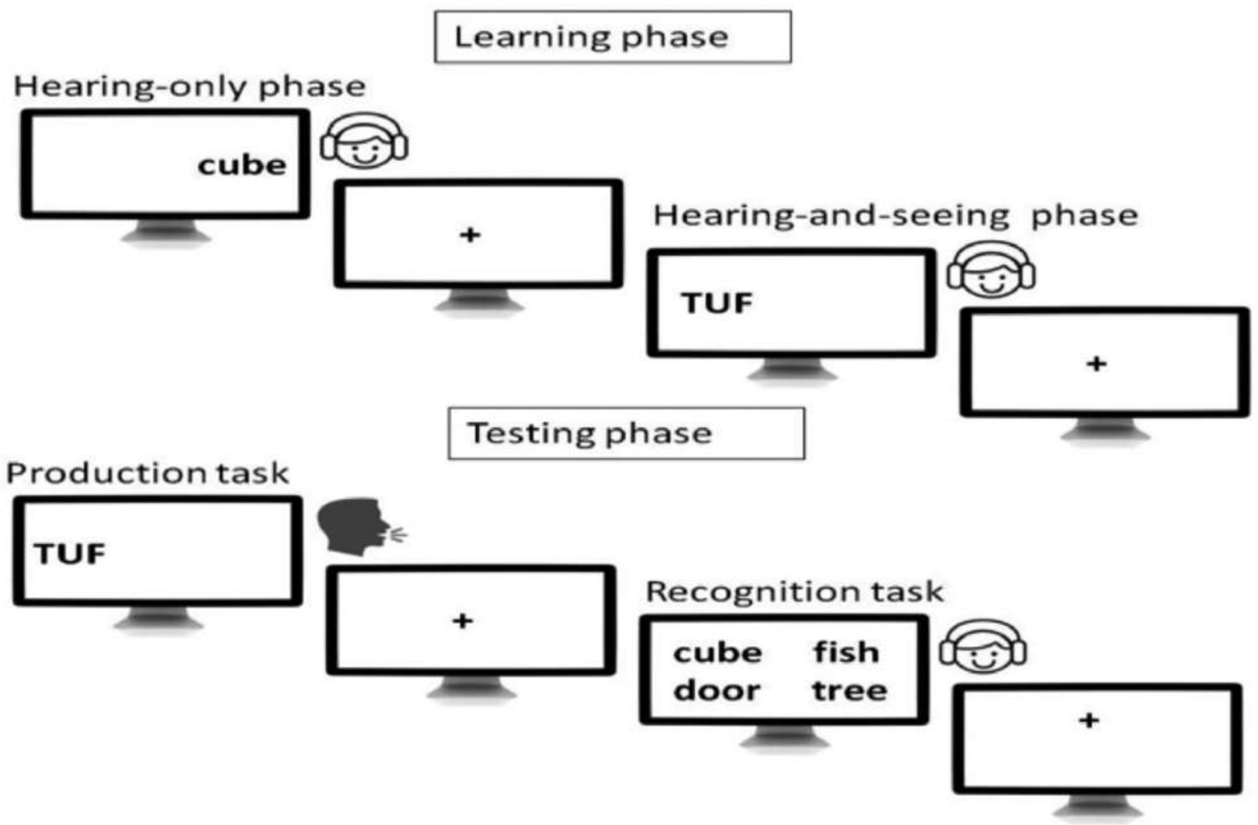


Figure 18. *A Word learning task adapted from (Kaushanskay & Marian, 2009). Vocabulary learning was divided into a hearing-only phase, in which participants heard the novel words and saw their written English translation, and the hearing-and-seeing phase, in which participants heard the novel word and saw the written form of the novel word. Vocabulary testing was divided into a production task, in which participants heard the novel word and had to pronounce its English translation, and a recognition task, in which participants heard the novel word and had to choose the correct English translation.*

As we saw in the previous chapter, orthographic markedness is an important cue for bilingual word processing when languages share the same script. It facilitates word recognition as well as language identification. Thus, we hypothesized that bilinguals, whose two languages share the same script, will also rely on their orthographic knowledge to learn new words; the characteristics specific to each language may affect how new pieces of information are learned. Previous research has shown that bilinguals prefer to learn words with low frequency letter sequences if they have to learn words through reading (*Lutjharms, 1994*) because these words are more salient in the text. However, when bilinguals are forced to learn new words, as is the case with a word learning task, bilinguals prefer to learn words that have high frequency letter sequences (*Ellis & Beaton, 1993; Speciale, Ellis, & Bywater, 2004*). Certainly, learning new phonological and orthographic sequences that also exist in one's native language(s) should be easier than learning completely different letter sequences. In line with this finding, Bialystok and colleagues (*2005*) showed that bilinguals whose two languages share the same print-to-sound principle and/or the same writing system (*i.e., Spanish-English*) performed better on a meta-phonological task (*counting the number of sounds in a word*) than bilinguals with two languages that follow different writing systems (*i.e., Chinese-English*).

Hence, the purpose of this chapter was to investigate whether simply being bilingual facilitates learning new words, or whether handling the different characteristics of the two languages, in this case different orthographic regularities, benefits learning of new words. To do so, we examined how learning new words depends on the sub-lexical characteristics in monolinguals and two groups of bilinguals, one whose languages were orthographically similar and the other whose languages were orthographically dissimilar. Novel words either violated or respected the orthographic rules of the languages they know. Legal words contained only bigrams that exist in the bilingual's languages, while illegal words included some bigrams that did not exist in either of their languages.

In the first series of experiments (*Experiments 3A and 3b*), we focused on older children (*as we saw in Chapter 2, around the age of 12 children may change the way they perceive sensitivity to markedness*), and then asked the same questions with young adult samples (*Experiment 4*) to explore the degree of generalization across age groups and the extent to which the effects may depend on a multilingual school environment.

Bilingual children attending a bilingual school need to deal with their two languages in printed materials within a single school context. They have to read in both languages and are constantly exposed to both written languages. The bilingual school scenario, in which two languages coexist, is markedly different from the scenario most adults experience, where concurrent exposure to two written languages is far less common. Thus, children may develop different strategies than adults to deal with their school context and its demands. In sum, we investigated if new vocabulary acquisition is easier for all types of bilinguals as compared to monolinguals (*see Margarita Kaushanskaya & Rechtzigel, 2012*), or if this benefit depends on the specific sub-lexical characteristics of the language combinations of these bilinguals, paying special attention to orthographic regularities.

In case we observed differences in novel word learning performance in a particular group that were not due to bilingualism per se, but rather because that group had experience handling two languages with different sub-lexical characteristics, we wanted to ascertain whether this effect was related to sensitivity to orthographic markedness. If so, we would conduct an extra experiment (*Experiment 5*) to observe the relationship between the novel word learning task and sensitivity to orthographic markedness in the critical group(*s*).

II. Experiment 3. Learning novel words with illegal orthographic regularities during late childhood

i. Rationale

As we saw in the previous section, the aim of this chapter is to investigate whether novel word learning with different orthographic regularities (*legal and illegal novel words*) is facilitated by being bilingual per se or relies on linguistic experience with the differences between the orthographic regularities in two languages. Thus, we focus on orthographic regularities as the main differences between two bilingual languages that share the same script.

In this first experiment (*Experiment 3*), we explored our research question in older children. As we saw in Chapter 2, at the age of twelve, children appear to change the way they perceive orthographic regularities. Thus, we selected children of this age in three different communities, Spanish-monolinguals, Spanish-Basque bilinguals and Spanish-Catalan bilinguals. We predicted that bilinguals who have to deal with languages that are more dissimilar at the level of orthographic regularities would benefit when learning novel words containing illegal bigram combinations since they might find it easier to deal with novel letter sequences due to their broader linguistic experience.

With this in mind, we focused on two bilingual communities with two language pairs, Spanish-Catalan and Spanish-Basque because while these three languages share the same Roman alphabet, their sub-lexical structures vary. Spanish and Catalan share most orthographic regularities, whereas Spanish and Basque are very dissimilar in their graphemic structure. Basque has many bigram combinations that are illegal according to Spanish (*and Catalan*) orthographic rules. In both the Spanish-Catalan and Spanish-Basque cases, participants are continuously exposed to both languages in their communities, both in daily life and through printed materials in the school context.

ii. Experiment 3A

01. Methods

Participants

Seventy-two children (*45 females; mean age=12.9 years, SD=0.8*) from three language communities (*Spanish-monolinguals, Spanish-Basque and Spanish-Catalan bilinguals*) took part in this experiment. Spanish and either Catalan or Basque occur within the same environments in specific bilingual areas in Spain. So, older children were recruited from three different schools that were located in different Autonomous Communities in Spain. First, Spanish monolinguals (*24 participants*) were recruited in Santander (*Cantabria*), which is a monolingual region located in the North of Spain. Second, Spanish-Catalan bilinguals (*24 participants*) were recruited in Barcelona (*Catalunya*), a bilingual community on the North East coast. And third, Spanish-Basque bilinguals (*24 participants*) were recruited in Vitoria (*Basque Country*), a bilingual community on the North coast. All participants were right-handed, and none were diagnosed with language disorders, learning disabilities, or auditory impairments. The protocol was carried out according to the guidelines approved by the BCBL (*Basque Center on Cognition, Brain and Language*) Ethics Committee in line with the Helsinki Declaration. Children's parents and the children themselves were appropriately informed about the experiments. Legal guardians signed consent forms before the experiment began. They were also asked to fill in a short parental questionnaire giving subjective ratings of the linguistic competence of their children and specifying their socioeconomic status.

The three communities selected represent markedly different language environments. Spanish monolingual children lived in a Spanish-only environment and attended a Spanish monolingual school (*daily percentage of exposure: to Spanish, M=93.7%, SD=1.56; to English, M= 6.3%, SD=2.43*). On the other hand, we ensure that both bilingual groups had acquired both their languages before the age of 6. Spanish-Catalan bilingual children were raised in a bilingual community and educated in a Spanish-Catalan bilingual school (*daily percentage of exposure: to Spanish, M=47.9%, SD=6.96; to Catalan, M=45.2%, SD=5.54; to English, M=6.9%, SD=3.48*). Spanish-Basque bilingual children were also attending a Spanish-Basque bilingual school (*daily percentage of exposure: to Spanish, M=52.8%, SD=2.54; to Basque, M=39.9%, SD=2.46 to English, M=7.3%, SD=2.79*). Note that although learning English is the norm in all schools in Spain, we controlled that all groups had similar competence in English as well as same amount of exposure to this language, to avoid any confound.

We controlled that older children had similar linguistic profiles. To do so, we assessed language proficiency, socioeconomic status, and IQ (*see 10*). Language proficiency was measured with one subjective scale and two objective scales. In the subjective scale, parents had to rate their child's language competence on a scale from 0 to 10. In the objective scales, we first administered a 20-item adapted version of a picture naming task (*de Bruin et al., 2017*), then a lexical decision task (*the LexTale cf., for the English version; Lemhöfer & Broersma, 2012; for the Spanish version; Izura et al., 2014; and the Basque version; de Bruin et al., 2017*), note that there is no Catalan version). In addition, we also made sure that, despite English being a mandatory subject in all Spanish

schools (*Age of Acquisition=8.67, SD= 2.14*), the participants' English level was relatively low as assessed by the English subjective scale, LexTale, and picture naming task (*see Table 10*). Furthermore, IQ was measured with a 6-minute abridged version of the K-BIT (*Kaufman, 2004*), in which older children had to complete as many matrices as they could in the time provided. Also, socioeconomic status was measured by asking parents to indicate on a scale from 1 to 10 how they perceived their socioeconomic situation as compared to other members of their community (*Adler & Stewart, 2007*).

Table 10. Descriptive statistics of assessments

	Monolinguals	Spanish-Basque bilinguals	Spanish-Catalan bilinguals	ANOVAs	
				F (df)	p
Age	13.13 (0.90)	12.71 (0.91)	13.08 (0.72)	F(2,69)=1.76	.179
Spanish competence	9.58 (0.97)	9.04 (0.91)	9.46 (0.72)	F(2,69)=2.05	.141
Basque competence	-	6.38 (0.88)	-	-	-
Catalan competence	-	-	9.25 (0.79)	-	-
English competence	3.54 (0.86)	3. 97 (0.61)	3.63 (0.92)	F(2,69)=2.94	.174
Spanish LexTale	84.44 (13.60)	88.15 (4.87)	82.74 (7.76)	F(2,69)=2.05	.141
Basque LexTale	-	70.71 (7.03)	-	-	-
English LexTale	45.44 (6.06)	49.55 (5.71)	45.80 (8.93)	F(2,69)=3.15	.320
Spanish picture naming	99.38 (1.69)	97.5 (2.95)	98.13 (3.23)	F(2,69)=2.36	.112
Basque picture naming	-	72.91 (2.80)	-	-	-
Catalan picture naming	-	-	96.25 (3.69)	-	-
English picture naming	50.38 (2.77)	51.57 (3.46)	50.89 (2.25)	F(2,69)=1.96	.192
Socioeconomic status	6.29 (1.12)	6.04 (1.60)	6.75 (0.85)	F(2,69)=2.05	.141
IQ	18.17 (4.43)	20.17 (3.45)	20.04 (3.63)	F(2,69)=2.02	.140

Note. Values reported are means and standard deviations in parenthesis of age (in years), subjective language competence (0-10 scale), LexTale (%), picture naming (% correct), socioeconomic status (1-10 scale), and IQ (correct answers). The last column shows the results from one-way ANOVAs comparing the three language groups on the different assessments.

The three groups were matched in terms of age, language proficiency in Spanish and English, socioeconomic status, and IQ. All groups showed no significant differences (*see Table 10 for stats*). As seen in Table 10, bilingual participants could not be matched on their second language competence (*i.e., Basque and Catalan*). Spanish-Basque bilinguals were less proficient in Basque than Spanish-Catalan bilinguals were in Catalan, and this may be due to the origin of the Spanish-Basque bilinguals, who came from and were tested in a city in which Basque is mainly used at school, while the Spanish-Catalan participants also used Catalan in daily life outside school.

Materials

Corpus of bigrams

Similar corpus data to that used in Experiment 1 and 2 was compiled, but in this case three languages were included, Spanish (*B-PAL; Davis & Perea, 2005*), Basque (*E-HITZ; Perea et al., 2006*), and Catalan (*NIM, Guasch, Boada, Ferré, & Sánchez-Casas, 2013*). Diacritics and words containing letters that do not exist in one of the other languages were removed. All words were broken down into bigram units and all bigram combinations were averaged in percent of appearance in the language. Then, bigrams were split into two lists: legal bigrams and illegal bigrams. Illegal critical bigrams did not appear in any of the three critical languages, Spanish, Basque and Catalan, such that frequency of use was 0. For example, the bigrams *bx* and *jl* do not exist in any of the three critical languages. On the other hand, legal bigrams were those that existed in the three languages and whose frequency of use did not differ statistically across languages [F(2, 22)= 0.697, p=.0.499, =.001]. For instance, the legal bigrams were plausible in Spanish, Catalan and Basque (*e.g., the consonant cluster sp appears in avispa, the Spanish for wasp, ispilu, mirror in Basque, and espai, which corresponds to space in Catalan; see Appendix 4 for a review*).

Novel word learning task

Thirty novel words were created for this experiment (*see Appendix 5*). Novel words were divided into legal and illegal novel words. They were created with the same orthographic structure: vowel, consonant bigram, vowel, consonant bigram, and vowel (*i.e., VCCVCCV*). The embedded consonant bigram (*CC*) was either a legal or illegal bigram, determining whether a novel word was legal or illegal. In total, twenty-three legal CC bigrams and nineteen illegal CC bigrams were selected (*see Appendix 4 for a list of selected CC bigrams*). In order to construct the novel words, we selected a second set comprising non-critical legal bigrams. These bigrams contained only one of the two letters from the legal CC bigrams and were either preceded or followed by a single vowel (*VC or CV*). These bigrams were selected to ensure that all non-critical bigrams used to compose novel words existed in all three languages. In total, seventy-nine non-critical legal bigrams were selected (*see Appendix 4 for a list of the selected non-critical bigrams*).

Novel words were presented both in written and auditory format. All novel words, legal (*e.g., aspilto*) and illegal (*e.g., ubxijla*), were pronounceable. They were fragmented into three pronounced syllables (*see Appendix 5 for the phonotactic clusters*) to control that all novel words were segmented equally. Legal and illegal novel words followed Spanish phonology, which is the common language for the three groups. The auditory format of the novel words was recorded in a soundproof room with a Marantz® professional PMD671. They were recorded by a native Spanish female with unmarked intonation.

Moreover, each of the 30 novel words was paired with a different video clip. The video clip was an invented 3D object that rotated on three axes (*see Antón, Thierry, & Duñabeitia, 2015*). Novel words were presented with an invented 3D object to facilitate learning because as some authors have demonstrated, children learn novel words better when they learn words with a referent (*Au & Glusman, 1990; Byers-Heinlein & Werker, 2013*).

Procedure

Participants were individually tested during school hours. The entire experiment lasted about one hour, including the initial assessment and the experimental phase. All visual stimuli were presented on a 13-inch MacBook® running Experiment Builder®. Auditory materials were presented to both ears simultaneously using Sennheiser® headphones.

The experiment was divided into learning and test phases and each phase had two tasks. In the learning phase, participants learned novel words by observation and then with a writing task (*see Figure 19*). First, participants saw and heard the thirty novel words in association with a 3D invented object. A fixation cross appeared for 500ms following by each word-object pair, which was presented for 6500ms on the screen. Each 3D invented object was visually presented together and aligned in time with the onset of the presentation of the visual (*written*) and auditory representations of the corresponding novel word. Participants did not have to press any key to move on to the next screen. Each object association was presented three times during the learning phase, leading to 90 trials that were presented in random order. After this observational task, participants were presented with a writing task. They had to type the name of the invented object on the keyboard. The object was presented with its written and auditory representation again, but this time a writing box appeared. Participants were instructed to write down the novel word paying attention to the written novel word that was still on the screen. They could only continue to the next trial if the novel word had been written correctly (*mean of incorrectly typed items= 2.46, SD=1.89*). Participants had to type each string-object pair twice. Pairs were presented in random order.

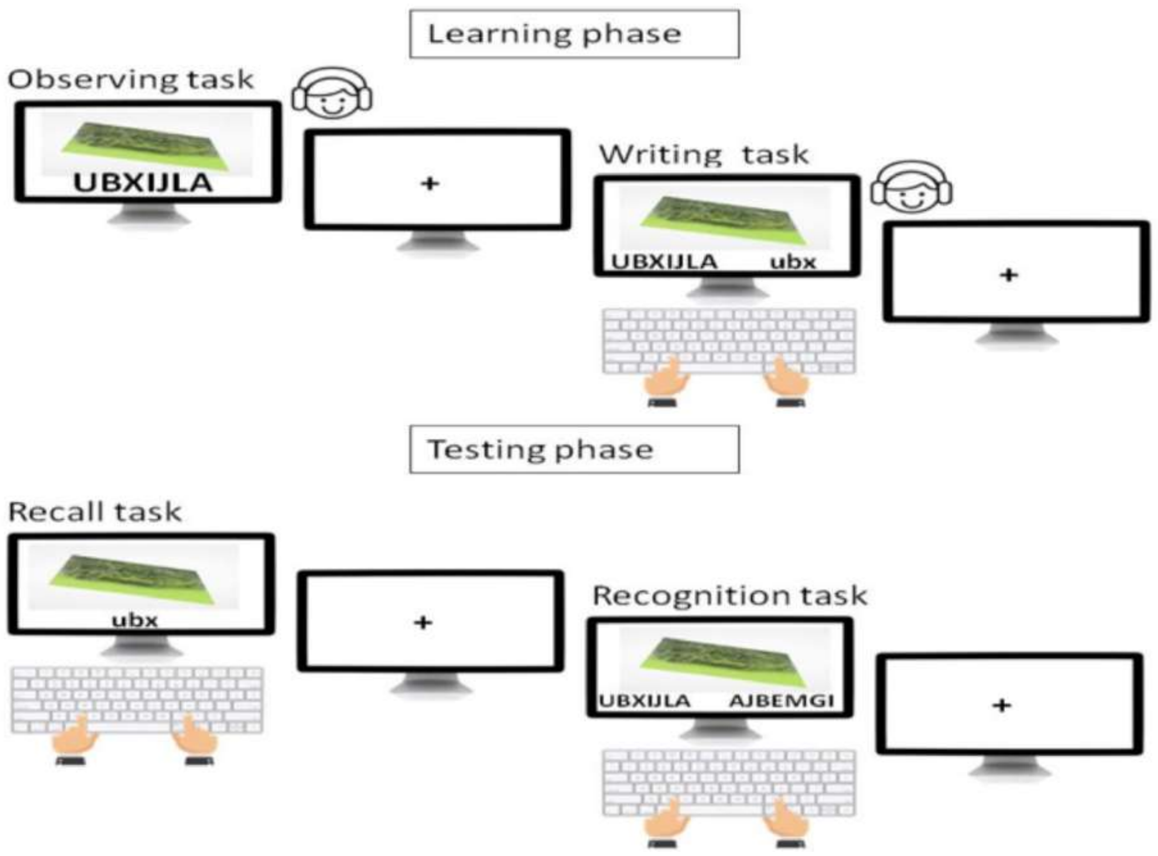


Figure 19. *Novel word learning task description of the learning and testing phases and its correspondent task.*

The testing phase also included two tasks: a recall and a recognition task (*see Figure 19*). First, participants saw each 3D invented object and had to write down the corresponding name that they had learned before. They were instructed to type the novel word. If they did not remember the whole novel word, they were told to approximate the novel word as much as possible. Objects were presented in random order. After the recall task, participants had to complete a recognition task (*2AFC task*). A fixation cross was displayed for 500 ms, then immediately followed by the centered presentation of the 3D invented object accompanied by two response options (*a correct and an incorrect novel word*) displayed at the lower right and left sides. Participants had to choose the correct novel word with its associated 3D invented object. To do so, they had to press the F key if they thought the correct novel word was the one on the right, or the J key if they thought it was the left one. Note, that the incorrect novel word option corresponded to a novel word that had been presented during the learning phase but was not matched with the correct object. In any case, response options included a legal and an illegal novel word. The location of correct and incorrect options was counterbalanced across trials. If no answer was given in 10000ms, the next 3D object was presented.

Data analysis

The two tasks of interest, the recall task and the recognition task, were analyzed separately in this experiment. Error rates and reaction times for correct responses were collected for both tasks (*see means in Table 11*). Before data analysis, outliers were excluded using R (*R core team, 2013*). Responses below 200 ms (*0.04%*) and timeouts above 10000 ms (*0.18%*) were initially excluded from the analyses. Also, responses above the 0.75 and below the 0.25 interquartile from the participant-based (*2.35 outliers*) and item-based (*1.08% outliers*) mean for all within-factors were excluded from the analyses (*in total, 3.05% of outliers*). Data analysis was conducted with Jamovi 0.9.6.7.

Two independent variables of interest were manipulated in this experiment: Orthographic Sequences (*Legal|Illegal*) as a within-subject factor and Group (*Spanish monolinguals|Spanish-Catalan bilinguals|Spanish-Basque bilinguals*) as a between-subject factor. A two-way ANOVA[3] were carried out, however, the dependent variables of interest were different for each task.

In the recall task, we analyzed two different dependent variables. First, we analyzed overall accuracy by considering the absolute number of correctly recalled items (*30 is the absolute number of novel words*). Taking into account that recall was clearly predicted to be low given the difficulty of the task and the number of items, we calculated the Levenshtein distance and used this as the dependent variable. This measure corresponds to the number of single-character substitutions, deletions, or insertions needed in each response to match the target novel word. A lower number of edits indicated that the response was closer to the target. Also, in the recognition task we used two dependent variables of interest: accuracy (*percentage of errors*) and reaction times of correct responses (*in milliseconds*).

In addition, and to support the absence and presence of an illegality effect in each of the language groups, we also conducted a Bayesian analysis. A Bayes factor (BF_{10}) shows the ratio of the probability that the data were observed under the alternative hypothesis versus the null hypothesis. For instance, BF_{10}=6 indicates that the observed data were six times more likely to have occurred under the alternative than the null hypothesis, and conversely, a BF_{10}=2 shows that the data were more likely to be observed under the null than the alternative hypothesis.

Note 3: Note that this data was also analyzed using mixed-effects models, accuracy with logistic mixed-effects models and reaction times with linear mixed-effects models (Baayen, Davidson & Bates, 2008; Jaeger, 2008; Barr, 2013), using the lme4 package for R (Bates, Maechler, Bolker & Walker, 2014), including subjects and items as random factors. Models converged with model simplifications, showing the same results as ANOVAs. However, we report ANOVAs to be consistent with the analyses reported in other chapters.

02. Results

Recognition task

Analysis of percentage of errors

Results of the accuracy (*percentage of errors*) analysis of the recognition task showed that there was a significant main effect of Orthographic Sequences (*see Table 11*), (*F1(1,69)=13.28, p<.001,* η_p^2 *=.148; F2(1,14)=3.96, p=.036,* η_p^2 *=.255*). Overall, participants were more accurate at recognizing the correct novel word for the object when it was a legal orthographic sequence than an illegal one. On the other hand, the main effect of Group was not significant (*F1(2,69)=0.121, p=.886* η_p^2 *=.003; F2(2,28)=.269, p=.766,* η_p^2 *=.019*) but the interaction between the two factors was significant (*F1(2,69)=3.61, p =.032,* η_p^2 *=.081; F2(2,28)=3.91, p=.032,* η_p^2 *=.218*). This interaction suggests that the illegality effect differs between the three groups.

