Manuscript for

Native language inhibition predicts more successful second language learning: Evidence of two ERP pathways during learning.

Manuel F. Pulido
Center for Language Science, Department of Spanish, Italian and Portuguese, The Pennsylvania State University

Mailing address: Department of Spanish, Italian and Portuguese, The Pennsylvania State University, 442 Burrowes Building, University Park, 16802, USA

Email: mpulido@psu.edu

ORCID: 0000-0002-8179-2593

Keywords
event-related potentials; ERP; learning; bilingualism; language control; multiword units

Author Contributions
M. F. Pulido is the sole author of this paper and is responsible for its content.

Declarations of Interest: None.

Abstract

How some individuals succeed in learning a second language as adults is still an unsolved question in cognitive neuroscience. At the brain level, adults' electrophysiological responses to input in a second language may differ after completing different types of training. However, there is limited understanding of what neural pathways are activated as learning unfolds, and which patterns of activation lead to successful learning. Using brain event-related potentials, this study explored whether individual brain responses to practice difficulty during second language learning predict learning outcomes. English-speaking learners of Spanish practiced completing newly learned phrases in their second language. For some learners, all the choices presented during practice were "easy" because non-target choices were unrelated distractors. In the more "difficult" practice mode, however, learners had to avoid choosing a competing word that would be acceptable based on their native language, but not in the second language being learned. Performance during practice was similar in both groups of learners. Critically, divergence in event-related potentials indicated alternative strategies to practice, based on the level of difficulty. At the group level, learners completing the easier practice revealed increased monitoring when making responses; in the difficult condition, learners showed inhibition of their native language (i.e., an N400 for phrases congruent with the native language) to avoid interference during word selection. Individual brain responses indexing the degree of native language inhibition predicted learning rates in tests.

Keywords

event-related potentials; ERP; learning; bilingualism; language control; multiword units

1. Introduction

In the last several decades, considerable energy has been devoted to investigating the factors that limit success in adult second language (L2) learning. Previous research has shed light on the role of age of acquisition (Hartshorne, Tenenbaum, & Pinker, 2018; Johnson & Newport, 1989), motivation (Csizér & Dörnyei, 2005; Grey, Williams, & Rebuschat, 2015), language learning context (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Tagarelli, Borges Mota, & Rebuschat, 2015), individual differences in cognitive ability (Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; Sunderman & Kroll, 2009) or cross-language similarity between the native and the non-native language (De Groot & Keijzer, 2000; Degani & Tokowicz, 2010a, 2010b; Pulido & Dussias, 2020). However, much less attention has been given to the question of *how* it is that some adults succeed in learning a second language.

It is now known that there are qualitative differences in how native and non-native speakers accomplish linguistic tasks such as reading, or processing normal and anomalous sentences (Arnon & Christiansen, 2017; Felser, Roberts, Marinis, & Gross, 2003; Foucart & Frenck-Mestre, 2011; Foucart, Martin, Moreno, & Costa, 2014; Hopp, 2010; Tokowicz & MacWhinney, 2005). While non-native speakers' brain responses are often native-like when processing linguistic aspects of the L2 that are similar to their own first language, they differ in other cases. In particular, linguistic aspects that only partly overlap between the first (L1) and the second (L2) language, causing conflict, tend to elicit different electrophysiological responses in non-native speakers during comprehension, relative to native speakers (Foucart & Frenck-Mestre, 2011, 2012; Tokowicz & MacWhinney, 2005).

These findings are in line with the well-established role of the native language in adult language learning. Aspects that are *incongruent* (i.e., differ) across two languages are harder to learn and to process, and are easier to forget (De Groot & Keijzer, 2000; Tokowicz & MacWhinney, 2005). For example, words whose meaning is only partly congruent across languages result in notorious learning difficulties (Degani & Tokowicz, 2010b, 2010a). Similar difficulties are found in so-called "false friends", which are words that have similar form but different meanings across languages (e.g., *librería* is not 'library' in Spanish, but 'bookstore') (Brenders, van Hell, & Dijkstra, 2011). A critical question is how individuals overcome such difficulties in learning.

