Bilingualism: Language and Cognition

cambridge.org/bil

Lexical and semantic training to acquire words in a foreign language: An electrophysiological study

Ana B. García-Gámez[1,2] and Pedro Macizo[3,4]

[1]Center for Health, Technology and Services Research (CINTESIS, Faro, Portugal); [2]University of Algarve (Faro, Portugal); [3]Mind, Brain and Behaviour Research Centre (CIMCYC, Granada, Spain) and [4]University of Granada (Granada, Spain)

Research Article

Cite this article: García-Gámez AB, Macizo P (2022). Lexical and semantic training to acquire words in a foreign language: An electrophysiological study. *Bilingualism: Language and Cognition* **25**, 768–785. https://doi.org/10.1017/S1366728921000456

Received: 12 November 2020
Revised: 14 July 2021
Accepted: 14 July 2021
First published online: 5 April 2022

Keywords:
L2 vocabulary learning; ERPs; N400; LPC

Address for correspondence:
Ana B. García-Gámez, Departamento de Neurociências Cognitivas, Center for Health, Technology and Services Research, University of Algarve, Gambelas Campus, Faro, 8005-139, Portugal. Email: abgamez@ualg.pt

Abstract

An event-related potential (ERP) study was conducted to evaluate the efficacy of two learning methods for the acquisition of vocabulary in a foreign language (FL). In the semantic method, FL words were presented with pictures denoting their meaning and the learners practiced with a semantic categorization task (to indicate whether FL words were exemplars of a semantic category). In the lexical method, FL words were paired with their translation in the first language (L1) and the learners practiced with a letter-monitoring task (to indicate whether L1-FL words contained a grapheme). A translation task and a picture-naming task were used to evaluate FL acquisition. ERP modulations associated with semantic processing were more evident and broadly distributed in the semantic versus lexical learning group. The pattern of results suggests that a single session of semantic learning favors the establishment of connections between semantics and the words learned in a new language.

1. Introduction

There are many questions about foreign language (FL) acquisition that need to be addressed. What is the efficient way to learn a new language? Are there strategies able to enhance this learning process? Nowadays, we are involved in multicultural societies and speaking different languages is becoming necessary. Not only children, but also adults have to face these new situations in their daily lives. It seems that immersion programs are the best way to learn a new language (Genesee, 2014) but this option is not always available. For this reason, it is important to look for learning tools or strategies able to facilitate FL acquisition in first language (L1) speaking contexts. It is necessary to implement efficient second language (L2) learning methodologies based on scientific evidence not only at school for children, but also for adults.

To design effective FL learning methods, it is first necessary to understand how novel and expert bilinguals manage linguistic processing across languages. According to the Revised Hierarchical Model (Kroll & Stewart, 1994), words in the first language (L1 lexicon), the second language (L2 lexicon) and the semantic system of L2 learners are interconnected. However, the weight and use of these connections depends on the fluency of participants in their L2 (De Groot & Poot, 1997). Thus, while novice learners make preferential use of L1-L2 connections, expert learners mainly rely on the connections between semantics and L2 words (see Kroll, Van Hell, Tokowicz & Green, 2010, for a review). However, semantic processing is also possible and desirable in novice learners of a FL (Ferré, Sánchez-Casas & Guasch, 2006; Talamas, Kroll & Dufour, 1999).

The word-translation task and the picture-naming task have been used to evaluate semantic and L1-L2 lexical processing during the acquisition of a foreign language. With regard to the word-translation task, behavioral studies have shown that the performance on this task depends on the translation direction and the L2 fluency of the learners. In particular, less fluent bilinguals translate more rapidly in the backward direction (from L2 to L1) than in the forward direction (from L1 to L2 – Kroll & Stewart, 1994). The better performance in the backward translation is explained by the use of direct lexical connections between the L1 and L2 lexicons while the forward translation would involve additional semantic processing which would slow down the translation process. However, this asymmetry in the translation task is attenuated in fluent bilinguals which seems to indicate that they use a semantic route of processing regardless of the translation direction (Kroll & Linck, 2007).

On the other hand, regarding the picture-naming task, many studies have confirmed that L2 naming is slower than L1 naming. In addition, these differences in the speed of processing depending on the language of the naming task are more evident in low vs. high fluency

CAMBRIDGE UNIVERSITY PRESS

bilinguals (Kroll, Michael, Tokowicz & Dufour, 2002). Less fluent bilinguals would have weak connections between the semantic system and the L2 lexicon, so they would use a route of processing through their first language which would slow down the retrieval of the picture names in L2. On the contrary, proficient bilinguals would employ a direct connection between meaning and L2 words which would produce faster naming times.

If we assume that expert bilinguals perform linguistic tasks (e.g., translation, naming) through semantic processing, it would be desirable to develop learning strategies that favor the establishment of connections between semantics and FL lexicon. In fact, the benefit of semantic vs. lexical learning programs has been confirmed in the past. To illustrate, Comesaña, Perea, Piñeiro, and Fraga (2009) compared the learning of L2 vocabulary in children through a word-word association method that reinforced the L1-L2 lexical connections, and a learning method based on the association between pictures and L2 words that favored the connections between the semantic system and L2 words. The results revealed that only children who received a picture-word association training showed effects derived from semantic processing (e.g., semantic interference effect in a translation recognition task); a pattern of results which is usually found in expert bilinguals (see also, Comesaña, Soares & Lima, 2010; Tonzar, Lotto & Job, 2009).

The learning benefits associated with semantic learning strategies do not only appear in children but also in adult L2 learners. In a recent study, García-Gámez and Macizo (2020) compared the acquisition of L2 vocabulary in adults through a semantic training (based on picture-L2 word association and semantic categorization) and a lexical training (based on L1-L2 word association and grapheme monitoring). After learning, the authors found that the semantic training group responded more slowly than the lexical training group in a forward translation task. These results suggested that the semantic training group made use of a route of processing that was conceptually mediated; while the lexical group made the translation task through direct connections between their lexicons. In addition, in a L2 naming task, the semantic training group was faster than the lexical training group suggesting again that the semantic training group efficiently used a direct connection between concepts and L2 words.

Thus, there is abundant behavioral evidence showing that L2 vocabulary learning based on semantic training favors the establishment of connections between the semantic system and new L2 words. This evidence has been obtained by enriching the L2 learning context with material and tasks that favor the semantic processing of the new words. These include: the use of semantic ratings about the new words (Barcroft, 2002); presenting L2 vocabulary in semantically grouped sets (Finkbeiner & Nicol, 2003); the use of pictures that denote the meaning of the word to be learned (e.g., Comesaña et al., 2009, 2010; Comesaña, Soares, Sánchez-Casas & Lima, 2012; Tonzar et al., 2009); pictures and listening/speaking exercises (e.g., Poarch, Van Hell & Kroll, 2014); gestures that represent the meaning of the new words (e.g., García-Gámez & Macizo, 2019; Tellier, 2008) etc. – see Rice and Tokowicz, 2020, for a review of laboratory studies of adult second language vocabulary training. However, the results of behavioral research sometimes diverge from the outcomes reported in electrophysiological reports. In studies on bilingualism, differences in the neural correlates of cognitive processes have been found in the absence of overt behavioral differences (Abutalebi et al., 2012; Bialystok, Craik, Grady, Chau, Ishii, Gunji & Pantev, 2005; Kousaie & Phillips, 2012). Concerning L2 vocabulary acquisition, behavioral studies sometimes fail to capture subtle changes in language processing that occur with minimal training in L2 learning (McLaughlin, Osterhout & Kim, 2004). For example, McLaughlin et al. showed that brain activity of adult L2 learners indexed by event-related potentials (ERPs) discriminated the processing of new L2 words compared to the processing of L2 pseudowords with only one session of L2 vocabulary training. However, no differences were found between L2 words and pseudowords when behavioral measures were considered. The authors concluded that some aspects of a new language may be overlooked by current behavioral assessments. To the best of our knowledge, there are no previous electrophysiological studies comparing the effect of semantic vs. lexical training on L2 vocabulary acquisition. In our study, we address this point directly. To be more specific, the goal of our study was to evaluate, from an electrophysiological approach (i.e., ERP data), the impact that a semantic vs. lexical L2 vocabulary learning could have on the posterior processing of the new words in language production tasks (i.e., translation and picture-naming).

Several electrophysiological components have been used to index vocabulary acquisition in L2 (Midgley, Holcomb & Grainger, 2009; Yum, Midgley, Holcomb & Grainger, 2014). The P2 component is a positive-going waveform with a peak latency ranging from 150 to 275 ms. It is thought to index mechanisms of selective attention (Hackley, Woldorff & Hillyard, 1990), and other sensory stages of item encoding (Dunn, Dunn, Languis & Andrews, 1998). When L2 learners perform word reading tasks, there are differences near the peak of the P2 component depending on the language in which they read the words. In particular, the P2 amplitude is more positive when reading in L2 compared to the reading of words in the native language (e.g., Midgley et al., 2009). These differences between languages would indicate a greater involvement of the attentional mechanism when learners process words in L2 vs L1. However, beyond this early ERP component, more relevant to the present study are two other electrophysiological indexes, the N400 and the late positivity component (LPC).

The N400 component is a negative-going waveform peaking at approximately 350-450 ms after stimulus onset, whose amplitude is sensitive to the processing of lexical-semantic information (Kutas & Hillyard, 1980). In bilingual studies, it has been shown that the processing of L2 words elicits smaller N400 than the processing of L1 words. Moreover, L2 words also produce larger N400 in highly proficient bilinguals compared with L2 learners with low proficiency (Midgley et al., 2009). Thus, modulations of the N400 amplitude when individuals process L2 words have been considered an index of L2 proficiency in previous studies (e.g., Pu, Holcomb & Midgley, 2016; Soskey, Holcomb & Midgley, 2016).

As indicated above, many behavioral studies have found that word retrieval is easier in backward translation than in forward translation due to the difficulty associated with semantic processing in L1-L2 translation (e.g., Cheung & Chen, 1998; Finkbeiner & Nicol, 2003; García-Gámez & Macizo, 2019, 2020; Kroll & Stewart, 1994; Poarch et al., 2014; Sholl, Sankaranarayanan & Kroll, 1995; for a critical review of asymmetry dependent on the translation direction, see Kroll et al., 2010). In electrophysiological terms, an easy retrieval of lexical information would be associated with an attenuation of the N400 component. For instance, word frequency is one of the main indicators of difficulty in lexical

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

access (e.g., Hudson & Bergman, 1985; Monsell, Doyle & Haggard, 1989), and this lexical factor produces a N400 attenuation (Rugg, 1990; Van Petten & Kutas, 1990) with reduced brain-wave negativity during the processing of high vs. low frequency words. However, electrophysiological studies with the translation task are very limited and the results are mixed. Christoffels, Ganushchak, and Koester (2013) observed a greater amplitude of the N400 component when Dutch (L1) – English (L2) bilinguals translated words in backward vs. forward. In contrast, Jost, Radman, Buetler, and Annoni (2018) did not find differences related to the translation direction in the N400 time-window.

The LPC component is a late-onset sustained positivity peaking between 500 and 900 ms. Although the function of the LPC is still not clearly determined, it has been related to long-term semantic memory (e.g., Coulson, Federmeier, Van Petten & Kutas, 2005) and episodic recollection (Rugg & Curran, 2007, for a review). This component overlaps both temporally and spatially with the P600, an ERP component that has been linked to different language sub-processes, especially at the syntactic level (Leckey & Federmeier, 2020, for a review). Some authors have proposed that the LPC/P600 are the same component or part of the same family of components which are not specific to syntactic processing but would be associated with reanalysis of information and response monitoring in language processing (Kolk & Chwilla, 2007).

In translation tasks, the LPC shows an inverse polarity between parietal and frontal regions and this component has been associated with the reprocessing of information between the input and the output language. For example, in a translation recognition task, a greater LPC amplitude appears in late time-windows regions when word pairs are not translations but are semantically related compared to unrelated word pairs (e.g., Guo, Misra, Tam & Kroll, 2012). Other authors have linked the LPC found in translation tasks to the process of establishing connections between words across languages (Jackson, Swainson, Cunnington & Jackson, 2001).

