Semantic richness, semantic context, and language learning

Candice Frances

Supervised by

Prof. Jon Andoni Duñabeitia & Prof. Clara D. Martin

2021

Universidad del País Vasco
Euskal Herriko Unibertsitatea

BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE

Semantic richness, semantic context, and language learning

Doctoral dissertation by

Candice Frances

Thesis submitted for the degree of

Doctor of Philosophy

Supervised by

Prof. Jon Andoni Duñabeitia & Prof. Clara D. Martin

Basque Center on Cognition, Brain and Language (BCBL)

University of the Basque Country (UPV/EHU)

Donostia-San Sebastián, 2021

This research was supported by a Predoctoral Grant (Ayudas para contratos predoctorales para la formación de doctores, ref. Number BES-2016-077169; project SEV-2015-0490-17-3) from the Spanish Ministry of Economy and Competitiviness and the Fondo Social Europeo.

Candice Frances

BCBL Basque Center on Cognition, Brain and Language

Paseo Mikeletegi, 69,

Donsotia-San Sebastián, Spain

2021

ACKNOWLEDGEMENTS

This research was supported by a Predoctoral Grant (Ayudas para contratos predoctorales para la formación de doctores, ref. Number BES-2016-077169; project SEV-2015-0490-17-3) for which I would like to thank the Spanish Ministry of Economy and Competitiveness and the Fondo Social Europeo.

I would like to thank all of my supervisors who have guided me throughout this process—Prof. Jon Andoni Duñabeitia, Prof. Clara D. Martin, Prof. Guillaume Thierry, and Prof. Angela de Bruin—as well as my committee for generously donating their time to assess my work. I should also mention the Speech and Bilingualism group that took me in and provided me with both support and much needed constructive criticism.

I would also like to thank the Admin team—Eider, Vanessa, Maider, and Ana—, you really are the ones that keep everything running smoothly and make sure we stay human throughout this difficult time, as well as the IT group, who keep things literally *running*. Of course, thank you Amets, Itziar, and Marisa as well as the rest of the lab crew for helping facilitate my data collection.

Naturally, I owe a thank you to all of my informal editors—Coco, Mónica, Jose, Sandra, Tere, Polina, Kevin, Euge, and Marc—as well as all of the predocs, whose unyielding support has got me through some very tough times.

Finally, I would like to thank my family—both biological and chosen, some close but most of them far away—for reminding me that there is a life beyond the PhD and that we are all more than our work—which, as silly as it sounds, is easy to forget.

RESUMEN EN CASTELLANO

Cuando aprendemos nueva información o contenido en una lengua extranjera, a la ya existente dificultad de adquirir y comprender nuevo material, se le suma el esfuerzo de llevar a cabo este proceso en un idioma que no dominamos completamente. El estudio del aprendizaje de nuevo material en una lengua no dominante es interesante, ya que ofrece información sobre cómo se estructura, configura y modula una segunda lengua en la mente de los bilingües. Sin embargo, la literatura existente sobre el aprendizaje de contenidos en una lengua extranjera en adultos es bastante limitada y la mayoría de estos estudios se centran en los posibles beneficios para el aprendizaje de idiomas (por ejemplo, Yang, 2014) y no necesariamente los efectos que la lengua extranjera tiene sobre la adquisición de nuevo contenido.

En los limitados estudios existentes comparando el aprendizaje en una lengua nativa y una extranjera, los resultados muestran que en las tareas de recuperación, la utilización de una lengua extranjera va en detrimento del rendimiento, es decir, cuando se lleva a cabo la tarea en una lengua extranjera, el rendimiento es peor que cuando la tarea se hace en la lengua materna; mientras que en las tareas de reconocimiento no se suelen encontrar diferencias entre lenguas (Dirix, Vander Beken, De Bruyne, Brysbaert, & Duyck, 2020; Vander Beken & Brysbaert, 2018; Vander Beken, Woumans, & Brysbaert, 2018). Por un lado, esta discrepancia entre los resultados de una tarea y otra podría estar relacionada con la mayor dificultad que implica la producción en un idioma extranjero como requieren las tareas de recuperación. Por otro lado, el efecto podría relacionarse con los niveles de procesamiento, es decir, que el procesamiento superficial (familiaridad en el reconocimiento) se mantiene intacto entre lenguas, pero el procesamiento más profundo (recuperación) se ve afectado por el idioma empleado en la tarea.

Estas dificultades adicionales, a su vez, se traducen en dificultades para aprender palabras nuevas, cuando para poder hablar o comprender un idioma, es necesario dominar una gran cantidad

de palabras y aprenderlas en un tiempo limitado. Dada esta necesidad, parte del vocabulario se aprende de forma incidental, es decir, el aprendizaje "pasivo" que tiene lugar sin hacer un esfuerzo voluntario y explícito para adquirir la nueva información. Incluso en el idioma nativo, se ha demostrado que una gran proporción del nuevo vocabulario que se aprende como adulto proviene del aprendizaje incidental en la lectura (Nagy, Herman, & Anderson, 1985). El mero hecho de visualizar una palabra un número reducido de veces ya genera un efecto sobre el vocabulario. Sin embargo, es cuando maximizamos el número de exposiciones a dicha palabra que mejoramos el aprendizaje (Hulme, Barsky, & Rodd, 2018). Esta relación entre el número de veces que vemos una palabra y el recuerdo no es lineal y se ve afectada por otros factores, como la distribución de estas exposiciones y el porcentaje del vocabulario del texto que se conoce previamente (Pérez-Serrano, Nogueroles-López, & Duñabeitia, 2021).

Los estudios sobre la adquisición de vocabulario de un idioma extranjero en textos han utilizado principalmente palabras en idiomas que los participantes desconocen, integradas en textos escritos en su idioma nativo (ver Horst & Cobb, 1998; Pitts, White, & Krashen, 1989 para una descripción general de *Clockwork Orange Studies*). Aunque esto brinda control experimental, difiere mucho de cómo normalmente aprendemos un nuevo idioma. De esta manera, existen pocos estudios sobre el aprendizaje incidental de vocabulario en una lengua extranjera que utilicen textos completos en esa lengua, pero los pocos que lo hacen sugieren que la incorporación de nuevos ítems léxicos es más lenta en estos casos que en la materna (Pellicer-Sánchez, 2016; Pellicer-Sánchez & Schmitt, 2010).

Una forma de cuantificar la facilidad con la que se aprende o se recuerda una palabra es mediante el concepto de riqueza semántica. La riqueza semántica engloba una serie de factores que determinan la facilidad con la cual una palabra será aprendida o recordada, dentro de los cuales podemos situar factores intrínsecos (como el número de significados que tiene una palabra o su nivel de ambigüedad léxica) y extrínsecos (como la densidad de su vecindario semántico o su

diversidad contextual) a la palabra (Duñabeitia, Avilés, & Carreiras, 2008; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Yap, Tan, Pexman, & Hargreaves, 2011). Este último grupo depende del contexto y de la relación de esta palabra con otras, lo cual hace que estas variables sean más fáciles de manipular que las cualidades intrínsecas de la palabra, convirtiéndolas en herramientas más accesibles para la enseñanza y para la comprensión de los procesos de aprendizaje. Por esta razón, nos centraremos en este segundo grupo de características y en dos factores en particular que forman parte del mismo: la emocionalidad del contexto semántico que rodea la información a aprender y la diversidad contextual.

En primer lugar, la emocionalidad —es decir, la información relacionada con las emociones y los sentimientos— proporciona dimensionalidad y características experienciales a una palabra de manera similar a las experiencias sensoriales en palabras concretas (Ferré, Ventura, Comesaña, & Fraga, 2015; Vigliocco, Meteyard, Andrews, & Kousta, 2009). Dicho de otro modo, la emocionalidad es una forma de enriquecer las representaciones de las palabras (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011). El beneficio a la hora de aprender palabras emocionales en comparación con el aprendizaje de palabras neutrales está ampliamente demostrada en el dominio monolingüe (por ejemplo, Kuperman, Estes, Brysbaert, & Warriner, 2014). Sin embargo, la literatura sobre este efecto en bilingües no es tan consistente y no está claro si este efecto estaría presente en un idioma extranjero, debido a que los resultados de estudios previos sobre la emocionalidad en la lengua extranjera muestran efectos dispares. Mientras que algunos estudios muestran una falta de efecto de emocionalidad en la lengua extranjera (Anooshian & Hertel, 1994), otros sí lo encuentran (Ayçiçeği-Dinn & Caldwell-Harris, 2009; Ayçiçeği & Harris, 2004; Ferré, Garcia, Fraga, Sanchez-Casas, & Molero, 2010).

En segundo lugar, la diversidad contextual se refiere al número de contextos en los que se encuentra una palabra. Se ha mostrado que al marcar una palabra como relevante en varios contextos, el acceso a la misma se ve facilitado (Adelman, Brown, & Quesada, 2006). Hasta hace

poco, la influencia de la diversidad contextual en el procesamiento y el aprendizaje de palabras no había recibido mucha atención, en parte, debido a su alta correlación con otro concepto que se ha estudiado en mucho más detalle: la frecuencia de palabras (número de veces en las que aparece la palabra en un corpus de textos). Algunos estudios en lengua materna han encontrado que las palabras que aparecen en más contextos se aprenden antes (Hills, Maouene, Riordan, & Smith, 2010) y con más facilidad (Pagán & Nation, 2019). El único estudio que ha abordado el tema de la diversidad contextual en el aprendizaje de vocabulario en una lengua extranjera encontró que presentar palabras nuevas en una mayor diversidad de oraciones mejoraba su posterior reconocimiento (Jones, Johns, & Recchia, 2012).

Muchas de las dificultades en el aprendizaje y uso de lenguas extranjeras pueden conceptualizarse en términos de riqueza semántica. Por ejemplo, la emocionalidad o información afectiva de una palabra contribuye a la riqueza semántica, pero los estudiantes de idiomas extranjeros son menos capaces de aprovecharlos debido a su reducida experiencia con el vocabulario en ese idioma, que también tiende a estar circunscripto a entornos académicos o afectivamente reducidos. De manera similar, la diversidad contextual es una característica que solo se puede aprovechar si el individuo está expuesto a la palabra repetidas veces y en un número de contextos proporcionales a aquellos en los que se usa normalmente, lo cual no suele ser el caso de los estudiantes de idiomas extranjeros.

Una forma de remediar estas limitaciones en el aprendizaje y el uso de lenguas extranjeras es centrarse en aumentar explícitamente la riqueza semántica de las palabras a medida que se aprenden. Por ello, esta tesis explora la cuestión de si manipular la emocionalidad del contexto semántico mejora el aprendizaje de vocabulario y contenido nuevo; y si aumentar el número de textos que contienen las palabras nuevas, mientras mantenemos constante el número de repeticiones, puede mejorar la memoria para esos ítems en una lengua extranjera. En esencia, mi objetivo es probar si es posible hacer extensivos a nuevas palabras los beneficios que tienen

naturalmente las palabras emocionales y contextualmente diversas, mediante la manipulación de las primeras exposiciones a estos nuevos elementos. El beneficio de esta investigación es, por un lado, ver si estos aspectos de la riqueza semántica pueden manipularse experimentalmente y, por el otro, si se pueden utilizar como estrategias para mejorar el aprendizaje.

En la primera publicación (Frances, de Bruin, & Duñabeitia, 2020b), se estudió si la valencia positiva transmitida a través del contexto semántico afectaba el aprendizaje de contenido nuevo. Nuestros resultados mostraron que los participantes recordaban la información aprendida en la lengua materna con mayor precisión que en la lengua extranjera, lo cual concuerda con los resultados presentes en la literatura (Nagy et al., 1985; Pellicer-Sánchez, 2016; Vander Beken & Brysbaert, 2018). Con respecto a nuestra principal variable de interés, los participantes desempeñaron mejor la condición positiva que la neutral en ambos idiomas, apoyando los hallazgos de la literatura previa que muestra que la emocionalidad tiene efectos positivos sobre la memoria (Kousta et al., 2011; Kuperman et al., 2014). Por último, el efecto de la emocionalidad fue de igual tamaño en ambos idiomas, lo que sugiere que la reducción de la emocionalidad en el idioma extranjero reportada en la literatura no redujo los efectos de la misma sobre el aprendizaje.

En la segunda publicación (Frances, de Bruin, & Duñabeitia, 2020a), se exploraron los efectos de la emocionalidad transmitida por el contexto semántico en el aprendizaje de vocabulario asociado a objetos desconocidos (pseudoobjetos) en una lengua extranjera. Se encontró un efecto de idioma en algunas tareas, mientras que en otras no (presente en la tarea de escritura del Experimento 1, pero no en el Experimento 2 y en la tarea de reconocimiento de significado del Experimento 1, pero no del Experimento 2). Es probable que esto se deba a las diferencias en la dificultad de la tarea de aprendizaje en cada caso, ya que en el Experimento 1, los participantes tenían que recordar el doble de elementos que en el Experimento 2. Con respecto a la emocionalidad, los participantes se beneficiaron de la afectividad positiva en ambos experimentos y en ambas tareas. Finalmente, los efectos de la emocionalidad resultaron ser iguales en ambos

idiomas, tanto para la tarea de recuperación como para la tarea de emparejar una palabra con su significado.

En la tercera publicación (Frances, Martin, & Duñabeitia, 2020), se exploraron los efectos de la diversidad contextual, definida como el número de historias en las que se presentó la palabra nueva, sobre el aprendizaje de vocabulario en un idioma extranjero y nativo. No se encontró ninguna diferencia entre idiomas para las tareas de recuperación y reconocimiento, pero sí en la precisión de la tarea de emparejamiento, de la misma manera que en la Publicación 2, lo que sugiere que la lengua puede afectar de diferente manera al recuerdo de la forma y el significado de una palabra. Con respecto a la diversidad contextual, los participantes recuperaron, reconocieron y emparejaron mejor aquellas palabras con mayor diversidad contextual que aquellas con menor diversidad. Tampoco aquí se encontró una interacción entre variables, de manera que la diversidad contextual tuvo un efecto similar y positivo en ambos idiomas.

En general, esta tesis en su conjunto muestra que manipular el contexto semántico en torno a la información por aprender no solo ayuda al aprendizaje, sino que lo hace tanto en la lengua extranjera como en la materna. Los resultados de la Publicación 1 sugieren que aprender contenido en un idioma extranjero es más difícil que en la lengua nativa, pero que la emocionalidad ayuda en este proceso y lo hace por igual en ambos idiomas. De manera similar, la Publicación 2 sugiere que esto también se aplica al aprendizaje de vocabulario. Estos resultados apoyan la idea de que el origen del efecto de lengua extranjera se relaciona con cómo adquirimos cada idioma, ya que no vemos la misma reducción en emocionalidad a la hora de aprender información nueva. La Publicación 3 sugiere que la diversidad contextual ayuda a aprender nuevos elementos de manera similar en ambos idiomas. Es importante destacar que este estudio establece un papel causal de la diversidad contextual, que a menudo se ha extrapolado de estudios correlacionales, pero rara vez se ha manipulado directamente.

Estos resultados sugieren que las dificultades para aprender vocabulario y contenido en un idioma extranjero no son tan generalizadas como podría pensarse intuitivamente. Además, estos estudios muestran que las diferencias en el rendimiento entre idiomas están moduladas por la información que se necesita aprender, tanto por el tipo de información como por la cantidad de la misma.

En resumen, los resultados en su totalidad sugieren que los mecanismos de aprendizaje y procesamiento de información en un idioma nativo y extranjero son más similares de lo que se pensaba inicialmente. Esto también proporciona evidencia de que las diferencias entre el aprendizaje de lenguas extranjeras y nativas son cuantitativas en lugar de cualitativas, posiblemente debido a una cantidad reducida de experiencias y de una dificultad adicional en esta lengua, no de una posición especial de la lengua materna.

En términos prácticos, estos resultados proporcionan herramientas sencillas, como la distribución del vocabulario nuevo en más textos o la inclusión de mayor cantidad de palabras emocionales, que se pueden extrapolar al aula para mejorar el aprendizaje de los estudiantes en una lengua extranjera.

ABSTRACT

As knowing a foreign language becomes a necessity in the modern world, a large portion of the population is faced with the challenge of learning a language in a classroom. This, in turn, presents a unique set of difficulties. Acquiring a language with limited and artificial exposure makes learning new information and vocabulary particularly difficult. The purpose of this thesis is to help us understand how we can compensate—at least partially—for these difficulties by presenting information in a way that aids learning. In particular, I focused on variables that affect semantic richness—meaning the amount and variability of information associated with a word. Some factors that affect semantic richness are intrinsic to the word and others pertain to that word's relationship with other items and information. This latter group depends on the context around the to-be-learned items rather than the words themselves. These variables are easier to manipulate than intrinsic qualities, making them more accessible tools for teaching and understanding learning. I focused on two factors: emotionality of the surrounding semantic context and contextual diversity.

Publication 1 (Frances, de Bruin, et al., 2020b) focused on content learning in a foreign language and whether the emotionality—positive or neutral—of the semantic context surrounding key information aided its learning. This built on prior research that showed a reduction in emotionality in a foreign language. Participants were taught information embedded in either positive or neutral semantic contexts in either their native or foreign language. When they were then tested on these embedded facts, participants' performance decreased in the foreign language. But, more importantly, they remembered better the information from the positive than the neutral semantic contexts.

In Publication 2 (Frances, de Bruin, et al., 2020a), I focused on how emotionality affected vocabulary learning. I taught participants the names of novel items described either in positive or neutral terms in either their native or foreign language. Participants were then asked to recall and

recognize the object's name—when cued with its image. The effects of language varied with the difficulty of the task—appearing in recall but not recognition tasks. Most importantly, learning the words in a positive context improved learning, particularly of the association between the image of the object and its name.

In Publication 3 (Frances, Martin, et al., 2020), I explored the effects of contextual diversity—namely, the number of texts a word appears in—on native and foreign language word learning. Participants read several texts that had novel pseudowords. The total number of encounters with the novel words was held constant, but they appeared in 1, 2, 4, or 8 texts in either their native or foreign language. Increasing contextual diversity—i.e., the number of texts a word appeared in—improved recall and recognition, as well as the ability to match the word with its meaning. Using a foreign language only affected performance when participants had to quickly identify the meaning of the word.

Overall, I found that the tested contextual factors related to semantic richness—i.e., emotionality of the semantic context and contextual diversity—can be manipulated to improve learning in a foreign language. Using positive emotionality not only improved learning in the foreign language, but it did so to the same extent as in the native language. On a theoretical level, this suggests that the reduction in emotionality in a foreign language is not ubiquitous and might relate to the way in which that language as learned.

The third article shows an experimental manipulation of contextual diversity and how this can affect learning of a lexical item, even if the amount of information known about the item is kept constant. As in the case of emotionality, the effects of contextual diversity were also the same between languages. Although deducing words from context is dependent on vocabulary size, this does not seem to hinder the benefits of contextual diversity in the foreign language.

Finally, as a whole, the articles contained in this compendium provide evidence that some aspects of semantic richness can be manipulated contextually to improve learning and memory. In

addition, the effects of these factors seem to be independent of language status—meaning, native or foreign—when learning new content. This suggests that learning in a foreign and a native language is not as different as I initially hypothesized, allowing us to take advantage of native language learning tools in the foreign language, as well.

Table of Contents

General Introduction

CONCEPTUAL FRAMEWORK

As foreign language learning becomes a necessity, it is common for people to sit in a classroom, struggling to incorporate copious amounts of vocabulary in one or two hours of weekly exposure to that language. Also, it has become increasingly common to attend courses or study in a foreign language, which presents its own set of difficulties. The purpose of this work is two-fold: ont the one hand, to study possible ways of compensating for the added effort in foreign language learning and, on the other, to improve our understanding of how this—meaning, foreign language learning—compares to learning in the native language.

On a theoretical level, I aim to understand the mechanisms underlying native and foreign language learning. More specifically, I investigate here how contextual factors can improve or hinder learning through reading and whether these effects vary by language status—meaning, native or foreign. I focus particularly on semantic context, as even in cases in which the content that needs to

be learned is set, one can modify how information is provided or presented in order to optimize learning. This will, in part, help us understand whether one's native language has a special status or whether differences found so far in the foreign language literature relate simply to the increased difficulty of using a foreign language. On a practical level, beyond the scope of this thesis, the ultimate hope is that this research can also be applied, providing practical ways of optimizing learning in a foreign language, and establishing whether the same tools can be used for native and foreign language learning.

Given the immensity of the scope of variables that affect language learning, in general, and foreign language learning, in particular; I have decided to narrow down the focus to variables that affect semantic processing. I will focus on texts—descriptions of objects, countries, and stories—rather than on word lists. This mimics a common way in which people naturally learn vocabulary, namely, incidentally from reading (Nagy et al., 1985). Given that new words are most commonly learned in context, manipulating this semantic context seems like a useful strategy to improve learning, which has not yet been explored in depth. More specifically, I have manipulated emotionality of the semantic context—positive versus neutral—and contextual diversity—from condensing exposure to one text to spreading it out maximally.

We know that semantic richness is reduced when using foreign languages. This is due to the reduction in amount and variability of exposure as well as vocabulary in that language (I elaborate on this in the section titled *Semantic richness*). I will mainly focus on how these two variables—namely, positive emotionality and contextual diversity—can help compensate for the reduced semantic richness in a foreign language. I will start by reviewing the difficulties that characterize foreign language learning through reading, with respect to both content and vocabulary (in the section titled *Difficulties in learning*

in a foreign language). I will then turn to describe and explain the concept of semantic richness as well as why and how it is reduced in the foreign language (in the section titled *Semantic richness*).

Then, I will explore in more depth the concepts of emotionality (*The effects of semantic emotionality on learning*) and contextual diversity (*The effects of contextual diversity on word learning*) and how they relate to foreign language learning and use. Finally, I turn to the main goals (*General and Specific Aims of the Current Work*) and studies (*Summary of Publications*) that comprise this thesis.

DIFFICULTIES IN LEARNING IN A FOREIGN LANGUAGE

Content learning in a foreign language

When it comes to learning new information or content in a foreign language, the difficulty becomes two-fold: acquiring and understanding new material and using a language that one does not have full dominion over. Nevertheless, mobility, study abroad programs, and education in languages other than the local one have become more popular, making it imperative that we understand how foreign language use affects learning. A key example of this is the growing popularity of content and language integrated learning (CLIL). These programs teach foreign languages through content classes (Dalton-Puffer & Nikula, 2014). Although the effectiveness of these programs is not entirely clear, their goal is to have students learn a foreign language by using it, simultaneously learning new content and language skills.

The literature on adult learning in a foreign language is quite limited, and most of the reported benefits are associated with language learning (e.g., Yang, 2014) and not necessarily with content acquisition. Within this literature, studies tend to focus on the efficiency of courses in general. In other words, these studies evaluate overall performance at the end of the course and often show no difference between the control (native language learning) and experimental (foreign language learning) groups on performance (e.g., Hernandez-Nanclares & Jimenez-Munoz, 2015). Although there is quite a bit of education literature of this kind, there are very few studies that examine the immediate understanding and learning of new content in a foreign language as opposed to the long term effects—and results—over the duration of—and at the end of—a course. Those that do report a difference find that instruction in a foreign language is detrimental, particularly without foreign language support, meaning, without added definitions and explanations

of the vocabulary (Roussel, Joulia, Tricot, & Sweller, 2017). These results have been accounted for in the context of cognitive load theory (Sweller, 1988), which suggests a working memory overload for individuals trying to learn content in a language they are not proficient in (Roussel et al., 2017). This literature as a whole paints a less than clear picture but suggests there might be negative effects in the short term, although it is not clear whether they persist in the long run.

As most of the research on content learning in a foreign language focuses on classroom performance, there are very few examples of controlled experiments comparing native and foreign language conditions. Of interest, Vander Beken and Brysbaert (Vander Beken & Brysbaert, 2018) had participants read texts either in their native or foreign language and found that participants' recall benefitted from reading in their native language, whereas recognition did not. They replicated the recognition results in another study using eye-tracking (Dirix et al., 2020). In an extension of this work, Vander Beken, Woumans, and Brysbaert (2018) retested using longer delays, showing a cost in the immediate Recall Task but no effect of language in the Recognition Task. There are two possible causes for this difference. On the one hand, the cost could have been related to the increased difficulty of production in a foreign language required by the Recall Task. On the other hand, the effect could relate to levels of processing, meaning that superficial processing (familiarity in recognition) was intact, but deeper processing (recall) was affected by language.

In sum, it seems that foreign language use might have some adverse short-term effects on learning content, but the current literature is insufficient to say for sure. Furthermore, the results we do have are somewhat inconsistent and, in some cases, can be explained by task demands rather than differences in learning. My first article addresses this gap in the literature, testing content learning while equating task demands between languages by using recognition rather than recall.

Vocabulary learning in a foreign language

A major aspect of language learning is incorporating new vocabulary. In order to speak or understand a language, a large number of words needs to be mastered. Even in the native language,

where we have much more oral exposure and practice than in a non-native language, a large proportion of new vocabulary learned in adulthood comes from incidental learning—meaning learning without making a voluntary effort to acquire the new information—in reading (Nagy et al., 1985). New vocabulary in the native language can be learned from reading with as little as two exposures and, importantly, learning improves with multiple exposures (Hulme et al., 2018). This is the case both for children (Jenkins, Stein, & Wysocki, 1984; Nagy et al., 1985) and adults (Hulme et al., 2018). In other words, even though with few exposures there is already an effect on people's vocabularies, it is essential to maximize experience with a word to improve learning. Nevertheless, this relationship is not linear and is affected by other factors such as how these exposures are distributed and what percentage of the text is understood (Pérez-Serrano et al., 2021).

Studies on the acquisition of foreign language vocabulary from context have mostly used foreign words from a language participants do not speak, embedded in a native language context (see Horst & Cobb, 1998; Pitts, White, & Krashen, 1989 for an overview of Clockwork Orange studies). Although this controls for the knowledge of the context and guarantees that the participants are unfamiliar with the vocabulary, it is quite different from the way in which we normally learn a new language. There are few studies using full texts in a foreign language that show incidental vocabulary learning in a foreign language (Pellicer-Sánchez & Schmitt, 2010) and they suggest that incorporating new lexical items is slower than in the native language (Pellicer-Sánchez, 2016).

Part of the reason for this added difficulty in foreign language vocabulary learning is that there are variables that affect word learning which are especially problematic in a foreign language. Incidental word learning correlates with vocabulary knowledge (Horst & Cobb, 1998). This means that the higher the percentage of words a person knows from a text, the higher the likelihood of acquiring the unknown words present in that text. Vocabulary size also correlates with reading comprehension (Laufer & Ravenhorst-Kalovski, 2010; Şen & Kuleli, 2015; Shen, 2013), which is a

necessary step in order to acquire new vocabulary in reading (Horst & Cobb, 1998). If the text is not understood, the focus moves away from the new vocabulary, making it difficult to deduce the meaning of the word from context. This is a great challenge for foreign language learners as they already have a limited vocabulary that limits comprehension in that language, so that incidental word learning is even more difficult for them.

Nevertheless, perhaps the greatest challenge for learning and remembering words in a foreign language is that they are not quite as "vivid" as they are in the native language (Dewaele, 2004). What I mean by this is that words in a foreign language have an impoverished representation due to reduced experience and exposure to them and the language. For one, this makes vocabulary in that language sometimes behave like low-frequency vocabulary in the native language (Francis & Gutiérrez, 2012). But, more importantly, there is evidence of poorer representations of foreign language words in studies of emotionality (Caldwell-Harris & Ayçiçeği-Dinn, 2009; Harris, 2004; Harris, Ayçiçeği, & Gleason, 2003). For example, emotional words in a foreign language show reduced effects in the emotional Stroop task (Winskel, 2013 but see Sutton, Altarriba, Gianico, & Basnight-Brown, 2007) as well as in physiological measures (Dewaele, 2004). In all of these cases, researchers have found that participants have less detailed representations of what they experience in a foreign language. One way of thinking about this is that representations of vocabulary in the foreign language are less rich than in the native language. In the next section, I will explore this idea in terms of semantic richness, which will be the main focus of this thesis.

SEMANTIC RICHNESS

What is semantic richness?

One way to quantify how easily a word is learned or remembered is through the concept of semantic richness. Semantic richness quantifies the amount and variability of information associated with a word which, in turn, improves the memory for this word. It is composed of several factors that can be divided into those intrinsic to the word or concept, such as the number of senses or lexical ambiguity, and those that relate to the usage of the word, such as the number of features, semantic neighborhood density, or contextual dispersion (Duñabeitia et al., 2008; Pexman et al., 2008; Yap et al., 2011). The first set of features—those intrinsic to the concept and the word—cannot be manipulated without modifying the word's meaning, whereas the latter group—those related to usage—can be exploited to improve learning and access to the word. We also know from studies using repetition priming that words that are more semantically rich improve learning of these relationships (Rabovsky, Sommer, & Rahman, 2012), which suggests an implicit learning advantage for semantically rich words.

Prior experimental studies on semantic richness have focused on enriching learning of items by manipulating the features that are intrinsic to the word. Some studies have manipulated the amount of information about the referent (Gladfelter & Goffman, 2018; E. Lund, Douglas, & Schuele, 2015) or the number of senses of a word (Taler, López Zunini, & Kousaie, 2016). Nevertheless, there are no studies manipulating semantic richness without restricting the meaning or information given about the word. This makes the research so far difficult to apply to real-world situations, as students always need to understand the meaning of new vocabulary as clearly as possible, and this meaning cannot be manipulated. Therefore, semantic richness cannot be modulated to improve learning based on aspects that are intrinsic to the word.

On the other hand, factors related to usage and experience with the word can more easily be manipulated in real-world learning situations. For this reason, I will focus on this second group of features. Of interest to the present work, affective information and contextual dispersion both can easily be manipulated to improve richness. Affective information—meaning, information related to emotions—provides dimensionality and experiential information to a word, much like concreteness does (Ferré et al., 2015; Vigliocco et al., 2009). Contextual dispersion—also referred to as contextual diversity (Adelman et al., 2006)—refers to the number of contexts in which a word is encountered. This can be defined as the number of content areas in which the word appears (Pexman et al., 2008) or, simply, the number of texts in which it appears in a corpus (Adelman et al., 2006). Higher contextual dispersion has been associated with faster naming and lexical decision times (Adelman et al., 2006) as well as spoken word recognition times (Johns, Gruenenfelder, Pisoni, & Jones, 2012) in the native language.

In the following sections, I will first address how foreign language use can affect the semantic richness and then how we can use emotionality and contextual diversity to counteract these negative effects.

How does "foreignness" affect semantic richness?

If we think about foreign language learning and use, we see that many of the difficulties can be conceptualized in terms of semantic richness. For example, the emotionality or affective information of a word contributes to semantic richness and is a stable quality of the item. Nevertheless, foreign language learners are less able to take advantage of these due to their reduced experience with vocabulary in that language, which also tends to be in academic or affectively reduced settings. Similarly, contextual diversity is a characteristic of the word itself but can only be taken advantage of if the individual is exposed to the word repeatedly and in a number of contexts proportional to those the word is normally used in, which is often not the case for foreign language learners. Usually, foreign language learners are exposed to this language in a very

limited number of contexts—mostly educational and/or work environment—and they are exposed to a limited number of texts and works.

How can we compensate for this reduction in semantic richness?

One way to remediate these limitations in foreign language learning and usage is to put the focus on artificially increasing the richness of words as they are being learned. In other words, we need to find specific strategies that increase a word's affective information and contextual diversity and apply these to foreign language learning. The benefit of this is two-fold: on the one hand, we will see if these usage aspects of semantic richness—from the point of view of the learner—can be manipulated and, on the other, we can use these strategies to improve learning.

As a first approach, I will take advantage of the variables I have discussed, which we already know improve semantic richness. Vigliocco, Meteyard, Andrews, and Kousta (2009) posited that affective information is fundamental to semantic representation. Therefore, emotional words can be said to also have higher semantic richness and thus carry with them a memory advantage (Ferré et al., 2015).

Similarly, the diversity of contexts in which one experiences a new word affects the learning of that word (Adelman et al., 2006; Brysbaert & New, 2009). In the case of students, this can be easily manipulated by changing the number of texts in which the item appears.

This thesis explores whether manipulating the semantic context's emotionality and increasing the number of texts containing novel words, while keeping repetitions constant, can improve memory for foreign language words. In essence, I aim to test whether one can imbue new words with the same facilitative effects that emotional and contextually diverse words naturally have by manipulating the first exposures the learner has to these new items.

Before presenting the thesis research, I will review current research on the effects of emotionality and diversity in word learning.

THE EFFECTS OF SEMANTIC EMOTIONALITY ON LEARNING

When speaking about emotionality, we can refer to words being emotional—i.e., whether the meaning and use of a word are emotional—or the visual or semantic context surrounding the word being emotional. I will first describe literature on emotional words since it represents the majority of the current work in this area, and then I will turn to emotional context, which is the focus of my work.

There are two basic aspects of emotionality: valence and arousal (Russell, 1980). Valence refers to whether something is positive (high valence) or negative (low valence), and arousal refers to how strong of an emotion that element evokes. For example, a word can have high valence and high arousal (e.g., winner) or high valence and low arousal (e.g., relaxation). Similarly, a word can be of neutral valence and high arousal (e.g., lightning) or neutral valence and low arousal (e.g., typewriter). The emotional valence of a word—meaning whether a word is positive or negative as opposed to neutral—facilitates encoding and retrieval, particularly in lexical decision tasks. In monolinguals, the advantage for emotional words relative to neutral words has been widely tested and quite consistently found (e.g., Dewhurst & Parry, 2000; Doerksen & Shimamura, 2001; Kuperman et al., 2014). Notably, the advantage of emotional words primarily correlates with valence and not arousal (Adelman & Estes, 2013; Brainerd, Stein, Silveira, Rohenkohl, & Reyna, 2008) and is more consistent for positive words (Kousta et al., 2011), which suggests a valence-specific mechanism.