Findings in the last decade have revealed that learning conditions may influence the pathways engaged by the brain to represent new information (Batterink, Reber, Neville, & Paller, 2015; Batterink, Reber, & Paller, 2015; Morgan-Short et al., 2012). In a learning study using an artificial language, Morgan-Short and colleagues found that the type of training influenced subsequent comprehension. Implicit (immersion-like) training conditions produced brain potentials during comprehension that were indistinguishable from those of native speakers, but this was not the case for participants trained in explicit (classroom-like) conditions (Morgan-Short et al., 2012). On the other hand, the same level of performance was achieved by learners in both groups at the behavioral level, irrespective of the type of training. That is, despite the promise of this approach, a clear association could not be established between brain activity and performance. Similar discrepancies between ERPs and behavior have been found in studies that employed finer-grained behavioral measures, e.g., real-time reading data from eye-tracking. In a study with L2 learners of French, Foucart and Frenck-Mestre (2012) found that L2 speakers showed non-native-like ERPs in response to gender agreement violations between nouns and predicative adjectives (Experiment 3), but the eye-tracking results were indistinguishable from reading patterns in native speakers (Experiment 4).

A later study on statistical learning by Batterink, Reber and Paller (2015) was able to establish a more direct association between the type of training and the mechanisms engaged during behavioral performance. In that study, all participants were exposed to an implicit statistical learning paradigm, in which a continuous string of syllables was auditorily presented. However, an explicit learning group was given an advantage, by being explicitly taught the words presented before exposure to the continuous stream of syllables for implicit learning. All participants then completed a recognition task and a target detection task while their EEG was recorded. The results showed that having explicit knowledge of the words affected processing in both tasks. In the recognition task, only learners in the explicit training group presented a late positive component (LPC), which was interpreted as an index of explicit recollection; in addition, the explicit group vastly outperformed the implicit learners (91.5% accuracy relative to 59.3%). The results established a more direct association between the neurophysiological activity and behavior following training.

The data available to date sheds light onto the cognitive mechanisms underlying second language processing as a result of different training conditions. However, previous studies have only assessed the effect of training on later processing and behavior. Therefore, an important question to be addressed is how the cognitive demands posed by different training conditions affect the cognitive mechanisms engaged, influencing learning outcomes. A potential approach, which is investigated in the present study, lies in using EEG to examine which cognitive pathways are activated by adults during training. One specific hypothesis tested here is that the level of difficulty experienced during training will impact the cognitive mechanisms engaged, with consequences on second language attainment. To investigate this idea, the current study compared retrieval practice conditions that differed solely in their difficulty level based on the distractors included, and examined their effect on neurophysiological activity and learning outcomes.

1.1. Native language regulation and second language attainment

Neuroimaging evidence has indicated that accessing information in a second language requires suppression of competing information in the native language (Abutalebi & Green, 2007; Calabria, Costa, Green, & Abutalebi, 2018; Mendez, 2019; Rodriguez-Fornells, Rotte, Helnze, Nösseit, & Münte, 2002). The languages of a speaker are now known to be highly interconnected, and even when only one language is being used, both languages have been found to become activated (Thierry & Wu, 2007). Given constant co-activation, speaking a second language is believed to hinge on the ability to inhibit the L1 (Abutalebi et al., 2008; Calabria et al., 2018; Rodriguez-Fornells et al., 2002). The role of inhibition during second language usage is well documented in the literature on bilingualism. For example, costs in switching back to the native language after using the second language are believed to reflect an inhibited L1. When interference from the other language must be avoided, bilinguals show increased activation of brain areas responsible for executive control and inhibition (e.g., the left-prefrontal cortex and the SMA) (Green & Abutalebi, 2013; Rodriguez-Fornells et al., 2005). On the other hand, bilingual speakers with impaired inhibitory skills due to Alzheimer's disease tend to have difficulty using or maintaining their second language, as they lose the ability to inhibit the dominant language (Mendez, 2019; Mendez, 1999). Additional evidence of long-term experience in regulating language activation comes from the literature on the consequences of bilingualism on cognitive control (Anderson, Mak, Keyvani Chahi, & Bialystok, 2018; Bialystok, Craik, Green, & Gollan, 2009; Grady, Luk, Craik, & Bialystok, 2014; Luo, Luk, & Bialystok, 2010).