On the other hand, LPC modulations have been observed when bilinguals perform picture-naming tasks (Martin et al., 2013). For instance, the LPC amplitude is greater when the difficulty in the retrieval of the picture names increases (e.g., naming tasks that involve language switching across trials). In general, studies about lexical processing in bilinguals seem to indicate that the more complex the processing of the stimuli (e.g., L2 naming vs. L1 naming) the greater the mean amplitude of the LPC component (Jackson et al., 2001; Kieffaber, Kruschke, Cho, Walker & Hetrick, 2013).

1. 1. The current study

The main objective of our study was to evaluate the impact of a semantic vs. lexical training on the acquisition of vocabulary in a foreign language. To this end, two groups of Spanish speakers (L1) learned words in an artificial language (Vimmi) under two types of training. The semantic training group was exposed to L2 words accompanied by a picture representing their meaning, and participants practiced during learning with a semantic categorization task. The lexical training group was exposed to the word in L2 along with its L1 translation and the practice task involved the identification of graphemes between languages (grapheme monitoring task). At the end of the training, electrophysiological data were collected while participants performed a translation and a naming task to evaluate the acquisition of L2 vocabulary.

As mentioned before, to our knowledge, there are no previous electrophysiological studies evaluating the impact that a lexical vs. semantic training would have on L2 vocabulary acquisition. Thus, the predictions of our study were based on preliminary behavioral research on the subject, the comparison between semantic vs. lexical L2 vocabulary acquisition trainings (e.g., Comesaña et al., 2012; García-Gámez & Macizo, 2020; Poarch et al., 2014), and electrophysiological studies regarding the evaluation tasks used in the current work (translation and picture-naming – e.g., Guo et al., 2012; Jackson et al., 2001; Jost et al., 2018).

Firstly, we expected to observe differences between the semantic and the lexical learning groups that would modulate the N400 amplitude. In particular, the N400 negativity would be larger in the semantic vs. lexical training. This prediction was based on the assumption that the semantic training would strengthen the connections between concepts and the new L2 words (e.g., Comesaña et al., 2012). These L2-concept connections are stronger in fluent bilinguals compared to less fluent bilinguals (e.g., Kroll & Stewart, 1994; Talamas et al., 1999) and the development of L2 fluency is accompanied by larger N400 amplitudes when bilinguals perform language tasks in L2 (Midgley et al., 2009; Moreno & Kutas, 2005; Neville, Mills & Lawson, 1992). Regarding the LPC amplitude, we did not anticipate differences between the learning groups. As already noted, The LPC has been related to the reanalysis of information (e.g., Kolk & Chwilla, 2007), and the LPC amplitude increases with the difficulty of the rechecking process (e.g., when it is difficult to determine if L1-L2 word pairs are correct translations, Guo et al., 2012). This comparison between L1-L2 lexical forms can be done through lexical connections between languages: these links would operate regardless of the type of training, and would remain even in fluent bilinguals (e.g., cross-language lexical activation persists for highly proficient bilinguals, Hoshino & Kroll, 2008). Thus, we expect that the process of rechecking lexical information underlying LPC modulations would be functional in the lexical and semantic training groups and, hence, we did not anticipate differences between L2 learning groups.

Looking at the translation task, we expected to found larger N400 amplitude in forward vs. backward translation. This prediction was not grounded in electrophysiological studies because there is mixed evidence regarding N400 modulations and the translation direction effect (Christoffels et al., 2013; but see Jost et al., 2018). However, the retrieval of lexical-semantic information is more difficult in forward than backward translation (e.g., Kroll & Stewart, 1994) which would result in larger N400 amplitudes related to the increased difficulty of lexical-semantic analysis in forward vs. backward translation (i.e., the processing of low vs. high frequency words, Rugg, 1990; the processing of semantically related vs. unrelated word pairs, Bentin, McCarthy & Wood, 1985). Furthermore, if we assume that the translation direction effect is due to differences in semantic processing (i.e., semantic mediation is larger in forward vs. backward translation – Kroll & Stewart, 1994), we anticipated that the magnitude of this effect would be greater in the semantic vs. lexical training group. On the other hand, the translation direction would modulate the LPC amplitude, since previous studies with the translation recognition task show that this ERP component is sensitive to task difficulty (e.g., Guo et al., 2012). However, as indicated above, we do not anticipate LPC amplitude differences between the learning groups.

Concerning the picture-naming task, we expect to observe an output language effect with larger N400 negativity and LPC positivity in L2 naming compared to L1 naming. The N400 modulation was predicted from previous studies that show a greater difficulty of picture-naming in the non-dominant language (Francis, Augustini & Sáenz, 2003; Radman et al., 2018). Similarly, the LPC amplitude would be larger as the difficulty of the naming task increases (Jackson et al., 2001; Kieffaber et al., 2013) – in our case, larger LPC amplitude in L2 vs. L1 naming. Finally, we did not anticipate between-group differences due to the output language since the greater difficulty in L2 vs. L1 naming is a robust effect that does not depend on different factors such as the age of the participants (e.g., Gollan, Montoya, Cera & Sandoval, 2008), or the bilingual fluency (i.e., fluent bilinguals show the effect – e.g., English–Spanish bilinguals, Sholl et al., 1995; Chinese–English bilinguals, Chen, Cheung & Lau, 1997).

2. Method

2. 1. Participants

Fifty-six native Spanish speakers, students at the University of Granada, participated in this study for course credits. At present, it is difficult to consider Spanish speakers as monolinguals due to their previous experience with other languages (e.g., L2 instruction at school and high school). However, we used different inclusion criteria to ensure that, when conducting the study, participants had as little contact as possible with languages other than Spanish. On a daily basis, they had to report that, (a) they had no contact with any other language different from Spanish, (b) their last contact with a FL had to be at high school, (c) they had never received any FL instruction apart from regular education and, (d) they had never obtained a FL certification. Importantly, the participants in the lexical and semantic training group were selected from the same pool (i.e., university students), and were assigned to the groups at random, reducing the possibility that the participants' previous linguistic experience would determine possible differences between the lexical and semantic training groups when acquiring new L2 words.

None of the participants reported history of language disabilities and they had normal or corrected-to-normal visual acuity. Twenty-eight participants were randomly assigned to the semantic training group (22 women, 6 men) and the rest of participants were assigned to the lexical training group (22 women, 6 men). The mean age of participants in the semantic ($M = 21.85$, $SD = 4.34$) and lexical group ($M = 20.79$, $SD = 3.18$) was equated, $t(54) = 1.05$, $p = .30$. Two participants in the semantic training and three participants in the lexical training were left-handed. The experiment was undertaken in accordance with the 1964 Helsinki declaration. The Ethic Committee at the University of Granada approved the experimental procedure used in the study (Number issued by the Ethical Committee: 86/CEIH/2015) and each participant provided written informed consent before taking part in the experiment. The required sample size was determined using the G*Power program 3.1.9.2 (Faul, Erdfelder, Lang & Buchner, 2007). To achieve a 95% statistical power at $\alpha = .05$ and a small effect size (0.42) computed based on a $\eta_p^2 = .15$, in a 2 x 2 repeated measures analysis of variance (ANOVA), the required sample size was $N = 22$. Thus, the sample used in this study was sufficient to capture the effects evaluated in the experiment.

2. 2. Design and materials

In the current study, participants learned a set of 60 Spanish (L1) – Vimmi (L2) words. The L2 vocabulary learning was conducted in a single session. After finishing the training part, the participants performed the evaluation tasks while recording the continuous electroencephalogram (EEG). The participants carried out two tasks to evaluate the acquisition of L2 words in the following order: a translation task and a naming task.

L2 Vocabulary learning tasks

Two L2 training methods were used. Half of the participants were subjected to a lexical training and the remaining half of participants performed the semantic training. When the L2 training phase was considered, a 2 x 10 mixed design was used with type of L2 training (semantic training, lexical training) as a between-participants factor and block of training (10 levels, from the first block to the last block of training) as a within-participants variable. In order to control several linguistic variables, Spanish words from six semantic categories were chosen (Battig & Montague, 1969). Three categories were from the living domain (four-footed animals, body-parts, fruits) and three from the non-living domain (kitchen utensils, musical instruments, vehicles). Within each semantic category, ten words were selected. Hence, the total amount of words to be learned was 60. In addition, for each word, a picture denoting its concept was selected (Pérez & Navalón, 2003; Snodgrass & Vanderwart, 1980). Each Spanish word and its corresponding picture were randomly paired with a Vimmi word. The complete set of stimuli is presented as Supplementary Materials (Supplementary Materials, Table S1). Statistical description of the material is detailed as Supplementary Materials (Supplementary Materials, Table S2): Lexical characteristics of Spanish words taken from Cuetos, Glez-Nosti, Barbón, and Brysbaert (2011), and Pérez and Navalón (2003 – length, word frequency, orthographic neighbourhood, age of acquisition, familiarity, manipulability, typicality, imageability, and concreteness of words), lexical properties of Vimmi words (length, orthographic neighbourhood, and shared graphemes with the Spanish words) and visual properties of pictures taken from Pérez and Navalón (visual complexity of pictures, image agreement, image variability, and picture-name agreement).

The 60 Spanish–Vimmi words were randomly grouped into 10 learning sets of 6 words. All participants performed ten blocks with the learning and practice tasks (i.e., L2 vocabulary learning tasks) described below. In each block, the 60 Vimmi words to be learned were presented once. Hence, the training involved a total of 10 exposures to each L2 word and the total amount of trials was 600. The number of expositions to the FL words was maintained constant across participants and no threshold was stablished to continue with the evaluation tasks. This way of proceeding was done to ensure that all participants were exposed to the new words an equal number of times and thus avoiding possible between-group differences in episodic memory associated to differences in the number of repetitions of the FL words in the lexical and semantic learning group.

In the L2 vocabulary learning phase, the task started with the message "word learning". Afterwards, participants pressed the space bar and 500 ms later the stimuli to be learned were presented (e.g., *plátano*, banana in Spanish). In the semantic training, a picture appeared in the middle of the screen (e.g., a picture

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

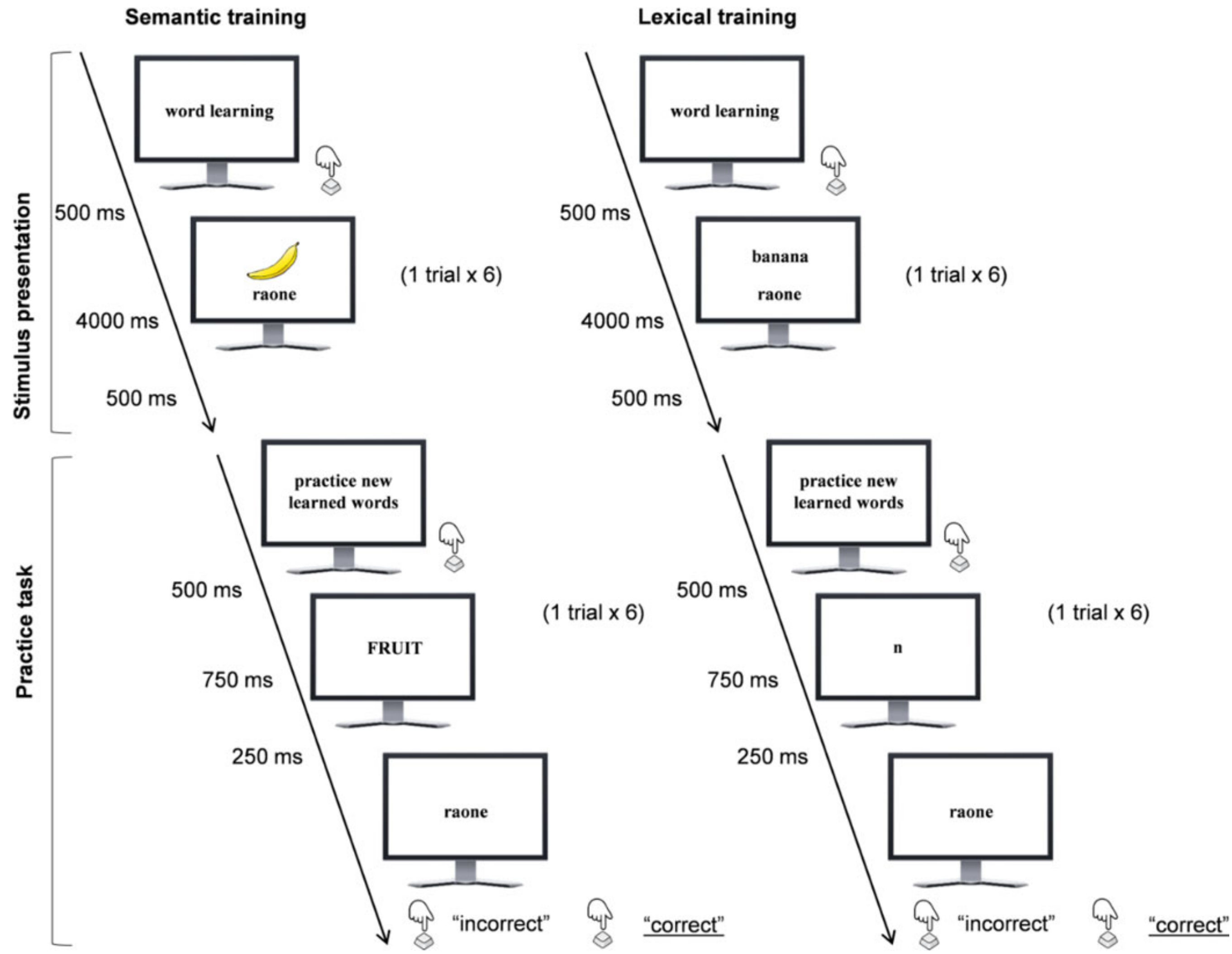

Figure 1. Description of the L2 vocabulary learning methods used in the current study.