Interestingly, the advantages of emotionality are not limited to emotional words but rather extend to neutral items in emotional contexts. Studies on the effects of emotional contexts on neutral words find that words presented in positive contexts are learned better (Erk et al., 2003; Erk, Martin, & Walter, 2005). Furthermore, these studies suggest that emotional contexts have such a

strong effect that the brain's activation for a neutral stimulus can be modulated by the emotionality of the stimuli surrounding it (Erk et al., 2005).

Emotionality as a form of semantic richness

Emotionality increases semantic richness by providing affective and experiential dimensions associated with the word as well as the number of features associated with that item. Emotionality interacts with other variables the same way words with dense semantic neighborhoods do (Duñabeitia et al., 2008; Syssau & Laxén, 2012) and perhaps can be perceived in words pre-attentively (Nielsen, Shapiro, & Mason, 2010).

Firstly, emotionality can provide additional semantic features to a word. Semantic features refer to characteristics or attributes that make up the meaning of a word (Pexman, Lupker, & Hino, 2002). The higher the number of features a word has, the faster the lexical decision times, naming latencies (Pexman et al., 2002), and semantic processing (Pexman, Holyk, & Monfils, 2003). In essence, emotionality provides additional semantic information that helps provide access to the word.

Finally, and perhaps more importantly, emotionality has been presented and conceived of as a way of enriching representations (Kousta et al., 2011). Kousta et al. (2011) suggest that the advantage observed for emotional words stems from a particular type of experiential information, specifically as being internally embodied through affective states—meaning, feeling the emotion. This presents the benefits of emotionality in parallel to those of concreteness, which are due to the experience through the senses—for instance, the experience of seeing or touching an object helps one remember it (Macedonia, 2014). In other words, one can feel an emotion, much like one can touch an object, and this "experience" helps create a richer representation for the word. This increases the number of features of a word and thus increasing semantic richness. Importantly, emotionality increases the semantic richness by adding experiential information to the word, even if this experience is internal (Vigliocco et al., 2009).

Although the effects and mechanisms of emotionality in lexical and semantic processing are not fully understood, it is clear that it provides additional and essential information (Vigliocco et al., 2009) and is one of the primary predictors of memory for words (Rubin & Friendly, 1986).

Effects of foreign language on emotionality

It is important to remember that not all of the effects in the monolingual literature apply to bilinguals and that some variables are affected by foreign language use. The effects of foreign language use on emotionality have been extensively researched and quite a complex image has emerged. Bilinguals self-report a reduction in emotionality when they use their foreign or second language, with a plethora of literature reporting this phenomenon in vastly varied areas from moral decision-making (see Costa, Vives, & Corey, 2017 for a review) and lying (Caldwell-Harris & Ayçiçeği-Dinn, 2009) to reacting to reprimands (Ayçiçeği & Harris, 2004) and jokes (Ayçiçeği-Dinn, Ayşe, Şişman-Bal, & Caldwell-Harris, 2018; Erdodi & Lajiness-O'neill, 2012). This, in turn, influences what effects we can expect from emotionality on learning. As a first step, let us summarize what we know of the effects of foreign language use on emotionality.

The origin of this effect is not clear, but there are several working hypotheses. There is evidence that both proficiency and age of acquisition are relevant factors, but neither is sufficient to explain the effect (Harris, Gleason, & Ayçiçeği, 2006). Furthermore, immersion (Dewaele, 2010a) and frequency of use (Degner, Doycheva, & Wentura, 2012) also modulate this effect. Some authors claim that it is increased cognitive load—or, in other words, a processing cost—is the cause (Hayakawa, Tannenbaum, Costa, Corey, & Keysar, 2017), suggesting that it is this depletion of resources that does not allow the cognitive system to incorporate emotional information. Support from this theory comes from studies that show that high proficiency reduces this effect (Harris et al., 2006). However, this hypothesis does not fully explain the aforementioned effects of immersion (Dewaele, 2010a) and use (Degner et al., 2012).

More importantly for the present work, others have claimed that the effect relates to experience and how we learn the language (Harris et al., 2006). More specifically, they point to the emotional neutrality of a classroom in comparison to learning a language at home or socially. In other words, the theory predicts that if the language is most frequently used in low arousal situations, the learner or user will experience that language as less emotional (Caldwell-Harris, 2014). This is in line with the idea that we incorporate vocabulary by accumulating experiences with that word, which adds up to creating its mental representation (Gleitman et al., 2005). If this is the case, the richness of the representation of a word can be affected by manipulating the experiences and contexts in which that word is encountered. In other words, if words are taught in an emotionally rich context, this reduction in emotionality in a foreign language might no longer apply.

Emotionality effects in foreign language learning

As I have discussed already, there is significant monolingual literature on the effects of emotionality in word processing and learning. Nevertheless, the literature on this effect in bilinguals is not as clear-cut, and it is not clear whether this effect would be present in a foreign language because of reduced emotionality in that language (as I discussed above).

Studies on the effects of emotionality on memory and processing in a foreign language have shown mixed results. Anooshian and Hertel (1994) found no effect of emotionality on memory in the second or foreign language, Ayçiçeği and Harris (Ayçiçeği & Harris, 2004) and Ayçiçeği-Dinn & Caldwell-Harris (2009) found a processing advantage for positive and taboo words as well as childhood reprimands. Ferré, García, Fraga, Sánchez-Casas, and Molero (2010) found emotionality effects in both languages and to the same extent. In a later study, Ferré, Anglada-Tort, and Guasch (2018) showed an interaction between language dominance and emotionality effects when participants were tested in a later classroom-learned language. They found that positive words had an advantage over neutral and negative words, with a higher emotionality effect for concrete versus abstract words—in contrast to Ferré et al. (2015) and Kousta et al. (2011).

The studies presented so far focus on testing words that were already known to participants. Within the foreign language literature, studies on the effects of emotionality that focus on new word learning are very few and quite distinct. For the purpose of this thesis, there are two major groups: those that manipulate the emotionality of the novel words and those that manipulate the emotional context of words.

Concerning emotional content (or the emotionality of the novel words), Altarriba and Basnight-Brown (2011) taught participants words in a novel language and showed an emotionality *disadvantage*. Ferré, Ventura, Comesaña, and Fraga (2015) also studied the effects of emotionality on word learning for the first words in a novel language. They found that the emotionality of their native language word aided learning for abstract words, but not concrete ones. They interpreted this as emotionality providing richness to abstract terms in much the same way that sensory representations do to concrete terms—in accordance with Kousta et al., (2011). In these cases, words were directly paired but it is less clear what the effects would be if participants did not have a counterpart for the word in their first language, as is the case when learning new content.

To my knowledge, aside from the present work, there is only one study on the effects of emotional context on foreign language learning. Brase and Mani (2017) taught participants new words in a native or a foreign language through the use of videos with definitions. They defined emotional context as the level of arousal or effusiveness that the actor showed in the video and crossed this with the emotionality of the novel words. They showed emotionality effects for new negative words in an emotional Stroop and a sentence completion task, but these differed between languages. In the foreign language condition, the emotionality effect was only present in the emotional context, whereas, in the native language, it occurred in both emotional and neutral contexts. It is worth noting that their recall results are in contrast with their Stroop task results as well as with prior studies: no emotionality effect in the native language and a restricted effect in the foreign language—only for negative words in the emotional context. Nevertheless, this study is

particularly relevant, as almost no other studies have examined how emotionality affects the learning of new words for new concepts rather than new labels for known objects or concepts.

The scarce literature, on the one hand, and the complex and contradictory results, on the other, point to a need to study the effects of emotionality on novel word learning in a foreign language further. Furthermore, these conflicting and inconsistent results may be explained by alternative accounts to reduced emotionality, such as the novelty effects suggested by Ayçiçeği and Harris (Ayçiçeği & Harris, 2004). Finally, the paradigms used thus far predominantly focus on emotionally-charged words in isolation rather than in context and are thus limited to single-word presentations, a gap I will address with the current work.

THE EFFECTS OF CONTEXTUAL DIVERSITY ON WORD LEARNING

One of the key variables in semantic richness is the number of different contexts a word is encountered in—in other words, contextual diversity. Contextual diversity generally refers to the number of texts—or contexts—in which a word can occur. This is commonly operationalized as the number of texts in which a word appears within a corpus (either the number of documents as defined by Adelman et al., 2006; or the number of films or TV series as defined by Brysbaert & New, 2009). If we apply this theoretical definition to the human experience, it is the total number of texts in which a person has seen or heard a word. In most cases, this is impossible to quantify and thus the count in a corpus is used as a proxy for this, assuming a proportional amount of exposure. Nevertheless, if we are talking about new word learning, this experience can easily be manipulated by changing the way in which words are presented—namely, by varying the number of texts a word appears in when the person is learning a new word. An important application of this is that it is less costly to distribute repetitions more (or differently) than it is to increase these repetitions, which makes contextual diversity an attractive tool for improving and economizing efforts in word learning, particularly in a foreign language.

Until recently, the influence of contextual diversity on word processing and learning had not received much attention, in part, due to its high correlation with another concept that has been studied in much more detail: word frequency. Word frequency refers to the number of times a word has been encountered or, in practical terms, how many times the word appears in a corpus. The most common definition of frequency is the number of instances per million words in a database. In these terms, it correlates positively with accuracy and negatively with response time in lexical decision tasks (e.g., Grainger, 1990). One of the issues with this measure is that it is unstable when the corpuses they are taken from are small and also are better predictors depending on the sources

of the corpora—namely, written works versus subtitles (Brysbaert & New, 2009). When it comes to word learning, the definition is in terms of the number of exposures. The number of times a person sees a new wordpositively correlates with memory performance, even in incidental learning and in both short- and long-term retention (Hulme et al., 2018). In other words, this verifies the intuition that the more one sees a word, regardless of whether one is making an effort or not, the better one will remember it.

Naturally, word frequency is highly correlated with contextual diversity—the more texts a word appears in, the higher the number of times it appears in total. As contextual diversity started receiving increased attention, there are studies that have found effects of it above and beyond those of frequency, even suggesting that it explains the frequency effect (Adelman & Brown, 2008). The effects of word frequency and contextual diversity have been found to have additive (Steyvers & Malmberg, 2003) or even opposite (Parmentier et al., 2017, but see Guitard, Miller, Neath, & Roodenrys, 2019 for a failure to replicate).

In general, contextual diversity has been found to influence word learning, word recognition—in adults (Adelman et al., 2006) and in children (Perea, Soares, & Comesaña, 2013)—, and serial recall performance (Parmentier et al., 2017) in the native language. More specifically, there are cases in which remembering the specific encounter with the word benefits performance—such as serial recall or recognition for recently presented word lists. In these cases, high contextual diversity seems to be detrimental, whereas the effects of frequency are inconsistent—sometimes beneficial (Parmentier et al., 2017) and sometimes detrimental (Steyvers & Malmberg, 2003). Importantly, when contextual diversity was controlled for, the effects of word frequency disappeared, whereas the converse did not happen (Adelman et al., 2006; Chen et al., 2017; Hsiao & Nation, 2018; Johns et al., 2012; Perea et al., 2013; Plummer, Perea, & Rayner, 2014). This highlights its relevance and need for further study. Although these data have the substantial limitation of using

naturally occurring frequency and diversity—as measured by word databases—these studies highlight that these two terms are not interchangeable and refer to qualitatively different concepts.

Studies using corpus-based data found that contextual diversity was a significant predictor of response times for word reading and lexical decisions (Adelman et al., 2006; Brysbaert & New, 2009; Hsiao & Nation, 2018) as well as for spoken word identification (Johns et al., 2012). These effects can also be found online using eye-tracking measures (Chen et al., 2017; Plummer et al., 2014), suggesting clear and early contextual diversity benefits. These effects have been found in a variety of languages such as Catalan (Boada, Guasch, Haro, Demestre, & Ferré, 2020), Chinese (Cai & Brysbaert, 2010; Sze, Rickard Liow, & Yap, 2014), Portuguese (Soares et al., 2015), Dutch (Keuleers, Brysbaert, & New, 2010), Greek (Dimitropoulou, Duñabeitia, Avilés, Corral, & Carreiras, 2010), and English (Brysbaert & New, 2009; van Heuven, Mandera, Keuleers, & Brysbaert, 2014). This suggests a universal effect. Furthermore, the effects of contextual diversity seem to be more resilient to the particulars of the database that is used than those of frequency, suggesting that it is a more robust measure (Brysbaert & New, 2009).

Contextual diversity in native language word learning

So far, I have discussed correlational effects, but within the word learning literature, there have also been causal effects of contextual diversity. First language studies looking at both correlational and causal analyses of contextual diversity—in terms of word elicitation—found that words that appear in more contexts are learned earlier (Hills et al., 2010). Johns, Dye, and Jones (2016) manipulated contextual diversity in a word learning study and found that when novel items in the native language are contained in texts on a variety of topics, they are then recognized faster and more accurately. Similarly, Rosa, Tapia, and Perea (2017) found that 3rd grade children benefitted from increased contextual diversity when learning new words. In this case, the words were presented in the same number of texts, but these were of different types—namely Spanish language, Natural Sciences, and Mathematics. In particular, they tested recall, recognition (in two

tasks), and picture matching. They showed a consistent benefit for high versus low contextual diversity.

A recent study by Pagán and Nation (2019) manipulated contextual diversity experimentally. Participants learned novel words (low-frequency unknown words) incidentally (reading for comprehension) that were presented either in repetitions of the same sentence or in different sentences. Words presented in repeated sentences were in the low diversity condition, and those in different sentences were in the high diversity condition. They found that diversity increased reading times during the learning phase and decreased them during the testing phase, suggesting a processing advantage during testing for words learned in diverse contexts.

Contextual diversity and foreign language

As in the native language, contextual diversity is thought to play a role in the foreign language, although the number of studies evaluating these effects is much more reduced in the foreign than the native language. In particular, it is a significant predictor of proficiency (Berger, Crossley, & Kyle, 2017; Monteiro, Crossley, & Kyle, 2020) and early foreign language word production, particularly for verbs (Crossley & Salsbury, 2010; Crossley, Subtirelu, & Salsbury, 2013) in a foreign language.

Only one study has approached the subject of contextual diversity in vocabulary learning in a foreign language. Using an artificial language learning paradigm, Jones et al., (2012) found that presenting novel words in different sentences affected the speed of a later pseudolexical decision task—meaning, a speeded task in which participants have to identify whether strings they are presented with are “words” from the learning set. In other words, participants identified the new words better when they had seen them used in different sentences rather than seeing the same sentence repeatedly. One issue with this study is the limited number of stimuli used as well as the limited vocabulary in the artificial language (12 words) and—in the case of low contextual diversity—

a repetition effect as participants were likely not rereading the sentences given that they were precisely the same and repeated 45 times.

To my knowledge, there are no other studies looking at this effect in the foreign language, and, given the importance of this factor in the native language, it is crucial to know its effects in foreign language learning.

GENERAL AND SPECIFIC AIMS OF THE CURRENT WORK

In the above introduction and literature review, I identified several research lines referring to various factors that affect learning. Accordingly, my main goal for this thesis, in general terms, is to investigate the relationship between semantic richness and foreign language learning. This aim has both theoretical as well as practical implications as it aims to both understand the mechanisms related to learning and identify possible ways of improving or optimizing it.

Within that broad intention of understanding how semantic richness affects foreign language learning, I have several specific aims. In my review of previous literature, I found the content's emotionality to be an influential factor both for memory and learning. I found a lack of literature on how using context to increase semantic richness—particularly by manipulating the emotionality of that context—affects learning, in general, and foreign language, in particular. This relates to my first and second specific aims, which are to investigate the relationship between emotional semantic context and content learning—meaning new information on a given topic or area—and vocabulary learning in a foreign language. In my review of the literature, I also found that contextual diversity is a highly influential factor in memory that has seldom been explored in learning. Furthermore, no studies characterize these effects in foreign language learning. This leads to my third specific aim, to investigate the relationship between contextual diversity and vocabulary learning in a foreign language.

Summary of Publications

PUBLICATION 1: THE INFLUENCE OF EMOTIONAL AND FOREIGN LANGUAGE CONTEXT IN CONTENT LEARNING

Overview of Methodology

For Publication 1, I carried out an experiment in which participants listened to the description of imaginary countries. The descriptions were approximately 1300 words long and participants listened to the descriptions in either their native language—Spanish—or their foreign language—English. Embedded in those descriptions were 50 items of information (e.g., national sport and population—see Supplementary Materials on page 86) that participants were later tested on. My main manipulation was changing the emotionality of the semantic context around the key information. To do so, I manipulated the surrounding filler sentences by including more positive

emotional words—in the case of the emotional condition—or more neutral words—as in the case of the neutral condition (see Table 4. of Publication 1 on page 90 for an example sentence from each condition).

I tested 76 native Spanish speakers (38 in each language group, 9 male, M_{age} 5 = 33.86, SD_{age} = 5 9.14; see Table 1 in Publication 1 on page 69 for a full list of demographics between groups). The descriptions were presented auditorily to diminish the fatigue of reading—particularly in the foreign language. Participants first listened to the two descriptions—one positive and one neutral as language was manipulated between participants—and then answered 50 self-paced multiple-choice questions about the countries (see Figure 2 from Publication 1 on page 90 for an example question). Then, participants completed a measure of English vocabulary—LexTALE (Lemhöfer & Broersma, 2012).

Hypotheses

I explored the effect of emotional semantic context in content learning in a foreign language—my first aim—in Publication 1. This experiment tested whether the inclusion of a positive semantic context—as opposed to neutral—improved learning of new information on a particular topic in a foreign language. I also compared the effects of a positive semantic context on content learning in a foreign and native language. As reviewed in earlier sections of this work, prior studies have been quite inconsistent in their results on the effects of foreign language use. Therefore, I tentatively expected an overall decrease in performance in the foreign language, regardless of context. I also expected a positive effect of emotionality, regardless of language. This relates to the quite consistent effects of positive emotionality improving memory. Finally, I expected a reduced effect of emotionality in the foreign language than in the native language. Although studies looking at this directly have not produced consistent results, based on the emotionality literature, I expected a reduction in the emotional response in the foreign language, leading to smaller emotionality effects. In short, these are my hypotheses:

- I expected a main effect of language, such that participants would have greater difficulty on the recognition test in their foreign language than in their native one.
- I expected a main effect of emotionality, such that participants would have greater difficulty on the recognition test in a neutral context than in a positive one.
- I expected an interaction between emotionality and language, such that there would be a smaller effect of emotionality on the recognition test in the foreign language than in the native language.

Results

In the first publication, I explored whether positive valence, as conveyed by semantic context, would affect content learning. I expected, based on prior literature (Nagy et al., 1985; Pellicer-Sánchez, 2016; Vander Beken & Brysbaert, 2018), that participants would have greater difficulties in their foreign than their native language. Our results showed exactly that: participants in the native language condition remembered the items from the learning phase more accurately than those who performed the task in the foreign language condition (see Table 3 on page 72 for accuracy data and Figure 1 of Publication 1 on page 73 for a graphical representation of the accuracy results). There is little literature comparing content learning between languages, and the results are quite heterogeneous. Nevertheless, tasks that require more language knowledge or that require making associations between items tend to cause more difficulty in the foreign language (Nott & Lambert, 1968). My task required that participants remember and associate to 50 items of information acquired within a continuous speech stream to a particular country. This made the task difficult and required that the participants not mix the information that corresponded to each country.

With respect to our main variable of interest, I expected participants in general to perform better in the positive than the neutral condition. This hypothesis was supported by prior literature that found that emotionality and, in particular, positive valence has positive effects on performance

(Kousta et al., 2011; Kuperman et al., 2014). That is exactly what I found, with participants performing better in the positive than the neutral condition in both languages (see Figure 1 of Publication 1 on page 73 for the summary data and the supplementary material figure on page 89).

Finally, I expected a smaller—or no—effects of emotionality on the recognition test in the foreign language compared to the native language. This was based on prior research that showed reduced emotionality in the foreign language (Ayçiçeği-Dinn, Ayşe et al., 2018; Ayçiçeği & Harris, 2004; Caldwell-Harris & Ayçiçeği-Dinn, 2009; Costa et al., 2017; Erdodi & Lajiness-O'neill, 2012). But, this expectation was not fulfilled. I found the same size effect in both conditions suggesting that the reduced emotionality in the foreign language did not modulate the influence of emotionality. I discuss this further in a later section (General Discussion on page 35).

PUBLICATION 2: THE EFFECTS OF LANGUAGE AND EMOTIONALITY OF STIMULI ON VOCABULARY LEARNING

Overview of Methodology

Publication 2 consisted of two experiments: the first with a between-subjects manipulation of language and the other fully within-subjects. Participants were shown the descriptions of 46 imaginary objects along with their descriptions (on average, 50 words long)—see Figure 1 of Publication 2 on page 98 for an example stimulus. In Experiment 1, each participant read these either in their first language—Spanish—or their foreign language—English. In Experiment 2, participants did 2 sessions one week apart: one in each language with half of the objects (6 items from Experiment 1 were removed so that 20 objects remained in each session). The key manipulation was modifying the emotionality of the semantic context. To modulate this, the description contained either more positive or more neutral words in the emotional and neutral condition, respectively.

In Experiment 1, 56 participants were tested and, in Experiment 2, 60 participants took part (see Tables 1 and 2 of Publication 2 on pages 96 and 97, respectively, for participant variables). Even though participants were not told they would be tested on the objects, after each description, participants were asked to type in the name of the object to make sure they had paid attention to it. After reading all descriptions—positive and neutral randomly mixed—participants were given a non-verbal Distractor Task—a Corsi blocks task (Berch, Krikorian, & Huha, 1998). In Experiment 1, this Distractor Task lasted 15 minutes; in Experiment 2, it lasted 7 minutes each day. After the Distractor Task, participants were asked to do a Recall Task—they were shown the image and had to recall the name—, an Old/New Recognition Task—only in Experiment 1, and it consisted of reporting whether the word they were presented with was one of the object names or not—, a Name Matching Task—

participants were shown the object and had to choose the correct name—, and an Attentional Check in which participants were asked to select the correct characteristic of the object they were presented (see Figures 2 and 4 of Publication 2 on pages 100 and 105, respectively, for example screens of each task).

Hypotheses

I explored the effect of emotional semantic context in vocabulary learning in a foreign language—my second aim—in Publication 2. These experiments tested whether the inclusion of a positive semantic context improves vocabulary learning in a foreign language as well as whether the effects of a positive semantic context on vocabulary learning in a foreign language are the same as in a native language. As in Publication 1, I expected greater difficulty in the foreign language overall than in the native one as well as a positive effects of emotionality, regardless of language, and an interaction between the two. These are my specific hypotheses:

- I expected a main effect of language, such that participants would have greater difficulty on the *recall test* in their foreign language than in their native one.
- I expected a main effect of emotionality, such that participants would have greater difficulty on the *recall test* in a neutral context than in a positive one.
- I expected an interaction between emotionality and language, such that there would be a smaller effect of emotionality on the *recall test* in the foreign language than in the native language.
- I expected a main effect of language, such that participants would have greater difficulty on the *recognition test* in their foreign language than in their native one.
- I expected a main effect of emotionality, such that participants would have greater difficulty on the *recognition test* in a neutral context than in a positive one.

- I expected an interaction between emotionality and language, such that there would be a smaller effect of emotionality on the *recognition test* in the foreign language than in the native language.

Results

In Publication 2, I explored the effects of emotionality—defined as positive valence and high arousal as conveyed by semantic context—on foreign language vocabulary learning for novel objects. I had similar hypotheses as in the first experiment. With respect to language, there was no clear consensus in the literature on its effects on memory for individual words. Nevertheless, I expected an effect in the Recall Task, with participants performing better in the native language because of the increased difficulty in producing the item, but not necessarily for the Recognition Task, which was less taxing (Vander Beken & Brysbaert, 2018). For the first part, I found an effect of language in the Recall Task of Experiment 1, but not of Experiment 2 (see Tables 5 and 8 of Publication 2 on pages 102 and 107, respectively, for summaries of the results). Similarly, I found an effect of language in the Matching Task of Experiment 1 (although only for accuracy), but not of Experiment 2 (see Figures 3 and 5 of Publication 2 on pages 101 and 105, respectively, for graphical representations of the results). The possible reasons for this will be explored in more detail in the next section, but, in short, the difficulty of the learning task might simply have modulated the effect of language—in Experiment 1, participants had to remember twice as many items as in Experiment 2.

With respect to emotionality, I again expected that participants' performance would benefit from positive valence. This is exactly what I found: participants benefitted from positive emotionality in both experiments in both tasks (although, note the effect was marginal in the Recall Task of Experiment 1).

Finally, for the interaction, I expected—as in Publication 1—that participants would show a differential effect of emotionality by language. This was in fact not the case: I found equal effects of

emotionality across languages quite consistently—for both Recall and Matching in both Experiments 1 and 2. This is quite a robust result that suggests that, at the moment of learning, emotionality has an equal effect between languages (more on this in the General Discussion on page 35).

PUBLICATION 3: THE EFFECTS OF CONTEXTUAL DIVERSITY ON INCIDENTAL VOCABULARY LEARNING IN THE NATIVE AND A FOREIGN LANGUAGE

Overview of Methodology

For Publication 3, I explored the effects of contextual diversity on vocabulary learning in a foreign language, my third aim. Stimuli consisted of 100-word stories using 8 high frequency words (hereafter, keywords) that were one of the most representative exemplars of their category group (Uyeda & Mandler, 1980): fruit (apple—manzana), vehicle (car—coche), furniture (table—mesa), animal (dog—perro), dwelling (house—casa), reading material (book—libro), beverage (water—agua), and toy (ball—balón). These words were then replaced with pseudowords to assess learning of the lexical item. Each story contained the keyword eight times (in 1 story), four times (in 2 stories), twice (in 4 stories), or just once (in 8 stories)—see Figure 1 of Publication 3 on page 120 for a schematic of how the stimuli were created. Half of the participants read the stories in their native language—Spanish—and the other half read them in their foreign language—English.

I tested 88 participants (44 in each language group, 25 males, M_{age} = 23.78, SD_{age} = 4.28; see Table 1 of Publication 3 on page 119 for a summary of the variables participants were matched on). Participants read 30 stories (2 keywords in each condition) at their own pace. After each story, they were asked a true/false comprehension question and then moved on to the next story (see Figure 2 of Publication 3 on page 121 for a schematic representation of the procedure). After reading all of the stories, participants completed a non-verbal Distractor Task—Corsi task (Berch et al., 1998)—for 10 minutes. They then completed a Recall Task in which participants were presented with the 8 sentences that originally contained the same key word, now replaced by a gap. They were asked to

type in the missing word that completed all 8 sentences. They were then asked immediately to choose the correct answer among 4 options (correct option, correct option with incorrect spelling, distractor, and distractor with incorrect spelling). Finally, participants were shown an image of an object paired with a pseudoword and had to say whether the name corresponded to the object or not.

Hypotheses

In Publication 3, I explored the third aim: to investigate the relationship between contextual diversity and vocabulary learning in a foreign language. This experiment tested whether increasing contextual diversity aided vocabulary learning in a foreign language. It also compared the effects of contextual diversity in a foreign language to those of the native language. Based on prior literature, I expected that contextual diversity would be beneficial in general. In addition, prior literature suggested that language learners have a reduced vocabulary in their second language which, in turn, makes inferring meaning from context more difficult (Pérez-Serrano et al., 2021). Given the added difficulty in inferring meaning, performance should be worse in the foreign language. Finally, given the added difficulty in incorporating lexical items in the foreign language, I hypothesized that clustering—meaning presenting more than one instance in one text—might be more beneficial than maximal diversity—meaning presenting only one instance of the word in each text—in the foreign language. I have the following hypotheses:

- I expected a main effect of language, such that participants would have greater difficulty on the recall test in their foreign language than in their native one.
- I expected a main effect of contextual diversity, such that participants would have higher rates of recall from greater levels of diversity.
- I expected an interaction between contextual diversity and language, such that the highest recall rates would be at the highest diversity in the native language, but in the foreign language recall would be highest for clusters of 2 or 4 encounters per text.

- I expected a main effect of language, such that participants would have greater difficulty on the recognition test in their foreign language than in their native one.
- I expected a main effect of contextual diversity, such that participants would have better recognition with greater levels of diversity.
- I expected an interaction between contextual diversity and language, such that the best recognition would be at the highest diversity in the native language, but in the foreign language recognition would be best for clusters of 2 or 4 encounters per text.
- I expected a main effect of language, such that participants would have greater difficulty on the matching test in their foreign language than in their native one.
- I expected a main effect of contextual diversity, such that participants would have better matching with greater levels of diversity.
- I expected an interaction between contextual diversity and language, such that the best matching would be at the highest diversity in the native language, but in the foreign language matching would be best for clusters of 2 or 4 encounters per text.

Results

In Publication 3, I explored the effects of contextual diversity—defined as the number of stories a word appeared in—on vocabulary learning in a foreign and native language. With respect to language, as in the case above, the literature did not show a clear consensus. Still, I expected an effect of language because learning was fully implicit in this experiment—unlike Publication 2 where, although participants were not told to specifically memorize the words, they were asked about them immediately afterwards—and the meaning of words had to be deduced from context. Nevertheless, I did not find a difference for Recall and Recognition Tasks (see Figures 3 and 4 of Publication 3 on pages 123 and 124, respectively, for graphical representations of the results and Table 2 on page 122 for a summary of means, standard errors, and confidence intervals). As I discuss later on (see General Discussion section), I believe this might be related to the difficulty of the learning phase and

the amount of to-be-learned information. Interestingly, I did find an effect of language on accuracy for the Matching Task—as I did in Publication 2 (see Figure 5 of Publication 3 on page 124 for a graphical representation of the results). Which suggests that remembering the form and the meaning of a word might carry different language effects.

With respect to contextual diversity, prior literature had shown quite consistently that contextual diversity—even in its various definitions—aided memory and learning. I expected that participants would perform better in the higher diversity conditions—defined as conditions in which pseudowords were presented in more texts—and this is exactly what I found. Participants recalled, recognized, and matched words with higher contextual diversity better than those with lower.

As before, I expected an interaction between language and contextual diversity. Although this effect was rather speculative. This expectation was based on eye-tracking literature showing that the number of exposures to a word has a different relationship between languages. More specifically, some eye tracking studies have shown that in the native language, participants start reading lexical items like known items starting from the second exposure, whereas for the foreign language this requires more exposures (Pellicer-Sánchez, 2016). This would suggest that clustering might provide a better option than fully dispersing encounters for word learning in a foreign language. Nevertheless, this was not the case. I found no interaction between variables suggesting that contextual diversity had an equal, monotonic, and positive effect on both languages.

General Discussion

RESULTS OF THE WORK AS A WHOLE

In four experiments divided into three publications, I explored different ways to improve learning in a foreign language—both of content and vocabulary. I focused on emotionality and contextual diversity and found that by manipulating the semantic context around this information and increasing semantic richness, learning could be improved. These results suggest that, on the one hand, the semantic richness of a word is not a stable feature of it but, rather, can be manipulated at the individual level. In addition, they suggest that manipulating the semantic context of new information can enhance its learning. Finally, on a practical level, this provides tools that could be applied to teaching.

The effects of language were somewhat mixed. I found effects of language in the cases where participants were tested using their vocabulary in that language—i.e., not the new words—as was the case in Publication 1, or when the learning task was more taxing, as was the case in

Experiment 1 of Publication 2 compared to Experiment 2. In the first case, it is quite reasonable that, when prior knowledge is necessary to carry out the task, language experience has an effect on performance. For the second task, it is possible that overloading the memory system during encoding had a greater effect on the foreign language, which is already more cognitively taxing. The difference in difficulty between the two experiments is quite clear in the Recall Task, where participants not only remembered a much lower percentage of words in Experiment 1 of Publication 2 (3% versus almost 10% in Experiment 2), but also a lower absolute number of words (1.39 versus 1.90 in Experiment 2). I also observed a similar pattern in the Matching Task, with an effect of language in Experiment 1 of Publication 2, but no effect in Experiment 2. This could be interpreted in one of two ways: as further evidence that the effect stems from the learning stage—as it affected all subsequent tasks—or as a spillover effect from the Recall Task. The second option is somewhat unlikely, as the difference between languages was approximately twice as large in the Matching Task (6.5%) as in the Recall Task (3.3%). Therefore, this suggests that the effect is likely to stem from the learning phase. In addition, the results of Publication 1 could also be interpreted in this way, as the learning phase required participants to incorporate 50 bits of information as well as sustaining attention for the duration of the recording (approximately 7 minutes each). Finally, Publication 3 showed the same results as Experiment 2 for memory of the word form (Recall and Recognition)—namely, no effect of language. Here, there were only 8 items to be learned. Although the amount of reading was closer to that of Experiment 1 of Publication 2 (approximately 3000 words versus approximately 2300), the amount of to-be-remembered information was much closer to each session in Experiment 2 (8 items versus 20). Importantly, this contrasts with the results of the Matching Task in Publication 3 where I did find an effect of language. It is possible that, since the meaning of these pseudowords could be directly tied to a real word in either language, the Matching Task relied on a direct association between the pseudoword and an object for which they had another word, creating associations more similar to those required in Publication 1. Nevertheless, this is somewhat speculative and the cause of this difference is not fully clear.

With respect to the contextual factors that I manipulated—namely, emotionality and contextual diversity—these showed quite consistent positive effects. In all three experiments I ran on emotionality, I found that positive valence and arousal aided memory and learning. In addition and given the lack of interaction with language, it seems that this is a language independent mechanism that is particularly helpful in the foreign language when new information is acquired—new vocabulary, in this case—independent of word knowledge in the native language. Similarly, contextual diversity showed very consistent results throughout all three tasks. This aligns with the literature word learning in the native language (Adelman et al., 2006; Hulme et al., 2018; Perea et al., 2013). Although this was not a main aim of the experiment, I was able to see that contextual diversity has a non-linear effect, such that there is a sharp effect when going from one to two contexts that then diminishes. This suggests that the effect is not additive, but rather more nuanced.