Therefore, we assessed this effect for participants in each group separately. Spanish-Catalan bilinguals (*t1(23)=3.10, p =.005, Cohen´s d=.633,* BF_{10}*=8.68; t2(14)=2.25, p =.041, Cohen´s d=.581,* BF_{10}*=1.79*) and monolinguals (*t1(23)=3.08, p =.005, Cohen´s d=.606,* BF_{10}*=8.27; t2(14)=2.33, p =.035, Cohen´s d=.602,* BF_{10}*=2.02*) showed a significant effect of illegality, meaning that they remembered legal novel words better than illegal words. In contrast, this effect was not observed for Spanish-Basque bilinguals (*t1(23)=0.099, p=.922, Cohen´s d=.020,* BF_{10} *=0.21; t2(14)=0.06, p =.953, Cohen´s d=.016,* BF_{10}*=0.26*), showing that they had learned illegal orthographic sequences to the same extent as legal ones (*see Figure 20*). To follow up on this interaction, we also looked at the

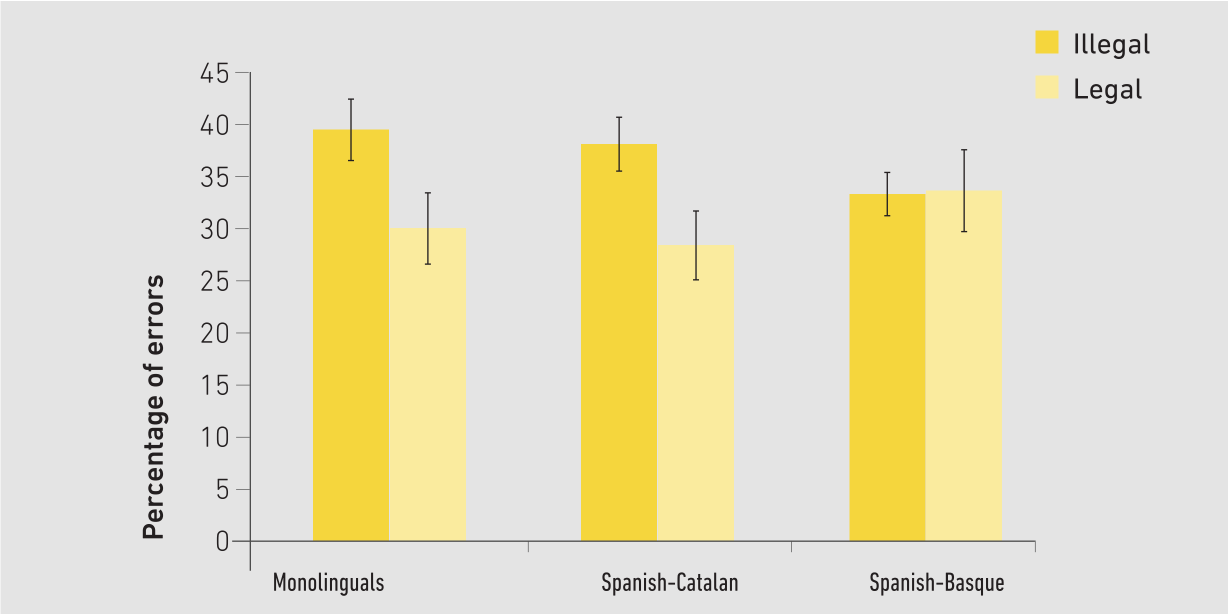


Figure 20. *Plot bars of the percentage of errors in the recognition task for legal and illegal orthographic sequences for each of the language groups (Spanish, Spanish-Basque, and Spanish-Catalan). Error bars refer to the standard error (SE) of each mean.*

simple main effects of Group on each level of Orthographic Sequence (*i.e., on legal and illegal orthographic sequences separately*). In a one-way ANOVA, we found no significant effect of Group for the legal (*F1(2,69)=.61, p=.545,* η_p^2 *=.017; F2(2,42)=.54, p=.586,* η_p^2 *=.025*) or the illegal orthographic sequences, (*F1(2,69)=1.63, p=.203,* η_p^2 *=.045; F2(2,42)=1.02, p=.371,* η_p^2 *=.046*). This means that the interaction between Group and Orthographic Sequences was not driven by the Spanish-Basque bilinguals performing better on the illegal sequences nor doing worse on the legal ones. Instead, it suggests that they perform similarly on legal and illegal orthographic sequences, whereas the other language groups perform worse on the illegal than on the legal sequences.

Analysis of reaction times

In terms of results of the reaction time (*RT*), responses to illegal orthographic sequences required slightly more time than legal ones (*see Table 11*) and this effect was marginally significant (*F1(1,69)=2.90, p=.063,* η_p^2 *=.040; F2(1,14)=2.29, p=.152,* η_p^2 *=.078*). The main effect of Group was not significant (*F1(2,69)=0.07, p=.932,* η_p^2 *=.002; F2(2,28)=0.28, p=.756,* η_p^2 *=.020*) and the interaction between Orthographic Sequences and Group was not significant either (*F1(2,69)=0.57, p=.567,* η_p^2 *=.016; F2(2,28)=0.01, p=.992,* η_p^2 *=.001*). This finding suggests that all groups spent the same amount of time in all responses, but they tended to need slightly more time to recognize illegal orthographic sequences.

iii. Justification for Experiment 3B

Experiment 3A aimed to examine whether novel word learning with different orthographic regularities (*legal and illegal novel words*) was facilitated by being bilingual or by linguistic experience with differences between orthographic regularities in children's two languages. To examine this, three different groups of children were tested: a bilingual group with similar (*Spanish-Catalan*), and a bilingual group with dissimilar (*Spanish-Basque*) orthographic regularities and a monolingual group. Participants had to learn thirty novel words; half were legal and half illegal. Results in the recognition task showed that the triple interaction between language group and illegality on the percentage of errors was significant, suggesting that Spanish monolinguals, Spanish-Catalan bilinguals, and Spanish-Basque bilinguals differed in the way they learned new legal and illegal sequences. While monolinguals and Spanish-Catalan bilinguals had more difficulty recognizing illegal sequences than legal ones, Basque-Spanish bilinguals did not show this effect. This result suggests that group differences in word learning were not due to bilingualism as such but rather related to the two specific languages that the bilinguals knew.

Spanish and Basque are more dissimilar (*e.g., in grammar, letter sequences, phonology*) than Spanish and Catalan. Therefore, the absence of a legality effect in the Spanish-Basque bilinguals could be due to linguistic experience with two distinct languages and the process of literacy acquisition (*having already acquired the two languages*). However, the lack of significance could be because the bilingual groups were not entirely matched on proficiency. As can be seen in Table 10, the Basque proficiency of the group of Spanish-Basque bilinguals was lower than the Catalan proficiency of the Spanish-Catalan bilinguals.

Thus, we wanted to replicate the null result of illegality in Spanish-Basque bilinguals. To do so, we ran the same task in Experiment 3B with two groups of Spanish-Basque bilinguals: one similar to the previous study and one group with higher Basque proficiency. If the absence of an illegality effect was only found in the group of Spanish-Basque bilinguals with a lower Basque proficiency level, the effect in Experiment 3A might have been driven by proficiency differences between the two bilingual groups. In contrast, if we did not observe an illegality effect in either group of Basque speakers in Experiment 3B, this would support our interpretation that the findings in Experiment 1 were related to linguistic experience.

iv. Experiment 3B

01. Methods

Participants

Forty-six Spanish-Basque bilingual children took part in this experiment (*34 females; mean age=12.9 years, SD=0.6*). Participants were recruited from two different Basque communities in the Basque Country. The first group (*24 participants*) came from Donostia-San Sebastian, a dense bilingual environment (*percentage of exposure to Spanish, M=39.7.8%, SD=5.47; percentage of exposure to Basque, M=53.6%, SD=7.38; percentage of exposure to English, M=6.7%, SD=3.27*). The other group (*24 participants*) was from Vitoria, as in Experiment 3A (*percentage of exposure to Spanish, M=51.64%, SD=3.54; percentage of exposure to Basque, M=40.76%, SD=2.87; percentage of exposure to English, M=7.6%, SD=2.26*).

As in Experiment 3A, all participants' parents received an information letter with a short questionnaire, in which parents had to rate their children's language proficiency and socioeconomic status. Parents had to return the signed written informed consent before testing. The study was approved by the BCBL (*Basque Center on Cognition, Brain and Language*) Ethics Committee. None of the children were left-handed, and none were diagnosed with language disorders, learning disabilities, or auditory impairments. We controlled for age of second language acquisition (*before the age of 6*) as in Experiment 3A. Also, participants were matched on their language proficiency in Spanish and English, their socioeconomic status, and their IQ (*see Table 12*). However, and critically for this experiment, the two Basque groups differed both in terms of their subjective measure of competence in Basque and their picture-naming performance in Basque (*see Table 12*).

Table 12. Descriptive statistics of assessments

	Highly proficient Basque bilinguals	Less proficient Basque bilinguals	T-test	
			T (*df*)	*p*
Age	13.05 (0.72)	12.79 (0.59)	t(44)=1.31	.197
Spanish competence	9.5 (0.86)	9.21 (0.59)	t(44)=1.35	.183
Basque competence	7.68 (1.09)	5.71 (1.37)	t(44)=5.38	<.001
English competence	3.95 (1.39)	3.91 (1.47)	t(44)=1.42	.209
Spanish Lextale	85.87 (5.59)	87.05 (5.17)	t(44)=0.74	.462
Basque Lextale	69.82 (7.49)	71.21 (8.60)	t(44)=0.58	.563
English Lextale	44.71 (6.13)	46.73 (5.42)	t(44)=0.98	.312
Spanish picture naming	87.73 (27.11)	97.71 (4.66)	t(44)=0.34	.729
Basque picture naming	77.45 (2.69)	67.83 (2.45)	t(44)=3.11	.003
English picture naming	50.29 (3.56)	55.48 (4.64)	t(44)=1.35	.183
Socioeconomic status	6.55 (1.14)	6.25 (1.03)	t(44)=0.92	.362
IQ	18.73 (2.12)	18.38 (3.03)	t(44)=0.45	.653

Note. Means and standard deviations in parenthesis of age (in years), subjective language competence (0-10 scale), LexTale (%), picture naming (% correct), socioeconomic status (1-10 scale), and IQ (number of correct answers in the timed test). The last column shows the results from the t-tests comparing the two Spanish-Basque groups on the different assessments.

Materials, Procedure and Data Analysis

Materials, procedure and data analysis were identical to those used in Experiment 3A.

02. Results

Recall task

We performed repeated measures ANOVAs with Group (*highly proficient Basque bilinguals|less proficient Basque bilinguals*) and Orthographic Sequences (*legal|illegal*) on accuracy and Levenshtein Distance in the recall task and percentage of error and reaction times with correct responses in the recognition task. In the recall task, participants recalled more legal than illegal words (*see Table 13*), (*F1(1,44)=13.57, p=<.001,* η_p^2 *=.077; F2(1,14)=4.45, p=.098,* η_p^2 *=.095*). Also, when the Levenshtein Distance was taken into account, the novel words recalled were closer to the target in the case of the legal as compared to the illegal novel words (*see 13*), (*F1(1,44)=26.97, p=<.001,* η_p^2 *=.090; F2(1,14)=13.37, p=.003,* η_p^2 *=.190*). However, there were no effects of Group on the accuracy (*F1(1,44)=1.18, p=.282,* η_p^2 *=.026; F2(1,14)=1.22, p=.288,* η_p^2 *=.017*) or Levenshtein Distance of the recalled items (*F1(1,44)=2.49, p=.122,* $_p^2$ *=.054; F2(1,14)=2.86, p=.113,* η_p^2 *=.086*). There was also no interaction between the illegality effect and Group on the accuracy (*F1(1, 44)=2.09, p=.155,* η_p^2 *=.035; F2(1,14)=1.75, p=.207,* η_p^2 *=.021*) or Levenshtein distance (*F1(1, 44)=2.95, p=.093,* η_p^2 *=.040; F2(1,14)=2.78, p=.119,* η_p^2 *=.045*).

Table 13. Descriptive statistics for the Recall task and the Recognition task.

	High proficient Basque bilinguals		Less proficient Basque bilinguals	
	Legal	Illegal	Legal	Illegal
Recall task				
Recalled items	3.03 (4.47)	1.21 (2.63)	5.56 (6.42)	1.39 (5.55)
LD	5.8 (0.50)	6.09 (0.43)	5.38 (0.92)	5.95 (0.69
Recognition task				
%error	30.61 (12.46)	33.94 (12.83)	29.72 (16.68)	31.67 (12.00)
RT	2043 (637)	2153 (785)	2031 (505)	2121 (546)

Note. Means and standard deviations in parenthesis of recalled items (absolute number) and Levenshtein Distance (LD) (Recall task) and percentage of errors and reaction times in ms (Recognition task) for legal and illegal orthographic sequences for the two language groups

Recognition task

Analysis of percentage of errors

In terms of accuracy (*percentage of errors*) in the recognition task, we observed that participants recognized legal and illegal words equally well (*F1(1,44)=0.86, p=.357,* η_p^2 *=.019; F2(1,14)=0.407, p=.534,* η_p^2 *=.005*) and no differences between groups were found (*F1(1,44)=0.19, p=.665,* η_p^2 *=.004; F2(1,14)=0.24, p=.626,* η_p^2 *=.017*), nor any interaction (*F1(1,44)=0.15, p=.699,* η_p^2 *=.003; F2(1,14)=0.22, p=.625,* η_p^2 *=.018*), showing that the lack of an illegality effect was similar for both groups of Spanish-Basque bilinguals (*see Figure 21*).

Analysis of reaction times

In terms of reaction times, participants tended to require more time to recognize illegal words than legal ones (*F1(1,44)=3.78, p=.078,* η_p^2 *=.211; F2(1,14)=3.27, p=.087,* η_p^2 *=.112*) but no differences between groups were observed (*F1(1,44)=1.12, p=.296,* η_p^2 *=.025; F2(1,14)=3.76, p=.098,* η_p^2 *=.112*), nor an interaction (*F1(1,44)=0.11, p=.742,* η_p^2 *=.002; F2(1,14)=0.87, p=.366,* η_p^2 *=.009*).

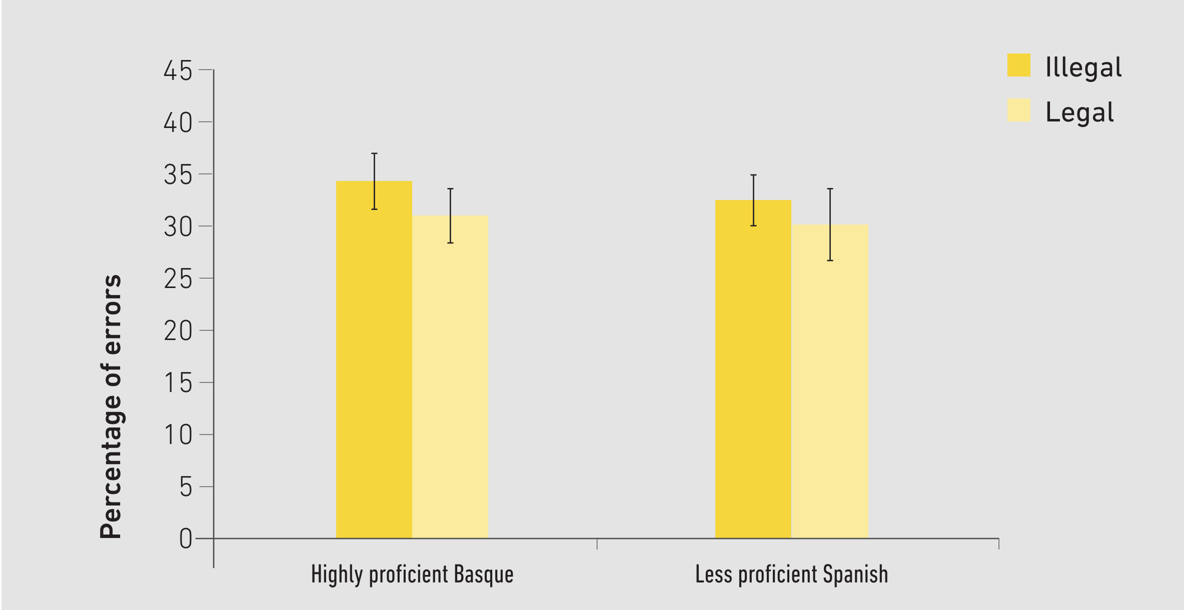


Figure 21. *Bar plots of the percentage of errors in the recognition task for legal and illegal orthographic sequences for each of the Spanish-Basque bilingual groups. Error bars refer to the standard error (SE) of each mean.*

iv. Summary of results

Experiment 3A aimed to examine whether older children (*around the age of twelve*) learned novel words with illegal and legal orthographic regularities and whether this learning was affected by being bilingual or rather by the experience of dealing with different orthographic regularities in their two languages. Therefore, we compared monolingual children's performance and two groups of bilinguals: a group of Spanish-Catalan bilinguals who speak two languages with similar orthographic regularities and a group of Spanish-Basque bilinguals who speak two languages with different orthographic regularities. Experiment 3B replicated the critical findings of Experiment 3A.

Results for the recall task, in both experiments 3A and 3B, showed that legal novel words were remembered better than illegal novel words. But, no differences between the three language groups were found in this regard. The recall task is not very informative due to the fact that children recalled a low percentage of novel words. This low performance could lead to a floor effect, in which we could not observe a difference between groups simply because performance did not allow for enough variability (*Baddeley, 1992*). In line with accuracy, the LD measure showed that responses to legal novel words were more similar to the original novel words than answers to illegal words. But, again no differences between the three language groups were found in this regard. A greater amount of practice might be needed to enhance recall performance and show further effects of language group.

However, the recognition task in Experiment 3A showed an interaction between language group and orthographic sequences in terms of accuracy, suggesting that Spanish monolinguals, Spanish-Catalan bilinguals, and Spanish-Basque bilinguals differed in the way they learned legal and illegal novel words. While monolinguals and Spanish-Catalan bilinguals were less able to recognize illegal sequences than legal ones, Basque-Spanish bilinguals did not show this effect. This result suggests that group differences in word learning are not due to bilingualism as such but rather related to the two specific languages that bilinguals know. Our first consideration was that this null result could have been due to differences in proficiency in both bilingual groups. Thus, we replicated the experiment with two new Spanish-Basque groups, one with high L2 proficiency and other less L2 proficiency (*the same as in Experiment 3A*) to observe whether the effects were due to differences in orthographic regularities between the languages or rather to the Basque proficiency of the children. Results again showed that Spanish-Basque children learned legal and illegal novel words to the same extent. But no differences were observed between these two groups regardless of their proficiency differences, suggesting that the (*absence of an*) illegality effect was not modulated by proficiency in Basque. These findings provided support for the results from Experiment 3A, suggesting that linguistic experience with languages that differ from each other at the level of orthographic regularities may modulate word learning in bilingual children.

III. Experiment 4. Learning novel words with illegal orthographic regularities in early adulthood

i. Rationale

As we saw in Experiment 3, Spanish-Basque bilingual children learn novel words in a different way than Spanish-Catalan bilingual and monolingual children. It should be noted that this effect does not

point to a general learning advantage, because the three groups learned the same number of novel words. However, it does indicate a modulation in novel word preference. Whereas monolinguals and Spanish-Catalan bilingual children prefer to learn legal novel words because these words follow their orthographic knowledge, Spanish-Catalan bilingual children learn the same number of legal and illegal novel words. These results suggested that dealing with language with differences in orthographic regularities may modulate learning novel words.

Next, we wanted to test whether adults learn novel words with legal and illegal orthography sequences. Specifically, we wanted to observe whether the same differential pattern observed in Spanish-Basque bilingual children, the absence of sensitivity to orthographic markedness during novel word learning (*e.g. flexibility in word learning*), was also present in Spanish-Basque bilingual adults. Or if, on the contrary, there was a developmental trajectory for these effects, specifically, that they might diminish as a function of age. Therefore, we conducted a study using the same methodology as in Experiment 3, comparing Spanish-Basque bilingual adults to Spanish monolingual adults. Note that the Spanish-Catalan bilingual group was excluded as a target group because the previous results with Spanish-Catalan children were in line with monolingual results. If experience with dealing different orthographic regularities between the two languages affects learning legal and illegal novel words across the lifespan, we would expect similar patterns to those reported in Experiment 3. However, considering that children are still in the process of acquiring and consolidating their two languages, they may exhibit different patterns than adults.

ii. Methods

Participants

Forty-eight adults took part in this experiment (*30 females; mean age=21.68 years, SD=2.8*). Half of them (*24 participants*) were Spanish monolinguals from the University of Cantabria and the other half (*24*) were Spanish-Basque bilinguals from the University of the Basque Country. As in Experiment 3, they were matched on their language proficiency in Spanish and English, socioeconomic status, and IQ (*see Table 14*). The study was approved by the BCBL Ethics Committee and all participants signed an informed consent form before the experiment and were compensated for their time. None of them were left-handed and none were diagnosed with language disorders, learning disabilities, or auditory impairments.

Table 14. Descriptive statistics of assessments

	Monolinguals	Spanish-Basque bilinguals	T-test	
			t (*df*)	*p*
Age	21.33 (3.12)	22.04 (2.47)	t(46)=0.84	.403
Spanish competence	9.41 (1.83)	9.12 (0.79)	t(46)=0.71	.408
Basque competence	0 (0)	9.29 (0.80)	-	-
English competence	6.45 (1.34)	6.97 (3.54)	t(46)=2.56	.149
Spanish LexTale	90.34 (6.70)	92.84 (3.84)	t(46)=2.05	.172
Basque LexTale	0 (0)	91.08 (4.45)	-	-
English Lextale	64.56 (6.49)	68.34 (4.65)	t(46)=1.32	.273
Spanish picture naming	94.15 (1.94)	97.71 (2.91)	t(46)=2.81	.138
Basque picture naming	0 (0)	89.75 (6.95)	-	-
English picture naming	62.83 (5.83)	65.38 (6.75)	t(46)=1.73	.265
Socioeconomic status	6.58 (0.88)	7 (0.78)	t(46)=2.72	.142
IQ	20.83 (1.83)	21.79 (2.48)	t(46)=2.31	.163

Note. Values reported are means and standard deviation in parenthesis on age (in years), subjective language competence (1-10 scale), LexTale (% correct), picture naming (% correct), socioeconomic status (0-10 scale), and IQ (number of correct answers in the timed test). The last column shows the results from the t-test comparing Monolingual and Spanish-Basque Bilingual groups on the different assessments.

Materials and Procedure

Materials and procedure were identical to those used in Experiment 3.

iii. Results

Recall task

Repeated measures ANOVAs were conducted with the factors Group (*Spanish monolinguals, Spanish-Basque bilinguals*) and Orthographic Structure (*legal, illegal*) on accuracy and Levenshtein distance in the recall task and percentage of errors and reaction times in the recognition task. In the recall task, participants recalled more legal than illegal orthographic sequences (*F1(1,46)=16.92, p<.001,* η_p^2 *=.262; F2(1,28)=7.62, p=.005,* η_p^2 *=.184*), but groups did not differ on performance (*F1(1,46)=0.47, p=.492,* η_p^2 *=.007; F2(1,28)=4.47, p=.440,* η_p^2 *=.017*), and there was no interaction between Group and Orthographic Sequence (*F1(1,46)=1.57, p=.217,* η_p^2 *=.024; F2(1,28)=4.47, p=.440,* η_p^2 *=.017*). The analysis of Levenshtein Distance showed that the recall of legal orthographic sequences was closer to the template than that of illegal sequences (*F1(1,46)=51.97, p=<.001,* η_p^2 *=.506; F2(1,28)=16.7, p<.001,* η_p^2 *=.211*). Also, a Group effect was found (*F1(1,46)=11.54, p=.001,* η_p^2 *=.201; F2(1,28)=11.67, p<.001,*

η_p^2 *=.32*), such that the Spanish-Basque bilinguals outperformed the monolinguals. The interaction between the two factors was significant (*F1(1,46)=4.83, p=.033,* η_p^2 *=.047; F2(1,28)=6.51, p=.016,* η_p^2 *=.020*), suggesting that Spanish-Basque bilingual adults' recall was more similar to the target items, especially in the case of the legal sequences (*see Table 15*).

Table 15. Descriptive statistics for the Recall task and the Recognition task

	Monolinguals		Spanish-Basque bilinguals	
	Legal	Illegal	Legal	Illegal
Recall task				
Recalled items	1.04 (1.90)	0.38 (1.44)	1.63 (2.26)	0.38 (0.58)
LD	5.69 (1.09)	6.18 (0.86)	4.58 (1.18)	5.51 (0.66)
Recognition task				
%error	25.83 (12.64)	35.83 (19.14)	11.11 (13.43)	24.72 (11.37)
RT	1974 (444)	2256 (637)	1899 (433)	2318 (642)

Note. Mean and standard deviation in parenthesis on recalled items (absolute number) and Levenshtein Distance (LD) (Recall task) and %error and reaction times in ms (Recognition task) for legal and illegal orthographicTT sequences for the three language groups.

Recognition task

Analysis of percentage of errors

In the recognition task, adults showed a main effect of Orthographic Sequence on the percentage of errors (*F1(1,46)=31.05, p=<.001,* η_p^2 *=.399; F2(1,28)=713.7, p<.001,* η_p^2 *=.189*), as well as Group (*F1(1,46)=12.91, p=<.001,* η_p^2 *=.219; F2(1,28)=32.94, p<.001,* η_p^2 *=.227*), but no interaction (*F1(1,46)=0.73, p=.399,* η_p^2 *=.009; F2(1,28)=0.64, p=.429,* η_p^2 *=.004*). This means that adults recognized the legal orthographic sequences more accurately than the illegal ones, and that Spanish-Basque bilingual outperformed monolinguals in overall learning (*see Figure 22*). However, the absence of interaction suggested that both language groups remembered the legal novel words better than the illegal ones to a similar extent.

Analysis of reaction times

The analysis of the reaction times in this recognition task showed that adults recognized legal orthographic sequences faster than illegal ones (*F1(1,46)=28.14, p=<.001,* η_p^2*=.374; F2(1,28)=10.7, p=.003,* η_p^2 *=.212*). However, there was no Group effect (*F1(1,46)=0.002, p=.965,* η_p^2 *<.001; F2(1,28)=0.02, p=.876,* η_p^2 *=.004*), nor an interaction (*F1(1,46)=1.09, p=.303,* η_p^2 *=.014; F2(1,28)=0.56, p=.457,* η_p^2 *=.007*).

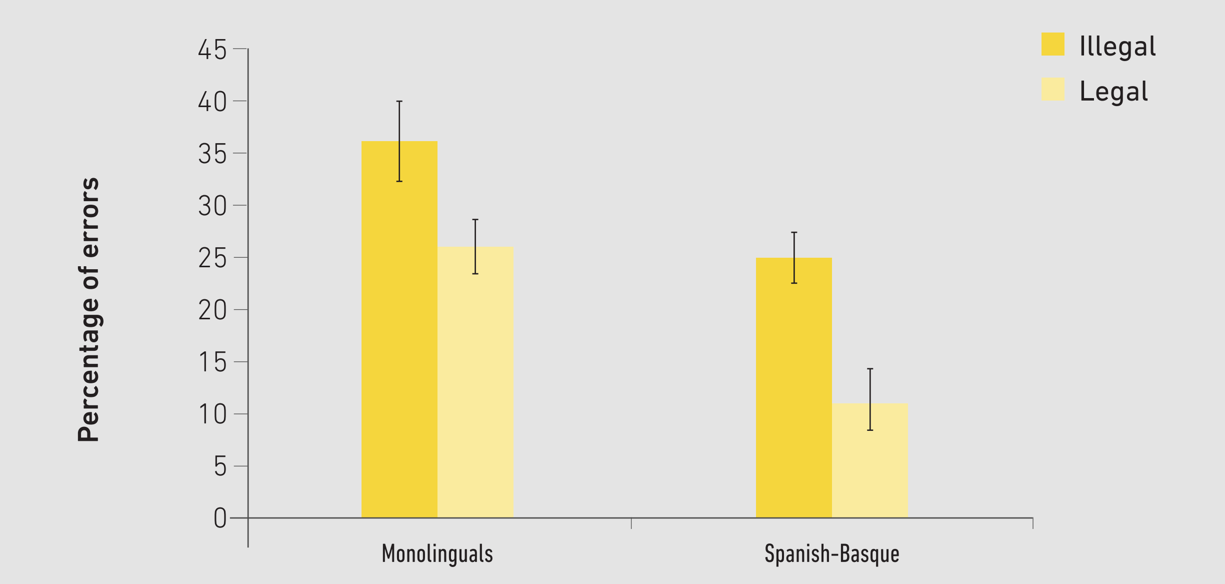


Figure 22. *Plot bars of the percentage of errors in the recognition task for legal and illegal orthographic sequences for Spanish-Basque bilingual and monolingual groups. Error bars refer to the standard error (SE) of each mean.*

iv. Summary of results

Experiment 4 assessed whether Spanish-Basque bilingual adults differed from Spanish monolingual adults in the way they learned legal and illegal novel words, as children did in the previous experiment. Similar to the results reported in Experiments 3A and 3B, the recall task showed that both language groups recalled more legal than illegal novel words. However, unlike Experiment 3, the recognition task showed a legality effect for both language groups. Both the Spanish monolinguals and the Spanish-Basque bilinguals remembered legal words better than illegal words. In terms of overall performance, Spanish-Basque bilinguals outperformed Spanish monolinguals. Bilinguals remembered more words than monolinguals in the recognition task and their responses in the recall task were closer to the target words.