depicting a banana – 7 x 7 cm average size) and the Vimmi word below (e.g., *raone* – Courier New, bold font, 18 point size). In the lexical training, the Spanish and the Vimmi words were presented one above the other (e.g., *plátano* and *raone* – Courier New, bold font, and 18 point size). The stimulus to be learned remained on the screen for 4000 ms and 500 ms later, the next stimulus appeared. After the presentation of 6 stimuli, the message "practice new learned words" appeared on the screen; participants pressed the space bar to continue and 500 ms later, they performed the practice task. In the practice task of the semantic training, a category name was presented in capital letters for 750 ms (e.g., fruit). After 250 ms, a Vimmi word appeared in the middle of the screen until the participant responded (e.g., *raone*). Participants pressed the Z and M key to indicate whether or not the Vimmi word denoted an exemplar of the category previously presented (e.g., "yes" in the *plátano-raone* pair). Within each set, on half of the trials the Vimmi word was an exemplar of the semantic category and on the remaining trials it was not. Across the entire task, there were the same number of "yes" and "no" responses in each of the six semantic categories. In addition, across participants, all Vimmi words were assigned to the "yes" and "no" responses. The assignment of Z and M keys to "yes" and "no" responses was counterbalanced across participants (see Figure 1).

The practice task in lexical learning consisted of the presentation of a grapheme in the middle of the screen for 750 ms (e.g., *n*). Following a 250 ms delay, a Vimmi word was presented until the participants' response (e.g., *raone*). Participants had to indicate by pressing the Z and M keys whether the Spanish translation of the Vimmi word contained the grapheme previously presented ("yes" in the *plátano-raone* pair). Each grapheme was one letter randomly selected from each Vimmi word. On half of the trials, this grapheme was present in the Spanish translation and on the remaining half of trials it was not. Across the task, half of the letters were vowels and the remaining graphemes were consonants. Across participants, all graphemes were assigned to "yes" and "no" responses. The assignment of Z and M keys to "yes" and "no" responses was counterbalanced across participants. The procedure used in the lexical and semantic training was taken from García-Gámez and Macizo (2020).

Translation task

In this task, the 60 Spanish–Vimmi word pairs learned in the L2 vocabulary-training phase were used. The 60 Spanish words were presented for translation into Vimmi (forward translation) and the 60 Vimmi words were presented for translation into Spanish (backward translation). The order in which the two translation tasks were presented was counterbalanced across participants. In addition, the words within each translation task were presented in random order. On each trial, a fixation point appeared for 1000 ms in the middle of the screen followed by the word to be translated (e.g., *raone*) for 500 ms (Arial, 30 point size, black font, white background). A white screen then remained until the participants' response. Participants were required to say aloud the translation of each word.

Picture-naming task

In this task, participants were presented with 60 pictures denoting the meaning of the words learned in L2. Participants performed the naming task in Spanish and in Vimmi. The order in which they performed these naming tasks was counterbalanced across participants. Within each naming task, the pictures were presented in random order. On each trial, a fixation point was presented for 1000 ms, after which the picture to be named (e.g., the picture of a banana – 14 x 14 cm average size) appeared in the middle of the screen for 500 ms, followed by a white screen that remained until the participants' response.

2. 3. Procedure

The study was conducted in a single day, with participants being tested individually in the EEG recording room. E-prime experimental software was used for stimulus presentation and data acquisition (Schneider, Eschman & Zuccolotto, 2002). Firstly, participants performed a familiarization task where all the 60 pictures used in the experiment were presented along with their Spanish names. The participants were instructed to see each picture and its more common name in Spanish and to press the space bar any time they wanted to see another picture.

After finishing the familiarization phase, the L2 vocabulary-learning task was introduced (semantic or lexical training). The mean time needed to complete the lexical and semantic training was similar, 89.54 minutes ($SE = 12.48$) and 83.84 minutes ($SE = 10.88$) respectively, t (54) = 1.82, $p = .07$. After completing the L2 acquisition phase, an elastic cap with 64 electrodes was placed on the head of the participants to record the EEG signal. Finally, participants continued with the evaluation tasks (translation and naming tasks). The evaluation tasks order was not counterbalanced across participants. To name a picture (regardless of the output language) it is always required to retrieve its meaning (Potter & Faulconer, 1975). However, the translation task can be performed without semantic processing though L1-L2 lexical connections (Kroll & Stewart, 1994). The reason underlying the fixed order in which participants received the evaluation tasks (the translation task followed by the picture-naming task) was to avoid the possible semantic effect that the naming of pictures (i.e., the retrieval of their meaning) could have in the subsequent translation task (i.e., in case of having presented the picture-naming task followed by translation task).

The complete session lasted approximately 150 minutes.

2. 4. Electrophysiological recording

After the learning phase, the continuous Electroencephalogram (EEG) was recorded from 64 scalp electrodes installed on an elastic cap (Quick-Cap, Neuroscan Inc.). The electrodes were arranged according to the extended 10-20 International System (Jasper, 1958). The EEG was initially recorded against an electrode placed in the midline of the cap (between Cz and CPz) and later off-line re-referenced to a common average reference. The decision to re-reference offline to the average of all electrodes instead of using the linked mastoids as a reference was based on previous research studies on the subject (e.g., ERPs studies on initial L2 vocabulary learning, Yum et al., 2014; see also Martin et al., 2013). In addition, we followed the recommendations provided by Makoto Miyakoshi's preprocessing pipeline (Swartz Center for Computational Neuroscience, 2019), which suggest re-referencing the data to average (see Yao, Qin, Hu, Dong, Vega & Sosa, 2019, for a comparison between the advantages and disadvantages associated with each type of reference used for EEG and ERP practice).

In order to control for blinks, a pair of electrodes was placed above and below the left eye. The horizontal and vertical eye movements were captured by another pair of electrodes located on the outer canthus in both eyes. The EEG signal was amplified by using the Neuroscan Synamps2 amplifiers (El Paso, TX) and filtered using a band pass of 0.01-100 Hz and digitalized at a 500 Hz sampling rate. The electrode impedance was kept below 5 kΩ. Digital tags were assigned to the stimuli of interest for each task. The EEG signal was analysed by using the open-source toolbox ERPLab (López-Calderón & Luck, 2014). Eye blinks and other artifacts components were identified and corrected by means of independent component analysis (ICA) and careful visual inspection of the recordings. Epochs were baseline corrected using the mean activity during the -100 to 0 ms pre-stimuli period. Based on previous studies, a low pass filter of 30 Hz was applied (Verhoef, Roelofs & Chwilla, 2009; Willems, Özyürek & Hagoort, 2008).

For ERP data analysis, a representative sub-array of nine channels was used (Blackford, Holcomb, Grainger & Kuperberg, 2012; Chauncey, Grainger & Holcomb, 2008; Grainger, Kiyonaga & Holcomb, 2006; Kuperberg, Delaney-Busch, Fanucci & Blackford, 2018). As Blackford and colleagues (2012) mentioned, following this design, a single analysis of variance (ANOVA) can be used for each time window analysis; and, hence, this electrode selection is a good meeting point between the use of a simple design and the correct description of the overall distribution of the effects. The selected electrodes formed three columns in the left (F3, C3, P3), central (Fz, Cz, Pz), and right (F4, C4, P4) sides extending from the front to the back of the head (see Supplementary Materials, Figure S3).

2. 5. Data analysis

The aim of the study was to examine electrophysiological data associated with the evaluation of learning tasks according to the type of L2 training. The behavioral data associated with the EEG recording session could not be analysed due to malfunctioning of the recording system. The stimulus presentation system (E-prime) correctly sent the identification codes for each type of stimulus and condition to the EEG signal recorded with the Neuroscan system. The EEG signal was correctly recorded but the system failed to record the behavioral responses provided by the participant because, when the experiment was programmed, an incorrect port was selected for sending signals from the E-prime software to the response box[1].

Fortunately, the current work was part of a research project in which different methods for FL vocabulary acquisition (learning by gestures, images, etc.) were evaluated (PSI2016-75250-P). Regarding the comparison between lexical and semantic trainings, before collecting the data reported in the current study, behavioral research was conducted (García-Gámez & Macizo, 2020). The authors examined the impact of lexical and semantic FL vocabulary trainings on the processing of the new acquired words within sentences and out-of-context tasks (lexical decision, picture-naming and translation). The participants in García-Gámez and Macizo did not take part in the current study but were selected from the same pool (students at the University of Granada). Regarding the translation task, the behavioral results revealed a training effect with slower translation times in the semantic group (1670 ms) than in the lexical group (1321 ms). The translation direction effect was also found with slower translation times in forward translation (1738 ms) than in backward translation (1252 ms). Furthermore, this effect (RT difference between forward translation minus backward translation) was larger in the semantic group (628 ms) than in the lexical group (342 ms). In the picture-naming task, the output language effect was obtained with slower naming time in the FL (484 ms) than in the L1 (1,076 ms). In addition, a training effect was observed which was modulated by the output language. There were no between-group differences when participants named the pictures in L1; however, when they named the pictures in the FL, the semantic group was faster (842 ms) than the lexical group (1311 ms).

When relevant, the electrophysiological data obtained in this study will be compared to the behavioral study conducted by García-Gámez and Macizo (2020). However, we would like to emphasize that this joint interpretation of empirical evidence has to be taken with caution. Although participants were part of the same pool, and the same training and evaluation procedures were used in both studies, data from two separate reports cannot be equated with the simultaneous recording of behavioral and electrophysiological data.

The analyses conducted with electrophysiological data according to the evaluation tasks are described below. In these analyses, two participants, one lexical learner and one semantic learner, were removed from data analysis due to the high number of rejected epochs (more than 100 out of 120) obtained in both evaluation tasks.

[1]We are grateful to an anonymous reviewer for suggesting that the loss of behavioral data in the current study could have been addressed by the inclusion of a second behavioral session with the same evaluation tasks (translation and picture-naming). The reviewer anticipated that, in this second session, the lexical group as well as the semantic group should show a better performance in L1 naming than in L2 naming. We agree with this prediction. In laboratory studies on L2 vocabulary acquisition, participants exhibit inferior performance on delayed versus immediate post-tests (Lawson & Hogben, 1998; Sagarra & Alba, 2006). This could imply an increased difficulty in accessing lexical/semantic information of new L2 words. On the contrary, no differences would be expected between evaluation sessions in L1 tasks (e.g., L1 picture-naming) since it is the participants' native language. Nevertheless, at the end of data collection in the current study, we did not come up with this possible post-test behavioral assessment. Moreover, the comparison between electrophysiological measures (obtained in an immediate evaluation) and behavioral measures (obtained in a delayed evaluation) would not be direct since additional consolidation processes would take place in the delayed post-test and not in the immediate post-test (see Bakker, Takashima, Van Hell, Janzen & McQueen, 2015; Liu & Van Hell, 2020, for novel word consolidation effects in novel word learning). In order to mitigate the loss of behavioral data in our study and with the aim of comparing electrophysiological and behavioral results, in this work, we considered the behavioral results reported by García-Gámez and Macizo (2020). Note that immediate post-test tasks were used in both studies.