Given the mixed results on the effects of emotionality on memory in a foreign language, the lack of interaction between factors in Publication 2 was not all that surprising. On the other hand, I was more surprised of the result in Publication 1. These results could have two theoretical interpretations. First, as a whole they suggest that the effects of positive valence on memory may rely on non-emotional mechanisms, such as making the texts more engaging and interesting or making the to-be-remembered information more interrelated (Goh & Hu, 2011; Talmi & Morsovitch, 2004; Talmi, Schimmack, Paterson, & Moscovitch, 2007). On the other hand, the effects might directly relate to emotional resonance but may happen specifically when learning new vocabulary—although this would only apply to Publication 2. If it is true that emotionality can be endowed to words at the moment of learning, it suggests a more interesting result in which we have more control over the semantic richness of words than if it was a stable feature of the word. If at the moment of learning, emotionality has the same effect between languages and is transferred to the to-be-learned item, then this would suggest that words are less arousing in the foreign language, not because of a characteristic inherent to the foreign language and its use, but rather because of the way they are taught. This relates strongly to prior claims that the foreign language effect is due to

learning in a classroom (Harris et al., 2006) and that it is modulated by experience in foreign language environments (Degner et al., 2012; Dewaele, 2010b).

The lack of interaction between contextual diversity and language has some theoretical ramifications. For one, it provides evidence that learning mechanisms do not differ as much between languages. This aligns with the idea that contextual diversity leads to richer representations of words (Adelman et al., 2006). To understand the origin of this effect, it is necessary to provide further studies, particularly some focusing on online processing.

CONTEXTUALIZING RESULTS IN THE LITERATURE

With respect to the effects of language, I mirror and somewhat explain the inconsistency in language effects in learning. On the one hand, I found that when participants are overloaded with information during the learning phase, they show an effect of language. This means that when the number of items that one must remember is too high, the added effort of reading and remembering in a language other than the native affects later performance. This is somewhat in line with the literature that finds that when participants are tested in more difficult tasks—e.g., recall—they show a language effect that disappears when they are tested using an easier task—e.g., recognition (Vander Beken & Brysbaert, 2018). Although I did not find the same asymmetry in our tasks, it is possible that this difference has to do with the instructions given, with participants learning explicitly in the prior cases and implicitly in the current studies.

In addition, when participants had to make associations between new concepts—as in Publication 1—or had to associate a new pseudoword with an image for which they had a readily available label—as in Publication 3—I did observe an effect of language. This is consistent with the literature showing that participants are able to benefit from deeper levels of processing—meaning semantic processing—in their native language but not their foreign one (Francis & Gutiérrez, 2012). Other research has also shown that categorization in learning is less helpful in the foreign language and that participants present reduced category clustering in their foreign language (Nott & Lambert, 1968). Taken together, these results suggest a shallower processing that is less supported by semantic associations.

One of the important theoretical points that this thesis raises relates to the foreign language effect (Costa et al., 2017) and reduced emotionality in a foreign language. The foreign language effect suggests that when people use a foreign language they behave differently than in their native

one because of a difference in emotionality between them. This difference in emotionality is supported by other studies (Caldwell-Harris, 2014, 2014), particularly those that use skin conductance (Caldwell-Harris & Ayçiçeği-Dinn, 2009; Harris et al., 2003). But, there are two important characteristics to keep in mind about these studies: (1) they predominantly use single words and (2) they measure reactions on words that are already known by participants. One of the hypothesis about the origin of the foreign language effect is that this effect is caused by learning the language in the emotionally impoverished environment provided by a language classroom (Harris et al., 2006). In fact, studies that use participants that are or were immersed in the foreign language do not find the same effects (Dewaele, 2010a). The current thesis suggests that, indeed, people are equally affected by the emotional context of the stimuli when learning. Although I did not test this directly, I suspect that emotional reactions to these words and concepts that were learned in emotionally rich contexts will remain more similar to those of the native language.

One of the consequence of this work is that I expand on emotionality and learning literature, in general, and in a foreign language, in particular, by showing ways of manipulating emotionality that lead to improvements in memory. This has theoretical consequences for our understanding of learning in a foreign language concerning which factors affect it and the fact that it is more similar to learning in a native language than perhaps originally thought at the outset of this thesis. Similarly, for contextual diversity, I again find that manipulating contextual diversity affects learning by improving memory. Importantly, I find again that learning in the native and foreign languages is affected in a similar way, showing that the difficulties of using a foreign language and the differences in reading time and vocabulary knowledge do not modulate these effects. This suggests very similar processes between languages.

On a more practical level, I find that simple manipulations of the context around the critical information can help students retain more information and get more out of their classes. Although further testing is necessary to say this with certainty, I do believe that modulating the positive

valence and arousal of the semantic context as well as distributing new vocabulary in several texts can help foreign language students.

This research adds to the question of what factors increase difficulty in foreign language learning. It does not seem to be task difficulty per se (Vander Beken & Brysbaert, 2018), but rather overloading verbal memory or perhaps difficulty during the learning phase. Still, this would require more direct comparisons with the same participants and further manipulations of the stimuli. Part of what mitigates the effects of language here is that the words are matched between languages—meaning that the to-be-remembered pseudowords are the same between languages. Intuitively, it is possible that by using language-dependent words, there might be more of an effect of the foreign language, not so much because of foreign language use, but rather because the items individually are more difficult to remember for participants—bigram frequency, orthotactics violations in the native language, weaker phonological representations (e.g.,Pérez-Serrano et al., 2021).

This also not only adds to the literature on the specific phenomena I was observing, but it also opens up the question of how context in general modulates learning. Furthermore, this opens up a whole new avenue of research into how the different aspects of semantic context affect learning and whether these can be profited from in order to improve learning. In order to fully understand this, further research is needed looking into exactly which contextual variables affect learning as well as how and why this happens.

Although I made an explicit effort to maintain learning implicit, it is worth exploring what would happen during explicit learning. In incidental learning, the participant is simply reading or listening for comprehension. It is easy to imagine that they might be making a greater effort to get a complete picture of whatever it is they are studying, and would probably give more attention and importance to context if they knew they would be tested on the material. On the other hand, in explicit learning, they are likely to hone in on the information that is relevant for later testing and ignore—or at least reduce the attention afforded to—the context. Under these circumstances,

would we observe the same effects? As I mentioned before, there is an added difficulty in understanding foreign language texts. Perhaps the added difficulty might modulate attention to different aspects of the texts, maybe making it less likely for people to filter out the context.

On a related note, I observed that participants compensated for the added difficulty of reading in a foreign language by spending more time reading each text. In the case of Publication 1, participants were not able to do this, as they heard each text read to them only once. There, I found an effect of language. Perhaps our results on the other experiments would have been different if I had created the same limitations on time allotted to studying the texts. In an immediate analysis, one might guess that more language effects would emerge. There are several studies that suggest that reading in a foreign language is both slower (Cop, Keuleers, Drieghe, & Duyck, 2015; Dirix et al., 2020; Whitford & Titone, 2012) and more taxing on working memory—much like performance under dual-task conditions (Sandoval, Gollan, Ferreira, & Salmon, 2010)—supporting this idea. Therefore, reducing the amount of time given to study the texts might also affect attention to context differently when resources are already taxed by foreign language use. This might expose the type of interactions I originally hypothesized would be present in my experiments.

Furthermore, all of the testing I did was on the immediate effects of these variables. It would be interesting—as well as highly theoretically relevant —to observe the long-term effects of these variables. Would these effects survive both time and consolidation processes? In particular, understanding whether the effects of emotionality are maintained equally in both languages would also help us understand the origins of the foreign language effect and the observed—as well as self-perceived—reductions in emotionality in a foreign language.

STRENGTHS AND LIMITATIONS

The current studies, in general, placed a high emphasis on variable control as well as on matching the two language conditions. For example, the same pseudowords—matched for acceptability between languages—were used in both conditions to make the testing tasks perfectly equitable between languages. Part of the success of this effort is evidenced by the fact that I rarely found an effect of language. By using the same pseudowords between conditions, I was also able to successfully and easily compare conditions, making sure that the to-be-remembered items were always equally difficult between conditions.

With respect to more practical applications, the fact that I focused on manipulating context rather than the to-be-remembered items makes this research easily applicable to the classroom, where there is often little control over what content needs to be taught but there is somewhat more freedom as to *how* it is to be taught.

The sematic context manipulation of emotionality I applied is somewhat innovative and provides a slightly different definition of emotionality that can also be easily manipulated—both experimentally and in the classroom. Similarly, I use a simple manipulation of contextual diversity which is quite rare, but can also be applied to classroom settings. What is more important on a theoretical level is that I manipulate contextual diversity in a causal way rather than correlating preexistent and estimated values of contextual diversity on measures of memory. This is not very common in the literature and provides strong evidence for a causal effect of contextual diversity.

Although these studies have many strengths, there are also some limitations that should be kept in mind when interpreting the results. One of the main limitations is that the tasks overall were quite difficult. I needed to use implicit learning and long texts for our purposes, but this also meant that recall performance was quite low overall. To complement these results, it might be useful for future studies to focus on more repetitions in order to improve recall memory.

Another limitation is that these results apply only to incidental learning. This was the focus of the current work as I aimed to understand what happens naturally, without an explicit effort to learn the new vocabulary. But, this means that results could be categorically different if explicit effort was employed. This is also a focus that future studies could take in order to understand better the modulating role of attention and effort.

With respect to the comparison between languages, I was faced with an important decision. Participants, given their increased difficulty with the foreign language, were likely to require extra time to read the texts and reach an equal level of understanding as in the native language. This meant that in order to equate understanding, extra time had to be given to the foreign language condition. This is what I opted for: giving participants free range to pace themselves as needed. The issue with this is that participants had more exposure to the words in the foreign language. Although this did not have an effect between subjects—meaning, there was no correlation between reading time and performance—it is likely that if participants were given the same amount of time for both conditions I would have observed a difference between languages, as I see in Publication 1. This limits—to some extent—our conclusions to self-paced reading, cases in which a similar level of understanding is achieved between languages, and cases matched for number of exposures, but the effect of time of exposure is not clear from our studies.

With respect to Publication 3, the influence of contextual diversity has also been linked to the benefit of spaced over mass practice (Verkoeijen, Rikers, & Schmidt, 2004). In our case, the amount of time between one exposure to the word and the next correlates with its contextual diversity, which relates to the concept of spacing. It is possible that spacing in our study contributed to the effects of contextual diversity. Nevertheless, it is important to point out that the two are conceptually different: whereas spacing aids memory, contextual diversity aids in creating a better mental representation of the item. If word meanings are created through the summation of experiences with a word and the words it co-occurs with (e.g., Bolger, Balass, Landen, & Perfetti,

2008; K. Lund & Burgess, 1996), then contextual diversity would improve these representations whereas spacing would not. This would suggest that one's concept of a word and its meaning would benefit more from encountering it in different contexts than seeing it repeatedly in one text, surrounded by a limited set of words. In addition, our conditions do not fit strict definitions of massed and spaced as words were never repeated consecutively—at most they were in consecutive sentences—and they were never spaced in separate sessions—each participant had only one session. Nevertheless, if spacing between instances would have been held constant, perhaps the effect of contextual diversity would have been smaller or different.

SUMMARY OF CONTRIBUTIONS

This work has made contributions on several levels. From a theoretical point of view, this dissertation provides support for the idea that the foreign language effect has its origin in the way we learn foreign languages, showing that when emotionality is manipulated at the moment of learning, the effects are the same in the foreign as in the native language. Furthermore, I have shown that emotionality affects content learning as well as vocabulary learning. With respect to semantic richness—particularly, contextual diversity—, I have shown that it is not a stable feature of a word, but rather depends on the individual's unique experience with it. This work also helps delineate the detrimental effects of foreign language use, showing that under equal conditions, the foreign language only shows added difficulty in performance when tasks are more demanding and when the amount of to-be-learned information is too great. Finally, and perhaps most importantly, this work shows that learning in a foreign language under these conditions is not qualitatively different from learning in one's native language.

With respect to methodology, this work provides examples of well-controlled experiments using texts, rather than single words or isolated phrases. In addition, these experiments show that it is possible to manipulate semantic richness through the semantic context surrounding the to-be-learned information. More specifically, I manipulate contextual diversity experimentally to show causal effects on learning, as well as manipulating emotionality experimentally through the semantic context as opposed to relying on the natural valence of the items the participants were taught.

In practical terms, I have shown that manipulating the valence and diversity of the context surrounding new vocabulary can improve learning. This, in turn, suggests that something as simple as spreading out exposures to new words into several contexts or embedding vocabulary into descriptions using more emotional terms can be new and useful tools for foreign language teaching.

General Conclusions

In the current work, I set out to explore how various contextual factors related to semantic richness affect learning in a foreign language, as well as in the native language. More specifically, I addressed how positive valence affects content learning—in Publication 1—and vocabulary learning—in Publication 2. In Publication 3, I addressed how contextual diversity affects foreign language learning. The implications of these studies are two-fold. On the one hand, they have theoretical implications for how we understand the influence of these factors on the process of learning and whether these differ according to the language status for the bilingual. On the other hand, they could have practical applications for foreign language classrooms.

The current work as a whole shows that manipulating the semantic context around to-be-learned information not only helps learning, but it does so just as much in the foreign language as in the native language. The results from Publication 1 suggest that learning content in a foreign language is more difficult and that emotionality aids this process and does so equally for both

languages. Similarly, Publication 2 suggests that this also applies to vocabulary learning. These results, as discussed before, also have consequences for our understanding of the origins of the foreign language effect. In Publication 3—as well as in the comparison between experiments in Publication 2—, I found that learning words in a foreign language is not more difficult when the number of items is reduced and these lexical items are matched between languages. This last study also suggests that contextual diversity helps learning new items, and that the effect is the same for both languages. Importantly, this study establishes a causal role of contextual diversity, which has often been extrapolated from correlational studies, but has seldom been manipulated directly.

These results suggest that the difficulties of learning vocabulary and content in a foreign language are not as pervasive as one might intuitively think. Furthermore, these studies show that differences in performance between languages are modulated by the information that needs to be learned. On the one hand, decreasing the number of items that need to be remembered in one session led to equal performance between languages in Experiment 2, when I had found a deficit in foreign language learning in Experiment 1 of Publication 2. On the other hand, I observed that content learning (as in Publication 1 and in the Matching Task of Publication 3) led to language effects whereas simply remembering lexical items did not. These results suggest very specific difficulties for the foreign language that relate more to an overwhelming cognitive load (Costa, Foucart, Arnon, Aparici, & Apesteguia, 2014) and difficulties in associating information in a foreign language (Nott & Lambert, 1968).

All in all, these results suggest that the mechanisms for learning and processing information in a native and foreign language are more similar than I hypothesized in the beginning of this thesis. This also provides evidence that the differences between foreign and native language learning is quantitative, not qualitative, suggesting that it comes from a reduced amount of experience and added difficulty in the language, not a special status of the native language. I learned that simple

manipulations of semantic context—manipulating emotionality and contextual diversity—affect native and foreign language learning in the same way and to the same extent.

In practical terms, these results provide simple tools that can be applied in a classroom to improve students' learning in a foreign language.

References

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, *115*(1), 214–227. https://doi.org/10.1037/0033-295X.115.1.214

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535. https://doi.org/10.1016/j.cognition.2013.08.014

Altarriba, J., & Basnight-Brown, D. M. (2011). The acquisition of concrete, abstract, and emotion words in a second language. *International Journal of Bilingualism*, *16*(4), 446–452. https://doi.org/10.1177/1367006911429511

Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition & Emotion*, *8*(6), 503–514. https://doi.org/10.1080/02699939408408956

Ayçiçeği-Dinn, Ayşe, Şişman-Bal, S., & Caldwell-Harris, C. L. (2018). Are jokes funnier in one's native language? *Humor*, *31*(1), 5–37. https://doi.org/10.1515/humor-2017-0112

Ayçiçeği-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, *12*(3), 291–303. https://doi.org/10.1017/S1366728909990125

Ayçiçeği, A., & Harris, C. L. (2004). Bilinguals' recall and recognition of emotion words. *Cognition and Emotion*, *18*(7), 977–987. https://doi.org/10.1080/02699930341000301

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and

theoretical considerations. *Brain and Cognition*, *38*(3), 317–338. https://doi.org/10.1006/brcg.1998.1039

Berger, C. M., Crossley, S. A., & Kyle, K. (2017). Using novel word context measures to predict human ratings of lexical proficiency. *Educational Technology & Society*, *20*(2), 201–212. Retrieved from http://www.jstor.org/stable/90002175

Boada, R., Guasch, M., Haro, J., Demestre, J., & Ferré, P. (2020). SUBTLEX-CAT: Subtitle word frequencies and contextual diversity for Catalan. *Behavior Research Methods*, *52*(1), 360–375. https://doi.org/10.3758/s13428-019-01233-1

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, *45*(2), 122–159. https://doi.org/10.1080/01638530701792826

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., & Reyna, V. F. (2008). How does negative emotion cause false memories? *Psychological Science*, *19*(9), 919–925. https://doi.org/10.1111/j.1467-9280.2008.02177.x

Brase, J., & Mani, N. (2017). Effects of learning context on the acquisition and processing of emotional words in bilinguals. *Emotion*, *17*(4), 628–639. https://doi.org/10.1037/emo0000263

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0010729

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, *5*(1055), 1–4. https://doi.org/10.3389/fpsyg.2014.01055

Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, *71*, 193–204. https://doi.org/10.1016/j.ijpsycho.2008.09.006

Chen, Q., Huang, X., Bai, L., Xu, X., Yang, Y., & Tanenhaus, M. K. (2017). The effect of contextual diversity on eye movements in Chinese sentence reading. *Psychonomic Bulletin and Review*, *24*(2), 510–518. https://doi.org/10.3758/s13423-016-1119-1

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin and Review*, *22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition*, *130*(2), 236–254. https://doi.org/10.1016/j.cognition.2013.11.010

Costa, A., Vives, M.-L., & Corey, J. D. (2017). On language processing shaping decision making. *Current Directions in Psychological Science*, *26*(2), 146–151. https://doi.org/10.1177/0963721416680263

Crossley, S., & Salsbury, T. (2010). Using lexical indices to predict produced and not produced words in second language learners. *The Mental Lexicon*, *5*(1), 115–147. https://doi.org/10.1075/ml.5.1.05cro

Crossley, S., Subtirelu, N., & Salsbury, T. (2013). Frequency effects or context effects in second language word learning. *Studies in Second Language Acquisition*, *35*(4), 727–755. https://doi.org/10.1017/S0272263113000375

Dalton-Puffer, C., & Nikula, T. (2014). Content and language integrated learning. *Language Learning Journal*, *42*(2), 117–122. https://doi.org/10.1080/09571736.2014.891370

Degner, J., Doycheva, C., & Wentura, D. (2012). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition*, *15*(1), 181–189. https://doi.org/10.1017/S1366728911000095

Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development*, *25*(2–3), 204–222. https://doi.org/10.1080/01434630408666529

Dewaele, J.-M. (2010a). *Emotions in Multiple Languages*. London: Palgrave Macmillan. https://doi.org/10.1057/9780230289505

Dewaele, J.-M. (2010b). Multilingualism and affordances: Variation in self-perceived communicative competence and communicative anxiety in French L1, L2, L3 and L4. *International Review of Applied Linguistics in Language Teaching*, *48*(2–3), 105–129. https://doi.org/10.1515/iral.2010.006

Dewhurst, S. A., & Parry, L. A. (2000). Emotionality, distinctiveness, and recollective experience. *European Journal of Cognitive Psychology*, *12*(4), 541–551. https://doi.org/10.1080/095414400750050222

Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: The case of Greek. *Frontiers in Psychology*, *1*(218). https://doi.org/10.3389/fpsyg.2010.00218

Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly*, *55*(3), 371–397. https://doi.org/10.1002/rrq.277

Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion*, *1*(1), 5–11. https://doi.org/10.1037//1528-3542.1.1.5

Duñabeitia, J. A., Avilés, A., & Carreiras, M. (2008). NoA's ark: Influence of the number of associates in visual word recognition. *Psychonomic Bulletin and Review*, *15*(6), 1072–1077. https://doi.org/10.3758/PBR.15.6.1072

Erdodi, L., & Lajiness-O'neill, R. (2012). Humor perception in bilinguals: Is language more than a code? *Humor*, *25*(4), 459–468. https://doi.org/10.1515/humor-2012-0024

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M., & Walter, H. (2003). Emotional context modulates subsequent memory effect. *NeuroImage*, *18*, 439–447. https://doi.org/10.1016/S1053-8119(02)00015-0

Erk, S., Martin, S., & Walter, H. (2005). Emotional context during encoding of neutral items modulates brain activation not only during encoding but also during recognition. *NeuroImage*, *26*, 829–838. https://doi.org/10.1016/j.neuroimage.2005.02.045

Ferré, P., Anglada-tort, M., & Guasch, M. (2018). Processing of emotional words in bilinguals: Testing the effects of word concreteness, task type and language status. *Second Language Research*, *34*(3), 371–394. https://doi.org/10.1177/0267658317744008

Ferré, P., Garcia, T., Fraga, I., Sanchez-Casas, R., & Molero, M. (2010). Memory for emotional words

in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition & Emotion*, *24*(5), 760–785. https://doi.org/10.1080/02699930902985779

Ferré, P., Ventura, D., Comesaña, M., & Fraga, I. (2015). The role of emotionality in the acquisition of new concrete and abstract words. *Frontiers in Psychology*, *6*(976), 1–10. https://doi.org/10.3389/fpsyg.2015.00976

Frances, C., de Bruin, A., & Duñabeitia, J. A. (2020a). The effects of language and emotionality of stimuli on vocabulary learning. *PLoS ONE*, *15*(10), e0240252. https://doi.org/10.1371/journal.pone.0240252

Frances, C., de Bruin, A., & Duñabeitia, J. A. (2020b). The influence of emotional and foreign language context in content learning. *Studies in Second Language Acquisition*, *42*(4), 891–903. https://doi.org/10.1017/S027226311900072X

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports*, *10*(13967). https://doi.org/10.1038/s41598-020-70922-1

Francis, W. S., & Gutiérrez, M. (2012). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory & Cognition*, *40*, 496–503. https://doi.org/10.3758/s13421-011-0163-3

Gladfelter, A., & Goffman, L. (2018). Semantic richness and word learning in children with Autism Spectrum Disorder. *Developmental Science*, *21*(2). https://doi.org/10.1111/desc.12543

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., Trueswell, J. C., Gleitman, L. R., … Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Goh, W. D., & Hu, W. (2011). Short-term memory for emotional words: Separating the effects of emotionality and semantic relatedness. In B. Kokinov, A. Karmiloff-Smith, & N. J. Nersessian (Eds.), *European Perspectives on Cognitive Science*. New Bulgarian University Press.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, *29*(2), 228–244. https://doi.org/10.1016/0749-596X(90)90074-A

Guitard, D., Miller, L. M., Neath, I., & Roodenrys, S. (2019). Does contextual diversity affect serial recall? *Journal of Cognitive Psychology*, *31*(4), 379–396. https://doi.org/10.1080/20445911.2019.1626401

Harris, C. L. (2004). Bilingual speakers in the lab: Psychophysiological measures of emotional reactivity. *Journal of Multilingual and Multicultural Development*, *25*(2&3), 223–247.

Harris, C. L., Ayçiçeği, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, *24*(4), 561–579. https://doi.org/10.1017.S0142716403000286

Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (Vol. 56, pp. 257–283). Clevedon, UK: Multilingual Matters.

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., & Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychological Science*, *28*(10), 1387–1397. https://doi.org/10.1177/0956797617720944

Hernandez-Nanclares, N., & Jimenez-Munoz, A. (2015). English as a medium of instruction: Evidence for language and content targets in bilingual education in economics. *International Journal of Bilingual Education and Bilingualism*, *20*(7), 883–896. https://doi.org/10.1080/13670050.2015.1125847

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Horst, M., & Cobb, T. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207–223. Retrieved from https://eric.ed.gov/?id=EJ577617

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*(August), 114–126. https://doi.org/10.1016/j.jml.2018.08.005

Hulme, R. C., Barsky, D., & Rodd, J. M. (2018). Incidental learning and long-term retention of new word meanings from stories: The effect of number of exposures. *Language Learning*, *69*(1), 18–43. https://doi.org/10.1111/lang.12313

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, *21*(4), 767–787. https://doi.org/10.1002/9781405198431

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin and Review*, *23*(4), 1214–1220. https://doi.org/10.3758/s13423-015-0980-7

Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B., & Jones, M. N. (2012). Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *The Journal of the Acoustical Society of America*, *132*(2), 74–80. https://doi.org/10.1121/1.4731641

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, *66*(2 SPL.ISSUE), 115–124. https://doi.org/10.1037/a0026727

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

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology General*, *143*(3), 1065–1081. https://doi.org/doi:10.1037/a0035669

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30. https://doi.org/10125/66648

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

Lund, E., Douglas, W. M., & Schuele, C. M. (2015). Semantic richness and word learning in children

with hearing loss who are developing spoken language: A single case design study. *Deafness and Education International*, *17*(3), 163–175. https://doi.org/10.1179/1557069X15Y.0000000004

Lund, K., & Burgess, C. (1996). Producing High-Dimensional Semantic Space from Lexical Co-Occurence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*(1467), 1–6. https://doi.org/10.3389/fpsyg.2014.01467

Monteiro, K. R., Crossley, S. A., & Kyle, K. (2020). In search of new benchmarks: Using L2 lexical frequency and contextual diversity indices to assess second language writing. *Applied Linguistics*, *41*(2), 280–300. https://doi.org/10.1093/applin/amy056

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233–253. https://doi.org/10.2307/747758

Nielsen, J. H., Shapiro, S. A., & Mason, C. H. (2010). Emotionality and semantic onsets: Exploring orienting attention responses in advertising. *Journal of Marketing Research*, *47*(6), 1138–1150. https://doi.org/10.1509/jmkr.47.6.1138

Nott, C. R., & Lambert, W. E. (1968). Free recall of bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *7*(6), 1065–1071. https://doi.org/10.1016/S0022-5371(68)80069-6

Pagán, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science*, *43*(e12705). https://doi.org/10.1111/cogs.12705

Parmentier, F. B. R., Comesaña, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology*, *70*(1), 1–17. https://doi.org/10.1080/17470218.2015.1105268

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart ? *Reading in a Foreign Language*, *22*(1), 31–55. https://doi.org/10125/66652

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology*, *116*(1), 37–44. https://doi.org/10.1016/j.jecp.2012.10.014

Pérez-Serrano, M., Nogueroles-López, M., & Duñabeitia, J. A. (2021). Incidental vocabulary learning with subtitles in a new language: Orthographic markedness and number of exposures. *PLoS ONE*, *16*(2), e0246933. https://doi.org/10.1371/journal.pone.0246933

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin and Review*, *15*(1), 161–167. https://doi.org/10.3758/PBR.15.1.161

Pexman, P. M., Holyk, G. G., & Monfils, M. H. (2003). Number-of-features effects and semantic processing. *Memory and Cognition*, *31*(6), 842–855. https://doi.org/10.3758/BF03196439

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recongnition. *Psychonomic Bulletin & Review*, *9*(3), 542–549.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*.

Plummer, P., Perea, M., & Rayner, K. (2014). The influence of contextual diversity on eye movements in reading. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, *40*(1), 275–283. https://doi.org/10.1037/a0034058

Rabovsky, M., Sommer, W., & Rahman, R. A. (2012). The time course of semantic richness effects in visual word recognition. *Frontiers in Human Neuroscience*, *6*(11), 1–9. https://doi.org/10.3389/fnhum.2012.00011

Rosa, E., Tapia, J. L., & Perea, M. (2017). Contextual diversity facilitates learning new words in the classroom. *PLoS ONE*, *12*(6), 1–12. https://doi.org/10.7910/DVN/JAVSUB.Funding

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction*, *52*, 69–79. https://doi.org/10.1016/j.learninstruc.2017.04.007

Rubin, D. C., & Friendly, M. (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Memory & Cognition*, *14*(1), 79–94. https://doi.org/10.3758/BF03209231

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology1*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism*, *13*(2), 231–252. https://doi.org/10.1017/S1366728909990514

Şen, Y., & Kuleli, M. (2015). The effect of vocabulary size and vocabulary depth on reading in EFL context. *Procedia - Social and Behavioral Sciences*, *199*, 555–562. https://doi.org/10.1016/j.sbspro.2015.07.546

Shen, Z. (2013). The effects of vocabulary knowledge and dictionary use on EFL reading performance. *English Language Teaching*, *6*(6), 77–85. https://doi.org/10.5539/elt.v6n6p77

Soares, A. P., Machado, J., Costa, A., Iriarte, Á., Simões, A., de Almeida, J. J., … Perea, M. (2015). On the advantages of word frequency and contextual diversity measures extracted from subtitles: The case of Portuguese. *Quarterly Journal of Experimental Psychology*, *68*(4), 680–696. https://doi.org/10.1080/17470218.2014.964271

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(5), 760–766. https://doi.org/10.1037/0278-7393.29.5.760

Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish-English bilingual speakers. *Cognition and Emotion*, *21*(5), 1077–1090. https://doi.org/10.1080/02699930601054133

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Syssau, A., & Laxén, J. (2012). L'influence de la richesse sémantique dans la reconnaissance visuelle des mots émotionnels. *Canadian Journal of Experimental Psychology*, *66*(1), 70–78. https://doi.org/10.1037/a0027083

Sze, W. P., Rickard Liow, S. J., & Yap, M. J. (2014). The Chinese Lexicon Project: A repository of lexical

decision behavioral responses for 2,500 Chinese characters. *Behavior Research Methods*, *46*(1), 263–273. https://doi.org/10.3758/s13428-013-0355-9

Taler, V., López Zunini, R., & Kousaie, S. (2016). Effects of semantic richness on lexical processing in monolinguals and bilinguals. *Frontiers in Human Neuroscience*, *10*(382), 1–11. https://doi.org/10.3389/fnhum.2016.00382

Talmi, D., & Morsovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*(5), 742–751. https://doi.org/10.3758/BF03195864

Talmi, D., Schimmack, U., Paterson, T., & Moscovitch, M. (2007). The role of attention and relatedness in emotionally enhanced memory. *Emotion*, *7*(1), 89–102. https://doi.org/10.1037/1528-3542.7.1.89

Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods & Instrumentation*, *12*(6), 587–595. https://doi.org/10.3758/BF03201848

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vander Beken, H., & Brysbaert, M. (2018). Studying texts in a second language: The importance of test type. *Bilingualism: Language and Cognition*, *21*(5), 1062–1074. https://doi.org/10.1017/S1366728917000189

Vander Beken, H., Woumans, E., & Brysbaert, M. (2018). Studying texts in a second language: No disadvantage in long term recognition memory. *Bilingualism: Language and Cognition*, *21*(4), 826–838. https://doi.org/10.1017/s1366728917000360

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(4), 796–800. https://doi.org/10.1037/0278-7393.30.4.796

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. https://doi.org/10.1515/langcog.2009.011

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin and Review*, *19*(1), 73–80. https://doi.org/10.3758/s13423-011-0179-5

Winskel, H. (2013). The emotional Stroop task and emotionality rating of negative and neutral words in late Thai-English bilinguals. *International Journal of Psychology*, *48*(6), 1090–1098. https://doi.org/10.1080/00207594.2013.793800

Yang, S.-W. (2014). *The study of variables influencing the effect of English medium instruction on academic content learning and English proficiency development*. Pennsylvania State University.

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin and Review*, *18*(4), 742–750. https://doi.org/10.3758/s13423-011-0092-y

Appendix: Published works

Publication 1:

THE INFLUENCE OF EMOTIONAL AND FOREIGN LANGUAGE CONTEXT IN CONTENT LEARNING

Frances, C., De Bruin, A., & Duñabeitia, J. A. (2020) The Influence of Emotional and Foreign Language Context in Content Learning. *Studies in Second Language Acquisition. 42*(4), 891-903. https://doi.org/10.1017/S027226311900072X [Version: author manuscript]

JOURNAL QUALITY INDICATORS: STUDIES IN SECOND LANGUAGE ACQUISITION

Scopus CiteScore 2019: 5.3

SCImago Journal Rank 2019: 3.238

SNIP 2019: 2.759

CiteScore rank 2019: D1 for Language and Linguistics (14/830), Linguistics and Language (17/884), and Education (51/1254).

H-Index (source: SCImago): 58

The influence of emotional and foreign language context in content learning

Candice Frances[1, 2*]

Angela De Bruin[1, 3]

Jon Andoni Duñabeitia[4,5]

[1] *BCBL, Basque Center on Brain, Language and Cognition; Donostia, Spain*

[2] *University of the Basque Country (UPV/EHU)*

[3] *Department of Psychology, University of York, York, United Kingdom*

[4] *Centro de Ciencia Cognitiva – C3, Universidad Nebrija*

[5] *Department of Language and Culture, The Arctic University of Norway*

*Contact information:

Candice Frances

c.frances@bcbl.eu

Basque Center on Cognition, Brain and Language (BCBL)

Paseo Mikeletegi 69, 2nd floor,

20009 Donostia - Spain

+34 943309300

Abstract

Prior research has found reduced emotionality with foreign language use, especially with single words, but what happens if emotionality is conveyed throughout a longer text? Does emotionality affect how well we remember and associate information, i.e., content learning? We played participants descriptions of two invented countries and tested how well they remembered facts about these countries. Each participant listened to one positive and one neutral description, which was read either in their native language (Spanish) or in their foreign language (English). Participants remembered facts they heard in positive semantic contexts better than those learned in neutral semantic contexts, and did better in their native than their foreign language. Importantly, there was no interaction between language and emotionality, suggesting that the previously reported decrease in emotionality in a foreign language might not extend to all areas of foreign language use. **Words**: 139

Keywords: emotionality; foreign language effects; non-native languages; learning; auditory modality

The Influence of Emotional and Foreign Language Context in Content Learning

As study abroad programs become more common, it is imperative that we understand how foreign languages (FL) affect our learning. For example, are we able to learn new content in a FL to the same extent as in our native language (NL)? There is a substantial amount of literature assessing this question in children, but there is little published research regarding adult learning. Furthermore, the current adult literature focuses mostly on memory for single words (e.g., Anooshian & Hertel, 1994; Ayçiçeği & Harris, 2004; Caldwell-Harris, 2009; Ferre, Garcia, Fraga, Sanchez-Casas, & Molero, 2010). One possible mechanism for improving content learning in an FL—drawing from the NL literature—is using emotionality to enhance memory. Emotional items are easier to remember in our NL than in our FL (see Caldwell-Harris, 2014 for a review). But, can this strategy be used to improve performance in an FL? Importantly, prior single word research has found reduced emotionality effects in an FL, but what happens if emotionality is conveyed throughout a longer text rather than in single words? The current study attempts to expand on these questions, testing memory for information embedded in an emotional context, to see whether this can boost content learning in an FL.