These results are in line with the literature showing that bilinguals have an advantage in word learning compared with monolinguals (*for a review, see Hirosh & Degani, 2018*), and also align with previous studies focused on adult word learning showing that bilinguals outperform monolinguals (*Bartolotti & Marian, 2012; Kaushanskay & Marian, 2009; Margarita Kaushanskaya & Rechtzigel, 2012; Kaushanskaya & Marian, 2009*). The results from Experiment 4 provide support for this view, suggesting that bilinguals can benefit from their previous experience with language (*e.g., learning vocabulary in various languages*) to achieve a higher level of performance in novel word learning tasks. However, we additionally showed that their experience with two languages differing in terms of orthographic structures did not affect performance.

In the last experiments (*Experiment 3 and 4*), we can see differences in performance between adults and children. While Spanish-Basque adults showed an overall better performance than their monolingual peers, Spanish-Basque children learned the novel words differently from the other groups. In this sense, this last group was in line with our prediction that experience with handling languages with different characteristics, rather than bilingualism per se, may support novel word learning. Thus, sensitivity to orthographic markedness may be related to novel word learning performance.

IV. Experiment 5. Relationship in older children between learning novel words with illegal orthographic regularities and sensitivity to markedness

i. Rationale

The aim of this chapter is to investigate whether experience with languages with different orthographic regularities may modulate learning novel words that follow or violate orthographic knowledge. As we saw in Experiments 3 and 4, children and adults tend to learn novel words differently. While children whose languages differ in terms of orthographic regularities (*Spanish-Basque*) learned the same number of both types of novel words (*legal and illegal*), monolinguals and Spanish-Catalan children, whose languages have similar orthographic regularities, were better able to learn legal than illegal novel words. Note that no group showed an advantage in learning. This instead shows that Spanish-Basque bilingual children were equally likely to learn words from both conditions, legal and illegal (*they showed different item preferences in learning*). However, Spanish-Basque adults showed an overall advantage in learning novel words compared with monolinguals.

As previous research has shown, bilinguals may be better than monolinguals at new word learning. This includes both child and adult populations (*Kaushanskaya et al., 2014; Kaushanskaya & Marian, 2009*). Surprisingly, Spanish-Basque bilingual children showed different patterns from other groups of children. Thus, we wanted to focus in this experiment on investigating whether Spanish-Basque bilinguals would show learning patterns that differed from the rest of the groups.

As we saw in Experiment 1, bilingual children at the age of twelve (*the same age as in Experiment 3*) may change their sensitivity to orthographic markedness. This change in sensitivity to orthographic markedness may be related to how they learn novel words. Hence, the aim of this experiment was to investigate whether learning (*il*)legal novel words related to how sensitive Spanish-Basque children are to markedness. To do so, we wanted to see how this tendency to learn the same amount of legal and illegal novel words correlated with sensitivity in recognizing marked (*or unmarked*) words from their known languages. The experiment consisted of two parts. First, children learned legal and illegal novel words, as in Experiment 3, so we could observe novel word learning and then, children were asked to recognize marked and unmarked words in the languages they know as fast as they could so we could observe their sensitivity to markedness as in Experiment 1.

Results from the correlation could reveal three possible scenarios: first, we might find that children who were least able to recognize words in the languages they knew were the ones that were equally able to learn both legal and illegal novel words. These children could be poor at perceiving differences between languages, so they would be insensitive to such differences when learning novel words. Second, children who were better at differentiating between languages might show little difference between learning legal and illegal novel words. These children might be good at perceiving differences between the orthographic regularities in their languages and thus good at learning things that are different, since they overcome their initial perception of illegality. Third, there could be no correlation between being better or worse at recognizing the differences between languages and learning legal and illegal novel words. This would suggest that being better at learning illegal novel words is not related to how well you perceive differences in orthographic regularities.

ii. Methods

Participants

Forty-five Spanish-Basque bilingual children from the Basque Country participated in this experiment (*21 females; Mean age=12.45 years, SD=0.59*). They were recruited from a bilingual school in Vitoria as in previous experiments. All participants had acquired Basque before the age of 6 and had been raised in a bilingual community and educated in a Spanish-Basque bilingual school.

All participants were right-handed and none were diagnosed with language disorders, learning disabilities, or auditory impairments. Children, and children´s families were appropriately informed and parents or legal guardians signed consent forms prior to the experiment. The protocol was carried out according to the guidelines approved by the BCBL Ethics Committee.

We assessed participants' language proficiency, socioeconomic status, and IQ to control their background. Three measures were used to evaluate language proficiency, subjectively and objectively. For the subjective measurement, parents had to rate their children´s languages competence on a scale from 1 to 10 comparing their performance to their peers (*Spanish M=9.4, SD=0,80; Basque M=6,10, SD=1,44; English M=5,2, SD=1,79*). For the objective measurements, participants did a lexical decision task (*LexTale*) in Spanish (*M=85.17, SD=8.55; Izura, Cuetos, & Brysbaert, 214*), in Basque (*M=66.40, SD=12,13; de Bruin, Carreiras, & Duñabeitia, 2017*), and English (*M=54.46, SD=8.18; Lemhöfer & Broersma, 2012*). Also, participants named twenty common objects from the adapted version of a picture naming task (*de Bruin et al., 2017*) in Spanish (*M=19.97, SD=0.15*), Basque (*M=13.68, SD=3.69*), and English (*M=12.68, SD=3.69*). The English measurement was not relevant for the task, but was included in order to make sure that the participants' English levels were relatively low and would not have an effect on their other two languages. Socioeconomic status was measured with a short questionnaire in which children´s parents had to rate on a scale from 1 to 10 how they perceived their economic situation as compared to other members of their community (*M=6.49, SD=1.59; Adler & Stewart, 2007*). Finally, IQ was measured with a 6-minute abridged version of the K-BIT (*Kaufman, 2004*) in which participants had to complete as many matrices as they could in the allotted time (*M=19.34, SD=2.75*).

Materials

Language decision task

The same language decision task as in Experiment 1 was run. See materials of Experiment 1 to review the task description.

Learning task

The same language decision task as in Experiment 1 was run. See materials of Experiment 1 to review the task description.

Procedure

The whole experiment lasted one hour, including the initial assessment, the language decision task and the novel word learning sections. Participants were individually tested during school hours. All visual stimuli were presented on a 13-inch MacBook® running Experiment Builder®. See Experiment 1 and 3 to review the procedures for the language decision task and novel word learning task.

Data analysis

Language decision task

Accuracy and reaction times were collected in this experiment (*see Table 16*). They were analyzed separately using a one-way ANOVA with the factor Marked Lexicality (*marked word|unmarked word|marked pseudoword*). Marked Lexicality was calculated averaging data from both languages (*Spanish and Basque*) in each condition (*see Table 16 for averages*). Note that there is no characteristic in the unmarked pseudowords that allows them to be differentiated as belonging to either language of interest. Therefore, we did not include unmarked pseudowords in the analysis because we could not infer the real accuracy of those responses.

Table 16. Descriptive statistics for the Language decision task.

	Marked words		Unmarked words		Marked pseudowords	
	Basque	Spanish	Basque	Spanish	Basque	Spanish
% errors	5.14 (4.38)	7.32 (7.66)	7.07 (6.04)	9.67 (6.49)	6.84 (7.60)	56.36 (10.20)
average	6.75 (8.32)		8.04 (6.26)		31.93 (9.45)	
RT	1069 (238)	948 (231)	1147 (275)	1041 (223)	1184 (313)	1629 (457)
average	1007 (234)		1087 (249)		1398 (388)	

Note. Values reported are means and standard deviation in parenthesis on accuracy (%errors), and reaction times (milliseconds) and on d' (Spanish compared with Basque)

Novel words learning task

Accuracy and reaction times were collected in this experiment (*see Table 17*). As in Experiment 3, the recall task (*accuracy and Lowenstein distance*) and recognition task (*% of errors and RT*) were analysed separately. To do so, a paired sample t-test was carried out to compare the factors Orthographic Sequences (*legal, illegal*) for each independent variable.

Table 17. Descriptive statistics for the Recall task and the Recognition task

	Legal	Illegal
Recall task		
Recalled items	5.30 (10.02)	1.21 (2.97)
LD	5.73 (1.22)	6.25 (0.84)
Recognition task		
%error	31.36 (16.55)	32.42 (13.68)
RT	1709 (532)	1721 (504)

Note. Mean and standard deviation in parenthesis on recalled items (absolute number) and Levenshtein Distance (LD) (Recall task) and %error and reaction times in ms (Recognition task) for legal and illegal orthographic sequences for the three language groups.

Linear regression between marked words/pseudowords and novel word learning task

Reaction times were the variable of interest in testing sensitivity to orthographic markedness in the language decision task. Unmarked words did not require extracting sub-lexical information for categorization, and individuals had to access their meaning in order to attribute the language (*see Casaponsa et al., 2014*). For this reason, only items that allowed for a categorization aided by fine-grained sub-lexical analysis (*i.e., marked words and marked pseudowords*) were chosen to test the relationship with learning novel words. Note that the percentage of word errors in the marked words was so low that there was little information (*and variability*) to correlate with novel word learning.

Novel word learning was calculated as the effect of learning novel words, corresponding to the difference between the percentage of errors in the recognition task in the novel word learning task for the legal minus the illegal items. Here, a positive number indicates that more errors were made with legal than illegal novel words, or put differently, that more illegal novel words were learned. Then, to better understand the relationship between the variables, a simple linear regression analysis was conducted, in which the effect of novel word learning was entered as the dependent variable and the response times to marked words and marked pseudowords as the covariates. This analysis showed how much variation in the effect of novel word learning could be explained by the sensitivity to marked words and marked pseudowords.

iii. Results

Language decision task

The analysis of sensitivity to orthographic markedness from reaction times showed a significant effect of Marked Lexicality (*$F(2,86)=16.5$, $p<.001$, $\eta_p^2=.391$*). To follow-up on this effect, a Tukey post-hoc was run. Marked words and unmarked words showed a significant difference in reaction time (*$t(86)=1.01$, $p=.038$*), indicating that children needed more time to decide on the language when words did not have any orthographic markers (*Casaponsa et al. 2014*). Marked words and marked pseudowords showed a significant effect (*$t(86)=80.1$, $p<.001$*), indicating that marked pseudowords required more time to process than words (*see Figure 23*).The analysis of percentage of errors showed a significant effect of Marked Lexicality (*$F(2,86)=15.2$, $p<.001$, $\eta_p^2=.261$*). Marked words and unmarked words did not differ in terms of errors (*$t(86)=1.29$, $p=.783$*), but marked pseudowords differed from marked words and unmarked words (*$t(86)=25.2$, $p<.001$ and $t(86)=23.9$, $p<.001$, respectively*) (*see Figure 23*).

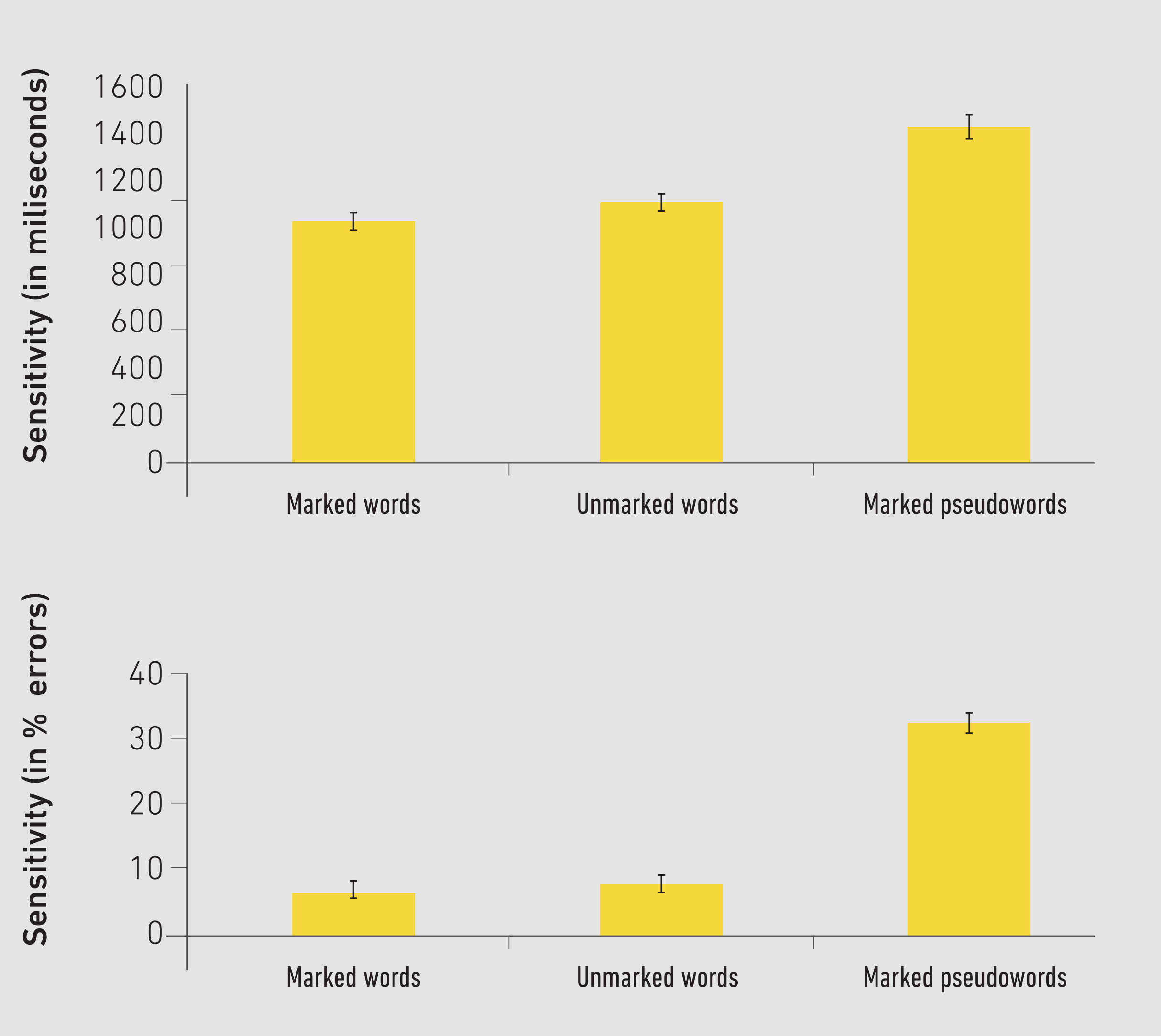


Figure 23. *Bar plots of the sensitivity of marked word, unmarked word and marked pseudowords in milliseconds and % of errors. Error bars refer to the standard error (SE) of each mean.*

<u>Novel Word learning task</u>

Results in the Recall task showed that children recalled legal novel words better than illegal ones (*t1(45)=3.04, p =.004, Cohen´s d=.454,* BF_{10}*=8.82; t2(15)=3.87, p <.001, Cohen´s d=.467,* BF_{10}*=9.22*). Also, in terms of the Levenshtein distance measure, legal novel words were closer to targets than illegal novel words (*t1(45)=4.32, p <.001, Cohen´s d=.647,* BF_{10}*=9.82; t2(15)=3.94, p <.001, Cohen´s d=.563,* BF_{10}*=9.46)*. In the recognition task, children learned legal and illegal novel words equally well (*t1(45)=0.68, p =.497, Cohen´s d=.102,* BF_{10}*=0.20; t2(14)=0.52, p =.609, Cohen´s d=.135,* BF_{10}*=0.296*). Also, reaction times showed that children recognized both types of novel words to the same extent (*t1(45)=0.13, p =.897, Cohen´s d=.019, BF10=0.16; t2(15)=1.36, p =.195, Cohen´s d=.351,* BF_{10}*=0.568*) (*see Figure 24*).

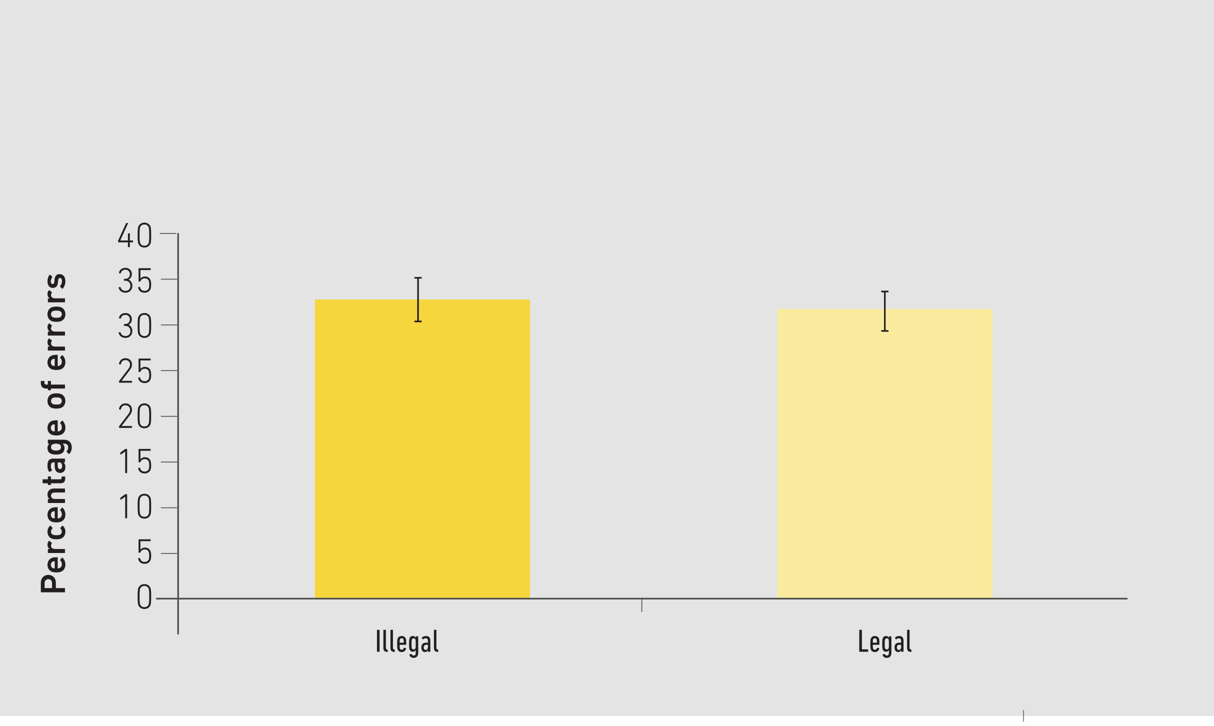


Figure 24. *Plot bars of the percentage of errors in the recognition task for legal and illegal orthographic sequences. Error bars refer to the standard error (SE) of each mean*

<u>Linear regression between marked words/pseudowords and novel word learning task</u>

The linear regression model showed a significant correlation between the three variables (R^2*=0.176, F(2,41)= 4.45, p=.018*), in which 18% of the variation in the marked words and/or marked pseudowords explained the variation in the effect of novel word learning. Marked words significantly predicted the effect of novel word learning (β*=-0.031, p=.026, SE=0.013*). However, the marked pseudowords did not contribute to the effect of novel word learning (β *=0.007, p=.412, SE=0.008*). These results suggested that only the RTs to marked words predicted the effect of novel word learning, and that the time needed to process marked pseudowords did not add any explanatory power to the model.

Looking specifically at each variable to understand why marked words explain the effect of novel word learning but not the marked pseudowords, two correlation analysis were conducted. Marked words significantly correlated negatively with the effect of novel word learning (*r=-0.402, p=.006; see Figure 25*), but the correlation of marked pseudowords was marginally significant (*r=-0.266, p=.077; see Figure 25*). This showed that children who were faster at detecting marked words were the ones who learned the illegal novel words better, since a higher positive number in the effect of novel word learning meant fewer errors for illegal novel words (*see Figure 25*). However, marked pseudowords may not predict the effects of novel word learning due to marginal correlation between the variables.

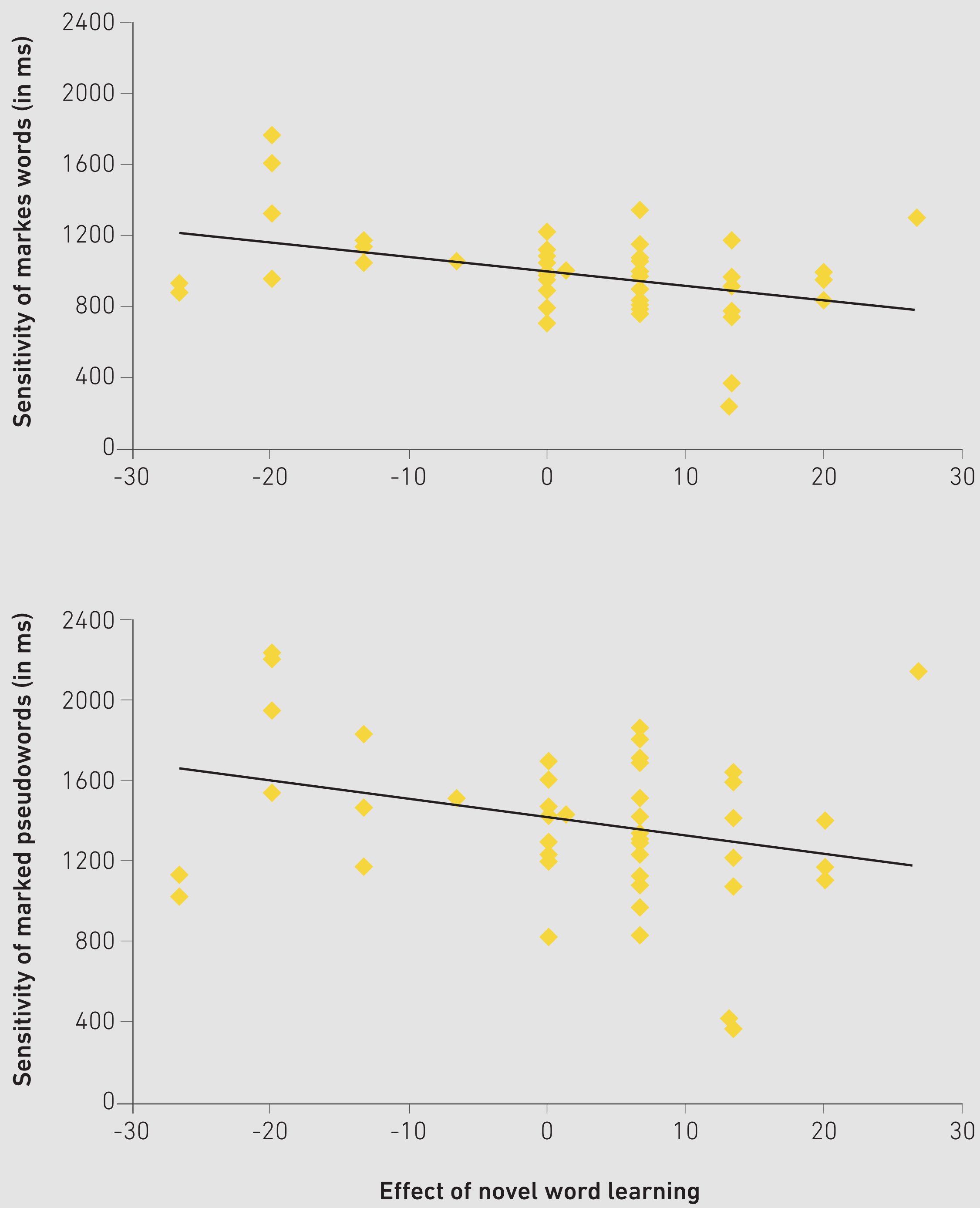


Figure 25. *Correlation plot between novel word learning effect (the difference between legal – ilegal percentage of errors of the recall task in the novel word learning task) and reaction times of marked words (in milliseconds)*

iv. Summary of results

The aim of this study was to investigate the relationship between learning novel words and the sensitivity to orthographic markedness in Spanish-Basque bilingual children (*at the age of 12*). Note that this was the only group in which learning appeared to be modulated by experience in handling different orthographic rules in two languages. To this end, a group of Spanish-Basque bilingual children, whose languages do not share many aspects and have different orthographic regularities, performed a language decision task and a novel word learning task to observe how orthographic sensitivity might predict novel word learning.

In the language decision task, sensitivity was measured by reaction times because this indicates how much time the participants needed to determine the language of the word/pseudowords using sub-lexical information (*the marked bigrams*). Shorter reaction times indicate better detection, therefore more sensitivity. Results indicated that children had longer reaction times when detecting the language for pseudowords than words. This is in line with previous research, which has found that pseudowords require more time to be read than words (*Price, Wise, & Frackowiak, 1996; Simos, 2002*). Also, results showed that marked words are processed faster than unmarked pseudowords, as in previous research (*Casaponsa et al., 2014*). On the other hand, in the novel word learning task, children had to learn legal and illegal novel words. Legal novel words followed the orthographic rules of both languages and illegal novel words contained illegal bigrams in all languages known by the bilinguals. Spanish-Basque children learned both orthographic sequences to the same extent, as shown by both percentage of errors and reaction times.

The linear regression analysis, which predicts the value of one variable based on the value of another, allowed us to investigate the relationship between sensitivity to marked words (*in milliseconds*), sensitivity to marked pseudowords (*in milliseconds*), and the effect of novel word learning (*legal-illegal % of errors*). In this case, the model showed a significant effect for marked bigrams, which predicted 18% of the variance in novel word learning. However, marked words explained most of the effect since marked pseudowords obtained a non-significant p-value in the model. This suggested that children were very good at detecting marked letter sequences in words and this implicit learning benefited children in picking up novel illegal novel words. However, we should not understand this as indicating an advantage in learning illegal words, but rather as showing that Spanish-Basque bilingual children can learn illegal and legal words equally well. Note that both children and adults could learn a maximum of 20 novel words on average from the 30 (*see Experiments 3 and 4*). In this sense, Spanish-Basque bilingual children who have more sensitivity to marked bigrams are better at learning illegal novel words. This shows that sensitivity to marked bigrams has a direct relationship to learning novel words.

To understand the model better, correlation analyses between marked words and the effect of novel word learning and marked pseudowords and the effect of novel word learning were conducted. The correlation results showed that the effect of learning novel words was negatively correlated with sensitivity to marked words. This indicated that children with higher sensitivity to marked bigrams in real words learned illegal novel words better. However, the correlation between learning novel words and marked pseudowords was marginal, suggesting that sensitivity to marked pseudowords is not an important factor. This result could be due to the fact children are still developing the ability to read pseudowords (*Van Bon & Van Der Pijl, 1997*), and even if they can benefit from marked sub-lexical information when determining the language of the pseudoword, they are not as good at detecting marked words. This would lead to a weak relationship between marked pseudowords and the effect of novel word learning.