Translation task

An ANOVA was conducted on ERP data with the Type of Training (semantic training, lexical training) as a between-participant factor and Translation Direction (forward, backward), Laterality (left, central, right) and Anterior-Posterior Electrode Distribution (frontal, central, parietal) as within participant variables. After eliminating trials in which the EEG signal was contaminated, the averages in each experimental condition (forward translation, backward translation) comprised a mean of 105 out of 120 trials in the lexical group of training trials and 97 trials out of 120 trials in the semantic group of training. The statistical analyses were conducted on three consecutive time windows time-locked to the onset of the words to be translated. These temporal windows were selected based on careful visual inspection and previous electrophysiological studies addressing translation tasks (Palmer, van Hooff & Havelka, 2010; Phillips, Klein, Mercier & de Boysson, 2006). The 150-300 ms window was used to index the P200 component associated with lexical access (Guo et al., 2012). The 300-500 ms time window was established to index the N400 component usually present in translation tasks (Phillips et al., 2006). Finally, the 500-700 ms time window was used to explore the LPC associated with the re-evaluation of information and semantic processing (Kolk & Chwilla, 2007).

Naming task

An ANOVA was conducted on ERP data with Type of Training (semantic training, lexical training) as a between participant factor and the Language in which participants named the pictures (L1 naming, L2 naming), Laterality (left, central, right) and Anterior-Posterior Electrode Distribution (frontal, central, parietal) as within participant variables. After eliminating trials in which the EEG signal was contaminated by eye movements or amplifier saturations, the averages in each experimental condition (L1 naming, L2 naming) were comprised of a mean of 106 out of 120 trials in the lexical group of training and 100 out of 120 trials in the semantic group of training. Three temporal windows were established time-locked to the presentation of the pictures to be named. These time-windows were selected based on visual inspection and previous studies on picture-naming (Blackford et al., 2012; Verhoef, 2009). The 150-250 ms window was used to index the N200 component usually found in picture-naming tasks (Christoffels, Firk & Schiller, 2007). The 250-400 ms time window was selected to index the N400 component (Blackford et al., 2012). Finally, the 400-700 ms time window was selected to evaluate late ERP modulations usually found when participants name pictures across languages (Christoffels et al., 2007).

In each time window of the translation and naming tasks, the main analyses are presented first. Afterwards, the results obtained in the lexical and semantic training group are presented separately. In all repeated-measure ANOVAs reported in the study, the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) for nonsphericity of variance was used for all *F*-ratios with more than one degree of freedom in the denominator; reported here are the original *df*, the corrected probability level, and the ϵ correction factor.

3. Results

Firstly, we report the behavioral results obtained during the training task. Afterwards, we present the electrophysiological results of the two evaluation tasks (translation, naming) that participants performed after the training session.

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

3. 1. L2 vocabulary learning tasks

Correct responses in the L2 lexical training and L2 semantic training phases were 71.18% ($SE = 2.63$), 95% CI [65.91, 76.46], and 84.95% ($SE = 2.63$) 95% CI [79.68, 90.22], respectively. The reaction times (RTs) associated with correct responses in the training tasks were trimmed following the procedure described by Tabachnick and Fidell (2001) to eliminate univariate outliers. Raw scores were converted to standard scores (z-scores). Following standardization, any data points that were 3 *SD* outside the normal distribution were considered outliers. After removing outliers from the distribution, z-scores were calculated again. The filter was applied in recursive cycles until no observations were outside 3 *SD*. The percentages of data excluded from the L2 lexical and semantic training phases were 4.39% and 10.00%, respectively. We conducted an analysis of variance (ANOVA) on recall percentages and RTs with type of L2 training (semantic training, lexical training) as a between-participants factor and block of training (10 levels, from the first block to the last block of training) as a within-participants variable.

The accuracy analyses revealed a significant effect of type of training, $F(1, 54) = 13.70$, $p < .001$, $\eta^2 = .20$. Participants recalled more L2 words in the semantic training (84.95%, $SE = 5.35$), relative to the lexical training (71.18%, $SE = 6.40$). The main effect of block of training was significant, $F(9, 486) = 46.16$, $p < .001$, $\eta^2 = .46$. The Type of Training x Block of Training interaction was significant too, $F(9, 486) = 2.07$, $p = .03$, $\eta^2 = .04$. Linear trend analysis was significant in the lexical training, $F(1, 27) = 58.45$, $p < .001$, $\eta^2 = .68$, and the semantic training, $F(1, 27) = 59.72$, $p < .001$, $\eta^2 = .69$. In the lexical training, participants were more accurate at the end of the training (80.30%, $SE = 2.97$) compared to the beginning of the training (58.75%, $SE = 1.90$). The same pattern of results was found in the semantic training; participants recalled more L2 words at the end of the training (89.17%, $SE = 3.00$) compared to the beginning of the procedure (72.62%, $SE = 1.90$). Thus, there was a practice effect with more correct responses at the end of the training relative to the beginning of the learning process (see Figure 2). However, the differences in recall percentage between the first and the last block of training were higher in the lexical training (21.55%) compared to the semantic training (16.55%).

The latency analyses revealed a significant effect of the type of training, $F(1, 54) = 46.95$, $p < .001$, $\eta^2 = .47$. RTs were faster during semantic training (1292 ms, $SE = 195$) compared to the lexical training (1964 ms, $SE = 294$). The main effect of block of training was significant, $F(9, 486) = 42.59$, $p < .001$, $\eta^2 = .44$. In addition, the Type of Training x Block of Training interaction was significant, $F(9, 486) = 3.10$, $p = .001$, $\eta^2 = .06$.

In the lexical training phase, linear trend analysis was significant, $F(1, 27) = 26.75$, $p < .001$, $\eta^2 = .50$, with faster responses at the end of training (1568 ms, $SE = 71.43$) relative to the beginning of training (2358 ms, $SE = 108.14$). Linear trend analysis was also significant during semantic training, $F(1, 27) = 107.02$, $p < .001$, $\eta^2 = .80$, with faster RTs at the end of training (1090 ms, $SE = 71.43$) relative to the beginning of the training (1695 ms, $SE = 108.13$). Thus, the more our participants trained, the faster they responded; however, the difference in RT between the first and the last block of training was larger in the lexical group (791 ms) compared to the semantic group (604 ms).

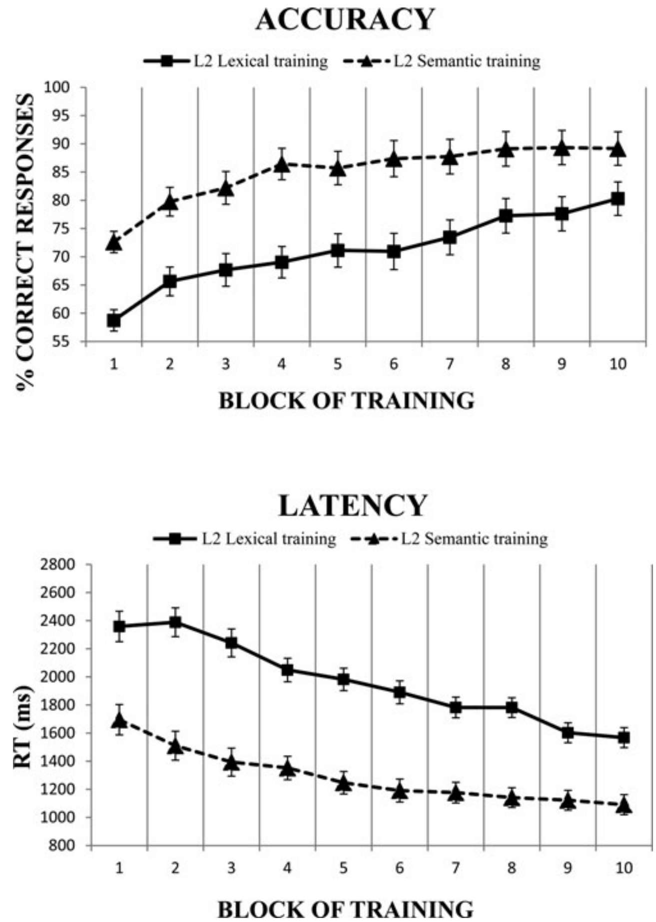

Figure 2. Percentage of correct responses and response times (in milliseconds) obtained in the L2 training task (lexical training, semantic training) across blocks of training. Error bars represent standard errors.

3. 2. Translation task

The results obtained in the translation task are presented in Figure 3. The complete pattern of statistical results obtained in the translation task for the lexical and semantic training groups is reported in Table 1.[2]

150-300 ms time window

The type of training effect was not significant, $F < 1$. No interactions including the type of training factor were significant (all $ps > .05$). No main effects or interactions were significant when these effects were evaluated separately in the lexical training and the semantic training group (all $ps > .05$).

300-500 ms time-window

The type of training effect was not significant, $F < 1$. The interaction between Type of Training x Laterality was marginal, $F(2, 104) = 2.89$, $p = .06$, $\epsilon = .99$, $\eta^2 = .05$. There were significant differences between types of training in the left hemisphere,

[2]False discovery rate (FDR) corrections were computed (Benjamini & Hochberg, 1995) to control for false positive results when multiple tests were conducted in our study. To this end, we first conducted pairwise comparisons in each spatial location and time window to examine the translation direction effect and the naming language effect in the lexical and semantic learning groups. The pattern of data after false discovery rate corrections for 9 pairwise comparisons in each time-window and learning group was similar to that reported in text. A detailed description of the results obtained in these analyses are available as Supplementary Materials (Supplementary Materials).

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

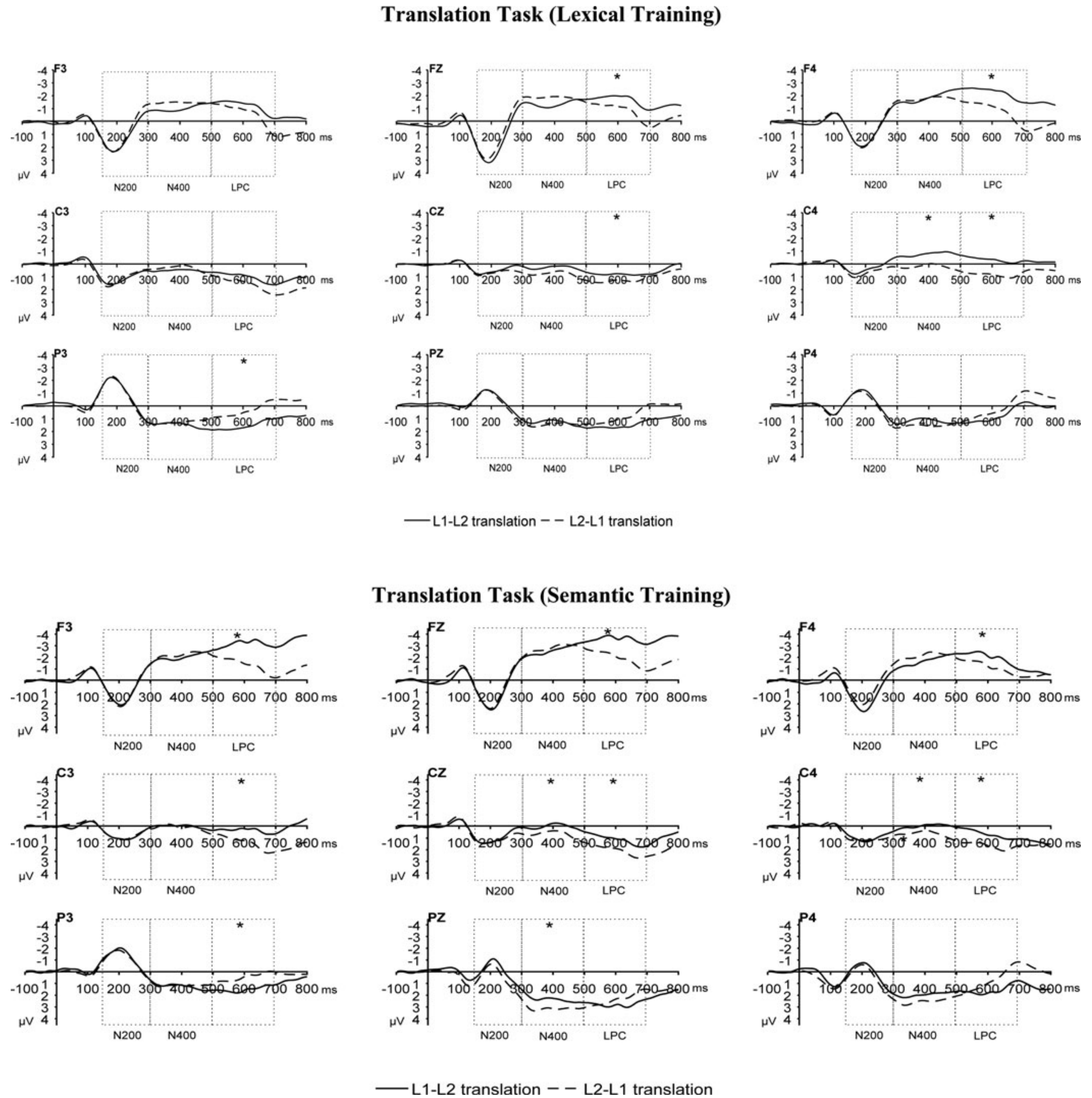

Figure 3. Grand Average ERPs for the forward (L1-L2) translation (solid lines) and backward (L2-L1) translation (dashed lines) obtained in the Lexical and Semantic Groups of Training. The time-windows analyzed in the study are framed by a dotted rectangle. * p < .05

$F(1, 52) = 5.69$, $p = .02$, $\eta^2 = .10$. Brain waves were more negative in the semantic training ($M = -.29$ µV, $SE = .14$) compared with the lexical training ($M = .20$ µV, $SE = .14$). The type of training effect was not significant in the midline areas, $F < 1$, or in the right region, $F(1, 52) = 2.00$, $p = .16$, $\eta^2 = .04$.