One of the most common types of programs that use FL to teach new information is content and language integrated learning (CLIL). CLIL refers to a curriculum-based approach used to teach content courses using a second language, in order to teach both content and language through immersion. Although research on the language learning aspects of CLIL quite conclusively shows an improvement in FL use and comprehension (Admiraal, Westhoff, & De Bot, 2006; Aguilar & Rodríguez, 2012; Bergroth, 2006; Dalton-Puffer, 2007; Jiménez Catalán & Ruiz de Zarobe, 2009; Ouazizi, 2016; Serra, 2007; Xanthou, 2011; although see Dallinger, Jonkmann, Hollm, & Fiege, 2016 for no improvement), the research on content learning is less clear-cut (Dalton-Puffer, 2011). There are studies that find positive effects (Day & Shapson, 1996; Jäppinen, 2005; Ouazizi, 2016; Pérez Cañado, 2018; Surmont, Struys, Van Den Noort, & Van De Craen, 2016; Van de Craen, Ceuleers, &

Mondt, 2007; Xanthou, 2011), while others find negative (Anghel, Cabrales, & Carro, 2016; Dallinger et al., 2016; Fernández-Sanjurjo, Fernández-Costales, & Arias Blanco, 2017) or null effects (Admiraal et al., 2006; Bergroth, 2006; Serra, 2007; Stohler, 2006). Consequently, these results paint a less than clear picture of how children learn new content in an FL.

The literature on adult FL-medium learning is more limited, with most of the reported benefits being associated with language (e.g., Yang, 2014) and not content. These studies often show no difference between the control and experimental group in overall performance at the end of the course (e.g., Hernandez-Nanclares & Jimenez-Munoz, 2015), but very few examine the immediate understanding and learning of new content in an FL. Those that do report a difference find that instruction in an FL is detrimental, particularly without FL support (Roussel, Joulia, Tricot, & Sweller, 2017). These results have been accounted for in the context of cognitive load theory, which suggests a working memory overload for individuals trying to learn content in a language they are not proficient in (Roussel et al., 2017). Importantly, contributing to this literature would influence and possibly improve teaching methods for adults studying in an FL.

Given the difficulties in learning new content in an FL, we need to find ways of compensating for or aiding in improving performance. One way of doing this is by applying what we know from the NL studies. Considering this literature, one of the variables that aids learning is emotionality, as learning emotional words (see Caldwell-Harris, 2014 for a review), or seeing neutral words in emotional contexts (Erk et al., 2003; Erk, Martin, & Walter, 2005), improves memory performance. However, several studies show that speakers are less emotional in an FL than in an NL context (Dewaele, 2010; Harris, Gleason, & Ayçiçeği, 2006; Pavlenko, 2002). One might extrapolate from these studies that using emotionality as a tool to boost learning would not be as efficient in an FL. Indeed, Anooshian and Hertel (1994) found that participants remembered emotional words better than neutral words in their NL, but not in their FL. This is in line with foreign language effect (FLE) research supporting a reduction in emotionality in an FL (Costa, Foucart, Hayakawa, et al., 2014;

Costa, Foucart, Arnon, Aparici, & Apesteguia, 2014; Costa, Vives, & Corey, 2017; Hadjichristidis, Geipel, & Savadori, 2015; Keysar, Hayakawa, & An, 2012, but see Vives, Aparici, & Costa, 2018). Conversely, other studies find the same effects of emotion on memory in both languages (Ayçiçeği & Harris, 2004; Caldwell-Harris, 2009; Ferré, Ventura, Comesaña, & Fraga, 2015; Ponari et al., 2015). Therefore, it is not clear how the effects of emotionality in an FL compare to those of the NL.

Nevertheless, these conflicting results may be explained by alternative accounts, such as a reduction in intuitive responses and depletion of cognitive resources (Geipel, Hadjichristidis, & Surian, 2015a, 2015b, 2016) or triggering of different cultural norms (Gawinkowska, Paradowski, & Bilewicz, 2013) in the FL. Gawinkowska et al. (2013) suggest that the FLE is due to a difference in social and cultural norms rather than a difference in emotional impact between languages. Regardless of the origin of the effect, it is not clear whether people respond similarly to emotional stimuli in their NL and FL, nor whether they benefit from the effects of emotionality on memory the same way in an FL as in an NL. Furthermore, the paradigms used thus far predominantly focus on emotionally-charged words in isolation rather than in context (e.g., Anooshian & Hertel, 1994; Ayçiçeği & Harris, 2004; Caldwell-Harris, 2009; Ferre et al., 2010) and are limited to using single-word auditory material. This is particularly relevant since, contrary to this approach, information taught in classrooms is most commonly conveyed in context.

The objective of this study is to investigate content learning and how it is affected both by an FL and an emotional context. There is little research directly comparing acquisition of new concepts and knowledge in a bilingual's NL and FL. Likewise, there is no research looking into the effects of emotionality in this context, nor listening to texts manipulating emotional context semantically. Understanding how these variables interact can contribute to classrooms that use an FL as the medium of teaching, improving methods and efficacy. To address this, we had participants listen to two descriptions of countries (one positive and one neutral) in either their NL (Spanish) or an FL (English), followed by a multiple choice test. Using longer texts than those used in prior

research, we aimed to create a more realistic replication of information processing and acquisition. Thus, participants were required to learn interrelated facts that made a coherent whole, rather than independent pieces of information disconnected from each other (see Frances, de Bruin, & Duñabeitia, n.d., for a similar study using vocabulary learning and non-related information). This would allow them to create more complex networks of meaning, which in turn would allow us to understand how semantic context can affect memory for individual facts within these larger conceptual networks. We hypothesized that despite the fact that their overall performance was likely to be poorer in the FL than in the NL contexts, bilinguals would not show an FLE, but instead would present similar emotionality effects in both languages. The rationale for this is that, if the FL affects responding by reducing reliance on intuition or simply requires more cognitive resources—as suggested before—, the effect of emotionality should remain the same.

Methods

Participants

Participants were 76 native Spanish speakers (38 in each language group, 9 male, M_{age} = 33.86, SD_{age} = 9.14), recruited through language schools and randomly assigned to either the NL or FL context. All participants completed a test of English vocabulary (LexTALE; Lemhöfer & Broersma, 2012) and had a minimum score of 60%. This is equivalent to a minimum of a B2 level according to the Common European Framework of reference for languages, with 50 participants at the B2 level range and 26 at the C1/C2 level (Lemhöfer & Broersma, 2012). Participants in the two language contexts were matched on age and education level (i.e., highest level of schooling achieved, in all cases at least high school) according to the sociodemographic information gathered, as well as multiple language variables. They were asked to rate their English level overall on a 1-to-10 scale as well as their listening, reading, speaking, and writing skills in that language. They also reported their estimated age of acquisition of English and the amount of time spent living in an English speaking country (M = 3.08 months SD = 4.65 months; all were living in Spain at the time of testing). Finally,

they were matched on English and Spanish vocabulary knowledge as assessed by LexTALE (Lemhöfer & Broersma, 2012) and the LexTALE-Esp (Izura, Cuetos, & Brysbaert, 2014). For a summary of these variables, see Table 1 and on-line supplementary materials for means, distributions, and Bayes factors. The study and protocols were approved by the ethics committee at the BCBL.

Table 1: Matched Means and Standard Deviations

		Self-Rated Level of English							
	Age	Listening	Reading	Speaking	Writing	Overall	AOA of English	Spanish LexTALE	English LexTALE
Foreign	33.07 (8.91)	7.07 (1.47)	8.21 (0.93)	6.81 (1.22)	7.31 (1.18)	7.15 (1.12)	9.81 (3.77)	0.94 (0.04)	0.76 (0.08)
Native	34.47 (9.63)	7.31 (1.69)	8.23 (1.26)	7.05 (1.52)	7.39 (1.53)	7.39 (1.26)	10.7 (6.71)	0.93 (0.05)	0.77 (0.09)
BF_{01}	3.49 (0.01)	3.50 (0.01)	4.19 (0.01)	3.30 (0.01)	4.09 (0.01)	3.05 (0.01)	3.32 (0.01)	3.20 (0.01)	4.09 (0.01)

Note: Numbers in parentheses refer to standard deviation for the FL and NL groups, except for in the final line (Bayes Factor) where they refer to error percentage. With BF_{01} a positive number above 1 supports no difference between the two groups, with 3 and above implying moderate evidence that the means are equal. Age and age of acquisition of English are in years, the self-ratings of level of English are on a scale from 1 to 10, and the LexTALEs are scored from 0 (chance) to 1 (perfect score).

[Return to Summary of Publication 1 on page 24]

Instruments

We created the description for two imaginary countries including 50 different items of information (e.g., national sport and population—see on-line supplementary materials for the list of test items). These two descriptions were then modified with filler sentences to include a more positive or neutral description of the country (e.g., neutral: “The population of Tecamer is defined politically as left wing, although they are considered generally quite moderate in their political, economic, and social opinions” and positive: “The population of Tecamer is defined politically as left wing and supports freedom, tolerance, and social inclusion as well as equal opportunity, leading many campaigns against discrimination”). The Spanish and English versions were created simultaneously and were matched on length. The texts were 50 to 56 sentences long and the average number of words in the English and Spanish versions were matched (1278.5 and 1317, respectively). The two emotional conditions were matched within languages on lemmatized word

frequency of the content words (Spanish using LEXESP database, Sebastián-Gallés, Martí, Carreiras, & Cuetos, 2000; English using the HAL database, Lund & Burgess, 1996—Table 2). Importantly, the positive and neutral versions of the texts significantly differed on the mean valence and arousal of the words used, according to the ANEW database (Bradley & Lang, 1999) (valence: BF_{01} = 2.42e+11, 5.22e-18; arousal: BF_{01} = 3.068e+10, 4.14e-17). The number of high arousal (arousal >5) and high valence (valence >5) words also varied by condition (6% of the neutral condition and 12% of the positive condition was high valence word—see Appendix).

Table 2: Average Word Frequency by Language and Emotional Condition

	Spanish	English
Neutral (M, SD)	616.48 (1306.08)	608.75 (847.98)
Positive (M, SD)	727.03 (1793.12)	641.29 (919.19)
Bayes Factor (BF01, %error)	7.29 (0.068)	14.12 (8.63 e-6)

Note: Numbers in parentheses refer to standard deviation for the FL and NL groups, except for in the final row (Bayes Factor) where they refer to error percentage. With BF_{01} a positive number above 1 supports no difference between the two groups, with 3 and above implying moderate evidence that the means are equal.

These four texts (2 countries, each with a neutral and a positive version) were read aloud and recorded by four female native Spanish speakers and four female native English speakers. Each recording lasted between 6.85 and 8.07 minutes ($M_{duration}$ = 7.51 minutes, $SD_{duration}$ = .333 minutes).

Procedure

Participants accessed the experiment through LimeSurvey (Schmitz, 2019). First, they filled out a demographics and language questionnaire and then listened to two audio files, one of each country in a given emotionality and different speakers (out of the 4 possible ones in that language). Each participant heard recordings in only one language and carried out the rest of the study in that same language. The order of the countries, emotional condition, and emotional condition/country matching were all randomized across participants to avoid any strategic or order effects. Once participants finished listening to the audio files, they proceeded to answer 50 multiple-choice

questions about the stimuli content. These questions had 4 answer choices and participants were asked to pick one for each of the countries.

Analysis

The size of the sample was determined using GPower (Faul, Erdfelder, Lang, & Buchner, 2007), assuming a small to medium size interaction (η_p^2 = .05) and 95% power.

We carried out a two-way mixed ANOVA exploring the effects of emotionality and language on performance in the test to address whether performance was better in the NL or FL, whether emotional semantic context affects performance, and whether there was an interaction between the two. A main effect of language would indicate whether participants perform better in one of their languages, whilst a main effect of emotionality would reveal whether the emotional manipulation affected performance. Finally, any interaction between language and emotionality would show whether the effect of emotionality is modulated by language—meaning, emotionality affects people differently in the FL than the NL. In all cases, assumptions of statistical tests were met.

We followed these tests up with Bayes factors (Jeffreys, 1961), which represent the likelihood of one model—in this case, the null hypothesis—over another—in this case, the alternative hypothesis. For example, a BF_{01} of 5 means that the null hypothesis is 5 times more likely to be true than the alternative one and a BF_{01} of .2 means that the alternative hypothesis is 5 times more likely to be true than the null. These Bayes Factors have become increasingly common as an alternative to frequentist models (Poirier, 2006), in particular for ANOVAs (Rouder, Morey, Speckman, & Province, 2012).

Results

First, we calculated the internal consistency between the questions of each country and found that the tests had good internal consistency (Mufelo α = .84; Tecamer α = .86).

We removed participants who were outliers, meaning 1.5 IQR away from the median in either condition (positive or neutral) for each language group. Using this procedure, we removed one participant from the English group and 3 from the Spanish group. The same tests were carried out with and without the outliers and the results were consistent between the two.

We carried out a two-way mixed ANOVA with emotionality and language on performance on the test (see Table 3 for means, standard deviations, and 95% confidence intervals). There was a significant main effect of emotionality, such that participants performed better in the positive (M = 69.00%, SD = 13.95%) than the neutral condition (M = 65.97%, SD = 14.71%), $F(1,70) = 8.54$, $p = .005$, $\eta_p^2 = .109$, $BF_{01} = .146$, error% = 1.26×10^{-6} (see Figure 1 and on-line supplementary materials). There was also a main effect of language, such that participants performed better in their NL (Spanish: M = 74.6%, SD = 11.2%) than in their FL (English: M = 60.3%, SD = 11.6%), $F(1,70) = 26.83$, $p<.001$, $\eta_p^2 = .277$, $BF_{01} = 1.40 \times 10^{-4}$, error% = 1.29×10^{-7}. There was no interaction between the two factors, $F(1,70) = .104$, $p = .748$, $\eta_p^2 = .001$. A Bayesian repeated measures ANOVA comparing the model with the interaction (emotionality * language) and without the interaction term confirmed that there was moderate evidence that the addition of the interaction term led to an equally likely model, $BF_{01} = 4.12$, error% = 3.15—namely, no interaction was over 4 times more likely than an interaction. We also ran an independent samples t-test on the emotionality effect—namely the score on the positive condition minus the score on the neutral one for each of the language conditions—and again found moderate evidence in support of the null hypothesis, $BF_{01} = 3.93$, error% = .012.

Table 3: Average Accuracy in Percent Correct by Condition

				95% Confidence Interval	
Language Condition	Emotionality	Mean	Standard Error	Lower	Upper
English	Positive	62.0%	2.10%	57.9%	66.1%
	Neutral	58.7%	2.10%	54.5%	62.8%
	Overall	60.3%	1.90%	56.5%	64.2%
Spanish	Positive	76.0%	2.10%	71.8%	80.1%
	Neutral	73.3%	2.10%	69.1%	77.4%
	Overall	74.6%	1.90%	70.7%	78.5%
Total	Positive	69.0%	1.50%	66.1%	71.9%
	Neutral	66.0%	1.50%	63.1%	68.9%
	Overall	67.5%	1.61%	64.3%	70.6%

Note: Participants showed no effect of order, $t(75) = .019$, $p = .891$, $BF_{01} = 7.85$, error% = 7.39×10^{-6}, showing moderate evidence that participants performed similarly regardless of order. Furthermore, there was moderate evidence that the two country descriptions were equally easy to remember, $t(75) = 1.23$, $p = .270$, $BF_{01} = 4.35$, error% = 5.15×10^{-6}.

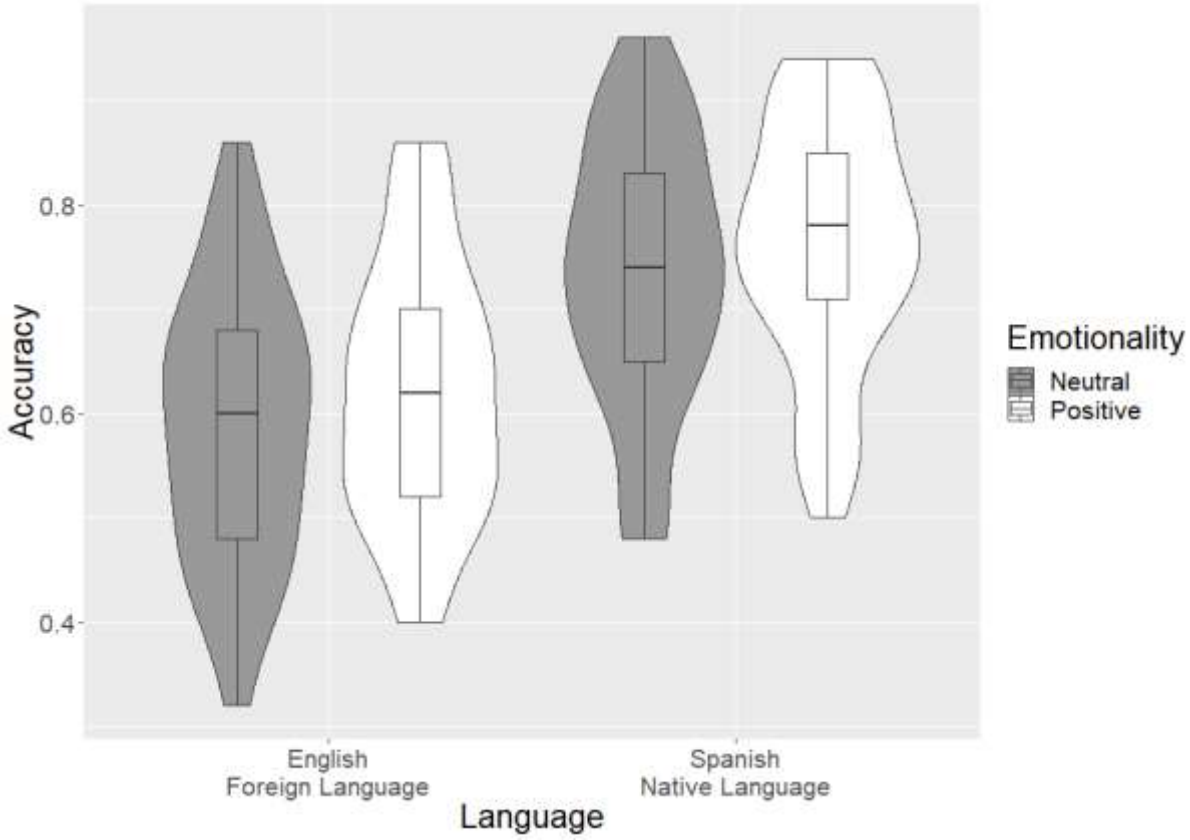

Figure 1: Distribution of average accuracy by language and condition. Participants in the native language condition (Spanish) did better on the task than those who carried out the task in their foreign language (English). In addition, participants did better when the information was presented in a positive rather than a neutral context. Nevertheless, the effect was the same in both languages.

[Return to Summary of Publication 1 on page 25]

Discussion

In the current study, we addressed the questions of whether learning new information in an FL could be improved using an emotional semantic context and whether this effect would be the same in the NL and FL. The main task of the study required participants to listen to descriptions of countries and answer questions about them. Although participants performed better in their NL, results suggested that they benefited equally from the positive emotional context in both languages.

Preceding studies on the effects of emotionality on memory have mainly used visual stimuli. In contrast, the current study emulates information transfer in classroom settings by focusing on aural stimuli. Results showed statistically reliable emotionality effects with auditory information in both NL and FL. The partial eta squared of this effect is considered to be of medium effect size, within the context of educational research (Richardson, 2011). This corresponds to 10.9% of the variance explained and a practical difference of 3% on the current test. Although relatively discrete, this effect could be the difference between passing and failing an exam for a student that is struggling in a class. In more general terms, this study suggests that emotionally loaded semantic contexts—not just emotional content—conveying new pieces of information can improve memory.

Given that there are no studies addressing the particular questions of the current study—namely, looking at the effects of emotional context on content learning—the results need to be understood within the wider literature. The effects found here (NL: 2.7%, FL: 3.3%) were smaller than those of single-word studies with known words. In particular, these studies show effects between 7 and 26% in the NL and between 9.5 and 18% in the FL (Anooshian & Hertel, 1994; Ayçiçeği & Harris, 2004; Caldwell-Harris, 2009; Ferré et al., 2010)—with one exception showing a non-significant effect in the FL (Anooshian & Hertel, 1994). Studies manipulating emotional context rather than emotional content have found larger effects than the current one in recall (12%) but not in recognition—no accuracy difference, only in response time (Erk et al., 2003, 2005). On the other hand, studies on new word learning show smaller effects (2 – 3.5%), more similar to the ones in the current study (Ferré et al., 2015). Overall, these results suggest that the effects of emotionality are reduced when only the context is manipulated and when there is learning of new content, rather than repeating information that is already known. Therefore, our results are in accordance with those reported by prior literature and are within the predictable effect size.

The key result in this study is that the effect of emotionality is the same in the FL and the NL. This result is consistent with many recent studies using emotionality in single-word processing

(Ayçiçeği & Harris, 2004; Caldwell-Harris, 2009; Ferré et al., 2015; Ponari et al., 2015), and suggest that this effect extends beyond individual word-learning to content learning. But, perhaps more importantly, this result challenges the view that the FL, in general terms, leads to emotional distancing (see Costa, Duñabeitia, & Keysar, 2018).

These results relate to the FLE and the theoretical issue of its origin. Hayakawa, et al. (2016) suggest that there are two main ways of explaining the FLE on moral decision-making: a reduction in emotional processing and increasing psychological distance. Both of these accounts would predict a reduced emotional effect in the FL compared to the native one. If emotionality is completely blocked, this described FLE would predict that emotionality and its effect on performance would be reduced or absent in the FL condition. With respect to psychological distance, the conclusion is the same: this would make the information seem more abstract, reducing the effect of emotionality. Therefore, neither of these ideas is consistent with our results—namely, an equal effect of emotionality in the NL and FL. On the other hand, if the FLE is circumscribed to only the manipulation of known information and its prior associations, it would explain why learning new information does not show the same effects. For example, learning the word “home” using neutral language would lead to more difficulty in learning it and a reduced emotional response for that word, whereas if it is presented using emotional language, perhaps it would be remembered better—showing an emotionality effect.

Looking at the results from this perspective, the current findings do not necessarily have to contradict the existence of the FLE. Instead, they suggest a possible mechanism for how it arises. Gawinkowska, et al.’s (2013) idea that the effect is due to social and cultural norm differences would suggest that emotionality should affect both language conditions equally in this case. This is consistent with our results, since if the FLE is circumscribed to differences in norms, it should not be present. Importantly, Geipel, et al.’s (2015a, 2015b, 2016) suggestion that the origin of this effect is a reduction of intuitive responses and a depletion of cognitive resources would imply a decrease in

performance overall in the FL, but not necessarily any difference in emotionality. This reduction of cognitive resource availability explains our data better, predicting our decrease in performance in the FL, as well as the consistency of emotionality effects between languages.

In other words, the results of the current study could suggest that, rather than emotionality being reduced overall in an FL context, learners' cognitive resources are taxed, affecting emotionality differently according to the task. Furthermore, if the reduction in emotionality is observed in cases where only already-known information is concerned, perhaps it is because they are lacking emotional associations within that language. These results suggest that providing FL learners with more emotional materials—as in this case—could help them learn these associations.

It is worth noting that, although we did not intend to manipulate interest—and effectively the content was the same between conditions—perhaps the positive condition could have also presented the information in a more interesting way than the neutral one, contributing to the effect we found (see Hidi, 1990 for a review on the effect of interest on learning). In future studies, the effect of emotionality could be contrasted with that of "interest" or engagement. In addition, the effect we observe here might be increased further by engaging the participants in an activity where they have to use this new content or by making the information to be remembered self-relevant. For example, with the current materials, engagement could be increased by asking participants to not only listen passively but also to actively decide if they would want to move to the described country. Nevertheless, the current results open way for a new way of looking at both emotionality effects and learning in a foreign language which, with further replications, could provide a useful tool for teaching in a non-native language.

Conclusion

The current study reports a well-controlled experiment in line with CLIL approaches, as participants learned the same content in either their NL or an FL and were then tested using exactly the same task and materials. Learning in an FL may sometimes hinder memory of new content as a

consequence of the difference in language knowledge and use with the NL. However, the use of emotional semantic contexts can be a short-term tool in the classroom, particularly during aural exercises or verbal transmission of new information in order to boost memory. Considering the emotional distancing or detachment that has been typically associated with FL contexts (see Costa et al., 2018), the use of emotionally loaded materials or activities in classroom settings could be useful for partially counteract existing FLEs.

Acknowledgements: This research has been partially funded by grants PGC2018-097145-B-I00, PSI2015-65689-P, and SEV-2015-0490 from the Spanish Government, and AThEME-613465 from the European Union. CF is supported by a MINECO predoctoral grant from the Spanish government (BES-2016-077169). AdB is supported by a Marie Skłodowska-Curie fellowship from the European Research Council (grant agreement number 743691). Thank you to Eneko Anton for his help in stimulus creation.

References

Admiraal, W., Westhoff, G., & De Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English 1. *Educational Research and Evaluation*, *12*(1), 75–93. https://doi.org/10.1080/13803610500392160

Aguilar, M., & Rodríguez, R. (2012). Lecturer and student perceptions on CLIL at a Spanish university. *International Journal of Bilingual Education and Bilingualism*, *15*(2), 183–197. https://doi.org/10.1080/13670050.2011.615906

Anghel, B., Cabrales, A., & Carro, J. M. (2016). Evaluating a bilingual education program in Spain: The impact beyond foreign language learning. *Economic Inquiry*, *54*(2), 1202–1223. https://doi.org/10.1111/ecin.12305

Anooshian, L. J., & Hertel, P. T. (1994). Emotionality in free recall: Language specificity in bilingual memory. *Cognition & Emotion*, *8*(6), 503–514. https://doi.org/10.1080/02699939408408956

Ayçiçeği, A., & Harris, C. L. (2004). Bilinguals' recall and recognition of emotion words. *Cognition and Emotion*, *18*(7), 977–987. https://doi.org/10.1080/02699930341000301

Bergroth, M. (2006). Immersion students in the matriculation examination three years after immersion. In S. Björklund, K. Mard-Miettinen, M. Bergström, & M. Södergard (Eds.), *Exploring Dual-Focussed Education: Integrating language and content for individual and societal needs* (pp. 123–134). Vaasan yliopisto. Retrieved from http://www.uva.fi/materiaali/pdf/isbn_952-476-149-1.pdf

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.* https://doi.org/10.1109/MIC.2008.114

Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, *12*(3), 291–303. https://doi.org/10.1017/S1366728909990125

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: Theoretical implications. *Frontiers in Psychology*, *5*(1055), 1–4. https://doi.org/10.3389/fpsyg.2014.01055

Costa, A., Duñabeitia, J. A., & Keysar, B. (2019). Language context and decision-making: Challenges and advances. *Quarterly Journal of Experimental Psychology.*, *72*(1), 1–2. https://doi.org/https://doi.org/10.1177/1747021818789799

Costa, A., Foucart, A., Arnon, I., Aparici, M., & Apesteguia, J. (2014). "Piensa" twice: On the foreign language effect in decision making. *Cognition*, *130*(2), 236–254. https://doi.org/10.1016/j.cognition.2013.11.010

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PLoS ONE*, *9*(4), 1–7. https://doi.org/10.1371/journal.pone.0094842

Costa, A., Vives, M.-L., & Corey, J. D. (2017). On language processing shaping decision making. *Current Directions in Psychological Science*, *26*(2), 146–151. https://doi.org/10.1177/0963721416680263

Dallinger, S., Jonkmann, K., Hollm, J., & Fiege, C. (2016). The effect of content and language integrated learning on students' English and history competences—Killing two birds with one stone? *Learning and Instruction*, *41*, 23–31. https://doi.org/10.1016/j.learninstruc.2015.09.003

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms. John Benjamins Publishing* (Vol. 20).

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. https://doi.org/10.1017/S0267190511000092

Day, E., & Shapson, S. (1996). *Studies in immersion education (Vol. 11)*. Multilingual Matters.

Dewaele, J.-M. (2010). *Emotions in Multiple Languages*. London: Palgrave Macmillan. https://doi.org/10.1057/9780230289505

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M., & Walter, H. (2003). Emotional context modulates subsequent memory effect. *NeuroImage*, *18*(2), 439–447. https://doi.org/10.1016/S1053-8119(02)00015-0

Erk, S., Martin, S., & Walter, H. (2005). Emotional context during encoding of neutral items modulates brain activation not only during encoding but also during recognition. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.02.045

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2017). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, *0050*, 1–14. https://doi.org/10.1080/13670050.2017.1294142

Ferré, P., Garcia, T., Fraga, I., Sanchez-Casas, R., & Molero, M. (2010). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition & Emotion*, *24*(5), 760–785. https://doi.org/10.1080/02699930902985779

Ferré, P., Ventura, D., Comesaña, M., & Fraga, I. (2015). The role of emotionality in the acquisition of new concrete and abstract words. *Frontiers in Psychology*, *6*(976), 1–10. https://doi.org/10.3389/fpsyg.2015.00976

Frances, C., de Bruin, A., & Duñabeitia, J. A. (n.d.). The effects of language and emotional context on vocabulary learning.

Gawinkowska, M., Paradowski, M. B., & Bilewicz, M. (2013). Second language as an exemptor from

sociocultural norms. Emotion-related language choice revisited. *PLoS ONE*, *8*(12), 1–6. https://doi.org/10.1371/journal.pone.0081225

Geipel, J., Hadjichristidis, C., & Surian, L. (2015a). How foreign language shapes moral judgment. *Journal of Experimental Social Psychology*, *59*, 8–17. https://doi.org/10.1016/j.jesp.2015.02.001

Geipel, J., Hadjichristidis, C., & Surian, L. (2015b). The foreign language effect on moral judgment: The role of emotions and norms. *PLoS ONE*, *10*(7), 1–17. https://doi.org/10.1371/journal.pone.0131529

Geipel, J., Hadjichristidis, C., & Surian, L. (2016). Foreign language affects the contribution of intentions and outcomes to moral judgment. *Cognition*, *154*, 34–39. https://doi.org/10.1016/j.cognition.2016.05.010

Hadjichristidis, C., Geipel, J., & Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: The role of affect. *Journal of Experimental Psychology: Applied*, *21*, 117–129. https://doi.org/10.1037/xap0000044

Harris, C. L., Gleason, J. B., & Ayçiçeği, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual minds: Emotional experience, expression, and representation* (Vol. 56, pp. 257–283). Clevedon, UK: Multilingual Matters.

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a foreign language changes our choices. *Trends in Cognitive Sciences*, *20*(11), 791–793. https://doi.org/10.1016/j.tics.2016.08.004

Hernandez-Nanclares, N., & Jimenez-Munoz, A. (2015). English as a medium of instruction: evidence for language and content targets in bilingual education in economics. *International Journal of Bilingual Education and Bilingualism*, *20*(7), 883–896.

https://doi.org/10.1080/13670050.2015.1125847

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research*, *60*(4), 549–571. https://doi.org/10.3102/00346543060004549

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*, 49–66.

Jäppinen, A.-K. (2005). Thinking and content learning of mathematics and science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland. *Language and Education*, *19*(2), 148–169. https://doi.org/10.1080/09500780508668671

Jeffreys, H. (1961). *Theory of Probability, Oxford*. Oxford, UK: Oxford University Press.

Jiménez Catalán, R. M., & Ruiz de Zarobe, Y. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In *Content and language integrated learning: Evidence from research in Europe* (pp. 81–92).

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, *23*(6), 661–668. https://doi.org/10.1177/0956797611432178

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

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, *28*(2), 203–208. https://doi.org/10.3354/ame01683

Ouazizi, K. (2016). The effects of CLIL education on the subject matter (Mathematics) and the target

language (English). *Latin American Journal of Content & Language Integrated Learning*, *9*(1), 110–137. https://doi.org/10.5294/laclil.2016.9.1.5

Pavlenko, Aneta. (2002). Bilingualism and emotions. *Multilingua* , *21*(1), 45–78.

Pérez Cañado, M. L. (2018). The effects of CLIL on L1 and content learning: Updated empirical evidence from monolingual contexts. *Learning and Instruction*, *57*, 18–33. https://doi.org/10.1016/j.learninstruc.2017.12.002

Poirier, D. J. (2006). The growth of Bayesian methods in statistics and economics since 1970. *Bayesian Analysis*, *1*(4), 969–980. https://doi.org/10.1214/06-BA132

Ponari, M., Rodriguez-Cuadrado, S., Vinson, D., Fox, N., Costa, A., & Vigliocco, G. (2015). Processing advantage for emotional words in bilingual speakers. *Emotion*, *15*(5), 644–652. https://doi.org/10.1037/emo0000061

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction*, *52*, 69–79. https://doi.org/10.1016/j.learninstruc.2017.04.007

Schmitz, L. P. T. / C. (2019). LimeSurvey: An open source survey tool. Hamburg, Germany: LimeSurvey Project. Retrieved from http://www.limesurvey.org

Sebastián-Gallés, N., Martí, M., Carreiras, M., & Cuetos, F. (2000). *LEXESP: Léxico Informatizado del*

Español. Barcelona: Universitat de Barcelona.

Serra, C. (2007). Assessing CLIL at primary school: A longitudinal study. *International Journal of Bilingual Education and Bilingualism*, *10*(5), 582–602. https://doi.org/10.2167/beb461.0

Stohler, U. (2006). The acquisition of knowledge in bilingual learning: An empirical study on the role of language in content learning. *VIEWS-Vienna English Working Papers*, *15*(3), 41–46. Retrieved from https://boris.unibe.ch/18531/1/views15_3_Stohler.pdf

Surmont, J., Struys, E., Van Den Noort, M., & Van De Craen, P. (2016). The effects of CLIL on mathematical content learning: A longitudinal study. *Studies in Second Language Learning and Teaching*, *6*(2), 319. https://doi.org/10.14746/ssllt.2016.6.2.7

Van de Craen, P., Ceuleers, E., & Mondt, K. (2007). Cognitive development and bilingualism in primary schools: Teaching maths in a CLIL environment. In *Diverse Contexts–Converging Goals. CLIL in Europe* (pp. 185–200). Frankfurt am Main: Peter Lang.