V. Discussion

The present chapter aimed to better understand the function of orthographic regularities in learning to process new words. Specifically, how sensitivity to orthographic markedness impacts word learning at different ages. Thus, we observed how orthographic markedness interacted with novel word learning at different ages.

To this end, we observed whether children and adults learned novel words with different orthographic sequences (*legal and illegal*) and whether this learning was driven by bilingual experience itself, or rather by specific experience with orthographic regularities in their languages. We tested two different age groups, older children and young adults, in order to ascertain whether these effects were only present during the process of language development.

Experiment 3 focused on whether greater language differences can affect novel word learning in late childhood. Three groups of children were tested in Experiment 3A: children with dissimilar orthographic sequences in their two languages (*Spanish-Basque*), children with orthographically similar languages (*Spanish-Catalan*), and a group of Spanish monolinguals. Results from Experiment 3A showed that monolinguals and Spanish-Catalan bilingual children were better at learning novel words. These results were in line with prior literature showing that it is easier to learn items that correspond with our prior

knowledge (*Ellis & Beaton, 1993*). However, Spanish-Basque bilingual children deviated from this, learning legal and illegal novel to the same extent. To attempt to replicate these null findings and control for the effects of proficiency, a follow–up study was carried out. Experiment 3B tested two additional Spanish-Basque groups of bilingual children with different proficiency levels in their second language. Again, results showed that Spanish-Basque bilingual children learned the same number of both types of novel words. These groups demonstrated that the absence of a legality effect in this population is a stable phenomenon that does not depend on the level of proficiency. These results are in line with previous research showing that early balanced bilingual (*Bartolotti & Marian, 2012; Kaushanskaya & Marian, 2009a*), early unbalanced bilinguals (*Kaushanskaya, Yoo, & Van Hecke, 2013*), and late bilinguals (*Nair, Biedermann, & Nickels, 2016*) all learn novel words differently from monolinguals. Although in our study bilingual children did not outperform monolinguals in terms of overall word learning (*the mean number of novel word learned=20*), Spanish-Basque bilinguals performed differently than the other two groups.

One important follow-up question was whether this pattern of results was only found in childhood, when language development was still ongoing, or was maintained through adulthood. In order to address this question, in Experiment 4 we tested Spanish monolingual and Spanish-Basque bilingual adults using the same task. Note that we did not include a Spanish-Catalan group because results with Spanish-Catalan children had shown the same performance as Spanish monolinguals in Experiment 3A. Results showed a different pattern of word learning compared to that of children, with Spanish-Basque bilingual adults clearly outperforming monolinguals overall (*bilinguals learned around 24 novel words and monolinguals around 20*), but with both language groups showing a comparable legality effect. Differing from the children, both the Spanish monolingual and the Spanish-Basque bilingual adults recognized legal words better than illegal words, and the former group had an overall poorer performance than the latter one.

It seems that all groups (*children and adults*) except Spanish-Basque bilingual children were better at learning legal novel words than illegal ones. This tendency to learn legal novel words better is in line with previous research (*Ellis, 2002; Ellis & Beaton, 1993*). When people intentionally learn new words, they prefer to learn orthographic sequences that are consistent with existing stored orthographic regularities storage. This makes sense because a priori what most closely resembles your language is easier to learn. However, Spanish-Basque bilingual children were equally likely to pick up both legal and illegal words. Lutjeharms (*1994; see Chapter 1. General introduction for review*) showed that bilingual readers who were tested on the extent to which they learned new words from a text preferred to choose words which had low frequency bigrams in their languages. He illustrated that bilinguals do not pay attention to high frequency letters because they are not usually salient. This leads us to think that, differing from the other groups, Spanish-Basque children prefer to pick up both high and null frequency bigrams. This may be due to their experience with two orthographically distinct languages.

In line with the results for Spanish-Basque bilingual children, we think it likely that experience managing different sets of rules for orthographic regularities in a bilingual's two languages may play an important role in learning new words. This was found to be the case in a study conducted by Van Gelderen and collegues (*2003*) with Dutch-Turkish, Dutch-Moroccan bilingual children and Dutch monolingual children, all on English reading tasks. Note that the age of these children was twelve, as in our target group. They did not observe a bilingual advantage in English reading because all groups performed equally well on tests of English word recognition, vocabulary and grammatical knowledge. The authors suggested that the lack of differences found between bilingual and monolingual children reflected the fact that bilingual participants were Dutch monoliterates, who had acquired literacy only in Dutch. Although these bilinguals had to deal with two languages, differences in performance may be specifically related to dealing with written orthographic sequences. Thus, these results are in line with the current findings because they suggest that dealing with different orthographic sequences within a language may induce a change in performance with new words.

However, this effect was only evident in Spanish-Basque bilingual children because adults did not show the same pattern of result. Spanish-Basque bilingual adults did overall learn more novel words than monolinguals. However, if we go back to the bilingual literature on new word learning, we see that the majority found that bilinguals were more efficient than monolinguals at new word learning, including both children and adults (*Kaushanskaya & Marian, 2009a, 2009b; Yoshida et al., 2011*). But our results did not reveal this pattern for children. It seems there may be differences in learning depending on age.

The null effect in learning novel words did not last across time because Spanish-Basque young adults no longer showed this same pattern of results. This suggests that differences between the age groups were due to changes in development. As we saw in Experiment 1 (*see Chapter 3. Changes in sensitivity to orthographic markedness for review*), children around the age of twelve showed changes in sensitivity to orthographic markedness within their languages. These changes in processing orthographic regularities may be influenced by the tendency to choose the orthographic sequences in novel words. However, Spanish-Basque children showed a stronger tendency to learn novel words than their peers. Thus, in Experiment 5, we attempted to observe the relationship between sensitivity to orthographic markedness and learning novel words in Spanish-Basque bilingual children. Our aim was to determine whether this critical group showed a benefit in learning both type of novel words. The only difference in this critical group compared with the other groups is their prior experience with handling two languages with different orthographic rules.

The regression analysis between the effect of novel word learning, sensitivity to marked words, and sensitivity to marked pseudowords showed that marked words explained 18 % of variance of the effect

of novel word learning but marked pseudowords did not contribute, showing a negative p value. Looking at the results in detail, the correlation analysis showed that sensitivity to marked words was negatively correlated with the effect of novel word learning, suggesting that higher sensitivity to (*less time to detect*) marked words leads to better learning of illegal novel words. However, the marked pseudowords showed only a marginal effect on the effect of novel word learning, suggesting that when all variables are accounted for, marked pseudowords do not contribute to novel word learning. In this regard, only sensitivity to marked words explained performance in learning novel words.

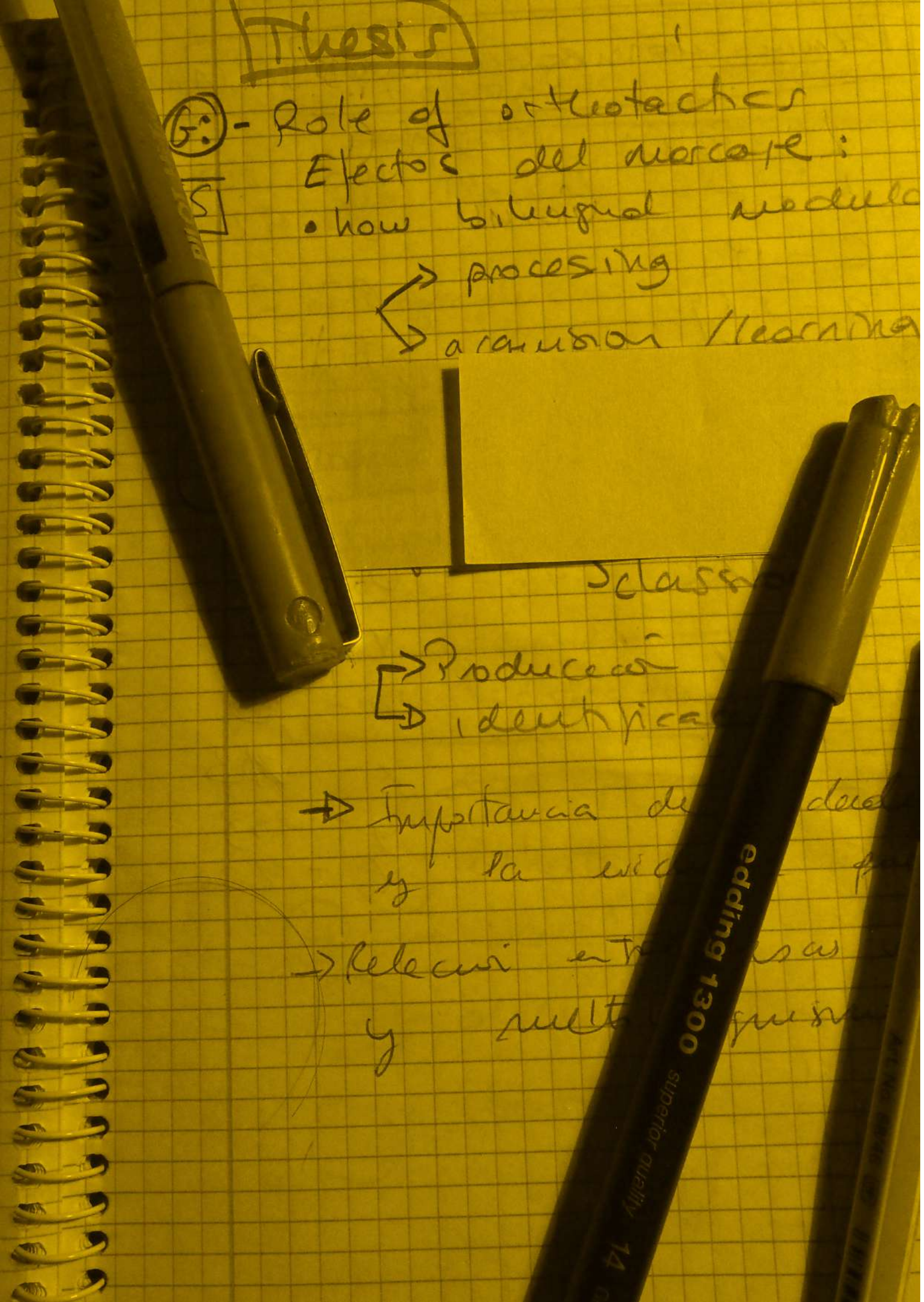
The fact that only marked words could predict learning novel words could be explained by automatic and implicit mechanisms that children adapt in order to learn novel words. Those automatic and implicit mechanisms are the same as those used to process real words (*with the goal of reading*), but not pseudowords. Coming back to reading process literature, most models describe the automaticity and fast detection of familiar words (*e.g., lexical route; Coltheart et al., 2001; Grainger & Ziegler, 2011*), which do not need letter-to-sound mediation but directly activate word semantics. However, reading pseudowords requires letter-to-sound conversion rules (*sub-lexical route; Coltheart et al., 2001; Grainger & Ziegler, 2011*). In this sense, when children saw the novel words for the first time they may have processed the novel words as pseudowords. But after a few repetitions, due to the experimental design, the novel words may become familiar allowing children to process them like familiar words. The repetition of the novel words in the task may encourage the rapid creation of a connection from the sub-lexical to the lexical level (*Martin & Gupta, 2004*). In order to memorize novel words, individuals need to process them as familiar words (*Ehri, 1995; Miller & Gildea, 1987*) because learning a novel word requires the sub-lexical to lexical connection to be encoded sufficiently strongly to resist decay over time. Thereby, it seems likely that only real words but not pseudowords explain novel word learning because novel words are processed as familiar words, using the same automatic and implicit mechanism.

Furthermore, this indicates that although children are good at capturing markedness in words in their own languages and therefore illegality in novel words, they tend to be more flexible and can learn both legal and illegal novel words. Flexibility in processing legal and illegal novel words could equally well arise from skill processing marked words because the bilingual child's two languages differ in terms of orthographic regularities. This experience handling different orthographic regularities and/or language capacity at this particular age could facilitate their learning both types of novel words to the same extent.

Following this same reasoning, associationist theories have hypothesized that children are sensitive to co-occurrences in language and that these regularities may be used to support the acquisition of novel words (*Plunkett, 1997; Samuelson & Smith, 1998; Smith, Jones, & Landau, 1996*). Through experience with the ambient language, children appear to learn the regularities of the language which in turn support word learning. These theories assume that experience with a language tunes the child's attention to its specific regularities. These theories might predict that children learn orthographic markedness probabilities and use this information to support word learning, as we observed. In this sense, as bilingual children deal with the differences in orthographic regularities in their languages, attention to regularities should become more strongly attuned, accounting for the observation that common and illegal letter sequences are learned equally well by this group.

In sum, results from the Spanish-bilingual children suggest that experience with languages that are orthographically very different leads children to become more sensitive to the illegality (*or legality*) of novel words with respect to their known languages (*Experiment 3*). Thus, they may learn both types of orthographic sequences equally well, because they cannot only learn letter sequences common to the languages that they know because they are easier to remember (*Ellis & Beaton, 1993*), but can also learn letter sequences distant from their prior knowledge because they are used to handling illegal orthographic sequences in their daily life. This demonstrates that the BIA models (*Casaponsa et al., 2014; Dijkstra & van Heuven, 2002; Van Kesteren et al., 2012*) do not account for the dynamism of the bilingual system. Even when participants tend to prefer to choose the orthographic sequences already stored in their lexicon, they use a feed-forward connection with the orthographic language node (*this finding will be discussed in Chapter 4. General discussion*). However, this pattern endures over time (*Experiment 4*). It seems that at the age of twelve, changes in sensitivity to orthographic markedness within known languages (*as Experiment 1 showed*) as well as other implicit learning changes occur (*this finding will be discussed in Chapter 4. General discussion*). These changes probably result in children learning in a different way. Although children are good at detecting markedness in words in their languages (*Experiment 5*), these changes make them more flexible in terms of picking up new orthographic sequences.

Chapter 3 Novel word learning and orthographic markedness



Chapter 4
General discussion

Previous research showed that orthographic markedness is very important to reduce lexical competition (*Chetail, 2015; Owsowitz, 1963; Rice & Robinson, 1975*), to speed up language attribution in bilinguals Oganian, Conrad, Aryani, Heekeren, & Spalek, 2016; Vaid & Frenck-Mestre, 2002; Van Kesteren, Dijkstra, & de Smedt, 2012), and in learning new words (*Ellis, 2002; Ellis & Beaton, 1993; Lutjeharms, 1994*). With this in mind, the present thesis aims to better understand the function of orthographic regularities through development in the word processing system (*in word processing and in learning to process new words*). Specifically, we wanted to understand better how the function of sensitivity to orthographic markedness in word recognition changes across the lifespan, and how the function of orthographic regularities impacts the learning of new words at different ages. A better knowledge of the impact of sensitivity to letter sequences would help us understand the impact they have on visual word recognition and allow us to determine whether the system is relatively static (*Dijkstra & van Heuven, 2002b*) or instead dynamic (*Kroll et al., 2014*).

We propose a twofold approach to investigate these research questions, through word processing and word learning. First, we aim to examine the function that sensitivity to orthographic markedness plays in the bilingual mind and whether it changes with age. Second, we hope to observe how the function of sensitivity to orthographic markedness impacts in novel word learning at different ages.

This PhD thesis provides novel evidence supporting a substantial role for sensitivity to orthographic markedness in the word processing system (*BIA+ extended model; Van Kesteren et al., 2012; Casaponsa et al., 2014*). This evidence was made possible by investigating the markedness effect and its implications across development (*Chapter 2*), normal early second language development (*Experiment 1*) after learning a second language late in life (*Experiment 2*), and in learning (*Chapter 2*) with children (*Experiment 3*) and adults (*Experiment 4*). Note that Experiment 5 was conducted to observe the relationship in the possible critical group between learning novel words and sensitivity to orthographic regularities. In this chapter, we first summarize the results obtained in each chapter, and then consider the implications of these findings for current bilingual interactive-activation models.

I. Summary of findings and conclusions

i. Changes to sensitivity to orthographic markedness

To address our first research questions, how the function of orthographic regularities in word recognition changes across development in the bilingual mind, we explored whether sensitivity to orthographic markedness was stable across the lifespan or whether it was subject to change. This approach was further developed in Chapter 2. We considered two possible scenarios: changes in early second language learners in normal development and changes as a consequence of learning a new language late in life.

To this end, we conducted two further experiments. In order to investigate sensitivity to orthographic markedness, we employed a language decision task choosing words in two conditions: marked (*at least one bigram does not exist in the other language*) and unmarked (*all bigrams exist in both languages*). Experiment 1 focused on the study of normal bilingual development. Therefore, four balanced Spanish-Basque bilingual groups of different ages (*younger children, older children, teenager and young adults*) performed the task. In this experiment, the task included words and pseudowords, since the pseudowords allowed us to observe better how the participants used sub-lexical cues to be able to attribute the language. On the other hand, Experiment 2 investigated changes in sensitivity to orthographic markedness as a consequence of learning a second language later in life. Thus, older monolingual adults learned a second language for a year and were tested in a language decision task just with pseudowords at three different moments during the learning process (*before learning, after learning and one year after the last test*). Results from Chapter 2 showed that sensitivity to orthographic markedness changed throughout development, both during normal bilingual development (*Experiment 1*) and as a consequence of second language learning late in life (*Experiment 2*).

It was known that even pre-readers are very sensitive to orthographic markedness in their language(*s*) even when they do not know how to read properly (*Chetail, 2015*). And even monolinguals who are not familiar with a second language are very sensitive to words from other language based solely in their orthographic markedness which violates the stored orthography (*sub-lexical*) in the language node (*Casaponsa et al., 2014*). So, it was expected that both bilinguals (*as we saw in Experiment 1*) as well as monolinguals (*as we saw in Experiment 2*) would be very sensitive to orthographic markedness in their second language.

But we also reported two novel findings. The first is that changes in sensitivity to orthographic markedness in the second language occur around the age of twelve as a result of normal development. And the second is that there are changes in sensitivity to orthographic markedness in the native language in adults (*in normal development and as a consequence of second language learning late in life*). As noted, in the BIA+ model proposed by Dijkstra and van Heuven (*2002*), the native language is considered to be stable across the lifespan and it is considered that only the second language can be modulated by the native language. But given our findings, we propose that the native language is also subject to modulation. In this case, the orthographic (*sub-lexical*) language node must have a feedback loop that produces changes in the language node.

Dissecting these two findings, the first novel finding is that sensitivity to orthographic markedness in the second language changes around the age of twelve due to normal bilingual development. As we reviewed in Chapter 2 (*see V. Discussion*), bilinguals use the sub-lexical information of the words (orthographic markedness) in order to speed up language attribution by restricting cross-language lexical activation (*BIA+ extended model, Casaponsa et al., 2014*). Although we could see that older children change the way they perceived native words, no advantage is found for their processing of marked words in the native language, as Vaid and Frenck-Mestre (*2002*) showed. In this sense, it seems that older children rely less on marked letter sequences to process words compared with their younger peers.

Children before the age of twelve may not be very competent in reading skills as they are still learning how to decode words. As we already reviewed in previous sections (*see Chapter 1. General introduction for a review*), orthographic regularities are acquired by statistical learning though exposure to written words, and even pre-readers are very sensitive to words with distinctive orthographic regularities (*Chetail, 2015*). Thus, younger children may show larger markedness effects than the rest of the groups, because they rely more on marked orthographic regularities when deciding on the language. However, at the age of twelve, older children may change the way they perceive distinctiveness because older children reach a competent reading level and they may rely less on orthographic cues to process words because they are already very skillful readers. It is interesting that a number of authors have described changes at this age in other domains, such as in sensitivity to implicit learning (*Janacsek et al., 2012*), in verbal fluency (*Sauzéon et al., 2004*), and in cerebral maturation (*Giedd et al., 1999*)

Janacsek and collegues (*2012*) described changes in sensitivity to implicit learning around the age of twelve. The authors tested individuals from 4 to 85 years of age in an Alternating Serial Reaction Time task (*ASRT task; Howard & Howard, 1997*), in which participants were instructed to respond to different stimuli that appeared in one of four empty circles arranged in a line on the computer. The probability of a specific stimulus sequence could be high or low, depending on its frequency of transitional probabilities within a sequence appearance. Results showed that older children displayed a rapid decrement in reaction times for both high and low probability events compared with younger groups who preferred high probability events. This suggests that individuals undergo a marked shift at this maturational point from strategies based on high probabilities to more complex interpretations of events. Together with our results this suggests that around the age of twelve children start to rely less on salient cues, whether for learning to process sub-lexical or other types of statistical information.

Changes in the way children rely on frequency information is also related to the cerebral maturation that Giedd and colleagues (*1999*) described in a neuroimaging study. The authors showed that most cerebral maturation in cerebral white and gray matter in the frontal cortices occurs around the age of twelve. This late frontal maturation may explain the fact that at this age, children change cognitive strategies. In other words, cognitive strategies may be dependent on the maturation of the frontal lobe. Behavioral studies corroborate this interpretation, showing that major improvements in cognitive strategies such as strategic search abilities and strategic retrieval and processing also change also around this age (*Guttentag, 1997; Passler, Isaac, & Hynd, 1985*).

Verbal fluency performance may rely on the same strategic retrieval function. Verbal fluency tasks require the participant to retrieve as many words as possible that begin with a given letter -excluding proper names and repetitions of the same word with different endings- in 60 seconds (*e.g., letter fluency task; Spreen & Strauss, 1998*). Sauzéon and collegues (*2004*) conducted a lifespan experiment with a letter fluency task. Authors showed that children around the age of twelve retrieved more words than their younger peers. It seems that before the age of twelve children prefer an orthographic clustering strategy characterized by a small number of large clusters. However, at the age of twelve children seem to use a more efficient clustering strategy consisting of numerous small orthographic clusters as well as small semantic clusters. Therefore, starting at this age, children are able to access a more extensive semantic network (*Blewitt, 1994*).

However, this early age trend differs in the case of pseudowords. The benefit of testing pseudowords is that participants need to trust sub-lexical cues in order to be able to attribute the language because these words do not have meaning. In this sense, again we saw that bilinguals had an advantage when processing pseudowords that were marked in the second language compared with the native language. And older children seem to change the way they perceive this orthographic markedness in the native language. So, the second novel finding is that adults change the way they

perceive orthographic markedness in the native language. Although it seems that bilinguals are not sensitive to orthographic markedness in their native language when they process words, it appears that around the age of eighteen this sensitivity changes in situations where bilinguals can only use sub-lexical cues to attribute the languages. It seems that by this age bilinguals are good enough at processing the words in their native languages that they do not need to rely on orthographic cues to reduce lexical competition. But in the absence of meaning, they have to rely sub-lexical cues in order to attribute language.

Preceding research on language categorization suggested a different trajectory of development in the bilingual linguistic system (*Segalowitz, 1991; Van Kesteren et al., 2012*). Researchers assumed that the native language was stable across time and that the second language should be the one that changes most throughout acquisition and consolidation. The native language should be the one that influences and modify the second language, and not the other way around (*see BIA+ model proposed by Dijkstra & van Heuven, 2002*). Evidence in support of this assumption came from studies showing that second language learners normally exhibited troubles with L2 accents and prosody, with spillover or transfer effects from the L1. This malleability of the L2 led some authors to characterize the native language as stable and resistant, and the L2 as weak and impressionable (*Frenck-Mestre & Pynte, 1997; Hernandez, Bates, & Avila, 1994*).

However, and not surprisingly, recent evidence shows that it is not only the L2 that changes during learning but also the L1 (*see, among many others, Baus, Costa, & Carreiras, 2013; Kroll, Dussias, Bice, & Perrotti, 2015*). These proposals are in line with our findings. Recent studies showed that total immersion in a second language context could lead to a process of losing the first language. One extreme example is children adopted from another country of origin. After stabilizing in the new country and being isolated from the native language, their native language seems to be erased from their minds (*Fillmore, 199*1). This is a very drastic example, but it has been demonstrated that gradual decline in the correct production and comprehension of a native language occurs after frequent exposure to and use of the second language (*Language attrition; see Schmid, 2008 for review*). The major and most noticeable changes in the native language are in vocabulary (*in their lexical access and their mental lexicon; Köpke & Schmid, 2004; Baus et al., 2013), grammar (desintegration of the structure; Seliger & Vago, 1991; Dussias & Sagarra, 2007*) and phonology (*native accent; Schmid, 2009; Chang, 2013*).

Our finding that L1 sensitivity can change with age and while learning a second language is in line with an adaptive view of the bilingual´s language system described by Kroll, Bob, & Hoshino (*2014, also see the review by Kroll, Dussias, Bice, & Perrotti, 2015*). These authors have argued that bilingual's language system is permeable to both languages, especially with high L2 proficiency. It seems that the native language changed in response to L2 use, showing that the bilingual language system is flexible and dynamic. So, it is worth noting to think that there are more aspects of the native language that change as a consequence of the second language acquisition. Our results showed a clear influence of the second language on the native language in orthographic processing.

The possible explanation regarding changes in sensitivity to native orthographic regularities is that readers may start to compare the letter sequences of both their languages. They may thereby become more sensitive to the orthographic differences between the languages. This seems to happen in normal bilingual development, which is linked with previous findings showing that rule extraction skills are part of the dynamic system that changes across the lifespan, and specifically adults change the way they learn statistical probabilities (*Campbell, Hasher, & Thomas, 2010; Campbell, Trelle, & Hasher, 2014*). Thus, age may play an important role in the way adults perceive markedness in their L1. But also, learning a second language influences how adults perceive markedness in the native language.

Learning implies integrating novel words with existing representations of the native language. While learning a second language, individuals may also learn about the similarities and distinctiveness between these novel words and our native vocabulary. Learners may check and compare the new letter sequences with those we already known (*native patterns*) in order to link new information with existing information. It seems plausible that as they compare the new letter sequences to the already known ones, they became more sensitive to characteristics of the already known ones. The old letter sequences seemed to be established and stuck and we had not even realized we used them, but, as they learn new letter sequences they may begin to be more aware of them. Hence, participants after learning a second language became more sensitive to native language distinctiveness and could detect that some strings contained native language specific cues when compared to the new language.