In the Lexical training, the Translation Direction x Lateral axis interaction was significant, $F(2, 52) = 4.77$, $p = .01$, $\epsilon = .95$, $\eta^2 = .16$. There were significant differences between translation directions in the right hemisphere, $F(1, 26) = 10.17$, $p = .004$, $\eta^2 = .28$. More negative brain waves were observed in the forward translation direction ($M = -.49$ µV, $SE = .20$) than in the backward translation direction ($M = .01$ µV, $SE = .19$). The differences between translation directions in the left hemisphere and the midline regions were not significant (all ps > .05).

In the Semantic training, the Translation Direction x Anterior-Posterior axis x Lateral axis interaction was marginal, $F(4, 104) = 2.27$, $p = .07$, $\epsilon = .87$, $\eta^2 = .08$. In frontal regions, the translation direction effect was not significant in the left area, $t(26) = 0.25$, $p = .80$, the midline area, $t(26) = 0.26$, $p = .79$, or the right area, $t(26) = 0.72$, $p = .47$. In central regions, the

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

Table 1. Statistical Analyses Performed on ERP Data. Translation Direction Effects and Interactions in the Translation task Performed by the Lexical Group of Training and the Semantic Group of Training

		Lexical Training		*Semantic Training*	
Time-window	Effects	*F*	*p*	*F*	*p*
150-300 ms	TD	.26	.61	2.23	.15
	TD x AP axis	1.99	.17	.62	.47
	TD x LM axis	1.51	.23	.46	.61
	TD x AP axis x LM axis	.84	.45	2.26	.09
300-500 ms	TD	1.27	.27	3.35	.07
	TD x AP axis	1.33	.26	1.57	.22
	TD x LM axis	4.77	.01*	2.72	.09
	TD x AP axis x LM axis	.92	.42	2.27	.07
500-700 ms	TD	12.05	.001**	12.82	.001**
	TD x AP axis	10.47	.001**	16.71	.001**
	TD x LM axis	4.46	.03*	.33	.63
	TD x AP axis x LM axis	.46	.72	1.78	.16

Abbreviations: TD: Translation Direction, AP: Anterior-Posterior, LM: Lateral-Medial. *$p \leq .05$, **$p \leq .001$

translation direction effect was not significant in the left area, $t(26) = 0.37$, $p = .72$, but brain waves were more negative in forward translation relative to backward translation in midline regions, $t(26) = 2.13$, $p = .04$, and right regions, $t(26) = 3.11$, $p = .004$. Finally, in parietal regions, the translation direction was not significant in the left area, $t(26) = 0.43$, $p = .67$, it was significant in midline regions, $t(26) = 2.47$, $p = .02$, and it was marginal in the right region, $t(26) = 1.95$, $p = .06$.

Thus, the translation direction effect in the 300-500 ms time-window was more broadly distributed in the semantic training group than in the lexical training group.

500-700 ms time-window

The Type of Training x Anterior-Posterior axis x Lateral axis interaction was significant, $F(4, 208) = 2.50$, $p = .04$, $\epsilon = .95$, $\eta^2 = .05$.

In the Lexical training, the Translation Direction x Anterior-Posterior axis interaction was significant, $F(2, 52) = 10.47$, $p = .001$, $\epsilon = .62$, $\eta^2 = .29$. Additionally, the Translation Direction x Lateral axis interaction was significant, $F(2, 52) = 4.46$, $p = .03$, $\epsilon = .66$, $\eta^2 = .15$. In frontal areas, the translation direction effect was not significant in the left region, $t(26) = 1.64$, $p = .11$, but it was significant in the midline region, $t(26) = 2.31$, $p = .02$, and in the right region, $t(26) = 3.60$, $p = .003$. In central areas, there were no significant differences between translation directions in the left region, $t(26) = 1.47$, $p = .15$, although they were significant in the midline region, $t(26) = 2.30$, $p = .03$, and in the right region, $t(26) = 4.03$, $p < .001$. Finally, in parietal areas, the differences between translation directions were significant in the left area, $t(26) = 3.56$, $p = .001$, and marginally significant in the midline, $t(26) = 1.76$, $p = .09$, and right areas, $t(26) = 1.73$, $p = .09$.

In the Semantic training, the Translation Direction x Anterior-Posterior axis interaction was significant, $F(2, 52) = 16.71$, $p < .001$, $\epsilon = 78$, $\eta^2 = .39$. There were differences between translation directions in frontal regions, $F(1,26) = 18.24$, $p < .001$, $\eta^2 = .41$, with more positive brain waves for the backward ($M = -1.41$ µV, $SE = .63$) than for the forward ($M = -2.97$ µV, $SE = .58$) translation direction. In central areas, the differences between translation directions were significant, $F(1, 26) = 16.46$, $p < .001$, $\eta^2 = .39$, with more positive amplitude for the backward ($M = 1.52$ µV, $SE = .27$) than for the forward translation direction ($M = .52$ µV, $SE = .23$). Finally, the difference between translation directions was significant in parietal regions, $F(1, 26) = 8.87$, $p = .006$, $\eta^2 = .25$. In this case, mean amplitude was more positive in the forward ($M = 1.83$ µV, $SE = .52$) than in the backward translation direction ($M = 1.07$ µV, $SE = .57$).

3. 3. Picture-Naming task

The results obtained by the lexical and semantic training groups are presented in Figure 4. Additionally, summaries of statistical analyses conducted for both training groups are reported in Table 2.

150-250 ms time-window

The type of training was not significant, $F(1, 52) = 1.51$, $p = .22$, $\eta^2 = .03$, nor did this variable interact with any other (all *ps* > .05). No main effects or interactions were significant when brain waves of the lexical training group and the semantic training group were analyzed separately (all *ps* > .05).

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

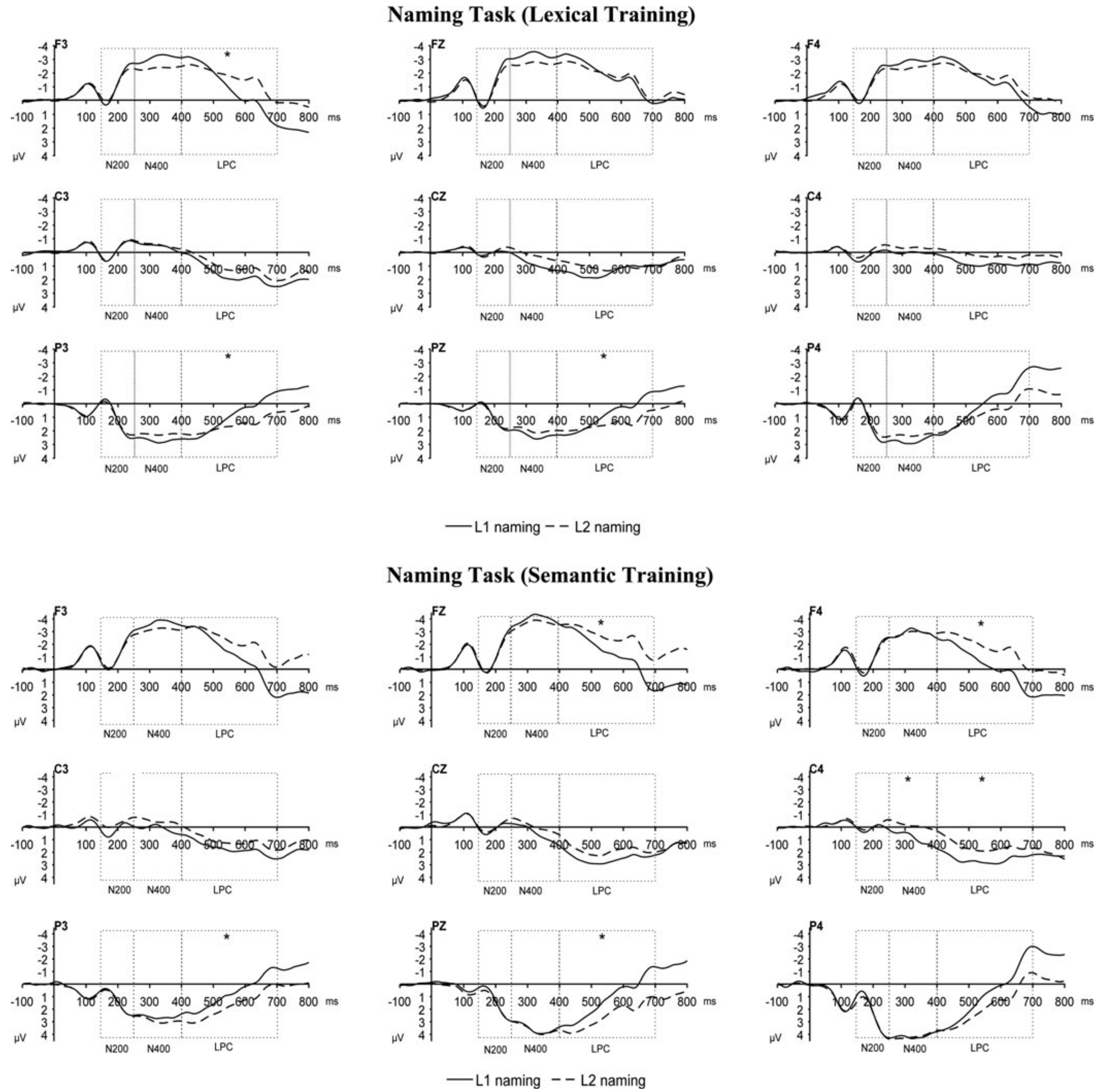

Figure 4. Grand Average ERPs for the Spanish naming (L1 – solid lines) and Vimmi naming (L2 – dashed lines) obtained in the Lexical and Semantic Groups of Training. The time-windows analyzed in the study are framed by a dotted rectangle. * $p < .05$

250-400 ms time-window

The main effect of type of training was not significant, $F < 1$, and this variable did not interact with any other (all *p*s > .05).

In the Lexical training, no main effects or interactions between variables were significant (all *p*s > .05).

In the Semantic training, the Naming Language x Anterior-Posterior axis interaction was significant, $F(2, 52) = 5.86$, $p = .007$, $\epsilon = .86$, $\eta^2 = .18$. In frontal areas, the naming language was not significant ($p > .05$). The effect was significant in central areas, $F(1, 26) = 9.86$, $p = .004$, $\eta^2 = .28$, with more negative amplitudes in L2 naming ($M = -.36$ Mv, $SE = .39$) than in L1 naming ($M = .33$ Mv, $SE = .39$). In parietal areas, the effect was not significant, $F < 1$.