Developing the ability to regulate the native language may be a key component of successfully learning a second language (Bogulski, Bice, & Kroll, 2019; Levy, McVeigh, Marful, & Anderson, 2007). Recently, Bogulski and colleagues proposed the L1 Regulation hypothesis, which suggests that experience in inhibiting interference from the L1 confers an advantage in language learning (Bogulski et al., 2019). Support for their proposal comes from data showing higher recall of foreign language vocabulary in bilinguals, relative to monolinguals; an advantage that they attribute to bilingual's constant need to suppress cross-language competition. Given this, a promising though counterintuitive approach lies in creating training conditions of "desirable difficulty", in which the two languages compete for selection (Anderson, Bjork, & Bjork, 1994; Bjork & Kroll, 2015). Conditions that induce interference from the native language may aid learning, because they afford practice in inhibiting the more dominant native language (Bogulski et al., 2019; Levy et al., 2007; Pulido & Dussias, 2020). In fact, a number of studies on L2 vocabulary learning have produced results consistent with the desirable difficulties approach (Bogulski et al., 2019; Potts & Shanks, 2014; Schneider, Healy, & Bourne, 2002). Specifically, these studies have underscored the importance of learners' ability to recall formed L2 representations.

Recently, a study by Pulido and Dussias (2020) tested the idea that second language practice conditions that induced L1 interference would lead to improve learning outcomes, relative to a non-interference practice condition. Spanish learners of English were tasked with learning conventional verb-noun phrases (termed "collocations") that were selected based on their degree of *congruency* with English (i.e., availability of a literal word-by-word equivalent; for example *rodar una película = 'roll a movie',* is equivalent to "shoot a movie" in English, and is not congruent across languages). As discussed, success in learning phrases that are incongruent across languages is lower because they require selecting word combinations that conflict with the native language (Nesselhauf, 2003; Peters, 2016; Pulido & Dussias, 2020; Yamashita & Jiang, 2010). The training level of difficulty was based on the need to suppress information from L1-like choices during practice. The results showed that higher performance in immediate and delayed recall tests was associated with inhibition of the native-language equivalents. However, not all participants were equally successful, and individual variability in the extent of L1 inhibition measures suggested that different participants might have engaged alternative mechanisms.

In summary, the evidence from various strands of the literature indicates that, in learning a second language, (i) adults may rely upon different cognitive mechanisms; (ii) the mechanisms engaged (i.e., neurological pathways) may depend upon training conditions; and (iii) training conditions impact behavioral learning outcomes, though with variable success. What is still lacking is a clear connection between these aspects. The present study aims to fill this gap by examining the association between types of training, the neurophysiological responses elicited under each condition, and individual behavioral outcomes.

2. The present study

The present study employed EEG to examine the effect of practice difficulty on brain activity in real time and to identify what brain responses predict successful second language learning. To do so, it takes direction from research suggesting that L1 regulation improves L2 learning outcomes. That is, language regulation is treated as a desirable difficulty upon which the manipulation for Practice conditions is based. ERPs will provide insight into previously proposed learning mechanisms under the L1 Regulation hypothesis (Bogulski et al., 2019), according to which inhibition of interference from L1-congruent information should be tied to improvements in L2 performance.