Vives, M.-L., Aparici, M., & Costa, A. (2018). The limits of the foreign language effect on decision-making: The case of the outcome bias and the representativeness heuristic. *PLoS ONE*, *13*(9), 1–14. https://doi.org/10.17605/osf.io/6mdvs.Funding

Xanthou, M. (2011). The impact of CLIL on L2 vocabulary development and content knowledge. *English Teaching*, *10*(4), 116–126.

Yang, S.-W. (2014). *The study of variables influencing the effect of English medium instruction on academic content learning and English proficiency development*. The Pennsylvania State University.

Appendix: Number of emotional words and the average rating overall by language and condition

	English				Spanish			
	High Valence Words		High Arousal Words		High Valence Words		High Arousal Words	
	N	M(SD)	N	M(SD)	N	M(SD)	N	M(SD)
Neutral	151	6.56 (1.03)	81	5.01 (0.91)	54	5.85 (1.71)	37	5.14 (1.07)
Positive	243	6.99 (1.01)	176	5.43 (0.95)	193	7.21 (1.06)	172	6.07 (1.08)

Note: N stands for the number of words with values >5. The means and standard deviations are overall on a scale from 1 to 9.

Supplementary Material: Average number of participants of each gender by group

	Female	Male	Total
Foreign	29	9	38
Native	27	11	38

Note: The Bayes Factor for this contingency table was BF_{01} = 2.80.

Supplementary Material: Average number of participants per educational level group

	High School	Associates Degree	Bachelors	Masters	Doctorate
Foreign	7	0	17	12	2
Native	3	3	18	11	3

Note: The Bayes Factor for this contingency table was BF_{01} = 3.16.

Supplementary Material: Average number of participants per time range living in an English speaking country

	Never	< 3 months	3 to 6 months	6 to 12 months	> 12 months
Foreign	21	10	4	3	0
Native	21	6	3	6	2

Note: The Bayes Factor for this contingency table was BF_{01} = 4.19.

Supplementary Material: English Version of the Test

1. What color was the flag of each country?
   Red/Blue/Yellow/Green
2. Who is the government in each country led by?
   King/Emperor/Chancellor/President
3. The population of each country is composed of...
   A proportion of four women to each man/Many more men than women/An equal number of men and women/Twice as many women as there are men
4. What is the most powerful economic sector in each country?
   Tourism/Metallurgy/Agriculture/Industry
5. What is the most commonly used mode of transportation between cities in each country?
   Ships/Cars/Plains/Trains
6. What is the national sport of each country?
   Baseball/Football/Basketball/Tennis
7. What is the currency of each country?
   The Franc/The Dollar/The Pound/The Peso
8. What continent is each country in?
   Asia/America/Africa/Oceania
9. What language do people speak in each country?
   English/French/Italian/Portuguese
10. What is the favorite pet in each country?
    Turtles/Dogs/Cats/Birds
11. How many children do people have in each country, on average?
    2/1/3/4
12. What is the most common drink in each country?
    Wine/Beer/Tea/Coffee
13. What is the most famous dish in each country made out of? *Meat/Fish/Pasta/Vegetables*
14. Which is the most common hair color among the population of each country?
    Black/Brown/Blonde/Red
15. What is the national anthem of each country about?
    War/History/Culture/Brotherhood
16. What does most of the population of each country define itself as, politically?
    Mainly left-wing/Mainly right-wing/Mainly center/Mainly apolitical
17. What was the biggest tragedy in each country?
    A fire/A hurricane/A bombing/An earthquake
18. What is the oldest monument in each country?
    The opera in the capital/The roman bridge/The national museum/The temple in the capital

19. What century was each country founded in*?*
    X. Century/V. Century/VII. Century/XIX. Century
20. What was the most famous historical character in each country?
    A scientist/An athlete/A singer/A politician
21. What is the average level of education in each country?
    University degree/High school/Elementary school/Technical certificates
22. In general, what is the climate of each country like?
    Tropical, warm all year round./Continental, with a large variation between summer and winter./Oceanic, with moderate temperatures all year round./Arid, with very little rain.
23. In each country, there is a famous company that produces...
    Computers/Cars/Watches/Phones
24. What is the most common profession in each country?
    Engineering/Medicine/Mechanic/Transporter
25. What is the predominant religious belief in each country?
    Christianity/Islam/Atheism/Buddhism
26. What is the traditional instrument of each country?
    The bagpipe/The flute/The drum/The guitar
27. What is the main attraction for tourists in each country?
    The beach/The mountains/The low cost/The culture
28. What is the legal driving age in each country?
    16 years old/21 years old/18 years old/20 years old
29. What is the unit of measurement for temperature in each country?
    Degrees Celsius/Degrees Fahrenheit/Degrees Réaumur/Degrees Kelvin
30. What does the traditional clothing in each country include?
    A sword/A cane/A sash/A hat/
31. What is the main ingredient in the traditional dessert of each country?
    Chocolate/Cream/Strawberries/Honey
32. What is the most common wild animal in each country?
    Boar/Wolf/Bear/Fox
33. What is the most salient geographical characteristic of each country?
    It has a volcano/It has numerous lakes/It has a desert/It has the highest mountain in the continent
34. These countries are...
    An island/An archipelago/A peninsula/In the interior of the continent
35. What is the most common sport in each country?
    Swimming/Hiking/Fishing/Mountain bike
36. What is the retirement age in each country?
    65 years old/70 years old/75 years old/72 years old

37. The healthcare in each country is…
    Private/Public/Recently nationalized/State assisted

38. What is the main health problem for the population of each country?
    Obesity/Stress/Alcoholism/Contamination

39. What is the population of each country?
    20 million/40 million/3 million/60 million

40. What is the most important event in recent history for each country?
    The Olympics/The discovery of ruins/A recent war with a neighboring country/The discovery of oil

41. What alphabet did the first people of each country use to write?
    Greek alphabet/ Japanese Kanjis/ Cyrillic alphabet/Egyptian hieroglyphics

42. In the schools of each country, all students are required to...
    Learn music/Learn a second language/Have a laptop/Do community service

43. What is the most common source of energy in each country?
    Nuclear energy/Wind power/Hydraulic energy/Solar energy

44. What is the immigration like in each country?
    There are more people leaving the country/There are less people leaving the country than arriving/There is practically no immigration/There are as many people leaving as there are coming in

45. What is the main export of each country?
    Textiles/Fruits and vegetables/Wine and oil/Gas and petroleum

46. Where are the most important universities in each country?
    In the capital/In a college town/In the two main cities in the country/Throughout the country

47. When the people in each country buy a car, how is it usually powered?
    Hybrid/Gas/Diesel/Electric

48. What is the preferred means of transportation within the cities of each country?
    Bicycle/Metro/Motorcycle/Bus

49. How many official languages are there in each country?
    1/3/4/2

50. Most of the public money in each country goes to...
    Culture/Healthcare/Armed forces/Research

[Return to Summary of Publication 1 on page 23]

Supplementary Material: Emotionality Effect by Participant

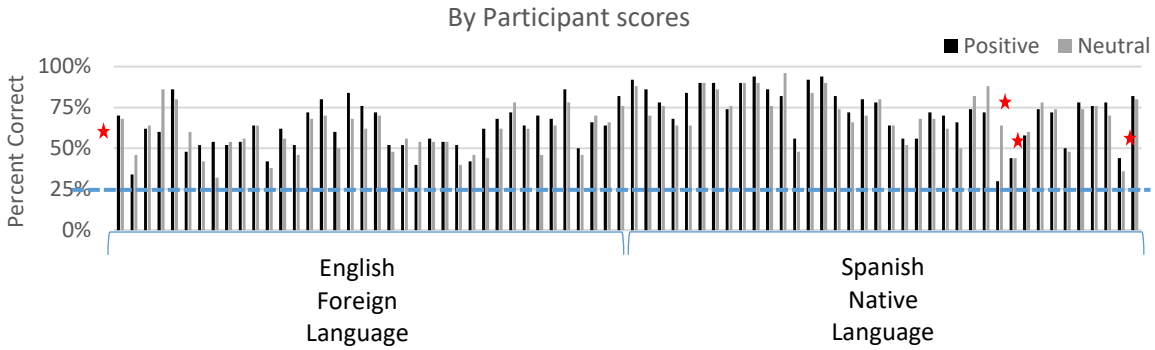

Fig. The following is the emotionality effect data by participant. The left half of the graph corresponds to the participants who carried out the task in their foreign language—namely, English—and the right side corresponds to those who carried it out in their native language—Spanish. Overall, 47 out of 76 participants (62%) showed an emotionality effect.

Note: Participants marked with a star are those that were excluded during outlier removal. The dotted line along the 25% marker shows chance performance.

[Return to Summary of Publication 1 on page 26]

Additional materials for the purpose of the present thesis

Table 4.Example stimuli in all four conditions.

	English	Spanish
Neutral	The population of Tecamer is defined politically as left wing, although they are considered generally quite moderate in their political, economic, and social opinions.	La población de Tecamer se define con inclinaciones políticas de izquierdas, aunque éstos se consideran por lo general bastante moderados en sus opiniones políticas, económicas y sociales.
Positive	The population of Tecamer is defined politically as left wing and supports freedom, tolerance, and social inclusion as well as equal opportunity, leading many campaigns against discrimination.	La población de Tecamer se define con inclinaciones políticas de izquierdas y apoya la libertad y tolerancia e inclusión social así como la igualdad de oportunidades, por lo que lideran muchas campañas contra la discriminación.

[Return to Summary of Publication 1 on page 24]

Figure 2. Example questions from the testing phase.

[Return to Summary of Publication 1 on page 24]

Publication 2:

THE EFFECTS OF LANGUAGE AND EMOTIONALITY OF STIMULI ON VOCABULARY LEARNING

Frances, C., De Bruin, A., & Duñabeitia, J. A. (2020) The effects of language and emotionality of stimuli on vocabulary learning. *PLOS ONE. 15*(10): e0240252. https://doi.org/10.1371/journal.pone.0240252

JOURNAL QUALITY INDICATORS:

PLOS ONE

Scopus CiteScore 2019: 5.2

SCImago Journal Rank 2019: 1.023

SNIP 2019: 1.205

CiteScore rank 2019: D1 (10/111) for Multidisciplinary.

H-Index (source: SCImago): 300

PLOS ONE

RESEARCH ARTICLE

The effects of language and emotionality of stimuli on vocabulary learning

Candice Frances[1,2]*, **Angela de Bruin**[3], **Jon Andoni Duñabeitia**[4,5]

1 Basque Center on Brain, Language and Cognition, Donostia, Spain, **2** Department of Social Sciences and Law, UPV/EHU, Donostia, Spain, **3** Department of Psychology, University of York, York, United Kingdom, **4** Centro de Ciencia Cognitiva–C3, Universidad Nebrija, Madrid, Spain, **5** Department of Language and Culture, The Arctic University of Norway, Tromsø, Norway

* candice.frances@ncl.edu, c.frances@bcbl.eu

OPEN ACCESS

Citation: Frances C, de Bruin A, Duñabeitia JA (2020) The effects of language and emotionality of stimuli on vocabulary learning. PLoS ONE 15(10): e0240252. https://doi.org/10.1371/journal.pone.0240252

Editor: Koji Miwa, Nagoya University, JAPAN

Received: March 16, 2020

Accepted: September 22, 2020

Published: October 7, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0240252

Data Availability Statement: All data used for this manuscript are available at https://osf.io/64b3e/?view_only=aac93c2fc41b42fb9cd3519466cd62af.

Funding: This research has been partially funded by the following entities: Spanish Government [https://ciencia.sede.gob.es/]: PGC2018-097145-B-

Abstract

Learning new content and vocabulary in a foreign language can be particularly difficult. Yet, there are educational programs that require people to study in a language they are not native speakers of. For this reason, it is important to understand how these learning processes work and possibly differ from native language learning, as well as to develop strategies to ease this process. The current study takes advantage of emotionality—operationally defined as positive valence and high arousal—to improve memory. In two experiments, the present paper addresses whether participants have more difficulty learning the names of objects they have never seen before in their foreign language and whether embedding them in a positive semantic context can help make learning easier. With this in mind, we had participants (with a minimum of a B2 level of English) in two experiments (43 participants in Experiment 1 and 54 in Experiment 2) read descriptions of made-up objects—either positive or neutral and either in their native or a foreign language. The effects of language varied with the difficulty of the task and measure used. In both cases, learning the words in a positive context improved learning. Importantly, the effect of emotionality was not modulated by language, suggesting that the effects of emotionality are independent of language and could potentially be a useful tool for improving foreign language vocabulary learning.

Introduction

Our current study addresses the question of whether emotionality—specifically, positive valence and high arousal—affects word learning in a native and a foreign language. In particular, the focus is on cases in which one's only experience with the to-be-learned object is a definition without having seen or touched that object in person. Although this question may seem odd, in reality, when we learn new information in academic settings, our experience with the content we learn is quite limited. If we take the example of biology or history class, the information we learn there, regardless of how concrete, will hardly ever be a part of our sensory experience. According to Kousta et al. [1], this should lead to poorer representations and more difficulty in acquisition. In cases like Content and Language Integrated Learning (CLIL), new information is also conveyed in a foreign language, adding to the complexity of learning and

I00 (JAD), RED2018-102615-T (JAD), SEV-2015-0490, and BES-2016-077169 (CF) Comunidad de Madrid [https://www.comunidad.madrid/]: H2019/HUM-5705 (JAD) The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

remembering information that is abstracted from sensory experience. Given this increased difficulty in learning, it is important to find tools that can improve or facilitate this process. One possible tool is emotionality, as it has been suggested to improve learning in other cases [1,2], but it is not clear whether it also be useful in this kind of experience. In particular, would having only verbal and not sensory experience with the content make emotionality useful even for acquiring new concrete nouns? Kousta et al.'s [1] theory would suggest that emotionality, given its independence from concreteness, would indeed play a facilitatory role in that case.

In this study, we therefore examined whether vocabulary learning in a native and foreign language can be facilitated by emotionality. We use the terms 'native' and 'foreign' rather than 'first language' and 'second language' because we focused on bilinguals living in a bilingual society who speak multiple languages in the community but also acquire another language (English) at school. In many (monolingual and bilingual) societies it is now common for children to acquire foreign languages at school. In some cases, the foreign language is taught through content courses in that language. Currently, most of the research addressing this type of learning in a foreign language comes from classroom studies on CLIL using children. This literature focuses on language outcomes more so than on content learning and has found mostly positive results in this area [3–11]. The literature on adult foreign language medium learning is more limited but also focuses on benefits associated with language [12] and not content. The research on content learning is both more restricted and less clear-cut, with studies showing positive, negative, and null results [13]. Overall, studies tend to focus on the efficiency of courses in general, evaluating overall performance at the end of the term. Very few studies compare the immediate understanding and learning of new content in a foreign and a native language. Those who do have found either detrimental or null effects of foreign language use, depending on the task [14,15], particularly without foreign language support [16]. These results have been accounted for in the context of cognitive load theory, which suggests a working memory overload for individuals trying to learn content in a language they are not proficient in [16]. Importantly, contributing to this literature would improve our understanding of whether foreign and native language learning differ as well as providing insights into possible improvements on foreign language teaching methods.

As mentioned before, we will focus on emotionality as a potential way of improving word learning. In particular, emotionality provides greater richness to lexical items, aiding their processing [1] and increasing familiarity at the time of retrieval. This process is facilitated by an enhancement of attention during encoding, leading to more durable memory [17]. This idea has been supported by studies on language processing in participants' native language [17]. These studies have shown that emotionality facilitates learning in the native language—with emotional words having lower ages of acquisition—as well as aiding processing and memory [1]. But this phenomenon is still poorly understood, especially in the case of bilinguals. In addition, research so far has focused mostly on how we process or respond to known emotional words, but we do not have a clear idea of how learning in general and vocabulary learning in particular are affected by the valence of the new word or concept. Therefore, it is unclear whether we learn words that refer to new emotional concepts better or easier than those referring to new neutral concepts.

In the foreign language, the story is not as simple. This is in part because of a decrease in emotionality in the foreign language [18]. Focusing on memory for known words, the effects of this decreased emotionality are very inconsistent. Some studies have found enhanced memory for emotional as compared to neutral words in a foreign language [19–23] while others have not [24]. In addition, others have even found inconsistent [25] or inverse effects [26].

The literature on emotionality and word acquisition in a foreign language is quite scarce. Most studies have focused on acquiring new lexical forms for known concepts. For example,

Ferré, Ventura, Comesaña, and Fraga [25] studied new lexical form learning in a foreign language through word paring—associating a new item with the native language word. Results showed an improvement in later word recognition and translation for abstract words that were emotional (positive or negative). This shows that new labels for known concepts carry the same emotionality effects as the original terms, but in that context, emotionality played a role only in the acquisition of new forms for abstract words.

The question of whether emotionality affects word learning in a non-native language relates to the foreign language effect: the idea that people are less biased and more emotionally distant in their foreign language [27]. It is hypothesized that this effect may be a consequence of learning the foreign language in an emotionally attenuated academic environment [28]. Therefore, understanding how emotionality affects new word learning in a foreign language—i.e., new concepts with their associated form, not just learning new labels for known words or remembering known words—has implications for understanding the foreign language effect. To test this explanation, we have to manipulate the circumstances or context of the first exposure to a word or concept. In the current study, we do exactly that: we teach participants new concepts for which they do not have an equivalent native language word. This allows us to see what the effects of emotionality are as it is conveyed only by the definition of a word, without any prior exposure to the word or concept, or knowledge of the object.

Most studies so far have examined how emotionality affects the acquisition of new labels for known objects rather than learning of new words for new concepts. One notable exception is the study by Brase and Mani [29], which showed emotionality effects for new negative words in an emotional Stroop task and a sentence completion task. In the foreign language condition, this occurred only in the emotional context, whereas in the native language, they showed an effect in both emotional and neutral contexts. But, it is worth noting that their recall results are in contrast with prior studies, with no emotionality effect in the native language and a restricted effect in the foreign language—only for negative words in the emotional context.

Frances et al. [30] took a similar approach to the one in the present study, where participants learn new information—in their case new content, instead of new vocabulary—that was embedded in a positive or a neutral semantic context. This contrasts with prior studies that focused either on memory for known words or on the acquisition of new lexical items for known concepts. They found that positively valenced content was remembered better than neutral content and, importantly, that this effect occurred in the native and foreign languages equally. As we are using a similar strategy of teaching new information—in this case, new objects and their names—in the foreign or native language, we expect similar results. Namely, we expect enhanced memory for positive items and a similar improvement in both languages.

Current study

The foreign language literature has shown inconsistent effects of emotionality and is unclear about whether emotionality helps increase vocabulary in that language. Therefore, the purpose of the current study is to understand new concept acquisition, removing the influence of prior experience with that concept or item. Importantly, in the current study, new information is taught implicitly through descriptive texts, which are more naturalistic than single, isolated words. This provides a more accurate understanding of what happens in more realistic learning situations.

Other studies have manipulated the emotionality of the context through association with emotional images [31] or paralinguistic information [29]. For the purposes of the present work, we have operationally defined emotionality as positive valence and high arousal

manipulated through the semantic context of the new word. Therefore, we manipulated emotionality semantically by conveying descriptions of objects in either a positive or neutral manner.

Given that one of the possible applications for this research is classroom contexts with students of different ages, only positive and neutral descriptions were used, as negative materials would not be advisable for educational settings. Furthermore, some studies have found that negative stimuli do not produce the same effects as positive stimuli and can hinder performance [19,21], particularly for the surrounding neutral words as they capture attention, unlike positive words which aid performance in these cases [32–34].

In short, this study aims to understand the effects of positive valence—manipulated semantically—on vocabulary learning. One of the priorities in the current study is to use naturalistic classroom-type stimuli than most prior research. The experiments here use novel words and concepts conveyed through paragraph-long definitions. We would like to emphasize that although the results of this study may have practical applications, our goal is to understand how learning can be influenced from a theoretical point of view. The ultimate goal is to understand how we process and incorporate this type of information in a foreign language, in particular as it compares to our native language.

Experiment 1

Methods

Participants. Participants were recruited from the Participa platform at the Basque Center on Cognition, Brain and Language. This database includes extensive demographic information, information on language history and use, and any physical or cognitive impairments. All eligible participants received an email inviting them to participate. Participants were typical young adults from the Basque Country.

Fifty-six participants were tested, three participants were removed for low accuracy (< 50%) in the old/new recognition task (two from the English condition and one from the Spanish condition) and 10 were removed for low accuracy (< 33%) in the name matching task (five from the English condition and five from the Spanish condition). The remaining participants were 43 native Spanish speakers (24.14 years old, SD = 3.99) with an intermediate to high level of English (BEST—Basque, English, and Spanish Test—Interview scores of four or five out of five [35]).

Language was a between-subjects factor. Participants did only one language condition: the native language condition (i.e., Spanish: 23) or the foreign language condition (i.e., English: 20 participants). Groups were matched for age (see Table 1), gender (13 females in the Spanish

Table 1. Variables matched between groups in Experiment 1 and values for Experiment 2.

			English Assessment		LexTALE		IQ		Corsi Task	
		Age	BEST	AOA	Spanish	English	Verbal	Non-verbal	Backward	Forward
Exp. 1	English group	24 (3)	4.2 (0.4)	6.0 (1.9)	95% (5%)	72% (10%)	108 (9)	113 (8)	6.30 (1.56)	6.50 (1.43)
	Spanish group	25 (4)	4.2 (0.4)	6.4 (2.5)	92% (7%)	70% (11%)	110 (7)	112 (9)	6.48 (1.65)	6.70 (1.55)
	Total	24 (4)	4.2 (0.4)	6.2 (2.2)	93% (6%)	71% (10%)	109 (8)	112 (9)	6.40 (1.59)	6.61 (1.48)
	T-value df	-.979 (41)	-.054 (41)	-.572 (41)	1.54 (41)	.528 (41)	-.724 (41)	.307 (41)	-.363 (41)	-.427 (41)
	p value	.333	.957	.571	.132	.601	.473	.761	.719	.671
Exp. 2	Total	26 (6)	4.2 (0.4)	6.4 (2.5)	94% (6%)	72% (10%)	100 (30)	105 (31)	6.54 (1.59)	6.72 (1.57)

Note: Values reported are means with standard deviations in parentheses. AoA stands for age of acquisition. BEST refers to the scores on the English interview portion of the BEST [35].

https://doi.org/10.1371/journal.pone.0240252.t001

[Return to Summary of Publication 2 on page 27]

Table 2. Education level by group and experiment.

	Language in Experiment 1			Experiment 2
Educational Level	**English**	**Spanish**	**Total**	**Total**
High School	6	8	14	2
Postgraduate	2	4	6	9
Professional Training	2	2	4	3
University	10	9	19	30
Total	20	23	43	54

Note: Figures represent the number of participants in each category.

https://doi.org/10.1371/journal.pone.0240252.t002

[Return to Summary of Publication 2 on page 27]

group and 12 females in the English group), education level (see Table 2), student status (17 students in the Spanish group and 16 in the English group), verbal and nonverbal IQ (Kbit) [36], age of acquisition of English, and Spanish and English LexTALE [37,38] (see Table 2). In addition, all participants were trilingual, with the exception of two participants in the Spanish condition who were only bilingual.

The experiment was approved by the Ethical and Scientific committees at the Basque Center on Cognition, Brain and Language (BCBL) and all participants provided written consent to participate in the study (Approval number: 7209). They were compensated 8€ for their time.

Stimuli. The stimuli consisted of 46 images of invented objects that had been used in another study [39] and 92 disyllabic pseudowords (46 for names of objects, 46 for fillers used in testing phase—see S1 Table for the full list of pseudowords and their corresponding bigram frequencies by language). Object names were equally likely to be an English or Spanish word (e.g., chefio) as measured by average token bigram frequency for each word (M_{SP} = 491, SD_{SP} = 411; M_{EN} = 594, SD_{EN} = 374; $t(45) = 1.22$, $p = .230$, $BF_{01} = 3.14$, *error* % = 8.24 x 10^{-6}) using B-Pal for Spanish [40] and N-Watch for English [41].

The object names were presented in 46 descriptions of made-up objects, their origin, and their use (see Fig 1 for an example). A positive and a neutral version of each description was created. The details changed between versions, but there was always at least one characteristic in common between the two descriptions (e.g., "it is a type of maze"). Each description was translated to create an English and a Spanish version. Descriptions were matched on length by language (M_{Spa} = 45.2, SD_{Pos} = 3.21; M_{Eng} = 45, SD_{Neu} = 3.07; $t(45) = .349$, $p = .728$, BF_{01} = 5.90, *error* % = 1.15 x 10^{-5}) and emotionality (M_{Pos} = 46, SD_{Pos} = 3.82; M_{Neu} = 44.8, SD_{Neu} = 3.52; $t(45) = 1.62$, $p = .111$, $BF_{01} = 1.85$, *error* % = 7.93 x 10^{-4}). Paragraphs had a length of 39 to 54 words per description, $M = 45.7$, $SD = 3.79$. Length was also matched using a 2-way ANOVA with language and emotionality on length of paragraph (p's > .05, see Table 3).

Each description was controlled for word frequency to match emotion conditions. Word frequency was computed by taking the average frequency of every adjective, noun, adverb, and verb (excluding auxiliary and modal verbs) in the paragraph. For Spanish, we used EsPal [42] and matched the two emotionality conditions ($t(45) = .507$, $p = .614$, $BF_{01} = 5.54$, *error* % = 1.12 x 10^{-5}, see Table 3 for means and standard deviations). For English, we used the frequency from the Hyperspace Analogue to Language [43] using the English Lexicon Project [44] and matched the two emotionality conditions ($t(45) = .525$, $p = .602$, $BF_{01} = 5.49$, *error* % = 1.11 x 10^{-5}, see Table 3).

Positive and neutral descriptions had significantly different valence (ANEW [45]) for both Spanish ($t(45) = 4.26$, $p < .001$, $BF_{01} = 218$, *error* % = 2.41 x 10^{-5}) and English ($t(45) = 7.86$, $p < .001$, BF_{01} = 1.96 x 10^{7}, *error* % = 8.24 x 10^{-13}), see Table 3. Positive and neutral descriptions also had significantly different arousal (ANEW—Bradley & Lang [45]) for both Spanish ($t(45)$

Neutral	This is a chefio. The chefio is a type of maze that is used in research laboratories. The chefio has both closed and open parts so that people can search inside and orient outside. The chefio is a laboratory object.	Esto es un chefio. El chefio es un tipo de laberinto para laboratorios de investigación. El chefio tiene partes cerradas y abiertas para que la gente busque por dentro y se oriente al salir. El chefio es un objeto de laboratorio.
Positive	This is a chefio. The chefio is a type of maze used at birthday parties for children. The chefio has both closed and open parts providing the enjoyment and thrill of independence as well as safety of seeing their mother. The chefio is a game for children.	Esto es un chefio. El chefio es un tipo de laberinto para fiestas de cumpleaños infantiles. El chefio tiene partes cerradas y abiertas para dar el goce y la emoción de la independencia, así como la seguridad de ver a su madre. El chefio es un juego infantil.

Fig 1. Example stimulus. All four versions of an example stimulus (neutral English, neutral Spanish, positive English, and positive Spanish) as well as the corresponding image. Due to copyright restrictions, the image is not one of the ones from our stimulus set, but it is similar to those we used.

https://doi.org/10.1371/journal.pone.0240252.g001

[Return to Summary of Publication 2 on page 27]

= 5.01, $p < .001$, $BF_{01} = 2.02 \times 10^{3}$, *error* % = 2.46×10^{-6}) and English ($t(45) = 7.38$, $p < .001$, $BF_{01} = 4.19 \times 10^{6}$, *error* % = 1.85×10^{-12}), see Table 3. In addition, and in order to validate the emotionality (both valence and arousal) of the descriptions, we carried out a norming study with 18 participants from the same population as the study evaluating the descriptions in Spanish (see S1 Fig for instructions), on a scale from 0 to 4. The by item analysis found that items were considered significantly higher valenced in the positive version ($M_{Pos} = 3.27$, $SD_{Pos} = .335$) than the neutral version ($M_{Neu} = 2.21$, $SD_{Neu} = .352$), $t(45) = 15.48$, $p < .001$, $BF_{01} = 3.96 \times 10^{16}$, *error* % = 5.76×10^{-21}). Similarly, items were considered to cause significantly higher arousal in the positive version ($M_{Pos} = 1.32$, $SD_{Pos} = .454$) than the neutral version ($M_{Neu} = .459$, $SD_{Neu} = .394$), $t(45) = 8.56$, $p < .001$, $BF_{01} = 1.86 \times 10^{8}$, *error* % = 6.17×10^{-12}).

Table 3. Means and standard deviations for control measures on the descriptions.

		Frequency (words per million)	Words per description	Valence	Arousal
Spanish	Neutral	153 (142)	44.61 (3.69)	5.86 (1.05)	5.10 (.62)
	Positive	139 (108)	45.80 (4.07)	6.69 (1.11)	5.71 (.71)
English	Neutral	640 (258)	44.89 (3.72)	6.02 (.76)	4.60 (.56)
	Positive	610 (334)	45.72 (3.90)	6.98 (.71)	5.01 (.57)

Note: Values in parentheses are standard deviations.

https://doi.org/10.1371/journal.pone.0240252.t003

Procedure. Participants were recruited from a database with some basic questionnaires: K-Bit—Kaufman Brief Intelligence Test, [36]—, LexTALE in English [37] and Spanish [38], linguistic background, socioeconomic background, and experience with bilingual education. These were used to select participants and match the groups. Then, they had a learning and a testing phase, which were either in English or Spanish (language was a between-subjects factor).

The learning phase consisted of seeing the picture of each object along with its description (containing the object name 4 times), one by one. Participants were instructed to read for comprehension. After seeing the object and reading the description for a minimum of 15 seconds (no Next button visible until the timeout) participants were allowed to move on to the next screen at their own pace. Participants took an average of 27.10 seconds (SD = 6.99 seconds) to read each description. The next screen asked them to type in the object's name. If they typed it incorrectly, they were taken back to the previous screen and then asked to type it again. This happened very infrequently (on average for two words for every three participants) and equally in every condition ($M_{Neu\text{-}Eng}$ = .186 words, $SD_{Neu\text{-}Eng}$ = .546; $M_{Pos\text{-}Eng}$ = .163 words, $SD_{Pos\text{-}Eng}$ = .374; $M_{Neu\text{-}Spa}$ = .163 words, $SD_{Neu\text{-}Spa}$ = .374; $M_{Pos\text{-}Spa}$ = .163 words, $SD_{Pos\text{-}Spa}$ = .374).

After seeing all 46 descriptions (23 positive and 23 neutral descriptions randomly mixed), they moved on to a filler task in order to reduce recency effects and assess working memory. The filler task consisted of a Corsi task (forwards and backwards) [46] which was timed to last 15 minutes. The purpose of the filler task was to decrease serial position effects [47,48] and reflect only information that had been more permanently registered in memory [49].

After the filler task, the testing phase started. The first task consisted of a cued recall in which participants were shown the image of an object and asked to recall its name (Fig 2A). This task was not timed. The second task was an old/new recognition task in which participants were presented the 46 object names along with 46 fillers one at a time and had 3000ms to say whether they had seen the word before or not (Fig 2B). The main task we were interested in, third task, consisted of a cued name-matching recognition task in which participants had 5000ms to select the correct name of the objects out of three choices (Fig 2C). For this task, the two distractors were pseudowords that had also been presented during the learning phase chosen randomly and matched for emotionality. This was our main task of interest because it was fully language independent—only the image and the name were presented, which were the same in both language conditions—and thus theoretically equally difficult. Finally, the last task was an attentional check that consisted of matching the correct characteristic that goes with the object (Fig 2D), with 5000ms to respond. The characteristic that had been held constant in both versions of the object description—positive and neutral—was used for this task. The two fillers were chosen at random from the other emotion-matched objects.

The order of tasks was chosen so that the earlier tasks would not affect the later tasks and held constant across participants. The name recall task was placed first because it was likely to be influenced by seeing the words again in the recognition tasks, but was unlikely to affect the other tasks. Then, in order to prevent increased familiarity with the learned items, we kept the Old/New task—which is purely based on familiarity—before the name matching task. Given that the attentional check was not of theoretical interest, we kept that task last.

Results and discussion

The following are the linear mixed effect model analyses for both experiments. In all cases, analyses were run using linear mixed-effects models in R, lme4 [50] and lmerTest packages [51]. Both two-level categorical predictors (emotionality and language) were coded as -0.5 and

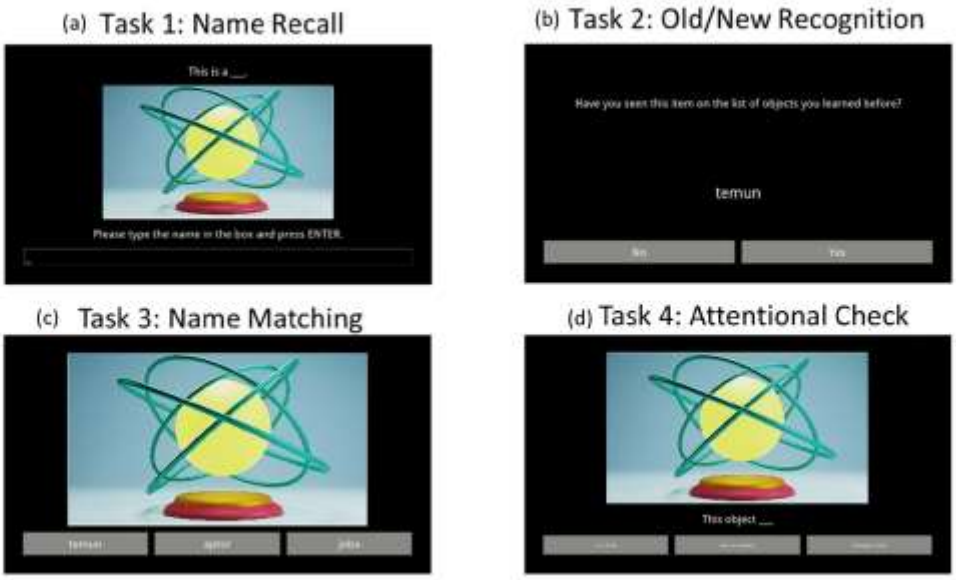

Fig 2. Example screen for each of the four tasks in Experiment 1. (a) Name recall task where participants were asked to type in the name of the object, (b) old/new recognition task where participants had to respond whether they had seen the word before or not, (c) name matching task where participants selected the correct name for the object, and (d) attentional check where participants selected the correct characteristic for the object. All responses were given using the keyboard (F and J keys for Old/New and A, F, and J keys for the matching tasks). Due to copyright restrictions, the image is not one of the ones from our stimulus set, but it is similar to those we used.

https://doi.org/10.1371/journal.pone.0240252.g002

[Return to Summary of Publication 2 on page 28]

0.5 (neutral/positive and English/Spanish). Subjects and items were included as random effects, but due to convergence issues, random slopes varied by analysis. Models on response times have the log transformation of response time as the dependent variable.