To sum up, results from this chapter help us to understand better how the function of sensitivity to orthographic markedness changes through development. We saw that sensitivity to orthographic markedness change through normal bilingual development (*Experiment 1*) and after learning a second language late in life (*Experiment 2*). This implies that the orthographic (*sub-lexical*) language node (*BIA+ extended model; Casaponsa et al., 2014*) has a feedback loop, making the system unstable across development. Thus, the present findings suggest an update to the BIA+ extended model (*Casaponsa et al., 2014*). The orthographic (*sub-lexical*) language node should include a feedback array in order to support the present findings (*see Figure 26, green feedback loop arrow*), showing that the orthographic language node can be modulated by both languages (*Kroll et al., 2014*)

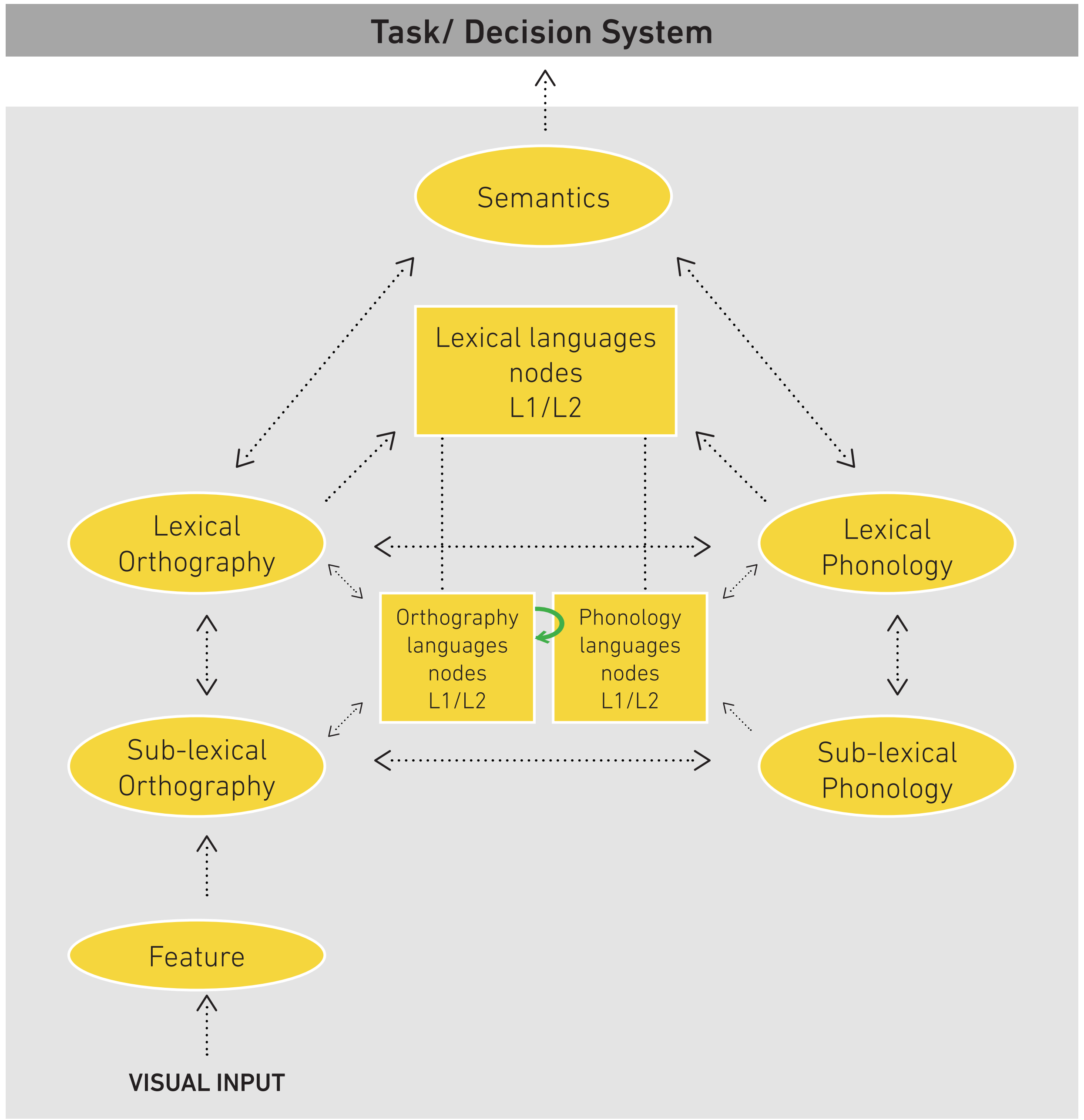


Figure 26. *Extension of Bilingual Interactive Activation model (BIA+ extended model; Casaponsa et al., 2014).*

ii. Changes to sensitivity to orthographic markedness

In the case of the second approach, how the function of orthographic regularities impacts the learning of novel words at different ages, we explored how sensitivity to orthographic markedness interacted with novel word learning in children and adults. This approach was developed in Chapter 3. Previous versions of the BIA models (*Casaponsa et al., 2014; Dijkstra & van Heuven, 2002b; Van Kesteren et al., 2012*) did not account to what extent the experience with different orthographic regularities in one mind could impact in learning novel words.

In this sense, we want to observe whether the sensitivity to orthographic markedness in the bilingual languages impact in storage new letter sequences in the orthography lexicon. To this end, we investigate whether novel words that have letter sequences that do not exist in any of the bilingual languages (*illegal e.g., bx is an illegal bigram because do not exist in Basque nor Spanish*) are learned, and thus process, differently than legal orthographic sequences (*all bigrams exist in the bilingual languages*). In this process, we consider also the role of bilingualism, in the sense of whether bilingual per se has an impact in learning novel words (*legal/illegal*) or whether the impact is modulated by the experience and handling with the differences between different orthographic regularities of the bilingual's languages. To do so, we conducted two experiments, one with children (*Experiment 3*) and the other one with adults (*Experiment 4*) performing a novel word learning task with legal and illegal novel words. Experiment 3A investigates whether children learn legal and illegal novel words, and three different groups of children participate in the experiment, one monolingual, one bilingual whose two languages had similar orthographic regularities, and one bilingual group whose two languages had different orthographic regularities. A follow-up study of the critical finding was conducted to replicate results (*Experiment 3B*). Then, Experiment 4 observed how adults learned novel words. Lastly, Experiment 5 aimed to observe the relationship between sensitivity to orthographic markedness and novel word learning in the critical group. Note that we conducted an extra experiment (*Experiment 5*) to examine a possible critical group, in which learning novel words was modulated by experience managing different sets of orthographic rules in their two languages. This experiment attempted to observe the relationship between learning novel words and sensitivity to orthographic markedness.

Results from Chapter 3 revealed that all tested groups, children and adults, except Spanish-Basque bilingual children preferred to learn legal to illegal novel words as previous research also showed (*Ellis, 2002; Ellis & Beaton, 1993*). The Spanish-Basque bilingual learned novel words with illegal orthographic sequences as well as legal ones. The BIA+ extended model (*Casaponsa et al., 2014*) only accounts for the bottom-up connection. But following our results, which suggest they prefer to learn legal novel words or legal and illegal equally, it seems that individuals rely on their previous knowledge in terms of sensitivity to orthographic markedness to pick up new letter sequences. Novel words are processed and then individuals decide, based on knowledge already stored in their orthographic language node, what type of letter sequences should be learned (*legal or illegal*). Thus, the BIA+ extended model (*Casaponsa et al., 2014*) should be updated and account for this feed-forward connection.

Thus, the first novel finding that Chapter 3 offered is that sensitivity to orthographic markedness plays an important role in the storage of new letter sequences. Specifically, experience with different orthographic regularities makes bilinguals, as is the case with Spanish-Basque bilinguals, storage new letter sequences in a different way than monolinguals or bilinguals whose languages share most orthographic regularities, as is the case for Spanish-Catalan bilinguals. We hypothesize that the driving factor leading to this differential effect could be linguistic experience with both bilingual languages. In particular, Spanish and Basque differ strongly in terms of their orthographic regularity rules, which is not so much the case with Spanish and Catalan. Thus, the experience of managing two different sets of orthographic rules is what sets this group of Spanish-Basque bilinguals apart from the other groups and may have allowed them to learn words equally regardless of whether their orthographic sequences violated rules in their already known languages. Then, it is worth mentioning that dealing with different orthographic regularities within the languages (*orthographic language nodes*) made bilinguals faster at integrating strange letter sequences in their orthographic (*sub-lexical*) lexicon. Thus, the BIA+ extended model (*Casaponsa et al., 2014*) should account for this finding that the whole system is malleable and the permeability of the levels depends on previous language-specific experience.

So, we tentatively could assume that it is not being bilingual per se that results in differences in novel word learning as previous research has suggested (*Kaushanskay & Marian, 2009; Margarita Kaushanskaya & Rechtzigel, 2012*), but rather the specific orthographic features of a bilinguals two languages what makes the system for storage of letter sequences permeable. The majority of previous experiments took bilingual learners with highly contrasting language combinations, such as Mandarin-English or Hebrew-English, French-English, etc. Some of these languages also do not share the same alphabet, but all compare bilinguals whose language differs with monolinguals. Thus, they found that bilinguals outperform monolinguals (*see Hirosh & Degani, 2018 for review*). We found the same pattern of results in our adult populations.

Bilingual adults overall outperformed monolinguals, performing equally well with both legal and illegal words. The effect found in Spanish-Basque bilingual children did not appear to endure across time, because adults did not learn legal and illegal novel words to the same extent. This is the second novel finding: differences in novel word learning with age. It is unclear what the underlying reason is for these differences between children and adults. What we know is that at the age of twelve children experience changes in sensitivity to implicit learning (*Janacsek et al., 2012*), in verbal fluency (*Sauzéon et al., 2004*), and in cerebral maturation (*Giedd et al., 1999*). So, we tentatively suggest that the changes we found are due to already acquired linguistic competence in the adults and the still ongoing language and cognitive development of children.

The ability to learn a new language changes across the lifespan and children may acquire new languages and new vocabulary in a different manner than adults (*Newport, 1990*). It has been shown that children who are still learning a language, learn more vocabulary and have a greater amount of exposure to new language elements (*e.g., rules, vocabulary, complex sentences*) than adults (*Shipley*

& McAfee, 2015). In contrast, adults undergo fewer changes in the language development of the languages they already speak. For instance, we saw in *Chapter 2* (*Changes in sensitivity to orthographic markedness*) that children at the age of twelve, which is the same age as our children group, showed major changes in sensitivity to orthographic markedness while adults showed minor changes.

Thus, we could consider that children are more flexible at learning novel words than adults. Thereby, one could tentatively account for the fact that only Spanish-Basque bilingual children, but not adults, learned legal and illegal novel words equally well. This group of children was still in the process of conforming to Basque and Spanish vocabulary, a process that could have made them less sensitive to illegality. On the other hand, adults who have already fully acquired both languages may have also developed sensitivity to differences between languages, something that they can use as a strategy for language attribution in some circumstances (*see Casaponsa et al., 2014 for a review*). Thus, unlike children, Spanish-Basque adults may be sensitive to legal versus illegal orthographic sequences and may consequently show better performance with legal words.

In order to address these questions and determine why Spanish-Basque bilingual children are equally able to learn legal and illegal words, we tested the relationship between learning legal/illegal novel words and sensitivity to orthographic markedness. We considered that Spanish-Basque bilingual children might learn both types of novel words to the same extent because they were blind to orthographic markedness, and thus picked words from both conditions. On the other hand, they could have a very high sensitivity to markedness in their languages and thus tended to pick both types of novel words. Thanks to Experiment 5, we were able to ascertain that the second scenario was the correct one. This should not be understood as an advantage for this group, but rather as a benefit that handling two languages with different orthographic regularities confers. In this sense, we suggest that flexibility in learning novel words with legal and illegal orthographic sequences benefits from experience with handling different orthographic rules.

We cannot forget to mention statistical learning. As we previously saw (*see Chapter 1. General lintroduction*) individuals have the ability to extract statistical regularities from the world around them to learn, and orthographic regularities are one form of these statistical regularities. Statistical learning is measured by a triplet learning paradigm, in which participants have to learn high and low transitional sequences. It is worth mentioning that statistical learning capacity is strongly related to better performance in learning novel words (*Frost, Siegelman, Narkiss, & Afek, 2013*) as well as greater vocabulary knowledge (*Evans, Saffran, & Robe-Torres, 2009*), but also with reading ability (*Arciuli & Simpson, 2012*). Statistical learning may help in learning novel words, which in turn indirectly supports reading ability by boosting a range of linguistic resources such as vocabulary that impact on reading. In this sense, Evans and colleagues (*2009*) conducted a study correlating children and adult performances. They showed statistical learning was a significant predictor of reading ability after age and attention were taken into consideration, demonstrating that higher statistical learning is related to increased vocabulary growth in primary-school children and vocabulary growth has been linked to reading ability (*see Gillon, 2004, for review*).

All in all, results from Chapter 3 helped us to understand better the function of orthographic regularities during the development of the word processing system, specifically in terms of learning to process novel words. We saw that sensitivity to orthographic markedness had a differential impact on learning novel words depending on age. Thus, the sensitivity to orthographic markedness had an impact on thee orthographic lexicon because when individuals were requested to learn novel words they relied on their previous orthographic knowledge to storee new letter sequences in the lexicon. The BIA+ extended model proposed by Casaponsa et al. (*2014*) should account for this finding by adding a feed-forward link. However, another novel finding was that this learning was modulated by experience with different orthographic regularities. Bilinguals whose languages differ in terms of orthographic regularities may develop a more flexible orthographic language node that allows them to learn letter sequences that are not yet in their lexicon with the same competence that the learn legal letter sequences. In contrast, monolinguals and bilinguals whose languages share most orthographic regularities do not have the benefit of this flexible system in learning new illegal letter sequences. Therefore, they prefer to pick novel words with letter sequences that are in their lexicon. We should note that this trend is only present in children because adults did not continue to show this pattern of learning. Bilingual adults outperform monolinguals overall. These differences in development could be understood as reflecting the fact that children are experiencing greater ongoing language and cognitive development compared to adults.

II. Outstanding questions

Our experimental designs attempted to address our research questions. In the experimental designs we tried to control as many factors as possible in order to observe in a lab environment results that would resemble a real environment. However, our experiments had some limitations. We would like to identify these so that future research can improve by taking them into account. On the other hand, the final goal of research is to implement what we can see in the lab in daily life. So, we want talk about possible application of our findings.

Our experiments had some limitations. In Experiment 1 we used age groups instead of treating age as a continuous variable. Van Walraven and Hart (*2008*) have explained the importance of accurately representing variable type but in this case the experimental design only allowed us to implement age as a between-subject variable. In order to accurately test a continuous variable we would have needed to test individuals across a distribution of different ages, not in a number of groups as we did. Recruiting children from many different classes would have complicated recruitment because of the school schedule. Thus, we decided to take groups of ages instead of random age participants. However, it is worth mentioning that we also carried out the analysis with age as continuous variable, and results did not differ from the results presented in Experiment 1.

Also, speaking about analyses, it is important to mention that recent research has pointed out that accuracy should be analyzed using mixed-effects models (*Baayen, Davidson, & Bates, 2008*). Accuracy is not a continuous variable because it is based on the response given by the participant, with incorrect responses coded as 0 and correct responses coded as 1. ANOVAs assume a normal distribution that depends on a result mean that stays exactly in the middle. For this reason, these authors argue that ANOVAs would not give a normal distribution in accuracy because of its binary nature. Given that the first studies of this thesis used ANOVAs, we decided to report ANOVAs in all experiments to ensure comparability. However, where possible we also ran mixed-effects models that showed similar findings.

Another limitation in our experiment was the learning phase in the novel word learning task (*Experiments 3 and 4*). As results from the recall task showed, there was a low percentage of words that were properly recalled. This low performance could lead to a floor effect, in which we cannot observe a difference between groups simply because performance does not allow for enough variability (*Baddeley, 1992*). In this case, more rehearsal might be needed to enhance recall performance and show further effects. But we have to consider that due to the difficulty of the task, recognition memory was more sensitive to showing the nuanced effects of language group because if participants had recalled the words better, we might not have been able to observe differences in recognition.

Future research should account for these limitations to improve data quality. Also, we suggest that future research should focus on studying these effects in other bilingual environments. This would allow the BIA+ extended model (*Casaponsa et al., 2014*) to be generalized to other kinds of bilingual populations. We have tentatively proposed, as our results showed, that the different levels in the word processing system may play different roles depending on the differences between a bilingual's two languages. Also, future research could focus on the study of the slight differences that being bilingual or learning a second language have on the word processing system at different ages. From our results we could see that bilingual adults and second language adult learners showed the same pattern of sensitivity to orthographic

markedness in the native language. However, as we found different patterns in bilingual children, we think it plausible that second language child learners may show different patterns as well.
Taken together, our findings have implications for implementations in both theory and daily life. We not only recommended updates to the Bilingual Interactive Activation model (*Casaponsa et al., 2014; Dijkstra & van Heuven, 2002b; Van Kesteren et al., 2012*) that improve our understanding of the bilingual mind, but also showed the importance of letter sequences while reading, specifically sensitivity to letter sequences in development. As we saw this sensitivity is learned by extracting regularities from the environment, which is related to how individuals learn. Tracking the development of sensitivity to letter sequences could be a significant way to assess learning and might also prove to be a marker for neurodevelopment. It would also be helpful for implementing new strategies for learning, specifically second language learning. Thus, these results are in strong agreement with everyday life experience. For instance, as our results shown, in accordance with previous findings, important changes occur around the age of twelve. So, it is worth considering if this might be a good age to start learning some types of sports, musical instruments, second languages, etc. since it could lead to higher levels of competence.

Understanding changes at this age could help us to understand the stages as well as modulations of learning ability across development. The importance of developmental and maturational knowledge is important because it could help us determine future skill levels. Thus, these results may have implications for the development of learning and memory, facilitating new skill training and sound pedagogical methods (*e.g. for teaching languages*). It may also contribute to our understanding of neurodevelopmental and age-related disorders (*e.g. autism, SLI, dyslexia and dementia*) and lead to appropriate treatment options.

Future research should also focus on better understanding the benefits associated with an ability to learn legal and illegal novel words to the same extent at this developmental age. In the near future, educational tools that offer training in different orthographic regularities could facilitate children around this age to learn novel words in a second or third language, for instance English. And perhaps, children with reading disabilities would also benefit from such a tool. Such children may detect their own known orthographic sequences yet still have trouble reading them.

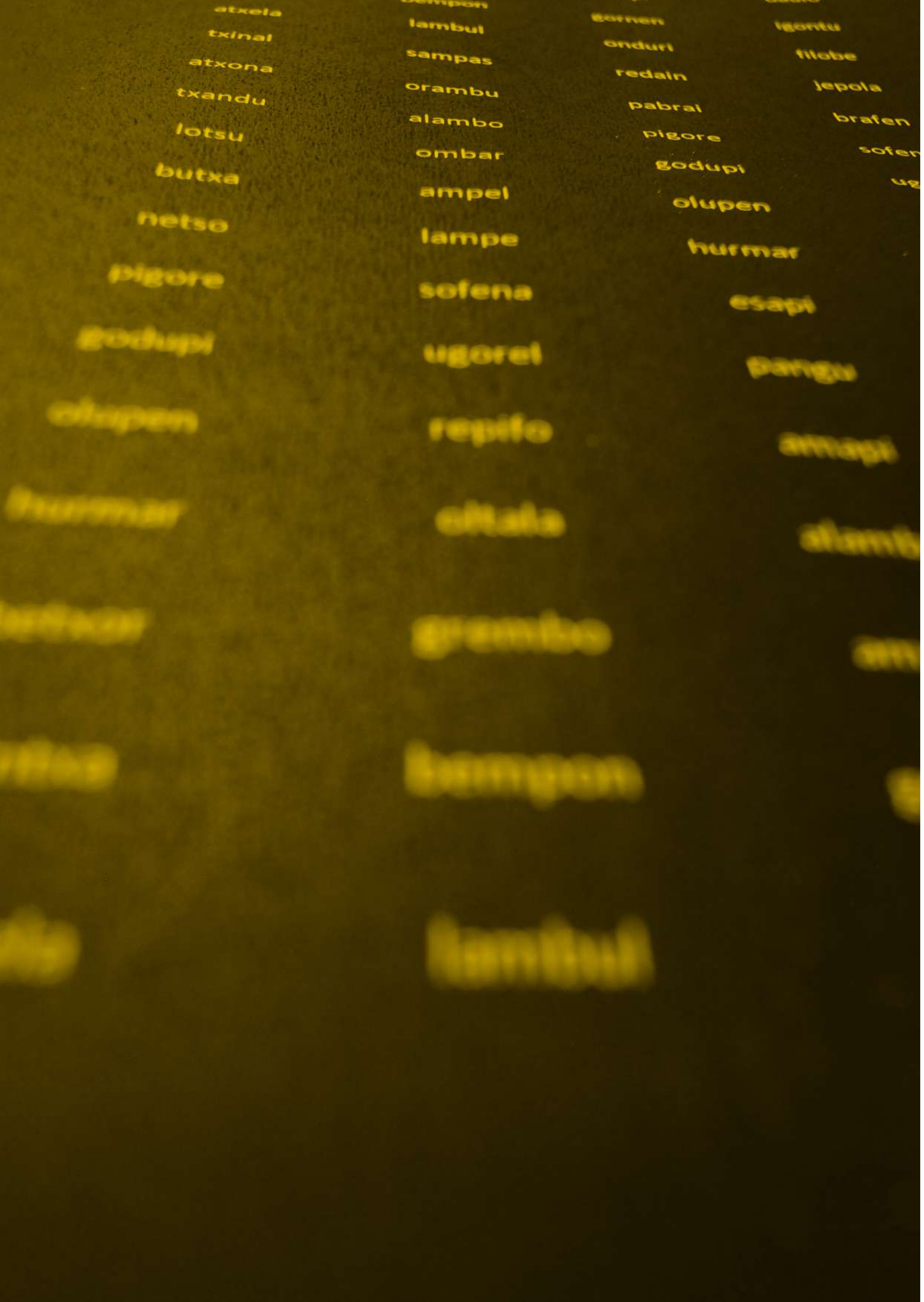
III. General Conclusion

To sum up, the present PhD thesis aimed to help us understand the function of orthographic regularities through development in the word processing system, specifically, in word recognition and in learning to process new words. To this end, we proposed a twofold approach to investigate the research questions. First, we aimed to investigate how the function of sensitivity to orthographic markedness changes through bilingual development (*Chapter 1*). Second, we tried to observe how the function of orthographic regularities impacted the learning of novel words at different ages (*Chapter 2*).

Results showed novel findings. First, sensitivity to orthographic markedness changed across development. Sensitivity to orthographic markedness in the second language changed around the age of twelve. It has been shown that at this age other changes occur, including changes in sensitivity to implicit learning (*Janacsek et al., 2012*), in verbal fluency (*Sauzéon et al., 2004*), and in frontal lobe cerebral maturation (*Giedd et al., 1999*). But it seems that not only does the L2 change as the BIA+ model assumed, but also sensitivity to orthographic markedness in the L1 changes with age and after learning a second language. This supports the idea of an adaptive system (*Kroll et al., 2014; Schmid, 2008*). These findings mean the BIA+ extended model (*Casaponsa et al., 2014*) needs to be updated to account for the fact that sensitivity to orthographic markedness in the orthographic language node is not stable thorough time and involves a feedback loop (*see 26*).

The second question was how the function of orthographic regularities impacts the learning of novel words at different ages. We could see that sensitivity to orthographic markedness modulated by the orthographic language node impacted storage of new letter sequences, whether or not letter sequences were already in the lexicon or were new. However, in the case of children at the age of twelve, we could see that experience with different orthographic regularities between their bilingual languages lead children to learn new letter sequences equally well as already known ones. However, this effect did not survive in adults. Again, developmental changes at this age should account as well as for language specific bilingual linguistic experience.

Chapter 4 General discussion



References

Acha, J., & Perea, M. (2010). On the role of consonants and vowels in visual-word processing: Evidence with a letter search paradigm. *Language and Cognitive Processes*, 25(3), 423–438. https://doi.org/10.1080/01690960903411666

Adams, M. J. (1994). *Beginning to read: Thinking and learning about print.* (M. Press, Ed.).

Adler, N., & Stewart, J. (2007). The MacArthur scale of subjective social status.

Andrews, S. (1992). Frequency and Neighborhood Effects on Lexical Access: Activation or Search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 802–814. https://doi.org/10.1037/0278-7393.15.5.802

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science,* 14(5), 255–259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Antón, E., Thierry, G., & Duñabeitia, J. A. (2015). Mixing languages during learning? Testing the one subject-one language rule. PLoS ONE, 10(6), 1–20. https://doi.org/10.1371/journal.pone.0130069

Apel, K. (2010). Kindergarten Children's Initial Spoken and Written Word Learning in a Storybook Context. *Scientific Studies of Reading,* 14(5), 440–463. https://doi.org/10.1080/10888431003623496

Apel, K. (2011). What is orthographic knowledge? *Language, Speech, and Hearing Services in Schools,* 42(4), 592–603. https://doi.org/10.1044/0161-1461(2011/10-0085)

Arciuli, J., & Simpson, I. C. (2012). Statistical Learning Is Related to Reading Ability in Children and Adults. *Cognitive Science,* 36(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Au, T. K., & Glusman, M. (1990). The Principle of Mutual Exclusivity in Word Learning : To Honor or Not to Honor ? Author (s): Terry Kit-fong Au and Mariana Glusman Published by : Wiley on behalf of the Society for Research in Child Development Stable URL : http://www.jstor.org/stable/. *Child Development,* 61(5), 1474–1490.

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

Baddeley, A. (1992). working memory. *Science*, 255(5044), 556–559. https://doi.org/10.1126/science.1736359

Baron, J. (1975). Successive stages in word recognition. *Attention and Performance V*, 563–574.

Bartolotti, J., & Marian, V. (2012). Language Learning and Control in Monolinguals and Bilinguals. *Cognitive Science*, 36(6), 1129–1147. https://doi.org/10.1111/j.1551-6709.2012.01243.x

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica,* 142(3), 402–409. https://doi.org/10.1016/j.actpsy.2013.01.010

Berndt, R. S., Reggia, J. A., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in english. *Behavior Research Methods, Instruments, & Computers*, 19(1), 1–9. https://doi.org/10.3758/BF03207663

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading*, 9(1), 43–61. https://doi.org/10.1207/s1532799xssr0901_4

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, 130(6), 858.

Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology*, 76(4), 610–622. https://doi.org/10.1037/0022-0663.76.4.610

Blewitt, P. (1994). Understanding Categorical Hierarchies : The Earliest Levels of Skill. *Child Development*, 65(5), 1279–1298.

Bordag, D., Kirschenbaum, A., Rogahn, M., & Tschirner, E. (2017). The role of orthotactic probability in incidental and intentional vocabulary acquisition L1 and L2. *Second Language Research*, 33(2), 147–178. https://doi.org/10.1177/0267658316665879

Broadbent, D. E., & Gregory, M. (1968). Visual perception of words differing in letter digram frequency. *Journal of Verbal Learning and Verbal Behavior,* 7(2), 569–571. https://doi.org/10.1016/S0022-5371(68)80052-0

Byers-Heinlein, K., & Werker, J. F. (2013). Lexicon structure and the disambiguation of novel words: Evidence from bilingual infants. *Cognition*. https://doi.org/10.1016/j.cognition.2013.05.010

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: A unique age effect. *Psychological Science*, 21(3), 399–405. https://doi.org/10.1177/0956797609359910

Campbell, K. L., Trelle, A., & Hasher, L. (2014). Hyper-binding across time: Age differences in the effect of temporal proximity on paired-associate learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(1), 300–305. https://doi.org/10.1037/a0034109

Casaponsa, A. (2016). *Bilingualism and Visual Word Recognition*.

Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: The impact of orthographic markedness. *Frontiers in Psychology*, 5(MAY), 1–10. 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(August), 153–166. https://doi.org/10.1016/j.brainres.2015.07.035

Casaponsa, A., & Duñabeitia, J. A. (2015). The Quarterly Journal of Experimental Psychology Lexical organization of language-ambiguous and language-specific words in bilinguals. https://doi.org/10.1080/17470218.2015.1064977doi.org/10.1080/17470218.2015.1064977

Casaponsa, A., & Duñabeitia, J. A. (2016). Lexical organization of language-ambiguous and language-specific words in bilinguals. *Quarterly Journal of Experimental Psychology*, 69(3), 589–604. https://doi.org/10.1080/17470218.2015.1064977

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, 89(4), 631–644. https://doi.org/10.1037/0022-0663.89.4.631

Cattell, J. M. (1886). The time it takes to see and name objecta. *Mind*, 11(41), 63–65.