The design of the current study is based on the practice procedure developed by Pulido and Dussias (2020). In this paradigm, beginner learners of a second language learn new conventional verb-noun phrases ("collocations"). While the words that make up the phrases are known by participants, it is their combination with a novel meaning in the L2 that is unfamiliar (e.g., participants know *gastar* 'spend' and *broma* 'joke' but not the phrase *gastar una broma*, equivalent to "play a joke"). Learners practiced in one of two conditions in which the difficulty level is manipulated. The level of difficulty is based on the need to suppress information from the native language from the choices presented. In the easier 'Unrelated' condition, some learners practiced selecting among two verbs which consisted of the target verb and an unrelated distractor, e.g., *gastar* ('spend') – *ordenar* ('order') – *broma* ('joke) (target: *gastar* 'spend'). In the difficult 'L1-Interference' condition, the distractors of critical items are acceptable according to English (*jugar* 'play'), but are inadequate in Spanish, e.g., *gastar* ('spend') – *jugar* ('play') – *broma* ('joke) (target: *gastar* 'spend'). Because the 'L1-Interference' distractors are acceptable for native English speakers, they must be more strongly inhibited in the more difficult learning condition (i.e., the verb 'play' will become inhibited). If inhibiting the native language is an important aspect of second language learning, and this is achieved in the difficult L1-Interference condition, ERPs should show inhibition of native-like verbs during the practice. The critical prediction is that ERP modulations induced by differences in practice will be associated with performance in learning tests, which are completed immediately after the practice, and again after a one-week delay. Further, it is expected that the effects of inhibiting the native language will be modulated by the type of test. In particular, L1 inhibition should be more apparent in a Translation test, in which L1-to-L2 translation begins with the direct presentation of an L1 equivalent, causing greater L1 interference (and thus requiring greater regulation); while L1 inhibition may be less powerful in a task completely entirely in the L2 (a gap-fill test).

2.1. Research questions and predictions

The study addresses four main questions. As discussed above, conditions of desirable difficulty have been found to produce higher rates of learning. First, a main question concerns what ERP components are elicited by the easy and difficult Practice conditions. The ERPs induced by congruent and incongruent items are expected to be significantly modulated under each of the two Practice conditions (i.e., Unrelated or L1-Interference); specifically, the difficult L1-Interference condition should lead to greater inhibition of L1-congruent items in ERPs. Guided by previous ERP studies, two components are anticipated. In psycholinguistic research, the N400 component is the best-studied neural correlate of language processing, and it has been shown to index the cost of lexical processing, due to lexical access (Kutas & Federmeier, 2010; Kutas & Hillyard, 1980, 1984) or integration (van Berkum, Hagoort & Brown, 1999; Hagoort & van Berkum, 2007) (for evidence of the independent contribution of these two factors, see Nieuwland et al., 2020). Based on processing studies on verb-noun collocations, where the first element (e.g., the verb) primes a frequently co-occurring collocate (e.g., the noun) (Pulido & Dussias, 2019; Wolter & Gyllstad, 2011, 2013), it is expected that phrases similar to the native language (i.e., congruent) should elicit an attenuated N400 in processing; on the other hand, L1-L2 incongruent collocations should elicit a greater N400 effect, indexing a processing cost.[1] Crucially, however, the described conditions of desirable difficulty should reverse this pattern, due to induced L1 inhibition. Recall that the difficult L1-Interference condition requires rejecting non-target verbs that are L1-like (e.g., *jugar una broma* 'play a joke'), to select the L2 targets (*gastar una broma* 'spend a joke'). In line with the L1 Regulation hypothesis, it is predicted that an inhibited L1 will be manifested in greater N400 amplitudes for the *congruent* (rather than incongruent) items during the Practice. A second prediction concerns ERP modulations due to degree of uncertainty in a response. In tasks that involve monitoring during response selection, a right frontal effect (RFE) of negative polarity is associated with increased uncertainty and need to monitor the selection of a response, overlapping with behavioral RTs (Finnigan, Humphreys, Dennis, & Geffen, 2002; Hayama, Johnson, & Rugg, 2008; Leynes & Kakadia, 2013; Pulido & Dussias, 2019; Rugg & Curran, 2007; Rugg, Herron, & Morcom, 2002). Previous evidence from lexical selection in collocations has reported a RFE in both native and non-native speakers; in a paradigm analogous to the one employed here, the RFE emerged between 600 and 800 ms (Pulido & Dussias, 2019). In the present paradigm, learners in the Unrelated group are expected to have more uncertainty during response selection. Therefore, collocations that are incongruent with the native language should generate greater uncertainty during response selection in the Unrelated group relative to the L1-Interference group, eliciting a RFE.