400-700 ms time-window

In this time-window, the Type of Training x Lateral axis interaction was significant, $F(2, 104) = 5.23$, $p = .007$, $\epsilon = .96$, $\eta^2 = .09$. Significant differences between types of training were obtained in the right hemisphere, $F(1, 52) = 13.11$, $p < .001$, $\eta^2 = .20$, with more positive amplitude in the semantic training ($M = .71$ Mv, $SE = .18$) than in the lexical training ($M = -.24$ Mv, $SE = .18$). The type of training was not significant in any other brain region (all *p*s > .05).

Table 2. Statistical Analyses Performed on ERP Data. Naming Language Effects and Interactions in the Naming task Performed by the Lexical Group of Training and the Semantic Group of Training

		Lexical Training		*Semantic Training*	
Time-window	Effects	*F*	*p*	*F*	*p*
150-300 ms	TD	.04	.83	.06	.80
	TD x AP axis	.35	.64	.64	.46
	TD x LM axis	.13	.78	.28	.73
	TD x AP axis x LM axis	.06	.86	.76	.52
300-500 ms	TD	.75	.39	.86	.36
	TD x AP axis	2.26	.13	5.86	.007*
	TD x LM axis	.58	.54	1.15	.30
	TD x AP axis x LM axis	2.20	.08	.56	.63
500-700 ms	TD	1.39	.25	2.65	.12
	TD x AP axis	4.05	.03*	6.02	.01*
	TD x LM axis	.81	.43	.52	.53
	TD x AP axis x LM axis	1.05	.37	.12	.93

Abbreviations: NL: Naming Language, AP: Anterior-Posterior, LM: Lateral-Medial. $^{*}p \leq .05$

In the Lexical training, the Naming Language x Anterior-Posterior axis interaction was significant, $F(2, 52) = 4.05$, $p = .03$, $\epsilon = .76$, $\eta^2 = .13$. The naming language effect was not significant in frontal and central regions (all *p*s > .05). However, in parietal areas, there were significant differences between naming languages, $F(1, 26) = 5.40$, $p = .03$, $\eta^2 = .17$, with more positive amplitude in L2 naming ($M = 1.26$ Mv, $SE = .41$) than in L1 naming ($M = .60$ Mv, $SE = .33$).

In the Semantic training, the Naming Language x Anterior-Posterior axis interaction was significant, $F(2, 52) = 6.02$, $p = .01$, $\epsilon = .71$, $\eta^2 = .19$. There were differences between naming languages in frontal regions, $F(1,26) = 5.90$, $p = .02$, $\eta^2 = .23$, with more positive brain waves in L1 naming ($M = -.95$ Mv, $SE = .50$) than in L2 naming ($M = -1.94$ Mv, $SE = .55$). In central regions, the differences between naming languages were significant too, $F(1, 26) = 5.45$, $p = .03$, $\eta^2 = .21$, with more positive amplitude in L1 naming ($M = 2.07$ Mv, $SE = .41$) than in L2 naming ($M = 1.34$ Mv, $SE = .28$). In parietal areas, the differences between naming languages were also significant, $F(1, 26) = 4.88$, $p = .03$, $\eta^2 = .19$. However, in this case more positive amplitude was found in the L2 naming ($M = 1.89$ Mv, $SE = .43$) than in L1 naming ($M = .97$ Mv, $SE = .49$).

The results obtained in the picture-naming task revealed a naming language effect in the LPC component. In posterior regions (i.e., parietal region), the LPC amplitude was more positive in the L2 naming compared to the L1 naming in both learning groups. However, the naming language only modulated the N400 component in the semantic group but not in the lexical group. In the semantic training group, the N400 amplitude was more negative in central regions when the learners named the pictures in L2 vs. L1. Thus, while the semantic learning group displayed a biphasic N400-LPC response (i.e., a negative modulation followed by a positive ERP modulation), the lexical group only showed a positive electrophysiological response in late time-windows. At this point, we conducted additional exploratory analyses. Firstly, we evaluated the possible relationship between the magnitude of the N400 and the LPC naming language effect and, secondly, we explored the distribution of the effects across participants and training groups (see Tanner & Van Hell, 2014, for a systematic analysis of the relationship between the magnitude of the N400/P600 effects and the evaluation of individual differences in morphosyntactic processing with monolingual individuals). To this end, we selected two representative electrodes in which the N400 and the LPC effects were mainly observed in our study (C4 and P3, respectively). For each participant, we computed the magnitude of the naming language effect on the N400 component (L1 naming minus L2 naming in the 250-400 ms time-window) and the LPC component (L2 naming minus L1 naming in the 400-700 ms time-window). The distribution of the N400 and LPC effects in the two learning groups are described in Figure 5. The correlation between the magnitude of the N400 effect and the LPC effect was not significant either in the lexical group, $r = .15$, $p = .47$, or in the semantic group, $r = .10$, $p = .63$. In other words, the magnitude of the N400 effect was not associated with an increase or reduction of the LPC effect in any of the learning groups.

4. Discussion

In the field of foreign language vocabulary acquisition, one of the main differences between novice and expert learners is the strength of the relationships between meaning and L2 words. Less fluent learners use their L1 lexicon when they process L2 words. In contrast, expert bilinguals rely on semantic analysis during L2 processing (Frenck-Mestre & Prince, 1997; Kroll & Stewart, 1994; Talamas et al., 1999). Many behavioral studies show that this increased semantic analysis is also possible in early stages of FL vocabulary acquisition when the learning tasks stimulate conceptual processing (picture-word association training) as compared to other learning methods (word-word association training – e.g., Comesaña et al., 2009). However, to our knowledge, there are no electrophysiological studies examining the consequences that the type of training (semantic vs. lexical) might have when individuals acquire vocabulary in a foreign language.

In our study, we compared the learning of L2 words from a lexical group that received pairs of L1-L2 words and practiced with a lexical task (grapheme monitoring) and a semantic group that received pictures-L2 words and practiced with a semantic task (semantic categorization). The results obtained at the end of 10 blocks of training revealed an overall learning effect with better performance (higher number of correct answers and faster response times) at the end of training compared to the beginning of the learning. Although the learning curve was independent of the type of learning, the performance of the training task was lower in the lexical group than in the semantic group

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

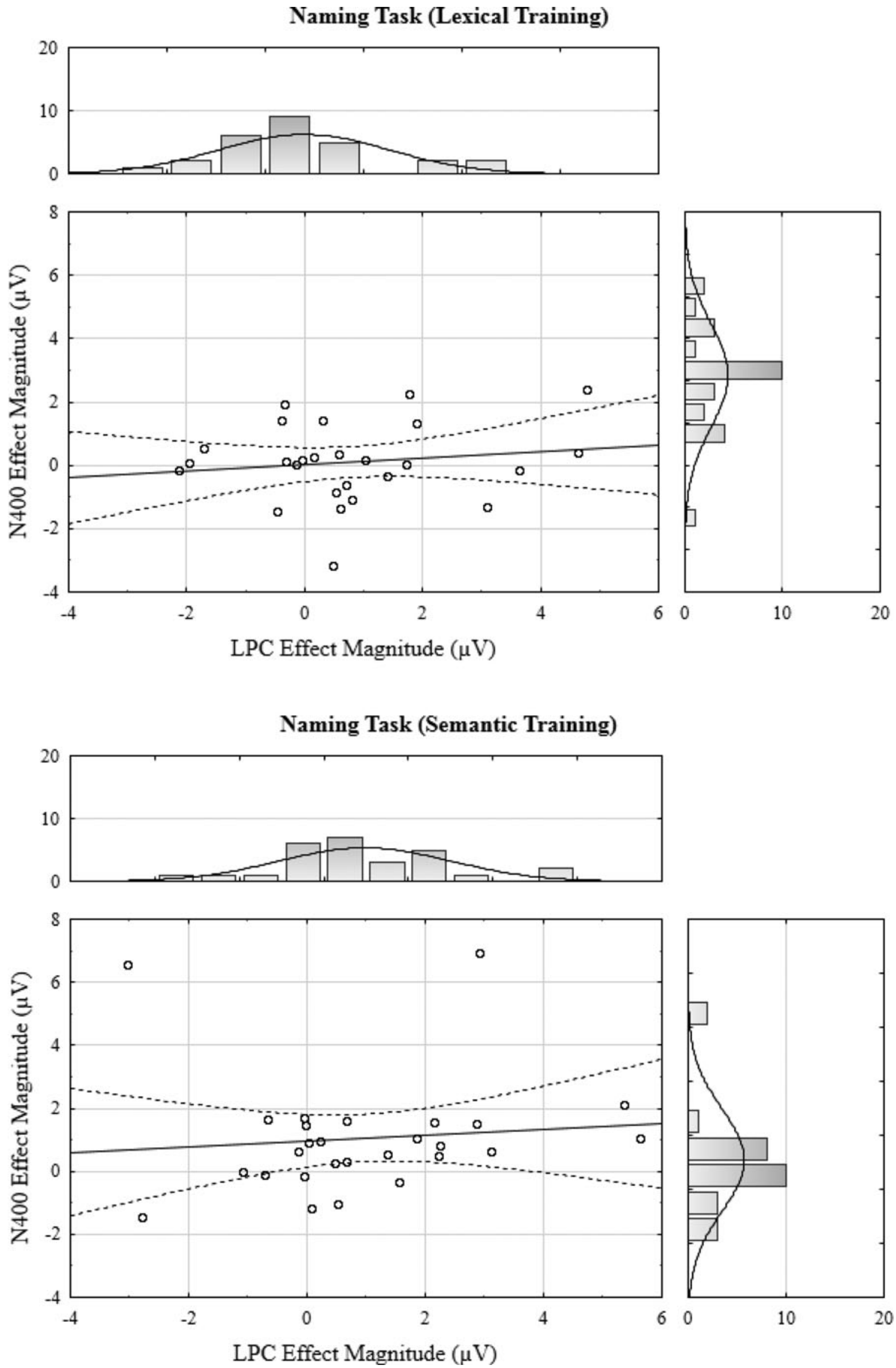

Figure 5. Scatterplots showing the relationship between the 400 and the LPC effect magnitudes in the picture-naming task across L2 learners in the lexical and semantic training group (top and bottom panels, respectively). The solid line indicates the best-fit line from the correlation analysis for each learning group. The dashed lines show the 95% confidence intervals. At the right and the top side of each graph are shown the histograms representing the distribution of the N400 and LPC effects, respectively.

https://doi.org/10.1017/S1366728921000456 Published online by Cambridge University Press

from the beginning of the acquisition process. This pattern of results was the same as that reported in a previous behavioral study (e.g., García-Gámez & Macizo, 2020) with the same training tasks as those used here. Less efficient performance in the lexical vs. semantic training group found in the current study seems to be due to the greater difficulty of the grapheme monitoring than the semantic categorization (see García-Gámez & Macizo, 2020, for a detailed discussion on this issue and for experimental data showing the greater difficulty of the lexical vs. semantic task). However, regardless of the differences between the two training groups, word recall was higher at the end of training (above 80%) compared to the beginning of learning.

After the learning phase, we considered electrophysiological measures associated with the translation and naming tasks to evaluate the impact of lexical and semantic learning on FL vocabulary acquisition. In line with the predictions provided in the introduction section, the results found in the translation task revealed greater brain-wave negativity in the 300-500 ms time-window in the semantic training group compared to the lexical training group. This pattern of results appears to indicate that the semantic training group was engaged in deeper semantic processing. In fact, previous studies show that the amplitude of the N400 component as an index of semantic processing depends on the fluency of participants in L2. Thus, L2 words produce larger N400 effects in highly proficient bilinguals compared with L2 learners with low proficiency (Midgley et al., 2009; Rodríguez-Fornells, Cunillera, Mestres-Missé & de Diego-Balaguer, 2009, for a review). Therefore, the acquisition of FL vocabulary based on semantic training would reflect a pattern of brain activity more similar to that observed in fluent bilinguals during the performance of linguistic tasks[3].

The electrophysiological data reported here can be compared to the behavioral evidence outlined by García-Gámez and Macizo (2020). The authors showed slower translation times in the semantic group than in the lexical group and this effect was taken as evidence of an increased conceptual processing in the semantic group which would slow down the translation task. In our study, the larger N400 amplitude observed in the semantic vs. lexical group would seem to indicate that the learners in the semantic group would retrieve the meaning of the words in the translation task to a greater extent than the lexical group. However, as noted in the method section, the comparison between the behavioral data from a previous study and the electrophysiological results reported here should be treated with caution because it is not comparable to the collection of behavioral and electrophysiological outcomes from the same sample of participants.