The models included all fixed effects of interest, as well as random intercepts for participants and items, unless otherwise stated. When models did not converge, all correlations between the random slopes and the random intercepts were removed [52]. If the model still did not converge after removing the random correlations, we built down the random effects structure by removing the item slopes that explained the least variance until convergence was reached and singular fits were fixed.

Only the LME results are reported here. For the ANOVA results and the comparison between the two, please see (S1 Appendix).

Name matching task. The first model had accuracy (0 incorrect; 1 correct) as the dependent variable and language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects (see Table 4 for means, standard deviations, and 95% confidence intervals and Fig 3A for percentage of errors by condition). In addition, we included the random intercepts for subject and item. In the model, there was a significant main effect of emotionality ($\beta = .217$, $SE = .092$, $z = 2.35$, $p = .019$), such that positive emotionality led to higher accuracy, and of language ($\beta = .245$, $SE = .092$, $z = 2.65$, $p = .008$), such that Spanish led to higher accuracy. There was no interaction between emotionality and language ($\beta = .020$, $SE = .183$, $z = .108$, $p = .914$).

Table 4. Means, standard deviations, and confidence intervals by participant for the name matching task by language and emotionality for Experiment 1.

	Accuracy		Response Time	
	Neutral	Positive	Neutral	Positive
English	0.42 (0.13) [0.36–0.47]	0.47 (0.11) [0.42–0.51]	2568.82 (669.92) [2275.21–2862.42]	2394.26 (604.78) [2129.2–2659.31]
Spanish	0.48 (0.11) [0.43–0.52]	0.54 (0.1) [0.49–0.58]	2514.84 (525.44) [2300.10–2729.58]	2358.78 (539.34) [2138.36–2579.20]

Note: Values in parentheses are standard deviations. The values in brackets are the 95% confidence intervals. Accuracy is in percent correct; response times are in milliseconds.

https://doi.org/10.1371/journal.pone.0240252.t004

The model with log transformed response time as the dependent variable had language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects (see Table 4 for means, standard deviations, and 95% confidence intervals and Fig 3B for average response times by condition). In addition, we included the random intercepts for subject and item and random slopes for emotionality by subject. In the model, there was no main effect of emotionality (β = -.042, SE = .027, t = -1.53, p = .137) or of language (β = -.005, SE = .025, t = -.215, p = .830) and no interaction between emotionality and language (β = -.023, SE = .047, t = -.496, p = .620). (See S1.2 Table in S1 Appendix for the 90% confidence intervals of the effect sizes.)

Given the null interactions, we followed up the analyses using Bayesian statistics on the by-subject averages. Analyses were run using JASP statistical software (version 0.12.2) [53]. We used the default settings for all model comparisons (the priors were r scale fixed effects of .5, r scale random effects of 1, and r scale covariates of .354). For the t-test the prior was a Cauchy distribution of scale .707.

First, we analyzed the effects on accuracy. We ran a comparison between a model predicting accuracy from Language, Emotionality, and Subject to one which also included the interaction (Language x Emotionality). The model with the additional factor—the interaction—showed evidence that the data was approximately three times more likely to be observed under the null model without the interaction compared to the one with it, BF_{01} = 3.33, *error* % = .485 (by item, BF_{01} = 4.65, *error* % = 2.26), suggesting that accuracy was not affected by an interaction between language and emotionality. In a similar vein, we ran a t-test comparing the emotionality effect (accuracy on the positive condition minus accuracy on the neutral condition) and found moderate evidence that the emotionality effect was in fact the same between languages, BF_{01} = 3.31, *error* % = .016 (by item, BF_{01} = 6.08, *error* % = 1.67 x 10^{-5}).

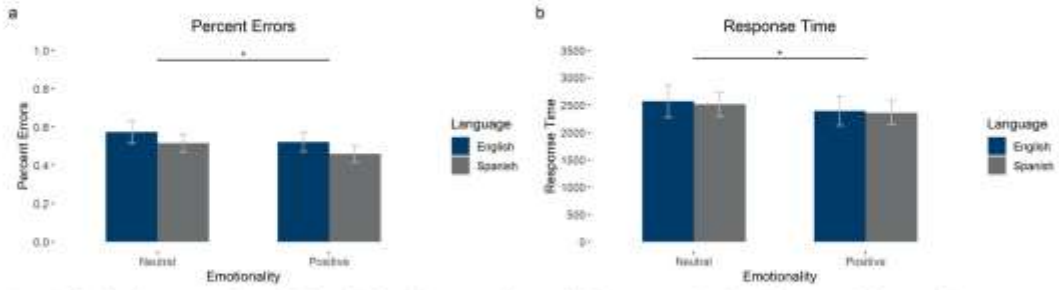

Fig 3. Results from the name matching task. The graphs show (a) percentage of errors and (b) response time. Error bars show 95% confidence intervals.

https://doi.org/10.1371/journal.pone.0240252.g003

[Return to Summary of Publication 2 on page 29]

We then ran the same analyses on response time. We ran a comparison between a model predicting response time from Language, Emotionality, and Subject to one which also included the interaction (Language x Emotionality). The model with the additional factor—the interaction—did not show an improvement, BF_{01} = 3.29, *error* % = .365 (by item, BF_{01} = 2.80, *error* % = 4.39), suggesting that the interaction did not have an effect on response time and that the data was moderately more likely to be observed under the model without the interaction rather than the one with it. In a similar vein, we ran a t-test comparing the emotionality effect (response time on the positive condition minus response time on the neutral condition) and found moderate evidence that the emotionality effect was in fact the same between languages, BF_{01} = 3.30, *error* % = .016 (by item, BF_{01} = 4.43, *error* % = 1.09 x 10^{-5}).

Name recall task. The name recall task was evaluated in two ways: exact correct recalls and Levenshtein distance (LD) [54] between the produced string and the original item. For the first analysis, the number of exact recollections (correct word with the correct object) were counted for each participant within each emotionality condition. As expected, given the difficulty of the task, recall was very low [14], on average fewer than two words (M = 1.91 words, SD = 2.75 words).

The first model had accuracy (0 incorrect; 1 correct) as the dependent variable and language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects (see Table 5 for means, standard deviations, and 95% confidence intervals by condition). In addition, we included the random intercepts for subject and item. In the model, there was no main effect of emotionality (β = .429, SE = .296, z = 1.45, p = .147) but there was one of language (β = 1.21, SE = .306, z = 3.94, $p < .001$), such that participants recalled more words in Spanish than in English, and no interaction between emotionality and language (β = .577, SE = .586, z = .985, p = .325).

Then, we calculated the normalized LD for each response. All responses that were shorter than 3 characters or contained a real word related to the object were removed as they were not considered real attempts. When calculating the LD, the number of insertions, deletions, and alterations needed to get from the produced word—the recalled string—to the original word—the studied string—is taken into account to calculate a standardized value. Then, these values are divided by the word length in order to normalize the value [55–57]. We also ran a model on the LD between the produced word and the correct response. The dependent variable was the normalized LD and language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction were the fixed effects. In addition, we included the random intercepts for subject and item. In the model, there was no main effect of emotionality (β = -.179, SE = .131, t = -1.30, p = .195) but there was one of language (β = -.374, SE = .142, t = -2.63, p = .008), such that participants produced strings closer to the correct one in Spanish than in English, and no interaction between emotionality and language (β = .082, SE = .249, t = .330, p = .742).

Old/new recognition task. The first model had accuracy (0 incorrect; 1 correct) as the dependent variable and language (English or Spanish) and emotionality (Neutral or Positive),

Table 5. Means, standard deviations, and confidence intervals by participant for the name recall task by language and emotionality for Experiment 1.

	Number of words		Normalized LD	
	Neutral	Positive	Neutral	Positive
English	0.45 (0.95) [0.04–0.86]	0.45 (0.69) [0.15–0.75]	0.76 (0.18) [0.68–0.83]	0.77 (0.18) [0.69–0.84]
Spanish	1.00 (1.48) [0.40–1.6]	1.78 (2.24) [0.87–2.70]	0.64 (0.23) [0.54–0.73]	0.60 (0.27) [0.49–0.71]

Note: Values are in number of words. Numbers in parentheses are standard deviations. The values in brackets are the 95% confidence intervals. Normalized LD values are in distance units with a range from 0 to 1, with 0 being the correct answer, identical to the target word, and 1 being a completely different word or no response.

https://doi.org/10.1371/journal.pone.0240252.t005

[Return to Summary of Publication 2 on page 29]

Table 6. Means, standard deviations, and confidence intervals by participant for the old/new task by language and emotionality for Experiment 1.

	D'		Response Time	
	Neutral	Positive	Neutral	Positive
English	0.78 (0.1) [0.74–0.83]	0.8 (0.06) [0.78–0.83]	1008.41 (189.65) [925.29–1091.52]	961.81 (191.79) [877.73–1045.87]
Spanish	0.81 (0.07) [0.78–0.84]	0.8 (0.08) [0.77–0.84]	1152.14 (392.57) [991.7–1312.57]	1098.78 (401.61) [934.64–1262.91]

Note: Values in parentheses are standard deviations. The values in brackets are the 95% confidence intervals. Response times are in milliseconds.

https://doi.org/10.1371/journal.pone.0240252.t006

as well as their interaction as fixed effects (see Table 6 for means, standard deviations, and confidence intervals by condition). In addition, we included the random intercepts for subject and item. In the model, there was no main effect of emotionality (β = -.085, SE = .103, z = -.829, p = .407) or of language (β = -.024, SE = .108, z = -.227, p = .820) and no interaction between emotionality and language (β = .115, SE = .204, z = .561, p = .575).

The model with log transformed response time as the dependent variable had language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects. In addition, we included the random intercept for subject but had to remove the random effects by item in order for the model to converge. In the model, there was a significant main effect of emotionality (β = -.046, SE = .018, t = -2.55, p = .011), such that positive emotionality led to faster response times, and of language (β = .079, SE = .020, t = 3.96, $p < .001$), such that Spanish led to higher accuracy. There was no interaction between emotionality and language (β = -.017, SE = .036, t = -.488, p = .626).

The effects of language and emotionality did not appear consistently in all measures—they were present for name matching accuracy, but not response time; old/new response time, but not accuracy; and name recall only showed an effect of language. Importantly, we consistently found no interaction between language and emotionality, but need further evidence to verify that this interaction in fact does not exist. In addition, the language comparison was between subjects and the task was quite difficult in general. This suggests that the results need to be conceptually replicated. The next experiment addresses this and provides a close conceptual replication of this experiment.

Experiment 2

The main goal of Experiment 2 was to replicate the results of Experiment 1 with a slightly different design. Besides that main goal, there were several minor goals. First, we wanted to have a fully within-subjects design in order to increase power, in case there was an interaction we were unable to detect before. Second, we wanted to improve the design and performance—in particular in the name recall task—by reducing the number of items to be learned per session. Although the total number of items was similar, participants learned only half in each of 2 sessions and the duration of the distractor task was reduced. Third, we wanted to remove the possible influence of the old/new task, which was not essential for answering our question. The attentional check also was not essential, but was maintained (at the end of the experiment) in order to (1) continue to assess whether participants had read the texts in their entirety and (2) to make sure that between session 1 and session 2 they were equally motivated to pay attention to the full texts, not just the object names.

Methods

Participants. Sixty participants from the same pool as Experiment 1 were tested. One participant was excluded for not following directions, four more were removed for low accuracy

in the name matching task (<50% accuracy in either language), and one more was removed for a technical error. The remaining participants were 54 (17 male) native Spanish speakers (25.98 years old, SD = 6.19) with an intermediate to high level of English (as in experiment 1). Participants were not significantly different (*p*'s > .05) from those in Experiment 1 on any of the following variables: age, gender, education level, student status, verbal and nonverbal IQ (Kbit) [36], age of acquisition of English, and Spanish and English LexTALE [37,38]. As in Experiment 1, all participants were at least trilingual, with the exception of three participants that were only bilingual.

The experiment was approved by the Ethical and Scientific committees at the BCBL and all participants provided written consent to participate in the study (Approval number: 11709). They were also compensated for their time with 12€ after finishing the second session.

Stimuli. Stimuli were the same as in Experiment 1, except for six items (bangel, dela, fortor, mova, testor, and trequi). These items were removed in order to reduce the stimuli to 10 items per condition (neutral Spanish, positive Spanish, neutral English, and positive English).

Procedure. Participants were recruited from the same database as in Experiment 1. Each participant completed two sessions: one in English (foreign language) and one in Spanish (native language). The two sessions were at least 5 days apart and the order of sessions was counterbalanced between participants. The learning phase was the same as in Experiment 1. Participants took an average of 32.07 seconds (SD = 8.57 seconds). Again, participants had to retype the words very infrequently and equally in every condition ($M_{Neu\text{-}Eng}$ = .63, $SD_{Neu\text{-}Eng}$ = 1.12; $M_{Pos\text{-}Eng}$ = .52, $SD_{Pos\text{-}Eng}$ = .86; $M_{Neu\text{-}Spa}$ = .50, $SD_{Neu\text{-}Spa}$ = 1.13; $M_{Pos\text{-}Spa}$ = .24, $SD_{Pos\text{-}Spa}$ = .61). After seeing the 20 descriptions (10 positive and 10 neutral descriptions randomly mixed) for that session, they moved on to a filler task—forwards Corsi in session 1 and backwards Corsi in session 2 [46]—which was programmed to last 7 minutes.

After the filler task, the testing phase started and concerned only the 20 items from that session. The first task consisted of a cued recall identical to that of Experiment 1 (Fig 4A). The second task consisted of a cued name-matching recognition task similar to that of Experiment 1, except that participants had 2500ms to select the correct name of the objects out of two choices (Fig 4B). The response options were reduced to 2 in order to increase the reliability of response times. For this task, the distractors were pseudowords that had also been presented during the learning phase, chosen randomly. For each item there were 2 trials: one with a distractor of equal emotionality and one with the distractor of opposite emotionality.

The last task, the attentional check, was modified to display the object name rather than the image when asking for the correct characteristic (Fig 4C). The reason behind this change is that some of the characteristics were quite visual, making the discrimination deducible by the image and not dependent on the description. In any case, this task was kept only as a check. Participants had only two response options and 3500ms to respond. As with the name matching task, there were 2 trials for each item, one with a distractor of equal emotionality and one with the distractor of opposite emotionality.

Results and discussion

All analyses were run as in Experiment 1. Only the LME results are reported here. For the ANOVA results and the comparison between the two, please see (S1 Appendix).

Name matching task. The first model had accuracy (0 incorrect; 1 correct) as the dependent variable and language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects (see Fig 5A and Table 7 for means, standard deviations, and 95% confidence intervals). In addition, we included the random intercepts for subject and item and the language slopes by participant (uncorrelated). In the model, there was a

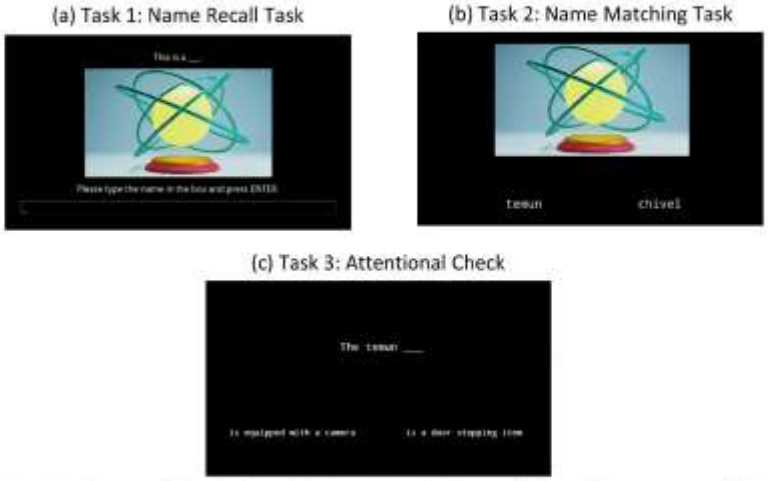

Fig 4. Example screen of each of the three tasks in Experiment 2. (a) Name recall task in which participants were asked to type-in the name of the object, (b) name matching task in which participants selected the correct name for the object, and (c) attentional check in which participants selected the correct characteristic for the object. In Tasks 2 and 3, participants responded using the keyboard keys F for left and J for right. Due to copyright restrictions, the image is not one of the ones from our stimuli, but it is similar to those we used.

https://doi.org/10.1371/journal.pone.0240252.g004 [Return to Summary of Publication 2 on page 28]

significant main effect of emotionality (β = .446, SE = .150, z = 2.98, p = .003), such that positive emotionality led to higher accuracy, but no effect of language (β = .170, SE = .135, z = 1.26, p = .209) and no interaction between emotionality and language (β = -.112, SE = .153, z = -.732, p = .464).

The model with log transformed response time as the dependent variable had language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects (see Fig 5B and Table 7 for means, standard deviations, and 95% confidence intervals). In addition, we included the random intercepts for subject and item and random slopes for emotionality by subject. In the model, there was a main effect of emotionality (β = -.065, SE =

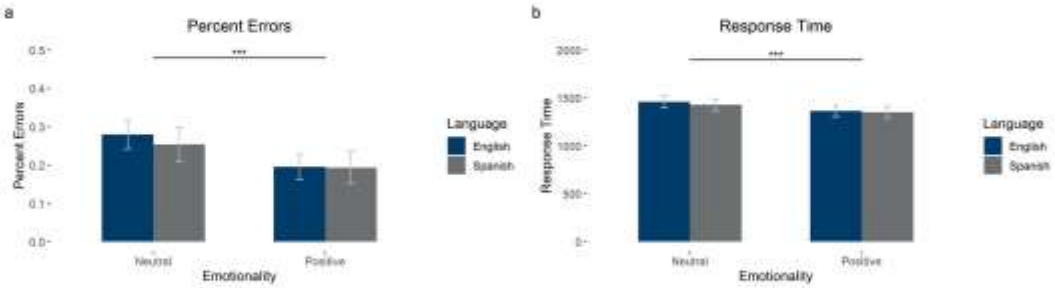

Fig 5. Results from the name matching task. The graphs show (a) percentage of errors and (b) response time in milliseconds. Error bars show 95% confidence intervals.

https://doi.org/10.1371/journal.pone.0240252.g005

[Return to Summary of Publication 2 on page 29]

Table 7. Means, standard deviations, and confidence intervals for the name matching task by language and emotionality for Experiment 2.

	Accuracy		Response Time	
	Neutral	Positive	Neutral	Positive
English	0.72 (0.14) [0.68–0.76]	0.81 (0.12) [0.77–0.84]	1605.37 (225.09) [1545.33–1665.41]	1501.73 (219.57) [1443.16–1560.29]
Spanish	0.75 (0.17) [0.70–0.79]	0.81 (0.16) [0.76–0.85]	1587.76 (250.04) [1521.07–1654.45]	1499.36 (255.88) [1431.12–1567.61]

Note: Values in parentheses are standard deviations. The values in brackets are the 95% confidence intervals. Accuracy is in percent correct; response times are in milliseconds.

https://doi.org/10.1371/journal.pone.0240252.t007

.020, t = -3.21, p = .003), such that positive items were responded to faster, and of language (β = -.026, SE = .009, t = -3.00, p = .003), such that participants responded faster in Spanish than in English, but no interaction between emotionality and language (β = .007, SE = .017, t = .403, p = .687).

As in Experiment 1, we followed up the null interactions using Bayesian statistics on the by-subject averages using the same software and priors.

First, we analyzed the effects of the interaction on accuracy. We ran a comparison between a model predicting accuracy from Language, Emotionality, and Subject to one which also included the interaction (Language x Emotionality). The model with the additional factor—the interaction—showed evidence that the data was almost four times more likely to be observed under the model without the interaction compared to the one with it, BF_{01} = 3.84, *error* % = .50 (by item, BF_{01} = 2.13, *error* % = 1.22), suggesting that accuracy was not affected by an interaction between language and emotionality. We ran a t-test comparing the emotionality effect (accuracy on the positive condition minus accuracy on the neutral condition) and found moderate evidence that the emotionality effect was in fact the same between languages, BF_{01} = 5.40, *error* % = .009.

We then ran the same analyses on response time. We ran a comparison between a model predicting response time from Language, Emotionality, and Subject to one which also included the interaction (Language x Emotionality). The model without the additional factor—the interaction—was almost four times more likely compared to the one with it, BF_{01} = 3.81, *error* % = 3.75 (by item, BF_{01} = 2.81, *error* % = .563), suggesting that the interaction did not have an effect on response time and that the data was moderately more likely to be observed under the model without the interaction rather than the one with it. We ran a t-test comparing the emotionality effect (response time on the positive condition minus response time on the neutral condition) and found moderate evidence that the emotionality effect was in fact the same between languages, BF_{01} = 4.69, *error* % = .019.

Name recall task. The name recall task was evaluated as in Experiment 1. For the first analysis, the number of exact recollections (correct word with the correct object) were counted for each participant within each emotional condition. As expected, given the difficulty of the task, recall was very low [14]; on average fewer than four words out of the 40 (M = 3.80 words, SD = 3.80 words) were correctly recalled. The first model had accuracy (0 incorrect; 1 correct) as the dependent variable and language (English or Spanish) and emotionality (Neutral or Positive), as well as their interaction as fixed effects (see Table 8 for means, standard deviations, and confidence intervals). In addition, we included the random intercepts for subject and item. In the model, there was no main effect of emotionality (β = .385, SE = .281, z = 1.37, p = .170) nor of language (β = .195, SE = .153, z = 1.27, p = .203), and no interaction between emotionality and language (β = -.107, SE = .305, z = -.349, p = .727).

We also ran a model on the Levenshtein distance between the produced word and the correct response. The dependent variable was the normalized Levenshtein distance and language

Table 8. Means, standard deviations, and confidence intervals for the name recall task by language and emotionality for Experiment 2.

	Number		LD	
	Neutral	Positive	Neutral	Positive
English	0.74 (1.03) [0.47–1.01]	1.00 (1.26) [0.66–1.34]	0.8 (0.14) [0.76–0.83]	0.77 (0.15) [0.73–0.81]
Spanish	0.93 (1.3) [0.58–1.28]	1.13 (1.39) [0.76–1.5]	0.76 (0.16) [0.72–0.8]	0.72 (0.17) [0.68–0.77]

Note: Values are in number of words. Numbers in parentheses are standard deviations. The values in brackets are the 95% confidence intervals. Normalized LD values are in distance units with a range from 0 to 1, with 0 being the correct answer, identical to the target word, and 1 being a completely different word or no response.

https://doi.org/10.1371/journal.pone.0240252.t008

[Return to Summary of Publication 2 on page 29]

(English or Spanish) and emotionality (Neutral or Positive), as well as their interaction were the fixed effects. In addition, we included the random intercepts for subject and item. In the model, there was no main effect of emotionality (β = -.030, SE = .023, t = -1.29, p = .204) but there was one of language (β = -.043, SE = .013, t = -3.23, p = .001), such that participants produced strings closer to the correct one in Spanish than in English, and no interaction between emotionality and language (β = -.009, SE = .026, t = -.338, p = .735).

Attentional check. Finally, the average score on the attentional check for English was 73.2% (SD = 11.5%) and for Spanish it was 73.5% (SD = 11.0%). We used both a frequentist and a Bayesian paired samples t-test of language on accuracy by subject to verify that participants read equally well in both language conditions. The test showed moderate evidence that the null model was more likely than the alternative model, $t_1(53)$ = .22, p = .829, Cohen's D = .029, BF_{01} = 6.59, *error* % = .001, $t_2(38)$ = .64, p = .525, Cohen's D = .203, BF_{01} = 5.72, *error* % = 9.91×10^{-6}.

Discussion. Overall, the results of Experiment 2 show a more consistent effect of emotionality in recognition, but not in recall. The effect of emotionality appears in the name matching task, both for accuracy and response time, but not in the name recall task. The effects of language are not very consistent, they appear in the name recall task—but, only for the Levenshtein distance measure—and for response time in the name matching task. Importantly, there was consistently no interaction between language and emotionality.

General discussion

The motivation for this study was to shed light on a highly common phenomenon: learning new content in a foreign language. This occurs for example in CLIL-based school environments or in university programs taught in language other than the official one in the region. In particular, the focus was on using naturalistic materials and tests and to understand whether learning in a foreign language differs from learning in a native language. We focused on the effects of emotional context—here defined as positive valence and high arousal—during the process of learning new words (in this case, pseudowords) attached to new content, here represented by novel objects.

Importantly, all of the strings that were learned had no prior associations for the participants, as they corresponded to concepts invented for the purpose of this study, and they were equally likely in both languages. This means that any emotionality effects were caused directly by the semantic context created in the experiment, removing the possible impact of a priori developed language-emotion connections. We focused specifically on the acquisition of lexical forms and their association with an object image. Furthermore, these new words or strings were embedded in descriptions or definitions of the objects they represented, much like how we normally acquire new information.

We hypothesized that new lexical items would be more difficult to learn in a foreign language—even though the target stimuli were exactly the same between language conditions—

and that positive emotionality would help compensate for this difficulty. This was addressed through two experiments in which participants carried out a name matching task between the images of the objects they studied with their names (among other tasks). This task tested the associations that were created during the learning phase. We focused on this task because it was language independent, and thus exactly the same for both language conditions and not as taxing for participants. In addition, participants did a lexical decision task (old/new recognition task) in Experiment 1 and a name recall task in both experiments.

In Experiment 1, we found foreign language effects in the three main tasks: name matching accuracy, old/new response time, and name recall (both accuracy and Levenshtein distance). The critical task—the name matching task—showed a language effect on accuracy in Experiment 1, when there was more information to be remembered. This effect was present in response times—but not accuracy—in Experiment 2, when the task was made easier and the to-be-remembered information was reduced. These results align with prior studies that found improved performance in the native language when different pieces of information need to be associated [58], but the diverging findings depending on the tasks and measures used also provide a possible explanation as to why other studies do not find the same language effects [3,5,9,59]. Although there were several differences between the two versions of the task, most differences related to making the task less demanding, for example by reducing the number of items to be remembered, the retention period, the number of tasks, and the number of answer options. Given that the language effects remained in the Levenshtein measure in recall—with recall being the most demanding of the tasks—, one could interpret that the difficulty of the task overall affected accuracy disproportionately in the foreign language. It is worth noting that, although the two groups in Experiment 1 were very well matched, Experiment 2 was within subjects, which could have also influenced the results. Nevertheless, given that the only task in Experiment 2 that showed a language effect was the most difficult one—namely, recall —, this would support our idea that difficulty drove the effect. In order to conclude this with certainty, this hypothesis needs to be explored further in future studies.

With respect to emotionality, we found that positive semantic contexts aided recognition memory and, in particular, in the name matching task. These data fit well with preceding studies showing that learning performance is better when the information they are tested on is positive [18]. These effects of positive semantic context were present in both experiments, but absent in the recall tasks—possibly due to a lack of power in this task. In this case, unlike with the effect of foreign language, it was not modulated by the amount of information that had to be learned. This means that it is a more stable effect that is robust to some manipulations of other variables—such as language (as we discuss below) or amount of information. This suggests that this manipulation—positive valence and arousal of the semantic context—can be used in several different circumstances to increase familiarity with learned items. It should be noted though, that these results do not inform our understanding of the effects of negative valence on word learning and are not definitive with respect to recall.

When analyzing the size of the observed effects, the partial eta squared of the effects found in this study are considered to be medium to large effects [60]. Depending on the task and experiment, 7 to 46% of the variance was explained by emotionality (see S1 Appendix). For accuracy in the name matching task, this was 13% in Experiment 1 and 34% in Experiment 2 —with the latter more than doubling the 14% benchmark of a large effect. For response times in that task, effect sizes were 15% for Experiment 1 and 46% in Experiment 2, both considered large effects. In practical terms, emotionality lead to a 5 (Experiment 1) and 7% (Experiment 2) difference in accuracy, which equated to roughly remembering 2 to 3 items more. Although a relatively discrete boost in performance, this could be the difference between passing and failing an exam for a student that is struggling in a class and could become more meaningful as

it accumulates. With respect to response times, there was a 165 ms (Experiment 1) and 98 ms (Experiment 2) difference for this task. In more general terms, this study suggests that emotionally loaded semantic contexts—not just emotional content—conveying new pieces of information can improve short term recognition.

With respect to the relationship between emotionality and foreign language, we found no interaction in either accuracy or response times in the name matching task—or in any of the other tasks or measures—in either experiment. This means that, although learning new words and concepts may be more demanding in a foreign language—depending on the task—, the effects of emotionality remain constant across languages. Similar results have been found by other studies in the context of memory for known words [19–23], but the current study is the first to demonstrate this in the acquisition of new words and concepts as well as using more naturalistic stimuli.

These results may seem surprising, as the foreign language effect suggests that people are less emotional—and thus less affected by emotional biases—in a foreign language [61]. Nevertheless, our results do not contradict this idea, but rather expand on possible causes for the effect as they suggest that these biases are not present during acquisition. In other words, in the first encounter with a word or concept, the emotional context around it can have the same effect in the foreign as in the native language. Therefore, the later observed effects might relate to the way in which this information or vocabulary was acquired, rather than with the use of a foreign language itself. In addition, our results have ramifications for our understanding of how we learn a foreign language. In particular, they support the idea that the reduced emotionality found in foreign language contexts has its origins in the way we learn a foreign language —namely, in a classroom and not in our more informal social environment—, rather than being a characteristic of how we process languages that are not our native one [62]. Similar claims have in fact been made by other authors [28]. Although practical applications of this research can be easily extracted, this is not our main goal. Therefore, we only provide the suggestion of looking into manipulating emotionality and semantic context as a way to improve learning, but do not provide practical advice for teachers.

It is worth noting that our study focuses on acquisition for the lexical item in the short term. Although we did not test indicators of word acquisition beyond the item itself and its association with a visual object, we provide evidence on the essential first steps towards word learning. Future research should expand our results to other areas of word acquisition such as semantics and word use. Despite the limitations of the experiments, this study provides ecological validity insofar as participants were taught new words in the same way they might learn them naturally in multilingual classroom contexts: by reading a textbook or asking for the definition of the word. Nevertheless, further research using even more naturalistic situations—e.g., a classroom setting, but maintaining the same level of experimental control—, is necessary in order to understand learning in these situations in more depth.

Conclusions

The current results get us one step closer to understanding the intricacies of language learning by non-balanced bilinguals and the complex relationship between emotionality and each of the languages they know. Our results point to an increased importance of chunking information—or splitting it up into multiple sessions—in the foreign language with respect to the native language. We highlight the importance of emotional semantic context when learning new words, equally in both the native and foreign language. And, finally, although not our main goal, this provides a practical teaching tool, where simply by adding a positive valence to texts, teachers might be able to help students improve their memory and learning of new vocabulary.

Supporting information

S1 Fig. Example trial from norming task. Each description was given followed by the questions "How positive or negative is the description of this object?" [¿Cómo de positiva o negativa es la descripción de este objeto?] and "How intense is the description of this object?" [¿Cómo de intensa es la descripción de este objeto?] with a clarification underneath saying "Please select how intense the emotional activation you feel is, regardless of whether the description is positive or negative" [Por favor selecciona cómo de intensa es la activación emocional que sientes, más allá de que la descripción sea positiva o negativa]. After each question there was a dropdown menu. In the valence question the answer options were very negative, somewhat negative, neutral, somewhat positive, and very positive [muy negativa, algo negativa, neutral, algo positiva, muy positiva]. For the arousal question, the options were: not at all intense, a little intense, somewhat intense, very intense, extremely intense [para nada intensa, un poco intensa, bastante intensa, muy intensa, extremadamente intensa].
(PDF)

S1 Table. Pseudowords and their average bigram frequencies. The following is a list of all the pseudowords used for the main tasks—learning and testing—by type, i.e. whether they were used as learn/test stimuli or as foils. Reported are the average token-type bigram frequencies for the word as reported by B-Pal [40] and N-Watch software [41].
(PDF)

S1 Appendix. ANOVAs and comparison with LME results.
(PDF)

Acknowledgments

We thank Julen Cristti© for the creation of part of the materials that were used in this set of experiments.

Author Contributions

Conceptualization: Candice Frances, Angela de Bruin, Jon Andoni Duñabeitia.

Data curation: Candice Frances.

Formal analysis: Candice Frances.

Methodology: Candice Frances, Angela de Bruin, Jon Andoni Duñabeitia.

Supervision: Angela de Bruin, Jon Andoni Duñabeitia.

Visualization: Candice Frances.

Writing – original draft: Candice Frances.

Writing – review & editing: Candice Frances, Angela de Bruin, Jon Andoni Duñabeitia.