The second question is whether the neurophysiological activity elicited during Practice will be predictive of recall rates in post-tests. It is expected that ERPs indexing inhibition (an N400 for L1-congruent items) should be correlated with performance in post-tests. This association is expected to be particularly strong in a Translation test, which requires more direct suppression of the L1 equivalents presented. By relying on known electrophysiological potentials, ERPs will provide an account of brain responses as they unfold under different experimental conditions. Importantly, ERPs may also reveal individual differences in brain responses during exposure to a non-native language (Tanner, Inoue, & Osterhout, 2014; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013). The possibility of individual variability makes particularly relevant the question of how practice difficulty affects individual neurophysiological responses, and their connection with L2 attainment.

A third question is whether ERPs that may index regulation of the L1, based on the predictions above, will be significantly correlated with different behavioral measures of assessment, which require regulating the L1 to a greater or lesser extent. In particular, it is expected that some types of language tests (such as translation) rely more directly on explicit L1-to-L2 mappings as well as more direct regulation of the L1 information to avoid interference during translation, while other tests do not rely on L1 regulation as much (e.g., context-based gap-fill). To address this question, the two types of test (translation and gap-fill) are considered. Specifically, the Translation test is expected to rely on language regulation to a greater extent than a gap-fill test, and to potentially elicit a stronger effect of Practice group.

Finally, a basic difference between the practice paradigm in the present study and in previous literature lies in the modality of response selection. A fourth question is therefore whether the magnitude of the effect may be affected by how responses are selected. While in the study by Pulido and Dussias (2020), participants provided vocal responses (i.e., by pronouncing the target), the present study relied on manual response selection through button presses. This was done for two reasons. First, pressing buttons rather than providing spoken language responses helps avoid motor artifacts derived from speech. While it has been shown that ERP data can be analyzed even in the presence of such artifacts (Ouyang, Sommer, Zhou, Aristei, Pinkpank & Abdel Rahman, 2016), relying on manual responses circumvents the potential methodological challenges in testing this experimental manipulation using EEG for the first time. Secondly, and perhaps more importantly, conducting the experiment with manual rather than vocal responses allows to investigate the role of mode of practice in replicating the results of Pulido and Dussias (2020). Research has suggested that engaging the production system may play an important role in learning, through what is known as the "production effect" (Forrin & MacLeod, 2017; Ozubko & MacLeod, 2010; Zormpa, Brehm, Hoedemaker & Meyer, 2018). Thus, the results will provide additional insight into the effect of response modality.

3. Method

3.1. Participants

A group of 61 undergraduate learners of Spanish was recruited at a large US university. Participants were native speakers of English completing a third semester university-level elementary Spanish course (roughly equivalent to the levels A2-B1 of the Common European Framework of Reference for Languages, Council of Europe, 2011). All participants gave informed consent and were paid 10 US dollars per hour of participation. To confirm their eligibility, participants completed baseline tests of vocabulary (Spanish-English translation) and collocations (Spanish multiple choice, described in Section 3.3.4) designed to ensure that they were knowledgeable of the individual words that composed the collocations, but were not familiar with the collocations per se. Eleven participants were excluded after the first session due to prior knowledge of the collocations. Data from four participants were excluded due to experimental error, and data from one additional participant due to excessive artifacts in the EEG recording. The remaining participants (N = 45) were randomly assigned to one of two learning conditions: a group in which unrelated distractors were presented during practice (henceforth, the 'Unrelated' group, N = 22; 64% female), and a group that saw L1-related distractors during recall of a list of incongruent collocations ('L1-Interference' group, N = 23; 52% female).[2]

3.2. Individual differences measures

Participants completed a number of measures of language proficiency in the L1 and L2, as well as additional tasks to ascertain individual differences in cognitive skills. This section briefly enumerates the measures and report the results below (see Table 1); additional details about the tasks in this section are available in the Supporting Information materials.