Furthermore, as expected, the results obtained in the current study revealed a translation direction effect with greater N400 amplitudes in the forward translation than in the backward translation. This pattern of data is consistent with other studies in which higher N400 is observed when the difficulty in retrieving lexical information increases (Rugg, 1990; Van Petten & Kutas, 1990). Moreover, this effect is consistent with the outcomes of many behavioral studies in which more difficulty (slower response time) is found when participants retrieve lexical forms in forward vs. backward translation (Kroll et al., 2010; Poarch et al., 2014). To illustrate, García-Gámez and Macizo (2020) observed this translation direction effect with the same language pair as those used in our study; slower translation times in L1-L2 direction (Spanish–Vimmi) than in L2-L1 direction (Vimmi–Spanish).

On the other hand, the N400 modulations due to the translation direction (greater negativity in forward than backward translation) were more widely distributed in the semantic training group (central and parietal regions) than the lexical training group (right central region – Figure 3). The broader distribution of the N400 component when the semantic learning group performed the forward translation task is in line with previous research in which highly fluent bilinguals show N400 modulations located in the same scalp locations (Deacon, Dynowska, Ritter & Grose-Fifer, 2004; Holcomb, 1993; Kerkhofs, Dijkstra, Chwilla & de Bruijn, 2006; Palmer et al., 2010). Thus, electrophysiological response observed in the semantic vs. lexical training group closely resembles that of highly fluent bilinguals.

In the late time window (500-700 ms), the translation direction modulated the LPC component. However, the magnitude and distribution of this component was similar in the semantic and lexical training groups. In translation tasks, the LPC has been associated with reprocessing the information and checking the degree of correspondence between the source and target language (Guo et al., 2012; Kolk & Chwilla, 2007). As we anticipated in the introduction section, we expected this rechecking mechanism would operate in the lexical and semantic training, so we did not predict differences between the learning groups in the LPC amplitude. Thus, the results obtained in our study suggest that, in translation tasks, the type of training (semantic vs. lexical learning) modulates the retrieval of words in the output language (modulation of the N400 component) but does not determine the evaluation of correspondences between words across languages.

With regards to the picture-naming task, we anticipated N400 modulations depending on the output language. In particular, we expect to find greater negativity in L2 naming compared to L1 naming due to the greater difficulty involved in retrieving lexical-semantic information when bilinguals perform the picture-naming task in their L2 language relative to their native language (Jackson et al., 2001; Kieffaber et al., 2013). In behavioral studies, this disadvantage in L2 naming (slower latency and reduced response accuracy in L2 vs. L1 naming) is a robust effect that seems relatively independent of several linguistic factors (see Hanulová, Davidson & Indefrey, 2011, for a review). Thus, in our study, we predicted this effect across FL learners regardless of the type of training that would have received. However, electrophysiological data revealed that the naming language effect was found in the semantic training group but not in the lexical training group. This was an unexpected pattern of results. In particular, the semantic training group showed greater negativity in central brain regions when the naming task was conducted in L2 compared to the L1 naming task. These data seem to indicate that the learners in the semantic group behaved similarly to fluent bilinguals during the L2 naming task. To illustrate, it has been observed in previous studies that expert vs. novice learners show greater sensitivity to conceptual processing when performing L2 linguistic tasks, which is reflected in a higher N400 (Midgley et al., 2009). On the other hand, in the semantic

[3]We would like to point out that in no case are we equating the performance of the participants in the semantic training group with fluent bilinguals. On the contrary, in the discussion section, we are suggesting that the performance of learners in the semantic group resembles the performance of fluent L2 learners within studies comparing less vs. more fluent students with laboratory tasks (e.g., translation task, translation recognition task, lexical decision, picture-naming, etc. – Comesaña et al., 2012; Kroll & Stewart, 1994; Sholl et al., 1995; Talamas et al., 1999). Furthermore, this comparison is limited to studies on L2 vocabulary leaving behind other linguistic components that would characterize the performance of fluent bilinguals (phonetic/phonological analysis, syntax, prosody, language control functions, etc.).

group, a greater N400 amplitude when naming in L2 vs. L1 would come from the weight of the connections between the semantic and the lexical systems in L1 and L2. It is widely demonstrated that, in L2 learners, the weight of the connections between the meaning and the words in their native language are stronger than the connections between the meaning and the words in L2 (De Groot & Poot, 1997; Kroll & Stewart, 1994; Poarch et al., 2014; Van Hell & De Groot, 2008). Thus, the participants in the semantic group would show greater ability in retrieving the name of the pictures in their native language than in their second language.

However, it may be asked why the participants in the lexical group did not show electrophysiological differences in the N400 time-window associated with the language used to name the pictures (L1 vs. L2). This outcome appears to suggest that, regardless of the naming language, these participants performed the task in the same manner by retrieving the name of the picture in their native language. Support for this conclusion comes from several behavioral studies showing that in less fluid bilinguals, L2 picture-naming involves the retrieval of lexical information in the L1 lexicon (Kroll et al., 2002). In the N400 time-window, the results obtained in the picture-naming task would indicate that the learners in the semantic training group made an early differentiation between their languages and were able to use a processing route from semantics to L1 words or to L2 words according to task (L1 naming or L2 naming, respectively). In contrast, the lexical training group seemed to use a common processing route via L1 lexical activation regardless of the language needed to perform the naming task.

In any case, as indicated above, we had no a priori reason to anticipate differences between the semantic and lexical learning groups in the naming language effect. In fact, we expected to find that the two training groups were sensitive to the greater difficulty of lexical-semantic retrieval when they named pictures in L2 vs. L1. We acknowledge that this pattern of outcomes has to be taken with caution for several reasons. Firstly, in order to offer a complete pattern of results, we analyzed the effect of each task (translation and picture-naming), factor (translation direction, naming language) and ERP component (N400 and LPC) separately for the semantic and lexical training groups. However, in the picture-naming task, statistical analyses did not reveal the type of training effect nor did this factor interact with the naming language in the N400 time window (250-400 ms). On the other hand, the naming language effect found in the semantic training group was very small and located in one topographical location only (central region, specifically, C4 electrode, see Figure 4).

Concerning the LPC component when participants performed the picture-naming task, in the late temporal window, a LPC modulation was found due to the output language. Specifically, in posterior regions, greater positivity was found when participants named the pictures in L2 than in L1. The polarity of the LPC found in our study was reversed in frontal regions; a pattern similar to that reported in previous work in which L2 learners show an opposite distribution of the LPC over the anterior scalp than the posterior scalp (e.g., Guo et al., 2012). The amplitude of the LPC component has been also related to the difficulty in rechecking information. For instance, Guo et al. reported greater amplitude of the LPC component in a translation recognition task when word pairs were not translation equivalents but were semantically related compared to unrelated word pairs. In the related word pair condition, further reanalysis of the information would be necessary to confirm that the words were not correct translations of each other. In our study, the greater amplitude of the LPC in the L2 vs. L1 naming task would imply the increased need to verify that the selected word actually designated the picture name in L2. Finally, the LPC modulation found in parietal regions was observed in the semantic and lexical training groups. As noted in the introduction section, we had no a priori reason to expect differences between the training groups in the LPC amplitude as an index of difficulty in rechecking information.

5. Conclusions

To conclude, the results obtained in this study seem to indicate that the type of training used for L2 vocabulary acquisition determines the strength and type of connections established between the semantic system and new FL words. Electrophysiological evidence seems to indicate a deeper semantic processing in the picture-L2 word association training (semantic learning) compared to the L1 word-L2 word association learning (lexical learning). Furthermore, the connections between the conceptual system and L2 lexicon seems to be available for the learners who acquired L2 words through a semantic training. These conclusions were mainly based on the translation task where it was observed that the translation direction effect due to the greater semantic processing in forward vs. backward translation was more widely distributed in the semantic training group. On the contrary, the pattern of N400 modulations obtained in the picture-naming task does not allow to establish clear differences between the learning groups, which could indicate that, at least in this study, word retrieval from a picture did not vary as a function of the learning method. Finally, it should be noted that the approach of our study was based on the comparison of the average performance of two groups of learners according to the type of training they received (semantic vs. lexical). However, previous electrophysiological evidence shows that the linguistic experience of individuals determines the way they manage several language sub-processes (e.g., morphosyntactic anomalies, such as agreement and tense violations, Tanner & Van Hell, 2014). Consequently, possible individual differences in L2 vocabulary acquisition with semantic and lexical procedures should be further explored.

Acknowledgements. This research was funded by the Spanish Ministry of Economy and Competitiveness (grant number PSI2016-75250-P) and the Spanish Ministry of Science and Innovation: Reference of the project: PID2019-111359GB-I00 / SRA (State Research Agency /10.13039/501100011033) awarded to Pedro Macizo. All procedures performed in this study involving human participants were in accordance with the ethical standards of the research ethical committee at the University of Granada (number issued by the Ethical Committee: 86/CEIH/2015). We thank Borja Molina Zea for computer support and software programming for the processing of EEG data reported in the current work.

Supplementary Material. For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1366728921000456

References

Abutalebi J, Della Rosa PA, Green DW, Hernandez M, Scifo P, Keim R, Cappa SF and Costa, A (2012) Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex* **22**, 2076–2086. doi: 10.1093/cercor/bhr287

Bakker I, Takashima A, Van Hell JG, Janzen G and McQueen JM (2015) Changes in theta and beta oscillations as signatures of novel word consolidation. *Journal of Cognitive Neuroscience* **27**, 1286–1297. doi: 10.1162/jocn_a_00801

Barcroft J (2002) Semantic and structural elaboration in L2 lexical acquisition. *Language Learning* **52**, 323–363. doi:10.1111/0023-8333.00186

Battig WF and Montague WE (1969) Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology* **80**, 1–46. doi: 10.1037/h0027577

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**, 289–300.

Bentin S, McCarthy G and Wood CC (1985) Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology* **60**, 343–355. doi: 10.1016/0013-4694(85)90008-2

Bialystok E, Craik FI, Grady C, Chau W, Ishii R, Gunji A and Pantev C (2005) Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage* **24**, 40–49. doi: 10.1016/j.neuroimage.2004.09.044

Blackford T, Holcomb PJ, Grainger J and Kuperberg GR (2012) A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition* **123**, 84–99. doi: 10.1016/j.cognition.2011.12.007

Chauncey K, Grainger J and Holcomb PJ (2008) Code-switching effects in bilingual word recognition: A masked priming study with event-related potentials. *Brain and Language* **105**, 161–174. doi: 10.1016/j.bandl.2007.11.006

Chen HC, Cheung H and Lau S (1997) Examining and reexamining the structure of Chinese-English bilingual memory. *Psychological Research* **60**, 270–283. doi: 10.1007/BF00419413

Cheung H and Chen HC (1998) Lexical and conceptual processing in Chinese-English bilinguals: Further evidence for asymmetry. *Memory and Cognition* **26**, 1002–1013. doi: 10.3758/BF03201179

Christoffels IK, Firk C and Schiller NO (2007) Bilingual language control: An event-related brain potential study. *Brain Research*, **1147**, 192–208. doi: 10.1016/j.brainres.2007.01.137

Christoffels IK, Ganushchak L and Koester D (2013) Language conflict in translation: An ERP study of translation production. *Journal of Cognitive Psychology* **25**, 646–664. doi: 10.1080/20445911.2013.821127

Comesaña M, Perea M, Piñeiro A and Fraga I (2009) Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology* **104**, 22–33. doi: 10.1016/j.jecp.2008.10.004

Comesaña M, Soares AP and Lima C (2010) Semantic representations of new cognate vs. noncognate words: Evidence from two second language learning methods. *Procedia-Social and Behavioral Sciences* **5**, 199–203. doi: 10.1016/j.sbspro.2010.07.072

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

Coulson S, Federmeier KD, Van Petten C and Kutas M (2005) Right hemisphere sensitivity to word-and sentence-level context: evidence from event-related brain potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **31**, 129–147. doi: 10.1037/0278-7393.31.1.129

Cuetos F, Glez-Nosti M, Barbón A and Brysbaert M (2011) Spanish word frequencies based on film subtitles. *Psicológica* **32**, 133–144.