References

1. Kousta ST, Vigliocco G, Vinson DP, Andrews M, Del Campo E. The representation of abstract words: Why emotion matters. J Exp Psychol Gen. 2011; 140(1):14–34. https://doi.org/10.1037/a0021446 PMID: 21171803
2. Kousta ST, Vinson DP, Vigliocco G. Emotion words, regardless of polarity, have a processing advantage over neutral words. Cognition. 2009; 112(3):473–81. https://doi.org/10.1016/j.cognition.2009.06.007 PMID: 19591976

3. Admiraal W, Westhoff G, De Bot K. Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English 1. Educ Res Eval. 2006; 12(1):75–93.
4. Aguilar M, Rodríguez R. Lecturer and student perceptions on CLIL at a Spanish university. Int J Biling Educ Biling. 2012; 15(2):183–97.
5. Bergroth M. Immersion students in the matriculation examination three years after immersion. In: Björklund S, Mard-Miettinen K, Bergström M, Södergard M, editors. Exploring Dual-Focussed Education: Integrating language and content for individual and societal needs [Internet]. Vaasan yliopisto; 2006. p. 123–34. Available from: http://www.uva.fi/materiaali/pdf/isbn_952-476-149-1.pdf
6. Dalton-Puffer C. Discourse in Content and Language Integrated Learning (CLIL) Classrooms [Internet]. John Benjamins Publishing. John Benjamins Pub.; 2007. (Language Teachning; vol. 20). Available from: https://books.google.es/books?id=Y-deXDJJ9-gC
7. Jiménez Catalán RM, Ruiz de Zarobe Y. The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In: Content and language integrated learning: Evidence from research in Europe. 2009. p. 81–92.
8. Ouazizi K. The effects of CLIL education on the subject matter (Mathematics) and the target language (English). Lat Am J Content Lang Integr Learn. 2016; 9(1):110–37.
9. Serra C. Assessing CLIL at primary school: A longitudinal study. Int J Biling Educ Biling. 2007; 10 (5):582–602.
10. Xanthou M. The impact of CLIL on L2 vocabulary development and content knowledge. English Teach. 2011; 10(4):116–26.
11. Dallinger S, Jonkmann K, Hollm J, Fiege C. The effect of content and language integrated learning on students' English and history competences—Killing two birds with one stone? Learn Instr [Internet]. 2016; 41:23–31. Available from: http://dx.doi.org/10.1016/j.learninstruc.2015.09.003
12. Yang S-W. The study of variables influencing the effect of English medium instruction on academic content learning and English proficiency development. The Pennsylvania State University; 2014.
13. Dalton-Puffer C. Content-and-language integrated learning: From practice to principles? Annu Rev Appl Linguist. 2011; 31:182–204.
14. Vander Beken H, Brysbaert M. Studying texts in a second language: The importance of test type. Bilingualism. 2018; 21(5):1062–74.
15. Dirix N, Vander Beken H, De Bruyne E, Brysbaert M, Duyck W. Reading Text When Studying in a Second Language: An Eye-Tracking Study. Read Res Q. 2019;1–27.
16. Roussel S, Joulia D, Tricot A, Sweller J. Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. Learn Instr. 2017; 52:69–79.
17. Bennion KA, Ford JH, Murray BD, Kensinger EA. Oversimplification in the study of emotional memory. J Int Neuropsychol Soc. 2013; 19(9):953–61. https://doi.org/10.1017/S1355617713000845 PMID: 24007950
18. Caldwell-Harris CL. Emotionality differences between a native and foreign language: Theoretical implications. Front Psychol. 2014; 5(1055):1–4.
19. Ayçiçeği A, Harris CL. Bilinguals' recall and recognition of emotion words. Cogn Emot [Internet]. 2004; 18(7):977–87. Available from: : http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2004-19292-006&site=ehost-live&scope=site%5Cnhttp://charris@bu.edu%5Cnhttp://aycicegi1996@hotmail.com
20. Ayçiçeği-Dinn A, Caldwell-Harris CL. Emotion-memory effects in bilingual speakers: A levels-of-processing approach. Biling Lang Cogn [Internet]. 2009; 12(3):291–303. Available from: http://journals.cambridge.org
21. Ferré P, Anglada-tort M, Guasch M. Processing of emotional words in bilinguals: Testing the effects of word concreteness, task type and language status. Second Lang Res. 2018; 34(3):371–94.
22. Ferré P, Garcia T, Fraga I, Sanchez-Casas R, Molero M. Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? Cogn Emot [Internet]. 2010; 24(5):760–85. Available from: http://www.informaworld.com/openurl?genre=article&doi=10.1080/02699930902985779&magic=crossref%7C%7CD404A21C5BB053405B1A640AFFD44AE3
23. Ponari M, Rodriguez-Cuadrado S, Vinson D, Fox N, Costa A, Vigliocco G. Processing advantage for emotional words in bilingual speakers. Emotion. 2015; 15(5):644–52. https://doi.org/10.1037/emo0000061 PMID: 25893450
24. Ancoshian LJ, Hertel PT. Emotionality in free recall: Language specificity in bilingual memory. Cogn Emot [Internet]. 1994; 8(6):503–14. Available from: http://digitalcommons.trinity.edu/psych_faculty
25. Ferré P, Ventura D, Comesaña M, Fraga I. The role of emotionality in the acquisition of new concrete and abstract words. Front Psychol. 2015; 6(976):1–10.

26. Altarriba J, Basnight-Brown DM. The Acquisition of Concrete, Abstract, and Emotion words in a Second Language. Int J Biling. 2011; 16(4):446–52.

27. Costa A, Vives M-L, Corey JD. On language processing shaping decision making. Curr Dir Psychol Sci [Internet]. 2017; 26(2):146–51. Available from: https://doi.org/10.1177/0963721416680263

28. Costa A, Foucart A, Arnon I, Aparici M, Apesteguia J. "Piensa" twice: On the foreign language effect in decision making. Cognition. 2014; 130(2):236–54. https://doi.org/10.1016/j.cognition.2013.11.010 PMID: 24334107

29. Brase J, Mani N. Effects of learning context on the acquisition and processing of emotional words in bilinguals. Emotion. 2017; 17(4):628–39. https://doi.org/10.1037/emo0000263 PMID: 27991815

30. Frances C, De Bruin A, Duñabeitia JA. The influence of emotional and foreign language context in content learning. Stud Second Lang Acquis. 2019;

31. Erk S, Kiefer M, Grothe J, Wunderlich AP, Spitzer M, Walter H. Emotional context modulates subsequent memory effect. Neuroimage. 2003; 18(2):439–47. https://doi.org/10.1016/s1053-8119(02)00015-0 PMID: 12595197

32. Dolcos F, Katsumi Y, Denkova E, Dolcos S. Factors Influencing Opposing Effects of Emotion on Cognition: A Review of Evidence from Research on Perception and Memory. In: Opris I, Casanova MF, editors. The Physics of the Mind and Brain Disorders Springer Series in Cognitive and Neural System, vol 11. Cham: Springer; 2017. p. 297–341.

33. Hurlemann R, Hawellek B, Matusch A, Kolsch H, Wollersen H, Madea B, et al. Noradrenergic Modulation of Emotion-Induced Forgetting. J Neurosci. 2005; 25(27):6343–9. https://doi.org/10.1523/JNEUROSCI.0228-05.2005 PMID: 16000624

34. Iordan AD, Dolcos F. Brain Activity and Network Interactions Linked to Valence-Related Differences in the Impact of Emotional Distraction. Cereb Cortex. 2017; 27(1):731–49. https://doi.org/10.1093/cercor/bhv242 PMID: 26543041

35. de Bruin A, Carreiras M, Duñabeitia JA. The BEST dataset of language proficiency. Front Psychol. 2017; 8.

36. Kaufman AS, Kaufman NL. Kaufman Brief Intelligence Test. Encyclopedia of Special Education: A Reference for the Education of Children, Adolescents, and Adults with Disabilities and Other Exceptional Individuals. En. 2013.

37. Lemhöfer K, Broersma M. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. Behav Res Methods. 2012; 44(2):325–43. https://doi.org/10.3758/s13428-011-0146-0 PMID: 21898159

38. Izura C, Cuetos F, Brysbaert M. Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. Psicológica. 2014; 35:49–66.

39. Antón E, Thierry G, Duñabeitia JA. Mixing Languages during Learning? Testing the One Subject—One Language Rule. PLoS One. 2015; 10(6):1–45.

40. Davis CJ, Perea M. BuscaPalabras: A Program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. Behav Res Methods. 2005; 37(4):665–71. https://doi.org/10.3758/bf03192738 PMID: 16629030

41. Davis CJ. N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. Behav Res Methods. 2005; 37(1):65–70. https://doi.org/10.3758/bf03206399 PMID: 16097345

42. Duchon A, Perea M, Sebastián-Gallés N, Martí A, Carreiras M. EsPal: One-Stop Shopping for Spanish Word Properties. Behav Res Methods. 2013; 45(4):1246–58. https://doi.org/10.3758/s13428-013-0326-1 PMID: 23468181

43. Lund K, Burgess C. Producing High-Dimensional Semantic Space from Lexical Co-Occurence. Behav Res Methods, Instruments, Comput. 1996; 28(2):203–8.

44. Balota D, Yap MJ, Cortese MJ, Hutchinson KA, Kessler B, Loftis B, et al. The English Lexicon Project. Behav Res Methods. 2007; 39(3):445–59. https://doi.org/10.3758/bf03193014 PMID: 17958156

45. Bradley MM, Lang PJ. Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. 1999.

46. Milner B. Interhemispheric Differences in the Localization of Psychological Processes in Man. Br Med Bull. 1971; 27(3):272–7. https://doi.org/10.1093/oxfordjournals.bmb.a070866 PMID: 4937273

47. Cohen RL. The Effects of Interference Tasks on Recency in the Free Recall of Action Events. Psychol Res. 1989; 51(4):176–80. https://doi.org/10.1007/BF00309145 PMID: 2616696

48. Postman L, Phillips LW. Short-Term Temporal Changes in Free Recall. Q J Exp Psychol [Internet]. 1965; 17(2):132–8. Available from: http://www.tandfonline.com/doi/abs/10.1080/17470216508416422

49. Atkinson R. C, Shiffrin RM. The Control of Short-Term Memory. Sci Am [Internet]. 1971; 225(2):1–43. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.2237&rep=rep1&type=pdf

50. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015; 67(1):1–48.

51. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. J Stat Softw. 2017; 82(13):1–26.

52. Barr DJ. Random effects structure for testing interactions in linear mixed-effects models. Front Psychol. 2013; 4(328):3–4.

53. JASP Team. JASP (Version 0.9). 2018.

54. Kruskal JB. An Overview of Sequence Comparison: Time Warps, String Edits, and Macromolecules. SIAM Rev [Internet]. 1983; 25(2):201–37. Available from: http://www.jstor.org/stable/info/2030214

55. Casaponsa A, Antón E, Pérez A, Duñabeitia JA. Foreign Language Comprehension Achievement: Insights from the Cognate Facilitation Effect. Front Psychol. 2015;

56. Duñabeitia JA, Ivaz L, Casaponsa A. Developmental Changes Associated with Cross-language Similarity in Bilingual Children. J Cogn Psychol [Internet]. 2016 Jan 2; 28(1):16–31. Available from: http://dx.doi.org/10.1080/20445911.2015.1086773

57. Duñabeitia JA, Dimitropoulou M, Morris J, Diependaele K. The Role of Form in Morphological Priming: Evidence from Bilinguals. Lang Cogn Process. 2013; 28(7):967–87.

58. Nott CR, Lambert WE. Free recall of bilinguals. J Verbal Learning Verbal Behav [Internet]. 1968; 7 (6):1065–71. Available from: https://search.proquest.com/docview/1297342710?accountid=17248

59. Stohler U. The acquisition of knowledge in bilingual learning: An empirical study on the role of language in content learning. VIEWS-Vienna english Work Pap [Internet]. 2006; 15(3):41–6. Available from: https://boris.unibe.ch/18531/1/views15_3_Stohler.pdf

60. Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. Educ Res Rev [Internet]. 2011; 6(2):135–47. Available from: http://dx.doi.org/10.1016/j.edurev.2010.12.001

61. Keysar B, Hayakawa SL, An SG. The foreign-language effect: Thinking in a foreign tongue reduces decision biases. Psychol Sci [Internet]. 2012; 23(6):661–8. Available from: http://pss.sagepub.com/https://doi.org/10.1177/0956797611432178 PMID: 22517192

62. Duñabeitia JA. Emotional Diglossia in Multilingual Classroom Environments: A Proposal. Psychol Cogn Sci Open J. 2017; 3(3):74–8.

Publication 3:

THE EFFECTS OF CONTEXTUAL DIVERSITY ON INCIDENTAL VOCABULARY LEARNING IN THE NATIVE AND A FOREIGN LANGUAGE

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020) The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports (Nature Publisher Group), 10*(13967). https://doi.org/10.1038/s41598-020-70922-1

JOURNAL QUALITY INDICATORS:

SCIENTIFIC REPORTS

Scopus CiteScore 2019: 7.2

SCImago Journal Rank 2019: 1.341

SNIP 2019: 1.365

CiteScore rank 2019: D1 (8/111) for Multidisciplinary. H-Index (source: SCImago): 179

www.nature.com/scientificreports

SCIENTIFIC REPORTS
natureresearch

OPEN

The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language

Candice Frances[1,2]✉, Clara D. Martin[1] & Jon Andoni Duñabeitia[3,4]

Vocabulary learning occurs throughout the lifespan, often implicitly. For foreign language learners, this is particularly challenging as they must acquire a large number of new words with little exposure. In the present study, we explore the effects of contextual diversity—namely, the number of texts a word appears in—on native and foreign language word learning. Participants read several texts that had novel pseudowords replacing high-frequency words. The total number of encounters with the novel words was held constant, but they appeared in 1, 2, 4, or 8 texts. In addition, some participants read the texts in Spanish (their native language) and others in English (their foreign language). We found that increasing contextual diversity improved recall and recognition of the word, as well as the ability to match the word with its meaning while keeping comprehension unimpaired. Using a foreign language only affected performance in the matching task, where participants had to quickly identify the meaning of the word. Results are discussed in the greater context of the word learning and foreign language literature as well as their importance as a teaching tool.

Vocabulary learning is an essential aspect of language that continues throughout the lifespan. To a large extent, the vocabulary we incorporate comes from incidental learning during reading[1,2] rather than explicit effort. This becomes particularly relevant when learning a new language, where a large amount of vocabulary must be acquired very quickly and partially without supervision. Following this reasoning, several studies have shown that it is possible to learn vocabulary implicitly through reading in our foreign language[3–7]. In addition, we know that people incorporate new lexical forms with as little as one exposure in their native language and as little as two exposures in their foreign language[4], but that learning improves with exposure to multiple instances of the word[8].

Implicit word learning in context differs somewhat between one's native and foreign languages. In particular, reading times for new words in our native language decrease significantly after the first exposure—suggesting some level of incorporation of the lexical item—whereas for the foreign language this happens only after two to four exposures[4]. This might relate to the fact that it is more difficult to extract the meaning of words from context in a foreign language. This is in part because these skills seem to be affected by knowledge of the language and experience in that particular task[9,10]. Although the number of times people encounter a word affects how well they learn and remember it[5,8], there is not much literature on how spreading these encounters across passages affects learning. One of the ways in which this spread is quantified is through contextual diversity—namely, the number of texts in which a word appears in a database[11,12]. This variable can be used to describe the influence of context beyond the mere number of occurrences or the frequency with which we encounter a given word. Context affects learning of new information[7], in general, and words[2], in particular. Contextual diversity specifically has a strong effect on learning[13,14], as well as on the processing of words, decreasing reaction time in word recognition[11,15]. The effects of contextual diversity have received increasing amounts of attention as they have been found in several domains including spoken word recognition[16] and serial recall performance with written words[17].

Word frequency refers to the number of times a word appears in a database, which naturally is highly correlated with the number of texts it appears in[11], and has been better studied than contextual diversity. Although word frequency has historically been considered a significant predictor of performance in various

[1]Basque Center on Cognition, Brain and Language, BCBL, Donostia, Spain. [2]Department of Social Sciences and Law, UPV/EHU, Donostia, Spain. [3]Centro de Ciencia Cognitiva – C3, Nebrija University, Madrid, Spain. [4]Department of Language and Culture, The Arctic University of Norway, Tromsø, Norway. ✉email: candice.frances@ncf.edu

language-related tasks, such as word learning[8], lexical access[18], and serial recall performance[17], recent studies have questioned this effect. These studies suggest that contextual diversity might be the factor responsible for some of the effects initially attributed to word frequency[12,18,19], as in some cases it explains more variance than word frequency, rendering the later a non-significant predictor.

Although word frequency and contextual diversity are highly correlated, they show different ERP signatures[19], suggesting different underlying brain processes, and in some cases, they show opposite behavioral effects. For example, there are cases in which recall is lower for words with high contextual diversity (showing a salience effect) but better for those with high word frequency (showing a benefit of multiple exposures to the word). Furthermore, word frequency—but not contextual diversity—predicts order error, with a higher number of errors for low frequency words, but not for low contextual diversity words[17]. This suggests that the effects of word frequency and contextual diversity may be differentiable in some contexts. In sum, the importance of contextual diversity above and beyond word frequency should not be diminished and its relevance in the context of language learning is still not fully understood.

A recent study by Pagán and Nation[20] manipulated diversity experimentally by presenting novel words (low frequency unknown words) in repetitions of the same sentence or in different sentences. They found that diversity increased reading times during the learning phase and decreased them during the testing phase. They interpreted this as a processing advantage during testing for words learned in diverse contexts. Although this provided a good first approach to the problem, there were several limitations. Perhaps the more salient one is that repetition, information about the word, and diversity were confounded. By providing the same sentence repeatedly, the reading times for the sentence overall probably decreased, not because of increased comprehension or incorporation of the term but simply due to a repetition effect. In addition, by providing different sentences in the high diversity condition, more information was provided about the meaning of the word. Similarly, they define contextual diversity as repeating (low diversity) or changing (high diversity) sentences when the main and most common definition of this concept is document count[12,21]. In reality, even if a word is repeated in a text, it is not the sentence itself that is repeated, but rather just the word in a new sentence. Another recent study by Rosa, Tapia, and Perea[22] manipulated contextual diversity by presenting 3rd grade students with different types of texts and found similar improvements in performance on a later test, with higher diversity. In particular, they tested recall, recognition (in two tasks), and picture matching. These tests focused on behavioral differences and showed a consistent benefit in performance for high versus low contextual diversity. All-in-all, these studies set a clear precedent for the importance of assessing the impact of contextual diversity in processing and performance, and the current study is set on these grounds. Here, we specifically tested the relative impact of contextual diversity while keeping frequency constant in foreign language vocabulary learning. To do so, we created several texts incorporating a group of novel words (real words replaced by pseudowords, in this case) to be learned in a nonnative language.

As a first approach to how repetitions should be spread out in our native and foreign languages—namely, the effects of contextual diversity—to improve learning, we had participants read short fictional texts, either in their native or a foreign language. In doing so, we manipulated the contextual diversity of each word, which we defined as the number of texts (or short stories) in which the novel item was encountered. These ranged from very low (eight times in 1 text) to very high (once in each of 8 texts). This allowed us to see not only the effects of contextual diversity on learning, but also whether this affected learning in the native and a foreign language differently.

We expected that distributing exposures in more texts would increase retention overall, as has been seen in other works[4].Participants were also likely to do better in their native language simply because the ability to derive meaning from context relates to the depth of vocabulary knowledge in that language[10]. This should make it easier for participants to extract the meaning of the pseudowords and incorporate the lexical form. But, if the stimuli are well matched for language difficulty and predictability from context, we may not see overall language effects. Finally, given the increased difficulty in incorporating and making associations between lexical items[23] in our foreign language, closer repetitions or repetition clusters could help participants extract meaning and incorporate the lexical form in that language[4]. Therefore, spreading might be more beneficial in the native language whereas clustering could be better in the foreign language. On the other hand, if participants are using the same mechanisms in both languages and these are not affected by language ability, we should observe the same contextual diversity effects in both languages.

Our results have practical applications for foreign language vocabulary learning. On the one hand, this study helps determine the importance of spreading practice into several sections (i.e., high contextual diversity) or clustering it (i.e., low or medium contextual diversity). On the other hand, the current research assesses the differences (or lack thereof) between learning in our native or a foreign language.

Methods

Participants. Using GPower[24], we ran an a priori power analysis based on prior studies[22,25] and a medium size effect ($\eta_p^2 = 0.06$) to establish sample size. We determined a minimum requirement of 80 participants to have 80% power.

Participants were 88 native Spanish speakers (44 in each language group, 25 males, $M_{age} = 23.78$, $SD_{age} = 4.28$). These were recruited through the internal database at the Basque Center on Cognition, Brain and Language (BCBL) and randomly assigned to either the native language (NL) or foreign language (FL) condition. All participants completed a test of receptive English and Spanish vocabularies (LexTALE[26] and LexTALE-Esp[27]). We only included participants with minimum score of 60% in English (80% for Spanish), which is equivalent to a minimum of a B2 level (C1 for Spanish) according to the Common European Framework of reference for languages[26]. Participants also completed a test of productive vocabulary (BEST picture naming task[28]) and had a minimum score of 40 out of 65 for English (61 out of 65 for Spanish). Participants were asked to rate their

Variable	Overall	English	Spanish	Statistic	Bayes Factor
Age	23.78 (4.28)	23.91 (4.70)	23.66 (3.87)	$t(86) = 0.27$, $p = 0.79$	$BF_{01} = 4.34$, Error % = 0.03
English LexTALE	71.28 (8.13)	70.43 (8.14)	72.151 (8.13)	$t(86) = -0.90$, $p = 0.37$	$BF_{01} = 3.14$, Error % = 0.03
Spanish LexTALE	95.20 (4.05)	95.61 (4.02)	94.79 (4.08)	$t(86) = 0.94$, $p = 0.35$	$BF_{01} = 3.04$, Error % = 0.03
English BEST	51.10 (6.87)	50.82 (6.574)	51.39 (7.21)	$t(86) = -0.39$, $p = 0.70$	$BF_{01} = 4.20$, Error % = 0.03
Spanish BEST	64.72 (0.68)	64.66 (0.81)	64.77 (0.52)	$t(86) = -0.79$, $p = 0.43$	$BF_{01} = 3.42$, Error % = 0.03
English level (overall)	6.99 (1.89)	7.00 (2.04)	6.98 (1.76)	$t(81) = 0.06$, $p = 0.96$	$BF_{01} = 4.37$, Error % = 0.03
English level (reading)	7.48 (2.18)	7.39 (2.20)	7.57 (2.19)	$t(81) = -0.38$, $p = 0.71$	$BF_{01} = 4.11$, Error % = 0.03
Spanish level (overall)	9.21 (1.67)	9.27 (1.62)	9.14 (1.73)	$t(86) = 0.38$, $p = 0.70$	$BF_{01} = 4.21$, Error % = 0.03
Spanish level (reading)	8.98 (2.38)	9.11 (2.22)	8.84 (2.54)	$t(86) = 0.54$, $p = 0.60$	$BF_{01} = 3.95$, Error % = 0.03
AOA English	5.81 (2.08)	5.82 (1.81)	5.80 (2.35)	$t(86) = 0.05$, $p = 0.96$	$BF_{01} = 4.48$, Error % = 0.03
AOA Spanish	0.17 (0.65)	0.14 (0.63)	0.21 (0.67)	$t(86) = -0.49$, $p = 0.62$	$BF_{01} = 4.03$, Error % = 0.03
Daily usage English	14.82 (9.67)	13.90 (8.02)	15.71 (11.08)	$t(81) = -0.85$, $p = 0.40$	$BF_{01} = 3.19$, Error % = 0.03
Daily usage Spanish	59.32 (16.39)	58.86 (16.17)	59.77 (16.77)	$t(86) = -0.26$, $p = 0.80$	$BF_{01} = 4.36$, Error % = 0.03
Verbal IQ	101.60 (22.50)	101.21 (24.25)	102.00 (20.88)	$t(86) = -0.17$, $p = 0.87$	$BF_{01} = 4.43$, Error % = 0.03
Non-verbal IQ	108.60 (18.58)	108.55 (18.63)	108.66 (18.75)	$t(86) = -0.03$, $p = 0.98$	$BF_{01} = 4.48$, Error % = 0.03
Compound IQ	103.50 (21.62)	102.61 (23.98)	104.39 (19.21)	$t(86) = -0.38$, $p = 0.70$	$BF_{01} = 4.21$, Error % = 0.03
Gender	88 (63)	44 (32)	44 (31)	$X^2(1) = 0.06$, $p = 0.81$	$BF_{01} = 4.13$
Handedness	88 (13)	88 (8)	88 (5)	$X^2(1) = 0.74$, $p = 0.39$	$BF_{01} = 3.75$
Educational level	88	88	88	$X^2(3) = 0.40$, $p = 0.94$	$BF_{01} = 28.47$
Student status	88 (67)	44 (32)	44 (35)	$X^2(1) = 0.56$, $p = 0.45$	$BF_{01} = 3.12$

Table 1. Means, standard deviations, and statistics for participants. Values in the Overall, English, and Spanish columns represent means and standard deviations (in parentheses). For gender, values represent count and number of females (in parentheses). For handedness, values represent count and number of left-handed people (in parentheses). For student status, values represent count and number of participants currently enrolled at a university (in parentheses). For educational level, only total count is presented because this was not a dichotomous variable (levels: high school, professional training, university, and graduate school).

[Return to Summary of Publication 3 on page 31]

English and Spanish levels overall on a 1-to-10 scale as well as their reading skills in that language. They also reported their estimated age of acquisition of each language and had a minimum age of 3 years for English and a maximum of 3 years for Spanish. Participants reported their daily exposure to each language, their educational level (highest level of schooling achieved, in all cases at least high school), and student status. And, finally, we collected measures of verbal, nonverbal, and compound IQ[29]. Participants were matched between groups on all of the above-mentioned variables. For a summary of these variables, see Table 1. All participants gave written informed consent and were compensated 8€ for their time. The study and protocols were approved by the ethics committee at the BCBL (approval number 11049) and were conducted in accordance with the Declaration of Helsinki.

Stimuli. Stimuli consisted of 100-word stories using 8 high frequency words (hereafter, keywords) that were one of the most representative exemplars of their category group[30]: fruit (apple—*manzana*), vehicle (car—*coche*), furniture (table—*mesa*), animal (dog—*perro*), dwelling (house—*casa*), reading material (book—*libro*), beverage (water—*agua*), and toy (ball—*balón*). These high frequency words would later be replaced by pseudowords.

Our choice of stimuli was motivated by several reasons. We needed stimuli that could (1) be easily understood and deduced from the sentences they were contained in, (2) apply to a variety of scenarios—as each one would appear in a set of 15 different stories, and (3) be easily identifiable from a picture. Given the current design, we needed that each sentence provided by itself enough information for participants to fully comprehend the critical

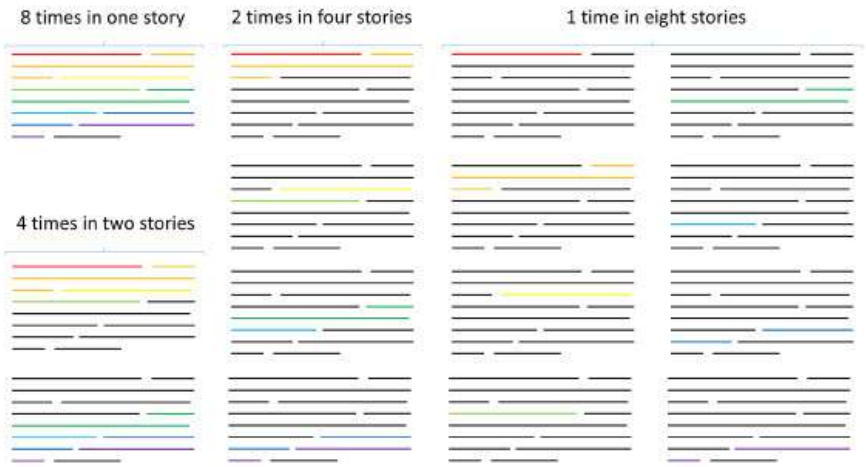

Figure 1. Schematic representation of how the stimulus stories were created. The color lines represent different sentences with the keyword and the black lines represent filler sentences (i.e., sentences that did not contain the keyword). Each text ended with a filler sentence and had a total of nine sentences each. First, the text with the term eight times was created. Then, this text was subdivided into two to create the two texts with the term four times. Each of these texts contained the term in four consecutive sentences that were exactly the same, in the same order and placement in the text as in the original story. The stories were then completed with filler sentences to reach nine sentences (respecting the original placement of the sentences containing the keywords). The original story was then subdivided in a similar fashion to create the four stories with the keyword twice and the eight stories with the keyword once. [Return to Summary of Publication 3 on page 31]

word without the need of a greater context. Besides, it should be kept in mind that the selected items should fit the native and foreign language conditions, and choosing medium or low frequency words would hardly represent a good option, since they would presumably be unknown to most or some participants in the foreign language, making the two language conditions unbalanced. Hence, by choosing very high frequency words, we could make sure that they were known in both languages, easily deduced from the sentence context, applicable to a variety of scenarios, and easily depicted by standardized images for the recognition tests.

Each story contained the keyword eight times (1 story), four times (2 stories), twice (4 stories), or just once (8 stories)—see Supplementary Table S1. The stories were created so that the story with the keyword eight times contained it in eight consecutive sentences and ended with a filler sentence (meaning a sentence without the keyword—see Supplementary Table S2 for a list of key sentences). Then, the sentences with the keyword were subdivided and filler sentences were added before and/or after in order to create the remaining texts (see Fig. 1 for a schematic representation of how the stimuli were created and Supplementary Table S3 for a worked out example). Although the other sentences were fillers, they did compose a cohesive paragraph. This way, the sentences containing the keyword were the same between conditions. These stories were then translated to create an English and a Spanish version, matched for word count. For each of the sentences containing the keyword, we carried out a norming study to assure that the predictability for the keywords was high (for Spanish [N = 9]: M = 81%, SD = 19%; for English L1 [N = 15]: M = 78%, SD = 22%; for English L2 [N = 9]: M = 64%, SD = 23%). Then, we replaced the keyword with a pseudoword of the form CVCVC, VCVCV, or VCVC, matched for bigram frequency (calculated using B-Pal[31] for Spanish and N-Watch[32] for English: bigram frequency mean token, $t(7) = 1.56$, $p = 0.16$; bigram frequency mean type, $t(7) = 0.96$, $p = 0.37$) and plausibility (rated from 1 to 5 by 14 native Spanish speakers, the average rating by item was not significantly different, $t(7) = 1.23$, $p = 0.26$) between languages. The pseudoword replacing the keyword in each story was the new target word to be learned during the task.

Procedure. Each participant was assigned either to the native language (Spanish; NL) or foreign language (English; FL) condition. Participants in each language condition were given all instructions in that language, both orally and on the screen, so as to avoid language switching effects. To assure comprehension, participants were given the instructions both orally and in written form. For the learning task, participants were given a practice trial and for the testing phase they were shown examples. All of the tasks were carried out using OpenSesame[33].

During the learning phase, participants were presented with texts and asked to read for comprehension. These texts contained eight novel pseudowords—two per diversity condition embedded in 30 stories (two with a pseudoword repeated eight times, four with a pseudoword repeated four times each, eight with a pseudoword repeated twice each, and 16 with a pseudoword only once per story). Participants were warned that there would be strange

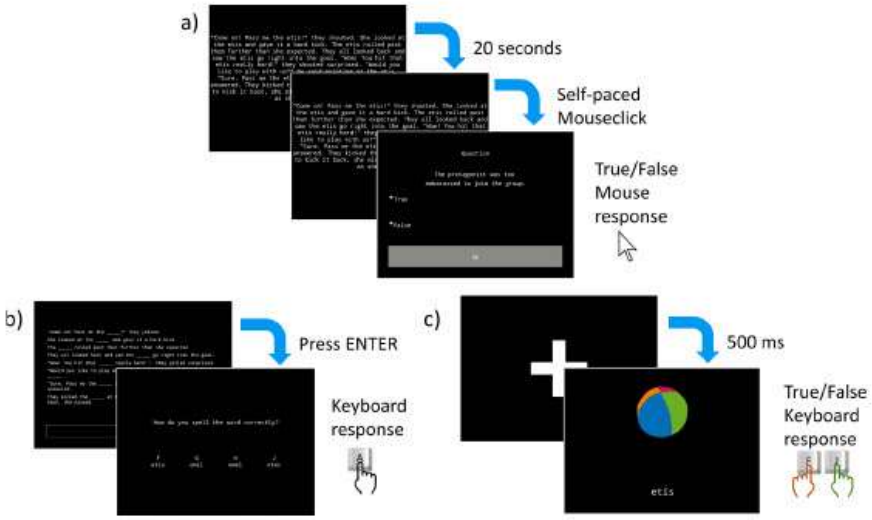

Figure 2. Schematic representation of the procedure. (**a**) Learning stage in which participants read each of the 30 texts. (**b**) Recall (fill in the blank) task which led to the recognition task, which required an untimed keyboard response. (**c**) Matching task in which participants were shown image-word pairs and were asked to determine whether they matched in meaning or not.

[Return to Summary of Publication 3 on page 31]

words in the texts and were advised to focus on reading for comprehension rather than focusing on those words. They then read one example text (before the 30 experimental texts) which contained a pseudoword they were not tested on, and were asked to answer two practice comprehension questions (see Fig. 2a and Supplementary Figure S4). The full list of texts (regardless of condition) was randomized. In order to avoid primacy effects on the conditions with the fewest number of texts (i.e., the condition with only one text with the pseudoword eight times and the one where there were two texts with the pseudoword four times) one of the eight texts from the highest diversity condition was put in first position. Similarly, to avoid recency effects we took one text from the other pseudoword in the highest diversity condition and placed it last as well as adding a distractor task before the testing portion. The other 28 texts were presented to each participant in a fully randomized order. While reading each story they were not allowed to continue until at least 15 s had passed. Once those 15 s had passed, an arrow appeared that the participant could press at any time (self-paced) to continue. After reading each text, participants answered one true–false comprehension question (not containing the novel pseudowords) to test both for attention and comprehension (comprehension check).