Chambers, S. M., & Forster, K. I. (1975). Evidence for lexical access in a simultaneous matching task. *Memory & Cognition*, 3(5), 549–559. https://doi.org/10.3758/BF03197530

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41(6), 520–533. https://doi.org/10.1016/j.wocn.2013.09.006

Chauncey, K., Grainger, J., & Holcomb, P. J. (2011). The role of subjective frequency in language switching: An ERP investigation using masked priming. *Memory and Cognition*, 39(2), 291–303. https://doi.org/10.3758/s13421-010-0006-7

Chauncey, K., Holcomb, P. J., & Grainger, J. (2008). Effects of stimulus font and size on masked repetition priming: An event-related potentials (ERP) investigation. *Language and Cognitive Processes*, 23(1), 183–200. https://doi.org/10.1080/01690960701579839

Chetail, F. (2015). Reconsidering the role of orthographic redundancy in visual word recognition. Frontiers in Psychology, 6(May), 1–10. https://doi.org/10.3389/fpsyg.2015.00645

Chetail, F., & Content, A. (2017). The perceptual structure of printed words: The case of silent E words in French. *Journal of Memory and Language*, 97, 121–134. https://doi.org/10.1016/j.jml.2017.07.007

Chetail, F., & Mathey, S. (2010). InfoSyll: A Syllabary Providing Statistical Information on Phonological and Orthographic Syllables. *Journal of Psycholinguistic Research*, 39(6), 485–504. https://doi.org/10.1007/s10936-009-9146-y

Coltheart, Davelaar, E., Jonasson, J. T., Besner, D., & Dornic, S. (1977). *Attention and performance*.

Coltheart, M. (2001). Assumptions and methods in cognitive neuropsychology. *Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind.*

Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education*, 4(1), 5–17. https://doi.org/10.1080/13603110600574322

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Comesaña, M., Soares, A. P., Sánchez-Casas, R., & Lima, C. (2012). Lexical and semantic representations in the acquisition of L2 cognate and non-cognate words: Evidence from two learning methods in children. *British Journal of Psychology*, 103(3), 378–392. https://doi.org/10.1111/j.2044-8295.2011.02080.x

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and Bigrams: Orthographic Redundancy and Syllabic Units Affect Visual Word Recognition at Different Processing Levels. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 461–479. https://doi.org/10.1037/a0013480

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Cottrell, D., & Mason, H. (2009). *Sublexical units in English word recognition. In HCSNet Summerfest 2009 Conference* (p. 142). Sydney, NSW, Australia.

Council of Europe, C. (2011). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Press Syndicate of the University of Cambridge.

Cummins, J. (1993). Contexts for Second Language Teaching. *Annual Review of Applied Linguistics*, 13, 51–70.

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing*, 14(5–6), 549–568.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934–945. https://doi.org/10.1037/0012-1649.33.6.934

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37(4), 665–671. https://doi.org/10.3758/BF03192738

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

de Groot, A. M. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. (P. Press., Ed.), עטונה וולע.

Denckla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. *Brain and Language*, 3(1), 1–15. https://doi.org/10.1016/0093-934X(76)90001-8

Dijkstra, Ton, & van Heuven, W. J. B. (2002a). Modeling bilingual word recognition: Past, present and future. *Bilingualism: Language and Cognition*, 5(3), 219–224. https://doi.org/10.1017/s1366728902283017

Dijkstra, & van Heuven, W. J. B. (2002b). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition.* https://doi.org/10.1017/S1366728902003012

Dijkstra, Van Heuven, W. J., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Phonology by itself: Masked phonological priming effects with and without orthographic overlap. *Journal of Cognitive Psychology*. https://doi.org/10.1080/20445911.2011.477811

Duñabeitia, J. A., Dimitropoulou, M., Uribe-Etxebarria, O., Laka, I., & Carreiras, M. (2010). Electrophysiological correlates of the masked translation priming effect with highly proficient simultaneous bilinguals. *Brain Research*. https://doi.org/10.1016/j.brainres.2010.08.066

Duñabeitia, J. A., Ivaz, L., & Casaponsa, A. (2016). Journal of Cognitive Psychology Developmental changes associated with cross-language similarity in bilingual children. https://doi.org/10.1080/20445911.2015.1086773doi.org/10.1080/20445911.2015.1086773

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish - English bilinguals. *Bilingualism*, 10(1), 101–116. https://doi.org/10.1017/S1366728906002847

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading,* 18(2), 116–125. https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

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

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic Determinants of Foreign Language Vocabulary Learning. *Language Learning*, 43(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Estes, W. K. (1975). Memory, perception, and decision in letter identification. In R. Solso (Ed.), *Information Processing and Cognition* (pp. 3–30). Hillsdale:la.

Estes, W. K., Allmeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, 19(1), 1–15. https://doi.org/10.3758/BF03199379

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 321–335. https://doi.org/10.1038/jid.2014.371

Fillmore, L. W. (1991). When learning a second language means losing the first. *Early Childhood Research Quarterly*, 6(3), 323–346. https://doi.org/10.1016/S0885-2006(05)80059-6

Flege, J. E. (1987). A critical period for learning to pronounce foreign languages? , *Applied Linguistics*, 8 (1987) p.162. Applied Linguistics, 8, 162–177. https://doi.org/10.1093/applin/8.2.162

Frankish, C., & Barnes, L. (2008). Lexical and sublexical processes in the perception of transposed-letter anagrams. *Quarterly Journal of Experimental Psychology*, 61(3), 381–391. https://doi.org/10.1080/17470210701664880

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology*, 50A(1), 119–148.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What Predicts Successful Literacy Acquisition in a Second Language? *Psychological Science*, 24(7), 1243–1252. https://doi.org/10.1177/0956797612472207

Garnett, K., & Fleischner, J. E. (1983). Automatization and basic fact performance of normal and learning disabled children. *Learning Disability Quarterly,* 6(2), 223–230. https://doi.org/10.2307/1510801

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A., & Martin, A. (1999). Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13(December 1997), 65–77.

Gathercole, S., Willis, C., & Baddeley, A. (1991). ifferentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology*, (82), 87–98.

Gathercole, Susan E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28(2), 200–213. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, Susan E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological Memory and Vocabulary Development During the Early School Years: A Longitudinal Study. *Developmental Psychology*, 28(5), 887–898. https://doi.org/10.1037/0012-1649.28.5.887

Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development & disorders: A handbook on bilingualism & second language learning.* Paul H Brookes Publishing.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study [2]. *Nature Neuroscience*, 2(10), 861–863. https://doi.org/10.1038/13158

Goswami, U. (2011). No A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15(1), 3–10.

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

Grainger, J., & Dijkstra, T. (1996). Visual word recognition: Models and experiments. In T. Dijkstra & K. de Smedt (Eds.), *Computational psycholinguistics: AI and connectionist models of human language processing* (pp. 136–265). Taylor & Francis.

Grainger, J., & Holcomb, P. J. (2010). Neural constraints on a functional architecture for word recognition. *The Neural Basis of Reading*, 3–32.

Grainger, J., & Van Heuven, W. J. (2004). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *Mental lexicon: "Some words to talk about words."* Hauppauge, NY, US: Nova Science Publishers.

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology,* 2(APR), 1–13. https://doi.org/10.3389/fpsyg.2011.00054

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(02), 67. https://doi.org/10.1017/S1366728998000133

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A Web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, 45(3), 765–771. https://doi.org/10.3758/s13428-012-0296-8

Guttentag, R. (1997). Memory development and processing resources. *He Development of Memory in Childhood, 247–274.*

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, 27(03), 689–705. Retrieved from http://journals.cambridge.org/abstract_S0305000900004414

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Harris, S. E., Houlihan, L. M., Corley, J., Gow, A. J., Starr, J. M., Penke, L., ... Rafnsson, S. B. (2009). Age-associated cognitive decline. *British Medical Bulletin*, 92(1), 135–152. https://doi.org/10.1093/bmb/ldp033

Henderson, & Chard, J. (1980). The reader's implicit knowledge of orthographic structure. Cognitive *Processes in Spelling*, 85–116.

Henderson, L. (1980). Is there a lexicality component in the word superiority effect? *Perception & Psychophysics*, 28(2), 179–184. https://doi.org/10.3758/BF03204346

Hernandez, A. E., Bates, E. A., & Avila, L. X. (1994). On-line sentence interpretation in Spanish–English bilinguals: What does it mean to be "in between"? *Applied Psycholinguistics*, 15(4), 417–446. https://doi.org/10.1017/s014271640000686x

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

Howard, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging*, 12(4), 634–656. https://doi.org/10.1037/0882-7974.12.4.634

Howard, J. H., & Howard, D. V. (2013). Aging mind and brain: is implicit learning spared in healthy aging? *Frontiers in Psychology,* 4(November), 1–6. https://doi.org/10.3389/fpsyg.2013.00817

Hu, R. (2016). The Age Factor in Second Language Learning. *Theory and Practice in Language Studies*, 6(11), 2164. https://doi.org/10.17507/tpls.0611.13

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, 43(4), 559–617.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental Science*, 15(4), 496–505. https://doi.org/10.1111/j.1467-7687.2012.01150.x

Juel, C. (1988). Learning to Read and Write: A Longitudinal Study of 54 Children From First Through Fourth Grades. *Journal of Educational Psychology*, 80(4), 437–447. https://doi.org/10.1037/0022-0663.80.4.437

Kahn-Horwitz, J., Kuash, S., Ibrahim, R., & Schwartz, M. (2014). How do previously acquired languages affect acquisition of English as a foreign language:The case of Circassian. *Written Language & Literacy*, 17, 40–61. https://doi.org/10.1075/bct.89.03kah

Kan, P. F., & Sadagopan, N. (2014). Novel word retention in bilingual and monolingual speakers. *Frontiers in Psychology*, 5(September), 1–10. https://doi.org/10.3389/fpsyg.2014.01024

Katz, L. (1977). Reading ability and single-letter orthographic redundancy. *Journal of Educational Psychology*, 69(6), 653.

Kaufman, A. S. (2004). KBIT: *Kaufman Brief Intelligence Test (KBIT, Spanish version)*. madrid: TEA Editions.

Kaushanskay, & Marian. (2009). Bilingualism Reduces Native-Language Interference During Novel-Word Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 35(3), 829–835. https://doi.org/10.1037/a0015275

Kaushanskaya, Gross, M., & Buac, M. (2014). Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science*, 17(4), 564–583. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Kaushanskaya, M., Yoo, J., & Van Hecke, S. (2013). Word learning in adults with second-language experience: Effects of phonological and referent familiarity. Journal of Speech, *Language, and Hearing Research.*, 71(2), 233–236. https://doi.org/10.1038/mp.2011.182.doi

Kaushanskaya, Margarita. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism*, 15(3), 470–489. https://doi.org/10.1017/S1366728911000472

Kaushanskaya, Margarita, & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin and Review*, 19(5), 935–941. https://doi.org/10.3758/s13423-012-0271-5

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

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

Kinoshita, S., & Norris, D. (2013). Letter order is not coded by open bigrams. *Journal of Memory and Language*, 69(2), 135–150. https://doi.org/10.1016/j.jml.2013.03.003

Köpke, B., & Schmid, M. (2004). *Language Attrition : The next phase To cite this version : and. First Language Attrition. Interdisciplinary perspectives on methodological issues*.

Krogh, L., Vlach, H. A., & Johnson, S. P. (2013). Statistical learning across development: Flexible yet constrained. *Frontiers in Psychology*, 3(JAN), 1–11. https://doi.org/10.3389/fpsyg.2012.00598

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two Languages in Mind: Bilingualism as a Tool to Investigate Language, Cognition, and the Brain. *Current Directions in Psychological Science*, 23(3), 159–163. https://doi.org/10.1177/0963721414528511

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism , Mind , and Brain. *Annual Review Linguistics*, 1(1), 377–394. https://doi.org/10.1146/annurev-linguist-030514-124937

Lehtonen, A., & Bryant, P. (2005). Doublet challenge: Form comes before function in children's understanding of their orthography. *Developmental Science*, 8(3), 211–217. https://doi.org/10.1111/j.1467-7687.2005.00409.x

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

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory and Cognition*. https://doi.org/10.3758/BF03195845

Lemhöfer, K., Koester, D., & Schreuder, R. (2011). When bicycle pump is harder to read than bicycle bell: Effects of parsing cues in first and second language compound reading. *Psychonomic Bulletin and Review*, 18(2), 364–370. https://doi.org/10.3758/s13423-010-0044-y

Lemhöfer, K., & Radach, R. (2009). Task context effects in bilingual nonword processing. *Experimental Psychology*, 56(1), 41–47. https://doi.org/10.1027/1618-3169.56.1.41

Levy, B. A., Gong, Z., Hessels, S., Evans, M. A., & Jared, D. (2006). Understanding print: Early reading development and the contributions of home literacy experiences. *Journal of Experimental Child Psychology*, 93(1), 63–93. https://doi.org/10.1016/j.jecp.2005.07.003

Lobo, A., Ezquerra, J., Gómez, F. B., Sala, J. M., & Seva, A. D. (1979). ognocitive mini-test (a simple practical test to detect intellectual changes in medical patients). *Actas Luso-Espanolas de Neurologia, Psiquiatria y Ciencias Afines*, 7(3), 189–202.

Lupker, S. J., Perea, M., & Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels and letter frequency. *Language and Cognitive Processes*, 23(1), 93–116.

Lutjeharms, M. (1994). Worterkennen beim Lesen einer Fremdsprache [Word recognition while reading in a foreign language]. In W. Borner & K. Vogel (Eds.), *ognitive Linguistik und Fremdsprachenerwerb* [Cognitive linguistics and foreign language acquisition] (pp. 144–161). Tübingen.

Lyon, G. R. (1998). Why reading is not. *Educational Leadership*, 55(6), 13–18.

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

Marslen-Wilson, W. (1989). *Lexical representation and process*. (M. Press, Ed.).

Martin, C. D., Strijkers, K., Santesteban, M., Escera, C., Hartsuiker, R. J., & Costa, A. (2013). The impact of early bilingualism on controlling a language learned late: An ERP study. *Frontiers in Psychology*, 4(NOV), 1–15. https://doi.org/10.3389/fpsyg.2013.00815

Martin, N., & Gupta, P. (2004). Exploring the relationship between word processing and verbal short-term memory: Evidence from associations and dissociations. *Cognitive Neuropsychology*, 21(2–4), 213–228. https://doi.org/10.1080/02643290342000447

Massaro, D. W., Jastrzembski, J. E., & Lucas, P. A. (1981). Frequency, orthographic regularity, and lexical status in letter and word perception. *Psychology of Learning and Motivation*, 15, 163–200. https://doi.org/10.1192/bjp.111.479.1009-a

Massaro, D. W., Venezky, R. L., & Taylor, G. A. (1979). Orthographic regularity, positional frequency, and visual processing of letter strings. *Journal of Experimental Psychology: General*, 108(1), 107–124. https://doi.org/10.1037/0096-3445.108.1.107

Massol, S., Duñabeitia, J. A., Carreiras, M., & Grainger, J. (2013). Evidence for Letter-Specific Position Coding Mechanisms. PLoS ONE, 8(7), 1–9. https://doi.org/10.1371/journal.pone.0068460

McClelland, J., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception & Psychophysics*, 22(3), 249–261. https://doi.org/10.3758/BF03199687

McClelland, & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375. https://doi.org/10.1016/S0022-0728(02)01421-3

Miller, G. A., Bruner, J. S., & Postman, L. (1954). Familiarity of letter sequences and tachistoscop icidentification. *The Journal of General Psychology*, 50(1), 129–139.

Miller, G. A., & Gildea, P. M. (1987). How children learn words. *Scientific American*, 257(3), 94–99.

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

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

New, B., & Grainger, J. (2011). On letter frequency effects. *Acta Psychologica*, 138(2), 322–328.

Newport, E. L. (1990). Maturational Constraints on Language learning. *Cognitive Science*, 14(1), 11–28. https://doi.org/10.1017/S0272263100009165

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language*, 62(3), 254–271. https://doi.org/10.1016/j.jml.2009.11.002

Oganian, Y., Conrad, M., Aryani, A., Heekeren, H. R., & Spalek, K. (2016). Interplay of bigram frequency and orthographic neighborhood statistics in language membership decision. *Bilingualism*, 19(3), 578–596. https://doi.org/10.1017/S1366728915000292

Ouellette, G., & Senechal, M. (2008a). Pathways to Literacy : A Study of Invented Spelling and Its Role in Learning to Read. *Child Development*, 79(4), 899–913.

Ouellette, G., & Senechal, M. (2008b). window into early literacy: Exploring the cognitive and linguistic underpinnings of invented spelling. *Scientific Studies of Reading*, 12, 195–219. https://doi.org/10.1080/10888430801917XXX

Owsowitz, S. E. (1963). The Effects of Word Familiarity and Letter Structure Familiarity on the Perception of Words. SantaMonica,CA: RandCorporation Publications.

Pacton, & Fayol, M. (2000). The Impact of Phonological Cues on Children's Judgements of Nonwords: The case of Double Letters. *Current Psychology Letters. Behaviour, Brain & Cognition*, (2000/1, 1).

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, 76(2), 324–339. https://doi.org/10.1111/j.1467-8624.2005.00848_a.x

Passler, M. A., Isaac, W., & Hynd, G. W. (1985). Neuropsychological Development of Behavior Attributed to Frontal Lobe Functioning in Children. Developmental Neuropsychology, 1(4), 349–370. https://doi.org/10.1080/87565648509540320

Peereman, R., & Content, A. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, and Computers,* 31(2), 376–379. https://doi.org/10.3758/BF03207735

Perea, M., & Carreiras, M. (2006). Do transposed-letter similarity effects occur at a prelexical phonological level? *Quarterly Journal of Experimental Psychology*, 59(9), 1600–1613. https://doi.org/10.1080/17470210500298880

Perea, M., & Carreiras, M. (2008). Do orthotactics and phonology constrain the transposed-letter effect? *Language and Cognitive Processes*, 23(1), 69–92. https://doi.org/10.1080/01690960701578146

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition*, 31(6), 829–841. Retrieved from https://link.springer.com/content/pdf/10.3758%2FBF03196438.pdf

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*, 38(4), 610–615. https://doi.org/10.3758/BF03193893

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383. https://doi.org/10.1080/10888430701530730

Pitchford, N. J., Ledgeway, T., & Masterson, J. (2008). Effect of orthographic processes on letter position encoding. *Journal of Research in Reading*, 31(1), 97–116. https://doi.org/10.1111/j.1467-9817.2007.00363.x

Pitts, White, & Krashen. (1989). Acquiring Second Language Vocabulary through Reading: A Replication of the Clockwork Orange Study Using Second Language Acquisition. *Reading in a Foreign Language*, 5(2), 271–271.

Plunkett, K. (1997). Theories of early language acquisition. *Trends in Cognitive Sciences*, 1(4), 146–153.

Polk, T. A., Lacey, H. P., Nelson, J. K., Demiralp, E., Newman, L. I., Krauss, D. A., … Farah, M. J. (2009). The development of abstract letter representations for reading: Evidence for the role of context. *Cognitive Neuropsychology*, 26(1), 70–90. https://doi.org/10.1080/02643290802618757

Price, C. J., Wise, R. J. S., & Frackowiak, R. S. J. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex*, 6(1), 62–70. https://doi.org/10.1093/cercor/6.1.62

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S. C., Shakeshaft, C., Seghier, M. L., & Price, C. J. (2013). Europe PMC Funders Group Verbal and nonverbal intelligence changes in the teenage brain, *479*(7371), 113–116. https://doi.org/10.1038/nature10514.Verbal

Rast, R., & Dommergues, J.-Y. (2003). Towards a characterisation of saliency on first exposure to a second language. *EUROSLA Yearbook*, 3(2000), 131–156. https://doi.org/10.1075/eurosla.3.09ras

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled lettres: There is a cost. *Psychological Science*, 17(3), 192–193. https://doi.org/10.1111/j.1467-9280.2006.01684.x

Reber, A. S. (1957). Reber67_Scanned.Pdf. *Journal of Verbal Learning and Verbal Behavior*. Retrieved from http://www.wjh.harvard.edu/~pal/pdfs/replaced_scanned_articles/reber67_scanned.pdf

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology,* 81(2), 275.

Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, 3(5), 513–518. https://doi.org/10.3758/BF03197523

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9(5), 487–494. https://doi.org/10.1016/S0022-5371(70)80091-3

Sagari, T., Nation, P., & Meister, G. (1978). Vocabulary learning and reading. *System*, (6), 70–78.

Samara, A., & Caravolas, M. (2014). Statistical learning of novel graphotactic constraints in children and adults. *Journal of Experimental Child Psychology*, 121(1), 137–155. https://doi.org/10.1016/j.jecp.2013.11.009

Samuels, S. J., LaBerge, D., & Bremer, C. D. (1978). Units of word recognition: Evidence for developmental changes. *Journal of Verbal Learning and Verbal Behavior*, 17(6), 715–720. https://doi.org/10.1016/S0022-5371(78)90433-4

Samuels, S. Jay, & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly*, 13(2), 107–121. https://doi.org/10.1080/1057356970130202

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: *An alternative account of Akhtar, Carpenter, and Tomasello. Child Development,* 69(1), 94–104.

Sauzéon, H., Lestage, P., Raboutet, C., N'Kaoua, B., & Claverie, B. (2004). Verbal fluency output in children aged 7-16 as a function of the production criterion: Qualitative analysis of clustering, switching processes, and semantic network exploitation. *Brain and Language*, 89(1), 192–202. https://doi.org/10.1016/S0093-934X(03)00367-5

Schmid. (2009). On L1 attrition and the linguistic system 1. *EUROSLA*, 9, 212–244.

Schmid, M. S. (2008). Defining language attrition. *Babylonia*, 2, 9–12.

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes*, 19(3), 333–367. https://doi.org/10.1080/01690960344000198

Segalowitz, N. (1991). Does Advanced Skill in a Second Language Reduce Automaticity in the First Language? *Language Learning*, 41(1), 59–83. https://doi.org/10.1111/j.1467-1770.1991.tb00676.x

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access Units or orthographic redundancy? In Attention and Performance XII: *The Psychology of Reading* (pp. 245–264). https://doi.org/10.4324/9781315630427

Seidenberg, M. S., & McClelland, J. L. (1989). The artificial intelligence and psychology project. *Psychological Review*, 96(4), 1–62.

Seliger, H. W., & Vago, R. M. (1991). First language attrition. *Cambridge University Press.*

Shattuck-Hufnagel, S. (1983). Sublexical Units and Suprasegmental Structure in Speech Production Planning. *The Production of Speech*, 109–136. https://doi.org/10.1007/978-1-4613-8202-7_6

Shipley, K. G., & McAfee, J. G. (2015). Assessment in speech-language pathology: A resource manual. *Nelson Education.*

Simos, P. G. (2002). Brain Mechanisms for Reading Words and Pseudowords: an Integrated Approach. *Cerebral Cortex*, 12(3), 297–305. https://doi.org/10.1093/cercor/12.3.297

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism? *Cognition*, 60(1), 143–171. https://doi.org/10.1353/mpq.2004.0034

Snow, C. E., & Kang, J. Y. (2006). Becoming bilingual, biliterate, and bicultural. In N. J. W. & Sons. (Ed.), *Handbook of child psychology, Volume 4, Child psychology in practice* (pp. 75–102). Hoboken.

Snowling, M. J., & Hulme, C. (2012). Annual research review: The nature and classification of reading disorders - A commentary on proposals for DSM-5. Journal of Child Psychology and Psychiatry and *Allied Disciplines*, 53(5), 593–607. https://doi.org/10.1111/j.1469-7610.2011.02495.x

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, 25(2), 293–321. https://doi.org/10.1017/s0142716404001146

Spreen, O., & Strauss, E. (1998). *compendium of neuropsychological tests: administration, norms, and commentary* (2nd ed.). NY: Oxford University Press.

Thomas, M. S. C., & Allport, A. (2000). Language Switching Costs in Bilingual Visual Word Recognition. *Journal of Memory and Language*, 43(1), 44–66. https://doi.org/10.1006/jmla.1999.2700

Torgesen, J. K., & Hudson, R. F. (2006). Reading fluency: Critical issues for struggling readers. *What Research Has to Say about Fluency Instruction.*, 130–158.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology*, 98(3), 642–652. https://doi.org/10.1037/0022-0663.98.3.642

Vaid, J., & Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French-English bilinguals. *Brain and Language*, 82(1), 47–53. https://doi.org/10.1016/S0093-934X(02)00008-1

Van Bon, W. H. J., & Van Der Pijl, J. M. L. (1997). Effects of word length and wordlikeness on pseudoword repetition by poor and normal readers. *Applied Psycholinguistics*, 18(1), 101–114. https://doi.org/10.1017/s0142716400009899

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 Kesteren, R., Dijkstra, T., & de Smedt, K. (2012). Markedness effects in Norwegian–English bilinguals: Task-dependent use of language-specific letters and bigrams. *The Quarterly Journal of Experimental Psychology*, 65(11), 2129–2154.

Van Walraven, C., & Hart, R. G. (2008). Leave'em alone-why continuous variables should be analyzed as such. *Neuroepidemiology*, 30(3), 138.