Deacon D, Dynowska A, Ritter W and Grose-Fifer J (2004) Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology* **41**, 60–74. doi: 10.1111/1469-8986.00120

De Groot AMB and Poot R (1997) Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning* **47**, 215–264. doi: 10.1111/0023-8333.71997007

Dunn BR, Dunn DA, Languis M and Andrews D (1998) The relation of ERP components to complex memory processing. *Brain and Cognition* **36**, 355–376. doi: 10.1006/brcg.1998.0998

Faul F, Erdfelder E, Lang AG and Buchner A (2007) G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* **39**, 175–191.doi: 10.3758/BF03193146

Ferré P, Sánchez-Casas R and Guasch M (2006) Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and nonproficient Spanish-Catalan bilinguals. *Language Learning* **56**, 571–608. doi: 10.1111/j.1467-9922.2006.00389.x

Finkbeiner M and Nicol J (2003) Semantic category effects in second language word learning. *Applied Psycholinguistics* **24**, 369–383. doi:10.1017/S0142716403000195.

Francis WS, Augustini, BK and Sáenz SP (2003) Repetition priming in picture naming and translation depends on shared processes and their difficulty: evidence from spanish-english bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**, 1283–1297. doi: 10.1037/0278-7393.29.6.1283

Frenck-Mestre C and Prince P (1997) Second language autonomy. *Journal of Memory and Language* **37**, 481–501. doi: 10.1006/jmla.1997.2526

García-Gámez AB and Macizo P (2019) Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics* **40**, 473–507. doi: 10.1017/S0142716418000656

García-Gámez AB and Macizo P (2020) The way in which foreign words are learned determines their use in isolation and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition* **46**, 364–379. doi: 10.1037/xlm0000721

Genesee F (2014) Second language learning in school settings: Lessons from immersion. In Reynols AG (ed.), *Bilingualism, multiculturalism, and second language learning*. New York: Psychology Press, pp. 203–222.

Gollan TH, Montoya RI, Cera C and Sandoval TC (2008) More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language* **58**, 787–814. doi: 10.1016/j.jml.2007.07.001

Grainger J, Kiyonaga K and Holcomb PJ (2006) The time course of orthographic and phonological code activation. *Psychological Science* **17**, 1021–1026. doi: 10.1111/j.1467-9280.2006.01821.x

Greenhouse SW and Geisser S (1959) On methods in the analysis of profile data. *Psychometrika* **24**, 95–112. doi: 10.1007/bf02289823

Guo T, Misra M, Tam JW and Kroll JF (2012) On the time course of accessing meaning in a second language: An electrophysiological and behavioral investigation of translation recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **38**, 1165–1186. doi: 10.1037/a0028076

Hackley SA, Woldorff M and Hillyard SA (1990) Cross-modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. *Psychophysiology* **27**, 195–208. doi: 10.1111/j.1469-8986.1990.tb00370.x

Hanulová J, Davidson DJ and Indefrey P (2011) Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes* **26**, 902–934. doi: 10.1080/01690965.2010.509946

Holcomb PJ (1993) Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology* **30**, 47–61. doi: 10.1111/j.1469-8986.1993.tb03204.x

Hoshino N and Kroll JF (2008) Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition* **106**, 501–511. doi: 10.1016/j.cognition.2007.02.001

Hudson PT and Bergman MW (1985) Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. *Journal of Memory and Language* **24**, 46–58. doi: 10.1016/0749-596x(85)90015-4

Jackson GM, Swainson R, Cunnington R and Jackson SR (2001) ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition* **4**, 169–178. doi: 10.1017/S1366728901000268

Jasper HH (1958) The 10-20 electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology* **10**, 371–375. doi: 10.1016/0013-4694(58)90053-1

Jost LB, Radman N, Buetler KA and Annoni JM (2018) Behavioral and electrophysiological signatures of word translation processes. *Neuropsychologia* **109**, 245-254. doi: 10.1016/j.neuropsychologia.2017.12.034

Kerkhofs R, Dijkstra T, Chwilla DJ and De Bruijn ER (2006) Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research* **1068**, 170-183. doi: 10.1016/j.brainres.2005.10.087

Kieffaber PD, Kruschke JK, Cho RY, Walker PM and Hetrick WP (2013) Dissociating stimulus-set and response-set in the context of task-set switching. *Journal of Experimental Psychology: Human Perception and Performance* **39**, 700-719. doi:10.1037/a0029545

Kolk H and Chwilla D (2007) Late positivities in unusual situations. *Brain and Language* **100**, 257-261.doi: 10.1016/j.bandl.2006.07.006
Kousaie S and Phillips NA (2012) Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research* **1446**, 71–90. doi:10.1016/j.brainres.2012.01.052
Kroll JF and Linck JA (2007) Representation and skill in second language learners and proficient bilinguals. In Kecskes I and Albertazzi L (eds.), *Cognitive Aspects of Bilingualism*. Dordrecht: Springer, pp. 237-269.
Kroll JF, Michael E, Tokowicz N and Dufour R (2002) The development of lexical fluency in a second language. *Second Language Research* **18**, 137-171.doi: 10.1191/0267658302sr201oa
Kroll J and Stewart E (1994) Category interference in translation and picture naming: evidence foe asymmetric connections between bilingual memory representations. *Journal of Memory and Language* **33**, 149-174. doi: 10.1006/jmla.1994.1008
Kroll J, Van Hell J, Tokowicz N and Green DW (2010) The revised hierarchical model: a critical review and assessment. *Bilingualism*: *Language and Cognition* **13**, 373-381. doi: 10.1017/S136672891000009X
Kuperberg GR, Delaney-Busch N, Fanucci K and Blackford T (2018) Priming production: Neural evidence for enhanced automatic semantic activity preceding language production in schizophrenia. *NeuroImage: Clinical* **18**, 74-85. doi: 10.1016/j.nicl.2017.12.026
Kutas M and Hillyard S (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* **207**, 287-301. doi: 10.1126/science.7350657
Lawson M J and Hogben D (1998) Learning and recall of foreign-language vocabulary: Effects of a keyword strategy for immediate and delayed recall. *Learning and Instruction* **8**, 179–194. doi:10.1016/S0959-4752(97)00016-9
Leckey M and Federmeier KD (2020) The P3b and P600 (s): Positive contributions to language comprehension. *Psychophysiology* **57**, e13351. doi: 10.1111/psyp.13351
Liu Y and van Hell JG (2020) Learning novel word meanings: An ERP study on lexical consolidation in monolingual, inexperienced foreign language learners. *Language Learning* **70**, 45-74. doi: 10.1111/lang.12403
Lopez-Calderón J and Luck SJ (2014) ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience* **8**, 1-14. doi: 10.3389/fnhum.2014.00213
Martin CD, Strijkers K, Santesteban M, Escera C, Hartsuiker RJ and Costa A (2013) The impact of early bilingualism on controlling a language learned late: an ERP study. *Frontiers in Psychology* **4**, 815, 1-15. doi: 10.3389/fpsyg.2013.00815
McLaughlin J, Osterhout L and Kim A (2004) Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience* **7**, 703-704. doi: 10.1038/nn1264
Midgley KJ, Holcomb PJ and Grainger J (2009) Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics* **22**, 281-300. doi: 10.1016/j.jneuroling.2008.08.001
Monsell S, Doyle MC and Haggard PN (1989) Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General* **118**, 43-71. doi: 10.1037//0096-3445.118.1.43
Moreno EM and Kutas M (2005) Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research* **22**, 205-220. doi: 10.1016/j.cogbrainres.2004.08.010
Neville HJ, Mills DL and Lawson D (1992) Fractionating language: Different neural systems with different sensitive periods. *Cerebral Cortex* **2**, 244-258. doi: 10.1093/cercor/2.3.244
Palmer SD, van Hooff JC and Havelka J (2010) Language representation and processing in fluent bilinguals: Electrophysiological evidence for asymmetric mapping in bilingual memory. *Neuropsychologia* **48**, 1426-1437. doi: 10.1016/j.neuropsychologia.2010.01.010
Pérez M.-Á and Navalón C (2003) Normas españolas de 290 nuevos dibujos: Acuerdo en la denominación, concordancia de la imagen, familiaridad, complejidad visual y variabilidad de la imagen. *Psicológica* **24**, 215-241.
Phillips NA, Klein D, Mercier J and de Boysson C (2006) ERP measures of auditory word repetition and translation priming in bilinguals. *Brain Research* **1125**, 116-131. doi: 10.1016/j.brainres.2006.10.002
Poarch GJ, Van Hell JG and Kroll JF (2014) Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism* **18**, 357-371. doi: 10.1017/S1366728914000558
Potter MC and Faulconer BA (1975) Time to understand pictures and word. *Nature* **253**, 437-438. doi: 10.1038/253437a0
Pu H, Holcomb PJ and Midgley KJ (2016) Neural changes underlying early stages of L2 vocabulary acquisition. *Journal of Neurolinguistics* **40**, 55-65. doi: 10.1016/j.jneuroling.2016.05.002
Radman N, Britz J, Buetler K, Weekes BS, Spierer L and Annoni JM. (2018) Dorsolateral prefrontal transcranial direct current stimulation modulates language processing but does not facilitate overt second language word production. *Frontiers in Neuroscience* **12**, 490. doi: 10.3389/fnins.2018.00490
Rice CA and Tokowicz N (2020) A review of laboratory studies of adult second language vocabulary training. *Studies in Second Language Acquisition* **42**, 439-470. doi: 10.1017/s0272263119000500
Rodríguez-Fornells A, Cunillera T, Mestres-Missé A and de Diego-Balaguer R (2009) Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**, 3711-3735. doi: 10.1098/rstb.2009.0130
Rugg MD (1990) Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory and Cognition* **18**, 367-379. doi: 10.3758/BF03197126
Rugg MD and Curran T (2007) Event-related potentials and recognition memory. *Trends in Cognitive Sciences* **11**, 251-257. doi: 10.1016/j.tics.2007.04.004
Sagarra N and Alba M (2006) The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *The Modern Language Journal* **90**, 228–243. doi: 10.1111/j.1540-4781.2006.00394.x
Schneider W, Eschman A and Zuccolotto A (2002) *E-Prime user's guide* (Version 1.1) Pittsburg: Psychology Software Tools.
Sholl A, Sankaranarayanan A and Kroll JF (1995) Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science* **6**, 45-49. doi: 10.1111/j.1467-9280.1995.tb00303.x
Snodgrass JG and Vanderwart M (1980) A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* **6**, 174-215. doi: 10.1037/0278-7393.6.2.174
Soskey L, Holcomb PJ and Midgley KJ (2016) Language effects in second-language learners: A longitudinal electrophysiological study of Spanish classroom learning. *Brain Research* **1646**, 44-52. doi: 10.1016/j.brainres.2016.05.028
Swartz Center for Computational Neuroscience (2019) *Makoto's preprocessing pipeline*. https://sccn.ucsd.edu/wiki/Makoto's_preprocessing_pipeline
Tabachnick, B., and Fidell, L. (2001) *Using multivariate statistics* (4th edition) Boston: Allyn and Bacon.
Talamas A, Kroll JF and Dufour R (1999) Form related errors in second language learning: A preliminary stage in the acquisition of L2 vocabulary. *Bilingualism: Language and Cognition* **2**, 45-58. doi: 10.1017/S1366728999000140
Tanner D and Van Hell JG (2014) ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* **56**, 289-301. doi: 10.1016/j.neuropsychologia.2014.02.002
Tellier M (2008) The effect of gestures on second language memorization by young children. *Gesture* **8**, 219-235. doi: 10.1075/gest.8.2.06tel
Tonzar C, Lotto L and Job R (2009) L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning* **59**, 623-646. doi: 10.1111/j.1467-9922.2009.00519.x
Van Hell JG and De Groot AM (2008) Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica* **128**, 431-451. doi: 10.1016/j.actpsy.2008.03.010
Van Petten C and Kutas M (1990) Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition* **18**, 380-393. doi: 10.3758/BF03197127

Verhoef K, Roelofs A and Chwilla DJ (2009) Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition* **110**, 84-99. doi: 10.1016/j.cognition.2008.10.013

Willems RM, Özyürek A and Hagoort P (2008) Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience* **20**, 1235-1249. doi: 10.1162/jocn.2008.20085

Yao D, Qin Y, Hu S, Dong L, Vega MLB and Sosa PAV (2019) Which reference should we use for EEG and ERP practice? *Brain Topography* **32**, 530-549. doi: 10.1007/s10548-019-00707-x

Yum YN, Midgley KJ, Holcomb PJ and Grainger J (2014) An ERP study on initial second language vocabulary learning. *Psychophysiology* **51**, 364-373. doi: 10.1111/psyp.12183