Once they had read all of the texts, participants completed a distractor task, which was the forwards and backwards Corsi Task[34], with a 10-min timer to assure that they all had equal-length breaks. After the distractor task, participants entered the testing phase, which consisted of a recall task, a recognition task, and a matching task (similar to those used by Rosa et al.[22]). For the recall task, they were presented with the eight sentences they had seen before in which the pseudoword appeared, except that blanks (lines) were placed where the pseudowords had been. Each sentence was presented on a separate line and in order, with the entirety of the text aligned left, but occupying most of the screen. Underneath, there was a rectangle in which participants were asked to type in the correct pseudoword that completed all of the sentences (see Fig. 2b and Supplementary Figure S5). Immediately after each fill-in-the-blank, they did the recognition task, which consisted of a multiple-choice question corresponding to the same pseudoword. They were presented with four options (the correct pseudoword, a competitor pseudoword, and two versions of these with middle consonants transposed) (see Fig. 2b and Supplementary Figure S5). Both of these tasks were self-paced and the order of words was fully randomized, while keeping the order of tasks constant (first recall and then recognition). After completing the recall and recognition for the first word, they proceeded similarly for the remaining seven pseudowords.

After participants had completed the recall and recognition tasks, they were asked to complete the matching task. They were presented with a drawing of a real object (centered horizontally but with its center on the one-third mark vertically), and a pseudoword (centered horizontally but with its center on the bottom one third, vertically) and were asked to say whether they matched (i.e., if the letter string meant the object) or not (see Fig. 2c and Supplementary Figure S6). The drawings were extracted from the MultiPic database[35] and depicted the high frequency words (with the exception of "water" for which we used the image for "faucet"). These images represented either the real object that was replaced by the pseudoword, a category competitor, a related word, or an unrelated image (i.e., the category competitor for a different pseudoword)—see Supplementary Table S1 for the full list and Supplementary Figure S6 for how it looked. They had 2500 ms to respond with the F and J

	8 texts	4 texts	2 texts	1 text
Comprehension				
Spanish	0.88 (0.02) [0.85; 0.92]	0.93 (0.02) [0.88; 0.98]	0.88 (0.02) [0.84; 0.92]	0.89 (0.04) [0.81; 0.96]
English	0.86 (0.01) [0.83; 0.89]	0.85 (0.02) [0.82; 0.89]	0.88 (0.02) [0.83; 0.93]	0.91 (0.03) [0.84; 0.98]
Recall (accuracy)				
Spanish	0.28 (0.05) [0.18; 0.39]	0.16 (0.05) [0.07; 0.25]	0.13 (0.03) [0.06; 0.19]	0.03 (0.02) [0.00; 0.07]
English	0.30 (0.05) [0.19; 0.40]	0.22 (0.05) [0.12; 0.32]	0.17 (0.05) [0.07; 0.27]	0.03 (0.02) [0.00; 0.07]
Recall (aline)				
Spanish	0.62 (0.05) [0.53; 0.71]	0.47 (0.05) [0.37; 0.57]	0.49 (0.04) [0.41; 0.58]	0.38 (0.04) [0.30; 0.46]
English	0.61 (0.05) [0.51; 0.72]	0.54 (0.05) [0.43; 0.65]	0.51 (0.05) [0.40; 0.61]	0.42 (0.04) [0.34; 0.50]
Recognition				
Spanish	0.82 (0.04) [0.74; 0.90]	0.75 (0.04) [0.66; 0.84]	0.82 (0.04) [0.73; 0.90]	0.63 (0.05) [0.53; 0.72]
English	0.89 (0.04) [0.81; 0.96]	0.74 (0.05) [0.63; 0.84]	0.80 (0.05) [0.70; 0.89]	0.59 (0.06) [0.48; 0.70]
Matching (accuracy)				
Spanish	0.77 (0.03) [0.71; 0.83]	0.71 (0.03) [0.65; 0.77]	0.75 (0.03) [0.69; 0.81]	0.71 (0.02) [0.66; 0.75]
English	0.70 (0.03) [0.64; 0.76]	0.66 (0.03) [0.60; 0.72]	0.64 (0.03) [0.58; 0.71]	0.62 (0.03) [0.56; 0.68]
Matching (response time)				
Spanish	1244.83 (40.87) [1,163.08; 1326.57]	1275.74 (35.13) [1205.47; 1346.00]	1289.63 (34.40) [1220.82; 1358.44]	1329.57 (35.39) [1258.78; 1400.35]
English	1285.50 (29.37) [1226.77; 1344.24]	1367.09 (38.48) [1290.12; 1444.06]	1366.06 (38.13) [1289.81; 1442.32]	1378.91 (41.19) [1296.53; 1461.29]

Table 2. Means, standard errors of the mean (numbers in parentheses), and 95% confidence intervals (values between brackets) by language for each of the tasks. Ranges of possible values for Comprehension, Recall, Recognition, and Matching accuracy are all 0 to 1. For Matching response time, ranges were 200 to 2500 ms.

[Return to Summary of Publication 3 on page 33]

keys on the keyboard for not-matching and matching decisions, respectively. Stimuli were presented in random order (see Fig. 2 for a schematic representation of the procedure).

Results

In all cases, alpha was set at 0.05. All t-tests reported are two-tailed. The number of participants (n) in all cases was 88. In all cases, the data was verified not to violate assumptions of normality. All analyses were run using JASP[36].

For the recall task, we also utilized the ALINE distance measure. ALINE distance is a measure of string alignment, which aligns phoneme strings, quantifying and standardizing the number of operations (insertions/deletions, substitutions, and expansions/compressions) necessary for going from one string to the other taking into account the features of the phonemes it compares[37]. This measures similarity between strings on a scale from 0 to 1, with lower scores showing increasing difference and one being exactly the same string. We calculated the ALINE similarity score between each item produced by the participant and the correct answer[38] using the alineR package[39] for R[40,41]. For these calculations, we removed any item that was shorter than 3 characters long and any items in which the participant produced the real word as opposed to the pseudoword, as these were not considered real attempts.

Comprehension check and reading times. The average accuracy score was 88% (SD = 8.4%). We carried out a two-way mixed ANOVA with Diversity (1, 2, 4, and 8 texts) and Language (foreign and native) on the performance on the comprehension test. There were no main effects of Language [$F_1(1,86) = 0.77$, $p = 0.38$, $\eta_p^2 = 0.01$, $BF_{01} = 4.65$, error% = 0.69; $F_2(1,7) = 1.82$, $p = 0.22$, $\eta_p^2 = 0.21$, $BF_{01} = 3.88$, error% = 1.21] or Diversity [$F_1(3,258) = 0.46$, $p = 0.71$, $\eta_p^2 = 0.01$, $BF_{01} = 43.88$, error% = 0.36; $F_2(3,21) = 0.07$, $p = 0.98$, $\eta_p^2 = 0.01$, $BF_{01} = 10.10$, error% = 0.72] and no interaction [$F_1(3,258) = 1.75$, $p = 0.16$, $\eta_p^2 = 0.02$, $BF_{01} = 3.88$, error% = 1.43; $F_2(3,21) = 0.24$, $p = 0.87$, $\eta_p^2 = 0.03$, $BF_{01} = 5.38$, error% = 3.73] (see Table 2).

The average time people took to read each paragraph was 38.05 s (SD = 9.84 s). We carried out a two-way mixed ANOVA with Diversity (1, 2, 4, and 8 texts) and Language (foreign and native) on reading times during learning. There was a main effect of Language [$F_1(1,86) = 28.74$, $p < 0.001$, $\eta_p^2 = 0.25$, $BF_{01} = 6.32 \times 10^{-5}$, error% = 1.69×10^{-7}; $F_2(1,7) = 275.13$, $p < 0.001$, $\eta_p^2 = 0.98$, $BF_{01} = 1.33 \times 10^{-14}$, error% = 2.95], with participants taking longer to read in the foreign language (M = 42.91 s; SD = 10.10 s) than in the native one (M = 33.19 s;

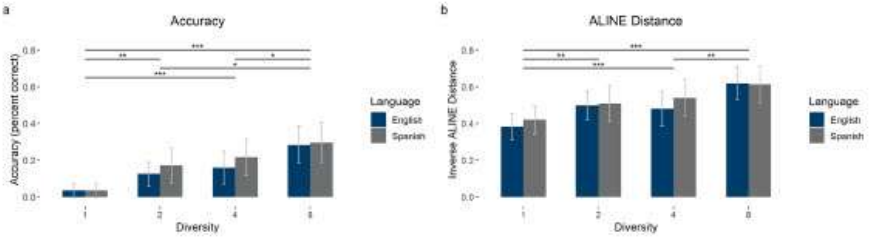

Figure 3. Recall task results. (**a**) Accuracy in the recall task by language and contextual diversity condition. (**b**) Average inverse Aline distance in the recall task by language and contextual diversity condition. Error bars represent 95% confidence intervals. Horizontal lines show significant pairwise comparisons for the main effect of diversity. *p < .05; **p < .01; ***p < .001. [Return to Summary of Publication 3 on page 33]

SD = 6.75 s). There was no main effect of Diversity [$F_1(3,258) = 0.98$, $p = 0.40$, $\eta_p^2 = 0.01$, $BF_{01} = 23.23$, error% = 3.03; $F_2(3,21) = 0.13$, $p = 0.94$, $\eta_p^2 = 0.02$, $BF_{01} = 11.148$, error% = 0.57] and no interaction [$F_1(3,258) = 0.32$, $p = 0.81$, $\eta_p^2 = 0.004$, $BF_{01} = 24.51$, error% = 2.29; $F_2(3,21) = 0.24$, $p = 0.87$, $\eta_p^2 = 0.03$, $BF_{01} = 6.08$, error% = 2.60]. For the foreign language, reading times correlated negatively with accuracy in the recall task ($r(43) = -0.43$, $p = 0.004$) and the recognition task (A': $r(43) = -0.33$, $p = 0.03$), as well as positively with response time in the matching task ($r(43) = 0.64$, $p < 0.001$) and the recognition task ($r(43) = 0.50$, $p < 0.001$), but not with ALINE distance in the recall task ($r(41) = -0.28$, $p = 0.07$) nor with A' in the matching task ($r(43) = -0.23$, $p = 0.13$). For the native language, reading times only correlated with response time in the recognition task ($r(43) = 0.32$, $p = 0.04$) and marginally with response time in the matching task ($r(43) = 0.27$, $p = 0.07$), but not with any of the other measures (p's > 0.4).

Recall task. The recall data (both accuracy and ALINE distance) was not normally distributed. To correct for the non-normality of the data, we carried out non-parametric tests—see Supplementary Table S4, but the results were the same as the frequentists and Bayesian tests. For homogeneity of analysis and for simplicity, here, we report the frequentist analyses.

For this part of the analysis, we only considered pseudowords that were correctly recalled—pseudowords for which the produced string matched perfectly with the target. On average, recall was fairly low (M = 17.8%, SD = 16.0%). We carried out a two-way mixed ANOVA with Diversity and Language on accuracy—namely, percent correct—in the recall task. There was no main effect of Language ($M_{Eng} = 15.1\%$, $SD_{Eng} = 14.7\%$; $M_{Spa} = 17.9\%$, $SD_{Spa} = 17.3\%$) [$F_1(1,86) = 0.69$, $p = 0.41$, $\eta_p^2 = 0.01$, $BF_{01} = 5.07$, error% = 1.43; $F_2(1,7) = 2.00$, $p = 0.20$, $\eta_p^2 = 0.22$, $BF_{01} = 2.92$, error% = 1.33], but there was a main effect of Diversity [$F_1(3,258) = 13.71$, $p < 0.001$, $\eta_p^2 = 0.14$, $BF_{01} = 6.72 \times 10^{-7}$, error% = 13.16; $F_2(3,21) = 10.67$, $p < 0.001$, $\eta_p^2 = 0.60$, $BF_{01} = 1.52 \times 10^{-5}$, error% = 0.46], such that items presented with greater diversity were recalled better (see Table 2). There was no interaction [$F_1(3,258) = 0.22$, $p = 0.88$, $\eta_p^2 < 0.01$, $BF_{01} = 25.53$, error% = 1.95; $F_2(3,21) = 0.32$, $p = 0.81$, $\eta_p^2 = 0.04$, $BF_{01} = 5.43$, error% = 5.81]. See Fig. 3.

Given the difficulty of the task, we also analyzed partial recall—pseudowords that were partially, but not completely correct. In order to quantify this partial recall, we used the ALINE similarity score (one minus the ALINE distance). Using those data, we carried out a two-way mixed ANOVA with Diversity and Language on Aline similarity in the recall task. The average Aline similarity score was 0.51 (SD = 0.23). There was no main effect of Language [$F_1(1,86) = 0.29$, $p = 0.59$, $\eta_p^2 < 0.01$, $BF_{01} = 3.82$, error% = 0.61; $F_2(1,7) = 1.29$, $p = 0.29$, $\eta_p^2 = 0.16$, $BF_{01} = 2.79$, error% = 1.40], but there was a main effect of Diversity [$F_1(3,258) = 13.65$, $p < 0.001$, $\eta_p^2 = 0.15$, $BF_{01} = 2.17 \times 10^{-6}$, error% = 1.13; $F_2(3,21) = 9.33$, $p < 0.001$, $\eta_p^2 = 0.57$, $BF_{01} = 3.04 \times 10^{-4}$, error% = 0.58], such that items presented with greater diversity elicited strings closer to the correct pseudoword (see Table 2 and Fig. 3). There was no interaction [$F_1(3,258) = 0.45$, $p = 0.72$, $\eta_p^2 = 0.01$, $BF_{01} = 18.43$, error% = 1.55; $F_2(3,21) = 0.34$, $p = 0.80$, $\eta_p^2 = 0.05$, $BF_{01} = 5.20$, error% = 1.86].

Recognition task. The recognition accuracy data was not normally distributed. To correct for the non-normality of the data, we carried out non-parametric tests—see Supplementary Table S4—, but the results were the same as the frequentists and Bayesian tests. For homogeneity of analysis and for simplicity, here, we report the frequentist analyses.

On the recognition task, the average correct recognition score was 75.28% (SD = 16.08%), with chance being 25%. We carried out a two-way mixed ANOVA with Diversity and Language on accuracy on the recognition task. There was a main effect of Diversity [$F_1(3,258) = 10.30$, $p < 0.001$, $\eta_p^2 = 0.11$, $BF_{01} = 5.77 \times 10^{-6}$, error% = 0.57; $F_2(3,21) = 9.54$, $p < 0.001$, $\eta_p^2 = 0.58$, $BF_{01} = 1.44 \times 10^{-4}$, error% = 0.73], but no main effect of Language [$F_1(1,86) = 9.14 \times 10^{-31}$, $p = 1$, $\eta_p^2 < 0.01$, $BF_{01} = 6.80$, error% = 1.61; $F_2(1,7) = 0.003$, $p = 0.96$, $\eta_p^2 < 0.001$, $BF_{01} = 4.01$, error% = 1.15] and no interaction [$F_1(3,258) = 0.49$, $p = 0.69$, $\eta_p^2 = 0.01$, $BF_{01} = 18.86$, error% = 1.97; $F_2(3,21) = 0.76$, $p = 0.53$, $\eta_p^2 = 0.10$, $BF_{01} = 4.16$, error% = 5.04]. The main effects showed that items presented with greater diversity were recognized better (see Table 2 and Fig. 4).

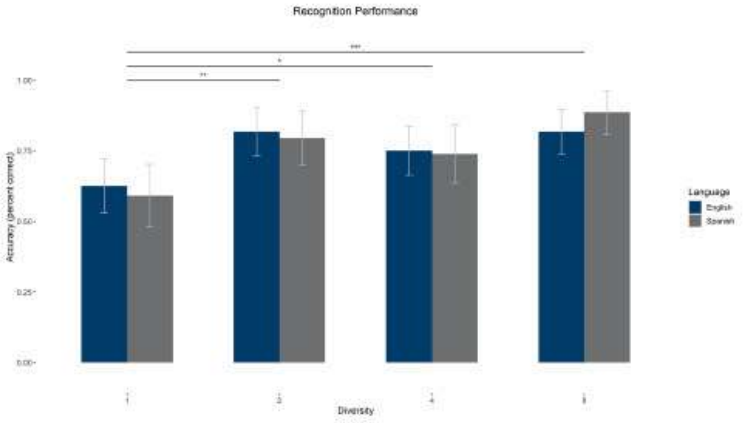

Figure 4. Recognition task results. Accuracy in the recognition task by language and contextual diversity condition. Error bars represent 95% confidence intervals. Brackets show significant pairwise comparisons for the main effect of diversity. *p < .05; **p < .01; ***p < .001.

[Return to Summary of Publication 3 on page 33]

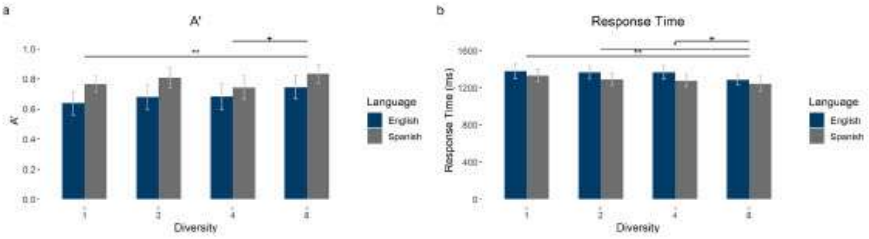

Figure 5. Matching task results. (**a**) A' in the matching task by language and contextual diversity condition. (**b**) Response times (RT) in the matching task by language and contextual diversity condition. Error bars represent 95% confidence intervals. Horizontal lines show significant pairwise comparisons for the main effect of diversity. *p < .1; *p < .05; **p < .01; ***p < .001.

[Return to Summary of Publication 3 on page 34]

Matching task. Overall, accuracy in the matching task was 69.97% (SD = 14.96), with chance being 50%. We calculated A'—a sensitivity index that takes into consideration hits and false alarms—using the Psycho package in R[42]. We carried out a two-way mixed ANOVA with Diversity and Language on A' on the matching task. There were main effects of Language [$F_1(1,86) = 6.75$, $p = 0.01$, $\eta_p^2 = 0.07$, $BF_{01} = 0.26$, error% = 0.42; $F_2(1,7) = 26.46$, $p < 0.001$, $\eta_p^2 = 0.79$, $BF_{01} = 0.005$, error% = 1.01] and Diversity [$F_1(3,258) = 3.51$, $p = 0.02$, $\eta_p^2 = 0.04$, $BF_{01} = 0.88$, error% = 0.25; $F_2(3,21) = 3.04$, $p = 0.05$, $\eta_p^2 = 0.30$, $BF_{01} = 0.91$, error% = 0.51], but no interaction [$F_1(3,258) = 0.63$, $p = 0.60$, $\eta_p^2 = 0.01$, $BF_{01} = 16.16$, error% = 0.96; $F_2(3,21) = 0.41$, $p = 0.75$, $\eta_p^2 = 0.06$, $BF_{01} = 4.79$, error% = 2.01]. The main effects showed that participants in the native language condition had better signal detection—i.e., a combination of more hits and fewer false alarms—than those in the foreign language condition and that items presented with greater diversity were matched with greater discrimination ability (see Table 2 and Fig. 5).

We also carried out a two-way mixed ANOVA with Diversity and Language on response time on the matching task. There was a main effect of Diversity [$F_1(3,258) = 5.11$, $p = 0.002$, $\eta_p^2 = 0.06$, $BF_{01} = 0.10$, error% = 2.86; $F_2(3,21) = 4.11$, $p = 0.02$, $\eta_p^2 = 0.37$, $BF_{01} = 0.38$, error% = 0.69], but no main effect of Language in the by participant analysis (although it does show up in the by item analysis, with response times in the foreign language being longer) [$F_1(1,86) = 2.21$, $p = 0.14$, $\eta_p^2 = 0.03$, $BF_{01} = 1.47$, error% = 0.35; $F_2(1,7) = 34.49$, $p < 0.001$, $\eta_p^2 = 0.83$, $BF_{01} = 0.04$, error% = 1.07] and no interaction [$F_1(3,258) = 0.50$, $p = 0.68$, $\eta_p^2 = 0.01$, $BF_{01} = 19.15$, error% = 1.73; $F_2(3,21) = 0.14$, $p = 0.93$, $\eta_p^2 = 0.02$, $BF_{01} = 5.58$, error% = 2.51]. The main effect of Diversity showed that items presented with greater diversity were responded to faster (see Table 2 and Fig. 5).

Discussion

The purpose of this study was to expand our knowledge of incidental learning. In particular, we tested whether the distribution of encounters with a new word in one or several texts affected learning, paying particular attention to the comparison between native and foreign language vocabulary learning. With this purpose in mind, we had participants read short stories with pseudowords replacing high frequency words. Those pseudowords could appear in fewer or more texts, being associated with lower or higher diversity, respectively. In addition, participants were performing the task either in their native (Spanish) or a foreign (English) language. Participants had to answer a question after every text and showed no effect of language or diversity in this comprehension check. We then had participants recall and recognize those pseudowords as well as match them with the objects they represented.

Overall, we found a main effect of contextual diversity in all tasks, with participants performing better—faster and/or more accurately—with pseudowords that they had seen in more contexts. This means that, in the full absence of comprehension problems, diversity only had a positive impact, making the pseudowords easier to recall, recognize, and match with their meaning. Our results are in line with prior studies that show effects of diversity above and beyond those of frequency[4,20,22]. This suggests that simply manipulating contextual diversity might be enough to improve performance without increasing frequency of exposure. It should be noted that in addition to increasing contextual diversity, our manipulation increased spacing between encounters, which might have also boosted the effects and had a positive effect on retention. Nevertheless, spacing literature refers to separate sessions, often carried out on different days. As an example, Sobel, Cepeda, and Kapler [43] used a 10 min task with just one minute between sessions in the massed condition, and with one week between sessions in the spaced condition. Whereas all of our conditions could be considered massed according to this view, it should be noted that our manipulations do not fit strict definitions of massed and spaced exposure, since words were never repeated consecutively—at most they were in consecutive sentences—and they were never spaced in separate sessions—each participant had only one session. Word meanings are created through the summation of experiences with a word and the words it co-occurs with [44,45]. Hence, while spacing and diversifying contexts can ultimately yield similar effects, they represent two conceptually different constructs: whereas spacing aids memory, contextual diversity aids in creating a richer mental representation of the item.

These results are particularly important for cases in which exposure to the language itself is limited—as for example, in foreign language classrooms—and increasing the number of instances of a word is very costly. In addition, we show that the effect of diversity is not simply binary, but rather a gradient where more diversity leads to better outcomes. Prior studies had mostly focused on an all-or-none definition of diversity which did not give a clear picture of whether the effect increased passed an initial benefit. The current study shows that if contextual diversity is increased further, the benefits increase as well (at least from 8 repetitions in 8 texts to 8 repetitions in 1 text, as tested here).

We found no effects of language on comprehension, suggesting that the texts were equally understandable in both languages. Language only affected performance on the matching task but did not affect recall or recognition. Even if participants performed equally on lexical access tasks in both languages, they had a greater sense of familiarity with the correct meaning of the pseudowords in their native language. This allowed them to recognize better whether the pseudoword matched the image presented in their native language than their foreign one. Interestingly, our study provides a more nuanced picture of some of the differences between learning in a native or a foreign language. We see here that when the lexical items are matched between languages, they are equally difficult or easy to learn. This is in contrast with some previous literature that found that memory tends to be worse in a foreign language[23,46,47], although these results are not very consistent[48]. This difference in results between memory for known vocabulary and new word learning suggests that either the effect is not very robust, or it does not extend to new vocabulary. Although not direct evidence, this is also somewhat in conflict with Pellicer-Sánchez's[4] findings that more exposures are necessary to reduce reading times in the foreign than the native language. Our results do suggest that perhaps their outcomes were partially caused by difficulties intrinsic to experience with the phonology or orthography of a language rather than to the language use itself. On the other hand, these effects could be influenced by the additional reading time in the foreign language context. Nevertheless, it should be kept in mind that this additional time within that condition correlated with worse performance. Also, even with the added reading time, we see that incorporating the item's meaning is somewhat more difficult in the foreign language. These results also relate to and support those of Nassaji[10] who found that the capacity to extract meaning from context relates to the knowledge of that language.

Importantly, there were no interactions between the main factors at study, showing that diversity had the same positive effect in both languages. Although against our initial hypothesis, this suggests that access to contextual information is enough to maintain the positive influence of diversity on word learning, despite the obvious difficulty of processing information in a non-native language. A prior study from the same authors found similar results with emotionality, where the effects of this variable were independent of language[49]. This supports the idea that the strategies for improving learning in the native language can apply to the foreign language, suggesting also that learning new vocabulary in one's native and foreign languages engage similar mechanisms. Nevertheless, it should be noted that these participants had upper intermediate (B2) and above levels of English, thus allowing for the possibility that results might differ with low proficiency bilinguals.

These results have several implications. First, they qualify the value of contextual diversity across languages as well as generalize its importance, even overcoming difficulties in processing fluency. Second, they show that incidental vocabulary acquisition occurs similarly in a foreign and a native language. This also gives us a tool for improving this process by making strategic use of context-based spreading of information. Here, we show that it is not necessary to increase the number of exposures in order to improve learning, highlighting the importance of context and pointing to a perhaps overstated importance of frequency. It is worth noting that these results

extend only to information recently learned and cannot speak to more long-term effects without further study. Nevertheless, this has important implications for education, where time and exposure are very limited and must be optimized. Future research should focus on possible mechanisms for these effects in order to understand their origin and extent. This paves the way for future studies focusing on how to affect context—or perhaps how novel words are presented in general—in order to improve incidental vocabulary learning.

Data Availability

All data, scripts, and stimuli are available at https://osf.io/7ks4f/?view_only=5364dfadf99a41c283fa8b0c3a094453.

Received: 11 March 2020; Accepted: 3 August 2020
Published online: 18 August 2020

References

1. Jenkins, J. R., Stein, M. L. & Wysocki, K. Learning vocabulary through reading. *Am. Educ. Res. J.* 21, 767–787 (1984).
2. Nagy, W. E., Herman, P. A. & Anderson, R. C. Learning words from context. *Read. Res. Q.* 20, 233–253 (1985).
3. Horst, M. & Cobb, T. Beyond a clockwork orange: acquiring second language vocabulary through reading. *Read. Foreign Lang.* 11, 207–223 (1998).
4. Pellicer-Sánchez, A. Incidental L2 vocabulary acquisition from and while reading: an eye-tracking study. *Stud. Second Lang. Acquis.* 38, 97–130 (2016).
5. Pellicer-Sánchez, A. & Schmitt, N. Incidental vocabulary acquisition from an authentic novel: Do things fall apart?. *Read. Foreign Lang.* 22, 31–55 (2010).
6. Pitts, M., White, H. & Krashen, S. Acquiring second language vocabulary through reading. *Read. Foreign Lang.* 5, 271–275 (1989).
7. Webb, S. The effects of context on incidental vocabulary learning. *Read. Foreign Lang.* 20, 232–245 (2008).
8. Hulme, R. C., Barsky, D. & Rodd, J. M. Incidental learning and long-term retention of new word meanings from stories: The effect of number of exposures. *Lang. Learn.* 69, 1–24. https://doi.org/10.17605/OSF.IO/7UYPB (2018).
9. Schmitt, N. Review article: Instructed second language vocabulary learning. *Lang. Teach. Res.* 12, 329–363 (2008).
10. Nassaji, H. The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *Can. Mod. Lang. Rev.* 61, 107–135 (2006).
11. Adelman, J. S., Brown, G. D. A. & Quesada, J. F. Contextual diversity not frequency determines word naming and lexical decision times. *Psychol. Sci.* 17, 814–823 (2006).
12. Brysbaert, M. & New, B. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990 (2009).
13. Kachergis, G., Yu, C. & Shiffrin, R. M. Frequency and contextual diversity effects in cross-situational word learning. *Proc. Cognit. Sci.* 31, 1–6 (2009).
14. Hills, T. T., Maouene, J., Riordan, B. & Smith, L. B. The associative structure of language: contextual diversity in early word learning. *J. Mem. Lang.* 63, 259–273 (2010).
15. Perea, M., Soares, A. P. & Comesaña, M. Contextual diversity is a main determinant of word identification times in young readers. *J. Exp. Child Psychol.* 116, 37–44 (2013).
16. Johns, B. T., Gruenenfelder, T. M., Pisoni, D. B. & Jones, M. N. Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *J. Acoust. Soc. Am.* 132, EL74–EL80 (2012).
17. Parmentier, F. B. R., Comesaña, M. & Soares, A. P. Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Q. J. Exp. Psychol.* 70, 1–17 (2017).
18. Grainger, J. Word frequency and neighborhood frequency effects in lexical decision and naming. *J. Mem. Lang.* 29, 228–244. https://doi.org/10.1111/lang.12313 (1990).
19. Vergara-Martínez, M., Comesaña, M. & Perea, M. The ERP signature of the contextual diversity effect in visual word recognition. *Cognit. Affect. Behav. Neurosci.* 17, 461–474 (2017).
20. Pagán, A. & Nation, K. Learning words via reading: contextual diversity, spacing, and retrieval effects in adults. *Cognit. Sci.* 43, 1–24 (2019).
21. Adelman, J. S., Brown, G. D. A. & Quesada, J. F. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychol. Sci.* 17, 814–823 (2006).
22. Rosa, E., Tapia, J. L. & Perea, M. Contextual diversity facilitates learning new words in the classroom. *PLoS ONE* 12, 1–12 (2017).
23. Vander Beken, H. & Brysbaert, M. Studying texts in a second language: The importance of test type. *Bilingualism* 21, 1062–1074 (2018).
24. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191 (2007).
25. Johns, B. T., Dye, M. & Jones, M. N. The influence of contextual diversity on word learning. *Psychon. Bull. Rev.* 23, 1214–1220 (2016).
26. Lemhöfer, K. & Broersma, M. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44, 325–343 (2012).
27. Izura, C., Cuetos, F. & Brysbaert, M. Lextale-Esp: a test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica* 35, 49–66 (2014).
28. de Bruin, A., Carreiras, M. & Duñabeitia, J. A. The BEST dataset of language proficiency. *Front. Psychol.* 8, 522 (2017).
29. Kaufman, A. S. & Kaufman, N. L. *Kaufman Brief Intelligence Test. Encyclopedia of Special Education: A Reference for the Education of Children, Adolescents, and Adults with Disabilities and Other Exceptional Individuals. En* (2013).
30. Uyeda, K. M. & Mandler, G. Prototypicality norms for 28 semantic categories. *Behav. Res. Methods Instrum.* 12, 587–595 (1981).
31. Davis, C. J. & Perea, M. BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behav. Res. Methods* 37, 665–671 (2005).
32. Davis, C. J. N-Watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behav. Res. Methods* 37, 65–70 (2005).
33. Mathôt, S., Schreij, D. & Theeuwes, J. OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324 (2012).
34. Berch, D. B., Krikorian, R. & Huha, E. M. The Corsi block-tapping task: methodological and theoretical considerations. *Brain Cognit.* 38, 317–338 (1998).
35. Duñabeitia, J. A. *et al.* MultiPic: a standardized set of 750 drawings with norms for six European languages. *Q. J. Exp. Psychol. (Hove)* 71, 808–816 (2018).
36. JASP Team. JASP (Version 0.9) (2018).
37. Kondrak, G. *Alignment of Phonetic Sequences*. https://www.cs.utoronto.ca/csri/reports.html (1999).

38. Downey, S. S., Hallmark, B., Cox, M. P., Norquest, P. & Lansing, J. S. Computational feature-sensitive reconstruction of language relationships: developing the ALINE distance for comparative historical linguistic reconstruction. *J. Quant. Linguist.* 15, 340–369 (2008).
39. Downey, S. S., Sun, G. & Norquest, P. alineR: An R package for optimizing feature-weighted alignments and linguistic distances. *R J.* 9, 138–152 (2017).
40. RStudio Team. RStudio: Integrated development for R (2016).
41. R Core Team. R: A language and environment for statistical computing (2018).
42. Makowski, D. The psycho package: an efficient and publishing-oriented workflow for psychological science. *J. Open Source Softw.* 3, 470 (2018).
43. Sobel, H. S., Cepeda, N. J. & Kapler, I. V. Spacing effects in real-world classroom vocabulary learning. *Appl. Cognit. Psychol.* 25, 763–767 (2011).
44. Bolger, D. J., Balass, M., Landen, E. & Perfetti, C. A. Context variation and definitions in learning the meanings of words: an instance-based learning approach. *Discourse Process.* 45, 122–159 (2008).
45. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208 (1996).
46. Francis, W. S. & Baca, Y. Effects of language dominance on item and order memory in free recall, serial recall and order reconstruction. *Memory* 22, 1060–1069 (2014).
47. Nott, C. R. & Lambert, W. E. Free recall of bilinguals. *J. Verbal Learn. Verbal Behav.* 7, 1065–1071 (1968).
48. Francis, W. S. *et al.* The effects of bilingual language proficiency on recall accuracy and semantic clustering in free recall output: evidence for shared semantic associations across languages across languages. *Memory* 26, 1364–1378 (2018).
49. Frances, C., De Bruin, A. & Dunăbeitia, J. A. The influence of emotional and foreign language context in content learning. *Stud. Second Lang. Acquis.* https://doi.org/10.1017/S027226311900072X (2019).

Acknowledgments
This research has been partially funded by grants PGC2018-097145-B-I00, RED2018-102615-T, PSI2015-65689-P, and SEV-2015-0490 from the Spanish Government, H2019/HUM-5705 from the Comunidad de Madrid, and AThEME-613465 from the European Union. This research is supported by the Basque Government through the BERC 2018-2021 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490. CF is supported by a MINECO predoctoral grant from the Spanish government (BES-2016-077169).

Author contributions
All authors conceived of the presented idea. C.F. carried out the experiment, developed the theory and performed the computations. J.A.D. and C.M. verified the analytical methods. J.A.D. and C.M. supervised the findings of this work. C.F. wrote the manuscript with support from J.A.D. and C.M. All authors discussed the results and edited the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-70922-1.

Correspondence and requests for materials should be addressed to C.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.