Velan, H., & Frost, R. (2007). Cambridge University versus Hebrew University: The impact of letter transposition on reading English and Hebrew. *Psychonomic Bulletin and Review*, 14(5), 913–918. https://doi.org/10.3758/BF03194121

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, 11(1), 3–32. https://doi.org/10.1207/s1532799xssr1101_2

Ware, C., Damnee, S., Djabelkhir, L., Cristancho, V., Wu, Y. H., Benovici, J., ... Rigaud, A. S. (2017). Maintaining cognitive functioning in healthy seniors with a technology-based foreign language program: A pilot feasibility study. *Frontiers in Aging Neuroscience*, 9(MAR), 1–10. https://doi.org/10.3389/fnagi.2017.00042

Werker, J. F., & Byers-Heinlein, K. (2008). *Bilingualism in infancy: first steps in perception and comprehension. Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2008.01.008

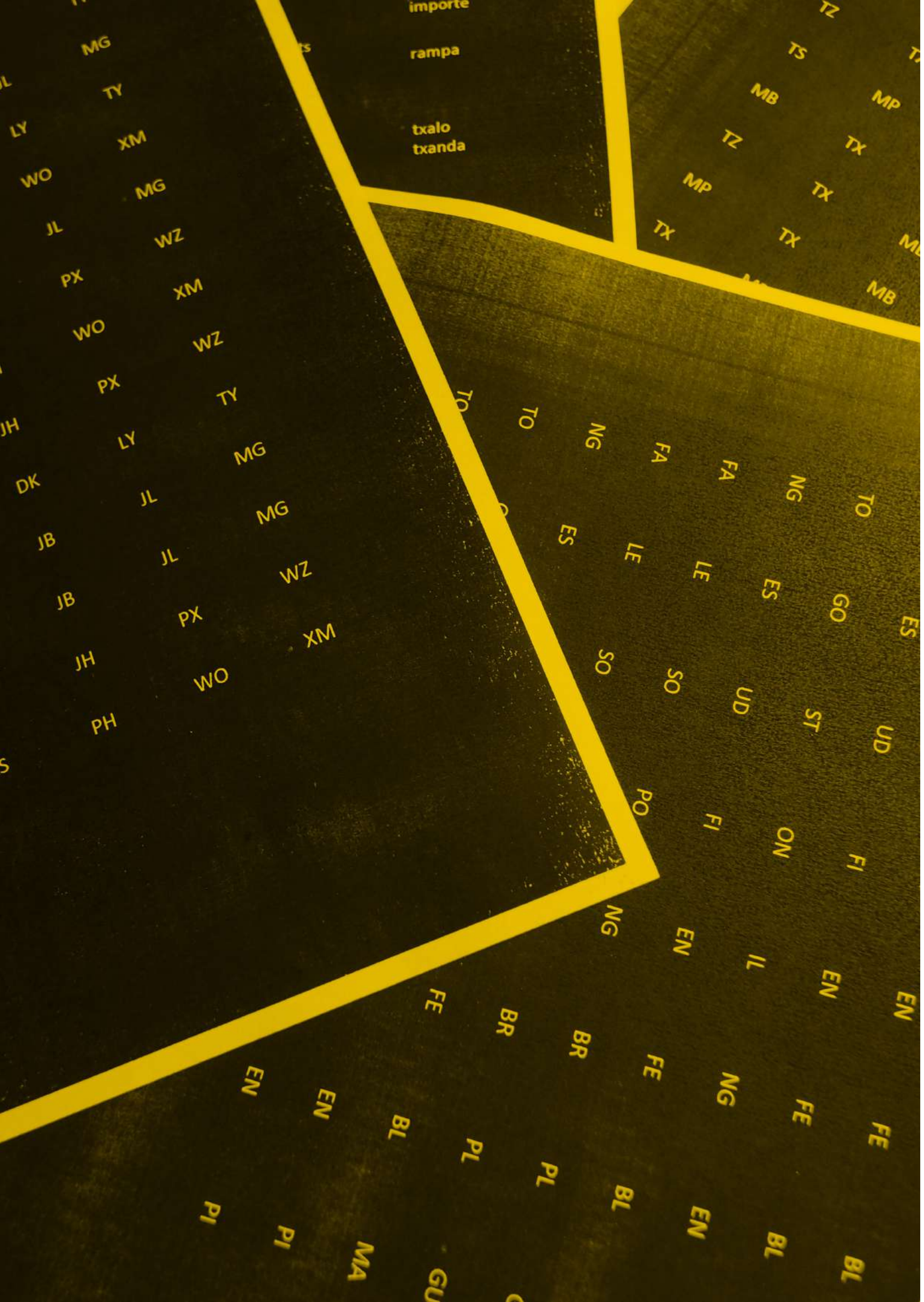
Westbury, C., & Buchanan, L. (2002). The probability of the least likely non-length-controlled bigram affects lexical decision reaction times. *Brain and Language*, 81(1–3), 66–78. https://doi.org/10.1006/brln.2001.2507

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, 1(1), 59–85.

Wolter, J. A., & Apel, K. (2010). Initial acquisition of mental graphemic representations in children with language impairment. *Journal of Speech, Language, and Hearing Research*, 53(1), 179–195. https://doi.org/10.1044/1092-4388(2009/07-0130)

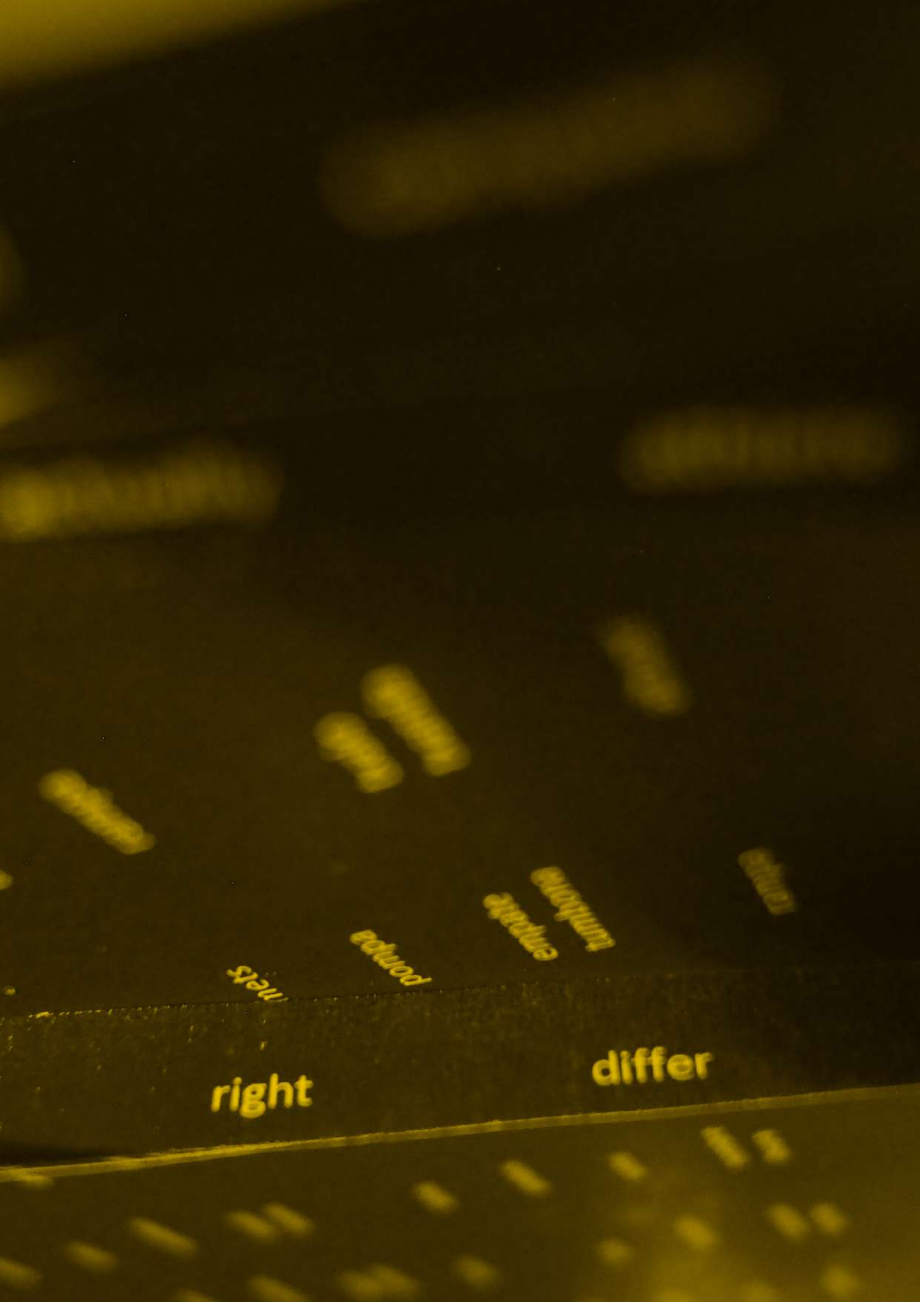
Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology*, 2(SEP), 1–14. https://doi.org/10.3389/fpsyg.2011.00210

References



Resulting publications and manuscripts

- Borragan, M., de Bruin, A., Duñabeitia, J.A. (*in prep.*). Changes throughout age in sensitivity to markedness.
- Borragan, M., Casaponsa, A., Anton, E., Duñabeitia, J.A. (*under review*). Incidental changes in orthographic processing in the native language as a function of learning new language in life. *Language cognition and neuroscience*. Special Issue.
- Borragan, M., de Bruin, A., Havas, V., de Diego Balaguer, R., Dimitrova-Vulchanova, M. D., Vulchanov, V., & Duñabeitia, J. A. (*under review*). Differences in word learning in children: bilingualism or linguistic experience?



Appendix

Appendix 1.

Words in the language decision task of Experiment 1

BASQUE		SPANISH	
MARKED	UNMARKED	MARKED	UNMARKED
hauts (powder)	hodei (cloud)	tumba (tomb)	bruma (mist)
lotsa (shame)	ipuin (story)	bombo (drum)	plazo (time limit)
amets (dreams)	samur (tender)	rampa (ramp)	feliz (happy)
bitxi (jewel)	mutil (boy)	rumbo (course)	jaula (cage)
etsai (enemy)	hegal (wings)	bomba (bomb)	baile (dance)
txalo (applause)	afari (dinner)	ambos (both of them)	pelea (fight)
txano (cap)	sagar (apple)	impar (odd)	lunes (Monday)
untxi (rabbit)	ispilu (mirror)	pompa (pomp)	fibra (fiber)
otsail (february)	biloba (grandchild)	embudo (funnel)	abuelo (grandfather)
txistu (whistle)	epaile (judge)	mimbre (wicker)	dureza (hardness)
etxola (cabin)	aldapa (cost)	empate (tie)	pedazo (piece)
altxor (treasure)	amorru (rage)	limpio (cleansed)	regazo (lap)
txanda (turn)	abendu (december)	amplio (large)	hierba (grass)
txerto (vaccine)	bidaia (trip)	hambre (hungry)	huerto (orchard)
itxura (shape)	igande (sunday)	sombra (shadow)	espina (thorn)
itsaso (sea)	aginte (power)	nombre (first name)	guante (glove)
atsegin (pleasure)	egungo (current)	importe (amount)	humilde (humble)
txosten (memory)	jelosia (envy)	tumbona (deck chair)	enemigo (enemy)
atseden (break)	hedapen (expansion)	fiambre (cold meet)	rigidez (rigidity)
txantxa (joke)	sumendi (volcano)	siembra (sowing)	deporte (sport)
ahaltsu (powerful)	ostegun (thursday)	ombligo (belly button)	gigante (giant)
jatetxe (restaurant)	gauerdi (midnight)	tumbado (lying down)	semanal (weekly)
mingots (bitter)	langile (employees)	temblor (tremor)	resumen (summary)
etsipen (despair)	iraupen (duration)	asombro (astonishment)	soltero (single)
tximino (monkey)	egongela (room)	sombrero (hat)	usuario (user)
zoritxar (problems)	amildegi (cliff)	empinada (steep)	humildad (humility)

BASQUE		SPANISH	
MARKED	UNMARKED	MARKED	UNMARKED
lainotsu (cloudy)	hiriburu (capital)	membrana (membrane)	paraguas (umbrella)
harritsu (rocky)	laburpen (summary)	temporal (temporary)	frialdad (coldness)
tximista (thunderbolt)	etorbide (avenue)	temprano (early)	detenido (deteined)
udaletxe (town hall)	omenaldi (tribute)	impuesto (tax)	heredero (inheritor)
gutxiegi (insufficient)	osotasun (integrity)	ambiente (ambient)	garganta (throat)
itsasalde (coast)	ibilaldi (walk)	empleado (employee)	plenitud (fullness)
itxaropen (hope)	argibide (instructions)	frambuesa (raspberry)	habilidad (ability)
lotsagabe (insolent)	gorespen (praise)	alumbrado (lighting)	peligroso (dangerous)
berdintsu (similary)	ondorengo (following)	ambulante (walking)	prometida (fiancee)
osasuntsu (healthy)	apaltasun (modesty)	tempestad (storm)	inestable (unstable)
gutxiengo (minority)	adeitasun (amiability)	imparable (unstoppable)	sobremesa (desktop)
tximeleta (butterfly)	lagunarte (company)	semblante (face)	resultado (result)
itsasadar (estuary)	iragarpen (prediction)	temporada (season)	siguiente (after)
igeltsero (builder)	abantaila (advantage)	limpiador (cleaner)	periodismo (journalist)

Appendix 2.

Pseudowords in the language decision task of Experiment 1

MARKED		UNMARKED	
BASQUE	SPANISH		
azots	dambu	sogen	falei
betsa	dempa	gamar	orbia
elets	ampes	ipola	antir
txisu	ompal	igore	aplur
lotsu	ombar	uduli	frola
butxa	ampel	esapi	nidru
netso	lampe	pangu	huiga
txosi	dombe	amapi	daulo
alitxo	tompal	gornen	igontu
betxor	grembo	onduri	filobe

MARKED		UNMARKED	
BASQUE	SPANISH		
bintxa	bempon	redain	jepola
atxela	lambul	pabrai	brafen
txinal	sampas	pigore	sofena
atxona	orambu	godupi	ugorel
txandu	alambo	olupen	repifo
daitsa	lampir	hurmar	oltala
etrutxo	simparu	harmile	ultorio
anditxo	nambrol	blodatu	nabalan
arotxun	empisor	enuarpe	errilta
lamatxa	liamban	lapurel	pugonel
ultatso	lampuso	esmabra	malurus
itsaton	arombio	erniepi	fablora
satxeta	sarampo	neprisu	dulaper
etxisan	dasampo	dafaina	luesmei
mirretxo	gestimbu	modurani	igergien
aistatxo	arrembon	sadutelo	manedari
berpitso	darombas	gegurone	prudarin
hungatso	onampegi	palorego	tomobrai
lutxandi	anambolo	urmatino	rusabrel
tsolasun	segampon	irubines	tagepiri
hastitxa	eresombi	nestagun	ilgarien
emotxeta	saleompo	goruimon	ruraiene
bitetxaba	pomboloti	ontapingo	ruirurpin
balutxeta	arempobes	anirpento	izomupero
lorintsol	pimbredol	surretimo	ramurdeta
aidetsolo	adempairo	errudaimo	darmongez
turontsus	usimbento	saimanede	rolganata
oraletsos	pampitros	birrepudo	femugeniz
eltoritso	ladarombe	badagaiso	gapifolas
putsielas	tampirela	ruromolta	osaumidar

Appendix 3.

Pseudowords in the language decision task of Experiment 2

SPANISH-MARKED	BASQUE-MARKED	UNMARKED
vipca	askaikzio	polal
cierca	meuts	girra
civugo	koitu	gaurro
deconcio	txarke	dindion
fuctor	irreko	baisia
ciel	arruzko	lusidor
incacia	bezkor	pazai
descisco	loku	tangaon
carco	korrako	espapta
cecita	txortik	erbe
cijel	zenkoi	derana
advocio	loke	gurrail
incañid	bruka	irruta
escicia	nenko	azipan
vierche	txuko	isefiraria
fevaon	txanak	henabigion
fevido	nutxa	beisa
esciciz	akomisako	zubo
cizcad	jausko	herta
voña	nitxa	olduri
cascido	baoizko	masida
paciña	txokonatu	pentije
guive	kozte	gerio
mavulchado	betxa	hirruto
guco	eukanekatu	basle
polco	eukuna	bizo
zucido	ultazko	miaza
olcaco	soken	musto
cacira	zuki	aduda

SPANISH-MARKED	BASQUE-MARKED	UNMARKED
asva	hungako	tensato
vacilla	gasko	tapor
guca	zuzka	dahaila
iveon	txenpan	fabrana
rucilla	kokasun	ebal
vacito	uzkinu	uritia
rucor	buzkir	neuratu
pamaño	zuke	gagur
fecino	hudeko	matia
clovena	ezets	sadra
plaña	amiozko	ziso
ompucloto	adezko	tirron
acirchea	iskotu	bemisidal
culcira	korraiko	pren
corcezo	azkostrako	eranjia
cimo	luts	huro

Appendix 4.

Bigrams for the novel word learning task of Experiments 3(3A, 3B) and 4

EMBEDDED CONSONANT BIGRAMS (Consonant-Consonant)							
	Average bigram frequency				Average bigram frequency		
LEGAL BIGRAM	SPANISH	BASQUE	CATALAN	ILLEGAL BIGRAM	SPANISH	BASQUE	CATALAN
BR	0,30	0,08	0,31	BX	0	0	0
BS	0,04	0,01	0,05	DX	0	0	0
DR	0,12	0,06	0,18	FD	0	0	0
FL	0,09	0,03	0,10	FJ	0	0	0
FR	0,14	0,07	0,14	FM	0	0	0
GL	0,04	0,02	0,08	JB	0	0	0

EMBEDDED CONSONANT BIGRAMS (Consonant-Consonant)							
	Average bigram frequency				Average bigram frequency		
LEGAL BIGRAM	SPANISH	BASQUE	CATALAN	ILLEGAL BIGRAM	SPANISH	BASQUE	CATALAN
GM	0,01	0,02	0,02	JD	0	0	0
GN	0,05	0,02	0,07	JL	0	0	0
LB	0,03	0,06	0,03	JM	0	0	0
LF	0,03	0,02	0,03	JN	0	0	0
LP	0,03	0,03	0,03	JS	0	0	0
LT	0,14	0,23	0,16	JT	0	0	0
NJ	0,04	0,01	0,05	MG	0	0	0
NT	1,37	1,20	1,76	MJ	0	0	0
PL	0,20	0,12	0,23	MX	0	0	0
PS	0,03	0,02	0,04	PJ	0	0	0
RB	0,09	0,14	0,12	PX	0	0	0
RD	0,19	0,31	0,19	XB	0	0	0
SF	0,03	0,03	0,04	XR	0	0	0
SM	0,23	0,12	0,26				
SP	0,24	0,18	0,26				
ST	0,97	0,84	1,03				
TR	0,74	0,39	0,75				

NO CRITICAL BIGRAMS (consonant/vowel and vowel/consonant)							
	Average bigram frequency				Average bigram frequency		
LEGAL BIGRAM	SPANISH	BASQUE	CATALAN	LEGAL BIGRAM	SPANISH	BASQUE	CATALAN
AB	0,50	0,53	0,41	LO	0,69	0,43	0,50
AF	0,16	0,10	0,21	ME	0,67	0,49	1,24
AG	0,27	0,46	0,28	MI	0,55	0,36	0,49
AJ	0,17	0,03	0,03	MO	0,66	0,36	0,40

NO CRITICAL BIGRAMS (consonant/vowel and vowel/consonant)							
	Average bigram frequency				Average bigram frequency		
LEGAL BIGRAM	SPANISH	BASQUE	CATALAN	LEGAL BIGRAM	SPANISH	BASQUE	CATALAN
AM	0,58	0,31	0,80	MU	0,14	0,21	0,16
AP	0,30	0,35	0,32	NI	0,56	0,34	0,62
AR	2,54	2,58	2,64	NU	0,11	0,08	0,11
AS	0,62	0,76	0,51	OB	0,20	0,15	0,20
BA	0,44	0,73	0,44	OD	0,16	0,12	0,15
BE	0,21	0,79	0,21	OF	0,09	0,05	0,11
BI	0,27	0,67	0,23	OJ	0,06	0,01	0,01
DA	1,31	0,72	1,09	OL	0,63	0,55	0,85
DI	0,72	0,73	0,68	OM	0,43	0,19	0,47
DO	1,40	0,31	0,64	OP	0,21	0,16	0,20
EB	0,11	0,09	0,11	OX	0,02	0,04	0,03
EF	0,11	0,04	0,14	PI	0,36	0,29	0,35
EG	0,26	0,52	0,37	PO	0,41	0,31	0,42
EJ	0,12	0,02	0,13	RA	2,04	2,24	2,12
EL	0,54	0,46	0,58	RE	1,44	1,04	1,74
EM	0,38	0,18	0,46	RI	1,42	1,66	1,44
EP	0,19	0,09	0,2	RO	1,13	0,75	0,88
ER	1,83	2,14	1,8	RU	0,22	0,35	0,22
ES	1,40	0,84	1,63	SA	0,78	0,58	1,01
ET	0,53	1,05	0,72	SE	0,51	0,31	0,48
EX	0,19	0,05	0,21	TE	1,37	1,10	0,9
FE	0,2	0,11	0,27	TO	1,01	0,51	0,6
GA	0,52	0,95	0,56	TU	0,32	1,71	0,31
GO	0,29	0,49	0,19	UB	0,12	0,07	0,11
IB	0,19	0,29	0,17	UD	0,17	0,11	0,16
ID	0,74	0,45	0,42	UG	0,06	0,09	0,08
IF	0,15	0,06	0,19	UJ	0,03	0,01	0,01
IJ	0,05	0,02	0,03	UM	0,18	0,12	0,18

NO CRITICAL BIGRAMS (consonant/vowel and vowel/consonant)							
	Average bigram frequency				Average bigram frequency		
LEGAL BIGRAM	SPANISH	BASQUE	CATALAN	LEGAL BIGRAM	SPANISH	BASQUE	CATALAN
IL	0,65	0,82	0,44	UN	0,25	0,84	0,25
IM	0,41	0,19	0,49	US	0,24	0,35	0,25
IN	1,31	1,43	1,25	UX	0,01	0,02	0,01
JA	0,20	0,21	0,23	XA	0,02	0,17	0,21
JE	0,15	0,08	0,03	XI	0,06	0,21	0,14
JO	0,13	0,06	0,07	XO	0,02	0,25	0,07
LA	1,20	1,09	1,38	XU	0,01	0,07	0,03
LE	0,86	0,80	0,89				

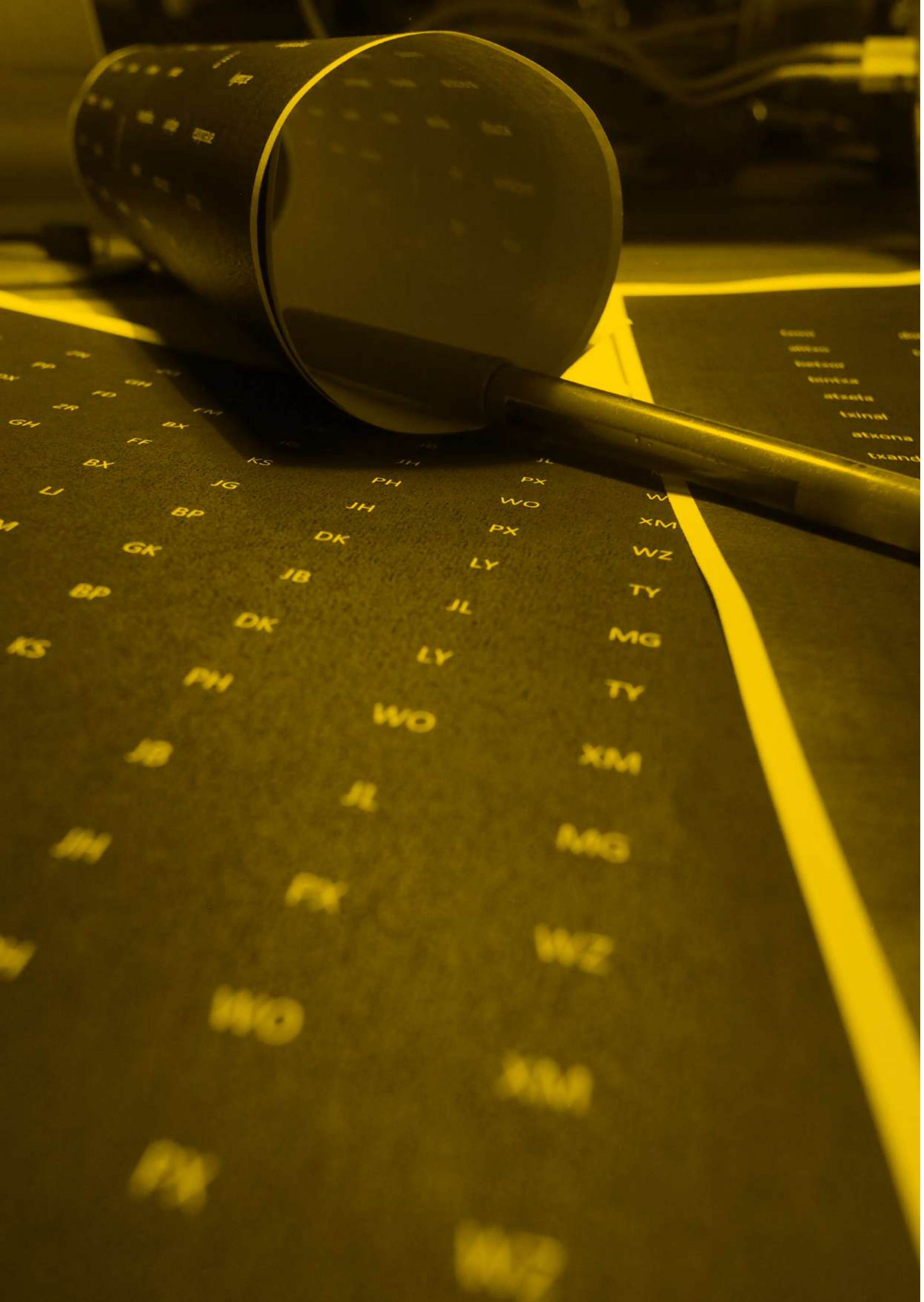
Materials: Hundred and two legal bigrams and nineteen illegal bigrams with their bigram frequency of use (appearance per percentage)

Appendix 5.

Novel words in the novel word learning task of Experiments 3(3A, 3B) and 4

	Average critical bigram frequency				Average critical bigram frequency		
LEGAL PSEUDO	SPANISH	BASQUE	CATALAN	ILLEGAL PSEUDO	SPANISH	BASQUE	CATALAN
AFLEGMO af/leg/mo	0.34	0.31	0.33	AJLEPXO aj/lep/xo	0	0	0
ASPILTO as/pil/to	0.50	0.46	0.39	AFDIJMO af/dig/mo	0	0	0
ABROFLE ab/rof/le	0.49	0.37	0.45	ABXOFJE ab/xof/je	0	0	0
EPSARDO ep/sar/do	0.85	0.65	0.79	EBXAMJO eb/xam/jo	0	0	0
ERBASMU er/bas/mu	0.56	0.68	0.55	EMJAPXU em/jap/xu	0	0	0
ETROBSA et/rob/sa	0.57	0.49	0.60	EXROJDA ex/roj/da	0	0	0
IDRUNJE id/run/je	0.35	0.46	0.33	IBXUJME ib/xuj/me	0	0	0
ILFESPO il/fes/po	0.49	0.38	0.51	IJBEMGO ij/bem/go	0	0	0
INTOPSE in/top/se	0.74	0.61	0.72	IMXOJTE im/xoj/te	0	0	0
ODRAGLE od/rag/le	0.58	0.62	0.62	OMGAPJE om/gap/je	0	0	0
OPLESTU op/les/tu	0.66	0.74	0.72	OXBEJNU ox/bej/nu	0	0	0
OFREGNI of/reg(ni	0.42	0.34	0.51	OJSEFMI oj/sef/mi	0	0	0
USFELPI us/fel/pi	0.23	0.21	0.25	UMJEPXI um/jep/xi	0	0	0
UBRIFLO ub/rif/lo	0.46	0.39	0.44	UXBIJTO ux/bij/to	0	0	0
UGMOLBA ug/mol/ba	0.31	0.30	0.30	UDXOJLA ud/xoj/la	0	0	0

Materials: thirty novel words with their orthographic form and phonotictics below with their average bigram frequency (appearance per percentage)



Resumen en castellano

El lenguaje permite compartir conocimiento y comunicarse con otras personas. Sin embargo, el lenguaje oral tiene el inconveniente de ser demasiado volátil. Por eso el ser humano inventó el lenguaje escrito, para cubrir la necesidad de trasmitir información y que esta perdure en el tiempo. Por lo tanto, el lenguaje escrito hace que se optimice el aprendizaje y las relaciones entre los seres humanos. Los humanos crearon diferentes formas de dejar por escrito el lenguaje oral usando sistemas estandarizados de escritura. Los sistemas de escritura pueden utilizar diferentes símbolos. Así, en un sistema logográfico, un símbolo representa una palabra (*p. ej.* 犬 *es el símbolo de la palabra perro en japonés*). Otra manera de representar las palabras orales consiste en asignar un símbolo (*grafema*) a cada sonido (*fonema*) de la lengua. Un ejemplo de este tipo de sistema es el alfabético, sistema que utilizan todas las lenguas oficiales en España (*español, euskera, catalán, aragonés y gallego*).

Durante la lectura, por el contrario, se produce el proceso opuesto: la extracción de información de las palabras escritas. Ya que la finalidad de la lectura es la comprensión de textos, el reconocimiento visual de las palabras es un estadio muy importante. Para poder acceder al significado de cada palabra escrita en un sistema alfabético se requiere decodificar las letras y asociarlas a un sonido (*decodificación grafema-fonema*). Esta asociación activa la palabra correcta en el lexicón, el almacén mental de palabras. Este proceso de decodificación se aprende a través de una enseñanza guiada que suele comenzar alrededor de los seis años de edad. A medida que el lector se vuelve más competente en la decodificación de letra-sonido y ha accedido a la fonología repetidas veces, las palabras se empiezan a leer globalmente en vez de tener que decodificar cada letra-sonido.

Los modelos centrados en la descripción del reconocimiento visual de palabras intentan describir el proceso de cómo se leen las palabras familiares y no familiares (*p. ej. "the dual route cascade model" [modelo de doble ruta], de Coltheart y colaboradores, 2001, y "the triangle model" [modelo triangular], de Seidenberg y McClelland, 1989; véase en Capítulo 1, sección I.i. Models of visual word recognition [Modelos de reconocimiento visual de palabras]*). Las unidades mínimas (*o subléxicas*) como son la letra y el sonido se procesan de forma diferente para cada tipo de palabra, para poder activar el léxico. Las palabras no familiares necesitan la decodificación de letra-sonido, en cambio las palabras familiares se leen de forma automática e implícita, de una forma más global, ya que no necesitan decodificación letra-sonido. La lectura global de las palabras agiliza el proceso lector, adquiriendo así fluidez en el reconocimiento visual de las palabras. Existe evidencia de que esta fluidez, junto con un buen reconocimiento visual de las palabras, está correlacionada con una buena comprensión de textos (*Adams, 1994; Juel, 1988; Vellutino et al., 2007*).

Sin embargo, si consideramos que muchos niños presentan dificultades en la lectura (*Bishop y Snowling, 2004; Goswami, 2011; Snowling y Hulme, 2012*), la decodificación no debe ser tan sencilla a priori, por lo que estos modelos no recogen toda la complejidad de los procesos lectores. Científicos de diferentes disciplinas como la psicología o la psicolingüística ha centrado en entender las causas que afectan a la fluidez del reconocimiento visual de palabras (*Coltheart et al., 2001; Harm y Seidenberg, 1999; Seidenberg y McClelland, 1989*), ya que es crucial para mejorar la intervención y la enseñanza de la lectura en las aulas. En las últimas décadas, los procesos fonológicos han recibido el foco de atención como procesos clave para el reconocimiento visual de las palabras (*Anthony y Francis, 2005; Torgesen y Hudson, 2006*), aunque estos en solitario no son suficientes para explicar las diferencias individuales. Pero, en la decodificación letra-sonido, no solo están implicados los procesos fonológicos, sino también los ortográficos. Por ello, en los últimos años los procesos ortográficos también han adquirido un papel importante en el estudio del reconocimiento visual de palabras (*Cunningham, Perry, y Stanovich, 2001*), ya que la singularidad de las letras da bastante información al lector.

Algunos modelos de reconocimiento visual de palabras muestran cómo los procesos ortográficos aceleran el acceso al léxico (*p. ej., "the interactive activation model" [modelo de interacción interactiva] , de McClelland y Rumelhart, 1981 y "the dual route model" [modelo de doble ruta] Grainger and Ziegler, 2011; véase en Capitulo 1, sección I.ii. Orthographic processing in visual word recognition [Los procesamientos ortográficos en el reconocimiento visual de palabras]*). El procesamiento de letras y/o de secuencias de letras ocurre en estadios tempranos del reconocimiento de palabras y estas unidades proporcionan bastante información al lector. Las regularidades ortográficas son patrones de secuencias de letras específicos que presentan diferentes frecuencias de uso y/o diferentes posiciones en las palabras en diferentes idiomas. Se ha demostrado que, después de poco tiempo de exposición a las palabras escritas, los individuos adquieren sensibilidad a las regularidades ortográficas, es decir, generan expectativas sobre estas secuencias (*véase Chetail, 2015 para una revisión; Chetail y Content, 2017; Samara y Caravolas, 2014*).

Se ha observado que incluso los prelectores, que aún no saben decodificar letra a sonido, pueden detectar secuencias de letras características en su lengua (*marcaje ortográfico*) simplemente por la exposición ambiental a palabras escritas (*p. ej. la palabra look en inglés tiene doble vocal; Ehri, 1995, 2005*) o a palabras que violan las regularidades ortográficas (*p. ej., la no palabra ffoge; Levy, Gong, Hessels, Evans, y Jared, 2006; Ouellette y Senechal, 2008*). Así, la sensibilidad a las regularidades ortográficas se adquiere muy rápido y mucho antes de saber decodificar letra-sonido. Por ello, la ciencia se ha encargado de estudiar el papel que juega esta sensibilidad para reconocer estas secuencias de letras y, específicamente, para las secuencias de letras características de cada lengua. Se ha demostrado que la sensibilidad al marcaje ortográfico es importante, ya que la percepción de secuencias de letras (*como es la identificación, procesamiento y codificación del orden de las letras*) reduce la competición léxica durante el acceso al léxico (*véase Capitulo 1, sección II.i. The role of orthographic regularities in visual word recognition [El papel de las regularidades ortográficas en el reconocimiento visual de palabras]*).

Resumen en castellano

En el caso de dos lenguas que comparten el mismo sistema alfabético y que coexisten en una mente, como es el caso de ciertos bilingües (*véase Capitulo 1, sección I.iii. Models of bilingual visual word recognition [Modelos de reconocimiento visual de palabras en bilingües]*), se ha visto que el marcaje ortográfico se utiliza también para agilizar la identificación de la lengua (*véase Capitulo 1, sección II.ii. Sensitivity to orthographic markedness in bilinguals [Sensibilidad al marcaje ortográfico en bilingües]). Una primera versión del modelo BIA+ (The Bilingual Interactive Activation Plus Model" [Modelo BIA+: modelo de activación interactiva bilingüe] de Dijkstra y Heuven, 2002*) describió que, en la mente de un bilingüe, el acceso al léxico tiene lugar en un único lexicón para ambas lenguas. Según este modelo, cuando un bilingüe lee una palabra que está formada por una cadena de letras, activa las palabras que contengan esas letras en ambas lenguas por similitud con la cadena de entrada. Luego, la selección de la lengua se lleva a cabo en el nodo del lenguaje léxico, donde la información de la entrada visual es integrada con la información contextual sobre la probabilidad de que la palabra visual provenga de una lengua en particular. Las representaciones de la otra lengua se inhiben en el nodo del lenguaje léxico. El nodo del lenguaje léxico es considerando como permeable, pero solo el procesamiento de la lengua nativa puede influir en la segunda lengua. Sin embargo esta primera versión del modelo no tiene en cuanto la información subléxica en el reconocimiento visual. Por eso, Van Kesteren y sus colegas (*2012*) propusieron una extensión del modelo, el modelo extendido BIA+ [BIA+extended model], demostrando una ventaja en la decisión de la lengua para las palabras que contienen marcaje ortográfico en uno de los idiomas, ya que reduce la competencia léxica con la otra lengua. Propusieron un enlace directo entre la información subléxica y la lengua, agregando un nodo del lenguaje subléxico. Este nodo del lenguaje subléxico ha sido corroborado por más autores (*p.ej. Casaponsa et al., 2014; Oganian et al., 2016*).

La mayoría de estudios que investigan cómo los lectores bilingües reconocen palabras a través del marcaje ortográfico utilizan tareas de decisión de lenguas donde los participantes leen palabras en dos idiomas y deben decidir a qué idioma pertenece. Estas palabras están marcadas o no marcadas. Para estas dos condiciones se escogen palabras dependiendo de la frecuencia de uso de combinaciones de dos letras consecutivas o bigramas. Los bigramas son considerados una unidad subléxica, ya que el conjunto de dos letras consecutivas es el mayor número de letras que un individuo puede procesar de una vez (*p. ej. "the Cambridge text" [el texto de Cambridge] de Velan y Frost, 2007 y "the transposed letter task" [tarea de* *trasposición de letras] de Perea y Carreiras, 2008*) y son la unidad mínima familiar de pronunciación (*p. ej. sílabas; véase capítulo 1, sección II. Orthographic regularities [Regularidades ortográficas]*). Así, definimos las palabras marcadas como aquellas que incluyen al menos un bigrama que no existe en la otra lengua (*la frecuencia de uso de este bigrama en la otra lengua es 0; p. ej., la palabra txakurra, que significa perro en euskera, es una palabra marcada en euskera porque tx no existe en castellano*). Por otro lado, las palabras no marcadas son aquellas cuyos bigramas existen en ambas lenguas (*p. ej., la palabra ardi, que significa oveja en euskera, y la palabra cerdo, ambas comparten todos los bigramas como rd que tiene una frecuencia de uso en porcentaje de 0,24 en castellano y de 0,30 en euskera*).

Estudios centrados en este tema han demostrado que los bilingües utilizan el marcaje ortográfico para acelerar la decisión consciente de la palabra (*o pseudopalabra*) a su lengua (*Oganian et al., 2016; Vaid y Frenck-Mestre, 2002; Van Kesteren et al., 2012*). Esto se ha demostrado debido a que los bilingües responden comparativamente más rápido y con mayor precisión cuando las palabras están marcadas. Incluso se ha demostrado que los monolingües que no conocen una de las lenguas en las que están las palabras de la tarea son capaces de decidir que las palabras marcadas no pertenecen a su lengua de forma rápida y correcta. Esto demuestra que las secuencias de letras marcadas ayudan a activar la lengua de la palabra, reduciendo la competición léxica innecesaria en el otro idioma (*Casaponsa et al., 2014; Oganian et al., 2016; Van Kesteren et al., 2012*). Así, se activan solo palabras con esa secuencia de letras en la determinada lengua, lo que permite que se agilice la identificación de la lengua. De esta manera, la decisión puede hacerse antes incluso de acceder al significado de la palabra, simplemente con información subléxica como es el caso de las secuencias de letras marcadas. Este hallazgo sugiere que hay un nodo del lenguaje ortográfico que agiliza la decisión de la lengua sin tener que acceder al nodo del lenguaje léxico.

Como vemos, el marcaje ortográfico ayuda a los bilingües en el reconocimiento visual de palabras; sin embargo, este no es el caso en el aprendizaje de palabras. Estudios previos han demostrado que los bilingües aprenden mejor palabras nuevas de forma intencionada cuando estas tienen secuencias de letras no marcadas o, en otras palabras, secuencias de letras que ya conocen (*Ellis, 2002; Ellis y Beaton, 1993*). Sin embargo, cuando el aprendizaje de nuevas palabras es accidental como es en el caso de la lectura de textos, los participantes prefieren aprender pseudopalabras con secuencias de letras diferentes a las conocidas porque destacan más en el texto que las pseudopalabras con secuencias de letras conocidas (*Lutjeharms; 1994) (véase capítulo 1, sección II.iii. The of orthographic regularities in novel words learning in bilinguals [El papel de las regularidades ortográficas en el aprendizaje de palabras nuevas en bilingües]*).

Por todo lo anterior, resaltamos la importancia del marcaje ortográfico para reducir la competición léxica, agilizar la identificación de lengua en bilingües y aprender palabras nuevas. Teniendo en cuenta esto, la finalidad de esta tesis es comprender mejor la función de las regularidades ortográficas en el sistema de procesamiento de palabras a lo largo de la vida de los individuos. Específicamente, comprender mejor la función de la sensibilidad al marcaje ortográfico en el reconocimiento visual de palabras y en el aprendizaje de palabras nuevas a través del desarrollo. Saber más sobre el papel que juega la sensibilidad a las secuencias específicas de letras nos ayudará a entender mejor su impacto en el reconocimiento visual de palabras, tema de interés para el campo de la psicolingüística y el bilingüismo. Por ello, proponemos dos enfoques: por una parte examinaremos la función de la sensibilidad al marcaje ortográfico en el reconocimiento visual de las palabras y, por otro lado, estudiaremos la función de la sensibilidad al marcaje ortográfico en el aprendizaje de palabras nuevas. Ambas perspectivas se aplicarán en varios grupos de edad para examinar los cambios durante el desarrollo de las personas. Por lo tanto, con el primer enfoque de investigación examinaremos cómo la sensibilidad al marcaje ortográfico cambia con el desarrollo y, en el segundo, observaremos cómo el procesamiento de las regularidades ortográficas produce un impacto en el aprendizaje de nuevas palabras.

En una primera aproximación (*véase capítulo 2. Changes in sensitivity to orthographic markedness [Cambios en la sensibilidad al marcaje ortográfico]*) exploramos si la sensibilidad al marcaje ortográfico es estable a lo largo de la vida o está sujeta a cambios en el sistema de procesamiento de palabras. El modelo BIA+ (*Dijkstra y van Heuven, 2002*) describe que la primera lengua es estable a través del desarrollo y es la segunda lengua la que está moldeada por la primera. En el caso de que la sensibilidad al marcaje ortográfico estuviera sujeta a cambios, hipotetizamos dos posibles escenarios: por un lado, que el cambio de la sensibilidad podría estar condicionado por el desarrollo normal de un bilingüe y, por otro, que podría estar condicionado por aprender una segunda lengua después de tener una primera lengua establecida en el sistema durante muchos años. Para investigar estos dos posibles escenarios, se realizan dos experimentos: el Experimento 1 se centra en investigar el posible cambio de la sensibilidad al marcaje a lo largo del desarrollo normal de bilingües que aprenden la segunda lengua a una edad temprana, y el Experimento 2 se centra en investigar el posible cambio de la sensibilidad al marcaje como consecuencia de aprender una segunda lengua. En ambos experimentos se utiliza una tarea de identificación de lenguas con palabras marcadas y no marcadas, siguiendo la literatura previa (*Casaponsa et al., 2014*), donde los participantes tienen que decidir si las palabras están en castellano o en euskera.

El Experimento 1 (*ver Tabla 1*) se centra en el primer escenario, y para ello se realiza la tarea de decisión de lenguas con palabras y pseudopalabras. Cabe señalar que se añaden las pseudopalabras para ver cómo afecta la sensibilidad al marcaje ortográfico cuando no hay acceso al significado. Esta tarea se realiza en cuatro grupos con edades diferentes (*niños entre 8-9 años, niños mayores entre 12-13, adolescentes entre 16-17 años y adultos mayores de 18 años*). Los participantes son bilingües simultáneos de castellano-euskera. El Experimento 2 (*ver Tabla 1*) investiga el segundo escenario. Para ello, un grupo de jubilados (*>65 años*) monolingües de castellano, residentes del País Vasco, participaron en este proyecto donde aprendieron euskera como segunda lengua. Este estudio longitudinal es una colaboración con el Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco, en donde los participantes reciben clases gratis de euskera a cambio de participar en diferentes estudios entre los que se encuentra el presente. Los participantes reciben clases formales por un periodo de nueve meses y antes de iniciar el curso, al acabar el curso y después de un año de finalizar el curso, realizan una tarea de decisión de lenguas esta vez solo con pseudopalabras.

En el segundo enfoque (*véase capítulo 3. Novel Word learning and orthographic markedness [Aprendizaje de palabras nuevas y marcaje ortográfico]*) exploramos cómo la sensibilidad al marcaje interactúa con el aprendizaje de nuevas palabras en diferentes grupos de edad. Queremos observar cómo la sensibilidad al marcaje en las lenguas conocidas por un bilingüe impacta en el almacenamiento de nuevas secuencias de letras. Para ello, investigamos cómo se aprenden palabras nuevas que contienen secuencias de letras que no existen en ninguna de las lenguas conocidas por los bilingües (*ilegales; p. ej., abxijmo, que contiene los bigramas bx y jm, no existentes en castellano, euskera ni catalán*) y de palabras nuevas que contengan secuencias de letras conocidas (*legales; p. ej., abrofle, br y fl existen en los tres idiomas mencionados anteriormente*). En este proceso consideramos también el rol del bilingüismo e hipotetizamos que el aprendizaje de estos dos tipos de palabras podría estar influenciado por ser bilingüe en sí o por la experiencia que los bilingües tienen al manejar las diferencias entre sus lenguas. Por esta razón, realizamos una tarea de aprendizaje de palabras legales e ilegales a niños y a adultos. El Experimento 3 (*ver Tabla 1*) se centra en el estudio de cómo los niños aprenden palabras legales e ilegales. En este estudio participan dos poblaciones de bilingües para responder a la pregunta de cómo afecta la experiencia de manejar las diferencias entre las lenguas. En concreto, participa una población bilingüe cuyas lenguas se parecen y comparten muchas de las regularidades ortográficas, los bilingües castellano-catalán, y otra población cuyas lenguas tienen regularidades ortográficas diferentes, los bilingües castellano-euskera. Además, participa un grupo de monolingües de castellano para comprobar si el aprendizaje de palabras está modulado por el bilingüismo. El Experimento 4 (*ver Tabla 1*) investiga cómo los adultos aprenden palabras nuevas en dos poblaciones: monolingües de castellano y bilingües castellano-euskera. Finalmente, nos planteamos un último experimento a llevar a cabo en caso de encontrar un grupo o grupos que mostraran un patrón de aprendizaje de palabras diferente al del resto. Este experimento fue el Experimento 5 (*ver Tabla 1*) y se añadió para ver si había relación entre la sensibilidad al marcaje ortográfico en las lenguas conocidas por el bilingüe y el aprendizaje de palabras nuevas.

Los resultados del Capítulo 2 nos muestran que la sensibilidad al marcaje cambia a lo largo del desarrollo de los bilingües, en ambos casos tanto a lo largo del desarrollo normal de los bilingües que aprenden la segunda lengua a una edad temprana (*Experimento 1*) como en el caso de los aprendices de una segunda lengua a una edad tardía (*Experimento 2*). Los niños a la edad de doce años muestran cambios en la sensibilidad al marcaje ortográfico en la segunda lengua. Estos resultados son acordes con hallazgos previos que mostraron que los niños a esta edad muestran cambios en el aprendizaje implícito (*Janacsek, Fiser, y Nemeth, 2012*), en la fluidez verbal (*Sauzéon, Lestage, Raboutet, N'Kaoua, y Claverie, 2004*), y en maduración del lóbulo frontal, implicado en el control atencional, la memoria de trabajo, o la flexibilidad mental (*Giedd et al., 1999*). Sin embargo, este primer experimento también muestra que los adultos cambian su forma de procesar el marcaje ortográfico en la lengua nativa, y que manifiestan más sensibilidad. Estos resultados se refuerzan con los resultados del Experimento 2, ya que los adultos que aprenden una segunda lengua a edades tardías también muestran más sensibilidad al marcaje de la lengua nativa. Esto demuestra que la sensibilidad al marcaje no es estable a lo largo de la vida y que, no solo la segunda lengua cambia a consecuencia de la primera, sino que también la primera se ve moldeada por la segunda. Esto apoya la hipótesis de un sistema adaptativo (*Kroll, Bobb, y Hoshino, 2014; Schmid, 2008*), e indicando que el modelo BIA+ extended necesita ser actualizado teniendo en cuenta que la sensibilidad al marcaje ortográfico no es estable a lo largo del tiempo.

Los resultados del Capítulo 3 revelan que la sensibilidad al marcaje ortográfico juega un papel importante en el aprendizaje de palabras nuevas y que esto depende de la edad. Observamos que los niños que manejan regularidades ortográficas diferentes en sus lenguas (*bilingües castellano-euskera*) aprendieron el mismo número de palabras legales e ilegales (*experimento 3A*) y que, sin embargo, los monolingües y los bilingües castellano-catalán aprendieron mejor las palabras legales. Para comprobar que estos resultados fueran producidos por la diferencia entre las lenguas, se replicó el experimento con dos nuevos grupos de bilingües castellano-euskera, uno con una competencia alta de euskera y otro con la misma competencia que el grupo anterior. Los resultados mostraron también que ambos grupos aprendían el mismo número de palabras legales e ilegales. Sin embargo, este efecto no está presente en los adultos, ya que muestran que ambos grupos aprenden mejor las palabras legales que las ilegales, aunque los bilingües aprenden mejor que los monolingües. Cabe mencionar que en este experimento no se incluyó el grupo de bilingües castellano-catalán debido a que en el experimento anterior mostraron un patrón de aprendizaje igual que el de los monolingües.

El efecto de aprender el mismo número de palabras legales e ilegales no debe ser entendida como una ventaja respecto a los otros grupos, ya que todos los grupos aprenden un número medio de palabras parecido (*en torno a 20 palabras*). Lo que cambia es la capacidad de aprender palabras de los dos tipos. Para intentar entender por qué los niños bilingües castellano-euskera aprendieron el mismo número de palabras se realiza el Experimento 5, donde un nuevo grupo de bilingües castellano-euskera realiza la tarea de decisión de lenguas en palabras y pseudopalabras y la tarea de aprendizaje de nuevas palabras. Calculamos la regresión entre el efecto de aprendizaje de palabras nuevas (*legales-ilegales*) y los tiempos de reacción de palabras marcadas y pseudopalabras marcadas. Los resultados muestran que la sensibilidad al marcaje ortográfico en las palabras explica el 18% del efecto del aprendizaje de palabras nuevas. Sin embargo, la sensibilidad al marcaje ortográfico en las pseudopalabras no explica el efecto ya que los resultados no son estadísticamente significativos. Los niños que detectan antes las palabras reales marcadas son los que aprenden mejor las palabras nuevas ilegales. El hecho de que únicamente las palabras marcadas puedan predecir el aprendizaje de palabras nuevas puede ser explicado por los mecanismos automáticos e implícitos que los niños utilizan para aprenden palabras nuevas. Los niños, para poder memorizar estas palabras, necesitan procesar las palabras nuevas como palabras familiares (*Ehri, 1995; Miller y Gildea, 1987*), ya que necesitan una conexión suficientemente fuerte entre el nivel subléxico y léxico para que resista en el tiempo. Este mecanismo, automático e implícito, es muy parecido al que usan los lectores para leer palabras familiares (*ruta léxica; Coltheart et al., 2001; Grainger y Ziegler, 2011*), pero no para leer palabras no familiares (*ruta subléxica; Coltheart et al., 2001; Grainger y Ziegler, 2011*). De ahí que tenga sentido que las palabras reales expliquen el aprendizaje, pero no las pseudopalabras.

Con estos resultados podemos ofrecer dos conclusiones (*véase capítulo 4. General discussion [Discusión general]*). Hemos descubierto que los cambios a la sensibilidad en la segunda lengua se producen principalmente alrededor de los doce años, y los cambios a la sensibilidad en la primera lengua a la edad adulta o como consecuencia de aprender una segunda lengua a una edad tardía. Esto muestra un sistema más adaptativo, ya que no solo la segunda lengua cambia a consecuencia de la primera, sino también la primera cambia a consecuencia de la segunda (*Kroll et al., 2014; Schmid, 2008*). Finalmente, concluimos que los niños bilingües que manejan diferencias entre las regularidades ortográficas de sus lenguas se ven beneficiados a la hora de aprender palabras nuevas, ya que tienden a aprender el mismo número de palabras legales e ilegales. Este hallazgo corrobora la relacion entre una mejor sensibilidad al marcaje ortográfico de sus palabras y un mejor aprendizaje de palabras ilegales.

Estas conclusiones aportan nueva información a los campos de la psicolingüística y el bilingüismo. Futuras investigaciones deberían centrarse en entender mejor los procesos de reconocimiento y aprendizaje de palabras y el papel del marcaje ortográfico, ya que el marcaje da mucha información de forma muy rápida al lector. Comprender mejor estos procesos ayudaría a mejorar las técnicas de aprendizaje lector y de nuevas lenguas en el aula, tanto en lectores normales como en lectores con dificultadores en la lectura.

Tabla 1. Resumen de los experimentos en cada capítulo.

Capítulo	Experimento	Paradigma	Lexicalidad	Marcaje	Lengua	Edad de los participantes	Lengua de los participantes
Capítulo 2. Changes in sensitivity to orthographic markedness	1	Tarea de decisión de lenguas	Palabras Pseudopalabras	Marcadas No marcadas	Castellano Euskera	Niños pequeños Niños mayores Adolescentes Adultos jóvenes	Bilingües Castellano-Euskera
	2	Tarea de decisión de lenguas	Pseudopalabras	Marcadas No marcadas	Castellano Euskera	Adultos retirados	Monolingües Castellano aprendiendo Euskera
Capítulo 3. Learning novel words with illegal orthographic regularities	3A	Tarea de aprendizaje de palabras	Palabras nuevas	Legal Ilegal	Nuevo	Niños mayores	Monolingües Castellano Bilingües Castellano-Catalán Bilingües Castellano-Euskera
	3B	Tarea de aprendizaje de palabras	Palabras nuevas	Legal Ilegal	Nuevo	Niños mayores	Bilingües Castellano-Euskera con alta competencia
	4	Tarea de aprendizaje de palabras	Palabras nuevas	Legal Ilegal	Nuevo	Adultos jóvenes	Bilingües Castellano-Euskera con baja competencia
	5	Tarea de decisión de lenguas	Palabras Pseudopalabras	Marcadas No marcadas	Castellano Euskera	Niños mayores	Bilingües Castellano-Euskera
		Tarea de aprendizaje de palabras	Palabras nuevas	Legal Ilegal	Nuevo		

cnrm cnr mcnr mcn rmcn rmc nrm cnrm cnr mcnr mcn rmcn rmc nrmc

nrm nr mnr mnr mnr mn rmn rm nrm nrm nrm nr mnr mnr mnr mn rmn

rm r mr m rm rm r mr m r mr mr mr m rm rm rm r mr mr m rm r mr

m m mm m mm m mm m mm m mm m mm mm mm m mm mm m mm mm m mm m m

jd jd jdj dj dj djd jd jd jdj djd jd jdj djd jdj djd jdj dj dj

dxt dxtd xtdx tdxt dxtd xtd xtdx tdxt dxt dxt dxt dxt dxtd xtd

xtbq xtbqx tbqx tbqx tbqx tbqx tbqx tbqx tbqxt bqxt bqxtb qxtb

qhlwp qhlwpq hlwpq hlwpq hlwpqh lwpqh lwpqh lwpqh lwpqhl wpqhl

wpzf wpz fwpz fwp zfwp zfw pzf wpzf wpz fwpz fwp zfwp zfw pzfw

pzf pz fpz fpz fpz fp zfp zf pzf pzf pzf pz fpz fpz fpz fp zfp

zf z fz f zf zf z fz fz fz fz f zf zf z f zf z fz fz f zf z fz

f f ff f ff ff f ff ff ff f ff ff f ff f ff f ff f ff f ff f f

sk sk sks ksk sks ks ksk sks ksk sks ks ks ksk sk sk sks ks ks

kvg kvgk vgk vgk vgkv gkv gkv gkv gkvg kvgk vgkv gkvg kvgk vgk

vgcn vgcnv gcnv gcnv gcnvg cnvg cnvg cnvg cnvg cnvg cnvgc nvgc