To assess linguistic profile and background in the L1 and L2, participants completed the LEAP-Q (Language Experience and Proficiency Questionnaire; Marian, Blumenfeld & Kaushanskaya, 2007), which includes questions on aspects such as onset of acquisition and frequency of exposure to each language. The selection of other cognitive and proficiency tasks was informed by measures found to be significant predictors in previous psycholinguistic research on learning of collocations employing a similar paradigm (Pulido & Dussias, 2020). The Flanker task (Eriksen & Eriksen, 1974) was used as a measure of cognitive control. Because interference inhibition is an important aspect of the experimental manipulation, the Flanker task allowed to check whether groups of participants were matched on inhibitory skill; no significant differences were found between the two groups ($t(36.82) = -0.94$, $p = .35$). The Nonword Repetition task (Baddeley, Papagno & Vallar, 1988) served as an index of phonological short-term memory (PSTM), which has been shown to predict learning of words and multiword units (Martin & Ellis, 2012; Pulido & Dussias, 2020).

Lexical knowledge served as a proxy for language proficiency, as it has been shown that vocabulary knowledge correlates with broader linguistic performance (Meara, 1996; Staehr, 2008). Lexical knowledge and fluency were measured through the Category Verbal Fluency. Two versions of the task were administered, one in English first and then in Spanish, each containing two categories (half animate, half inanimate). Participants also completed a multiple-choice test to assess knowledge of the individual words employed in the experiment, which they were expected to know. In addition, they completed a baseline collocations pre-test (reported in Section "Test administration and scoring") to ensure that participants were not familiar with the collocations included in the experiment. All participants were right-handed, and groups were matched across all measures (*p*-values shown in Table 1).

Table 1

Summary of Cognitive and Proficiency Measures

Measure	Unrelated group (N=22)			L1-Interference group (N=23)			*p*
	Valid *N*	*M*	*SD*	Valid *N*	*M*	*SD*	
Age (in years)	22	19	1.18	23	18.6	0.72	.19
L2 AoA (in years)	22	12.55	3.1	23	11.74	3.35	.34
Weekly exposure to L2 (%)	22	5.21	4.62	23	6.74	5.69	.42
English (L1) Verbal Fluency	22	30.08	6.46	23	31.96	5.42	.34
Spanish (L2) Verbal Fluency	22	9.5	4.06	22	11.4	4.6	.15
Baseline Vocabulary Test (/100)	22	82.3	5.33	23	84.7	6.47	.21
PSTM: Nonword repetition (/100)	21	60.63	9.34	23	60.43	9.31	.80
Flanker effect (ms)	22	49.44	27.27	20	41.19	18.68	.35

3.3. Materials and Procedure

Three types of materials were created: materials for (1) the Familiarization phase, for the (2) Practice phase and for the (3) Testing phase (the timeline of the experimental procedure is summarized in Table 2). The materials for Familiarization phase and the Testing phase were identical for all participants, while the materials used for Practice differed across the two groups. First the list of conventional phrases is described, followed by the materials and procedure for the Familiarization, Practice and Testing phases.

Table 2

Testing protocol

	Week 1		Week 2
	Session 1	Session 2	Session 3
Tasks	Baseline cognitive and proficiency measures	(1) Familiarization phase (2) Practice phase	

	(3.1) Immediate Tests	(3.2) Delayed Tests
Tests		

3.3.1. List of conventional phrases

Forty-two conventional phrases (i.e., "collocations") were extracted using the web-based version of Corpus del Español (Davies, 2016), to create three lists, described below. *(i)* 1/3 L1-L2 Incongruent collocations: In the critical experimental list of L1-L2 incongruent collocations only the noun, but not the verb, was equivalent across Spanish and English. For example, the Spanish collocation "gastar bromas" is conventionally expressed in English as "play jokes", although it would literally translate as 'spend jokes". *(ii)* 1/3 Congruent collocations: The related list of congruent collocations was derived from the first list and matched item-by-item. First, the idiomatic English equivalents were identified for the first list ("play jokes" in the example above) and the literal translation of the verb was used to select a matched L1-L2 congruent collocation (e.g., for "play" – "jugar", "jugar partidos" was selected, which means 'play matches'). Thus, the collocations in the first two lists were associated by virtue of this manipulation. *(iii)* 1/3 fillers: Finally, a third list consisting of filler collocations was created (e.g., "marcar un número" 'dial a number'). Fillers were also incongruent items but, importantly, they differed from the experimental condition in that no L2 verb that served as a literal translation of the L1 was provided, or could be easily identified[3]. Including more incongruent items than just the experimental condition served the purpose of distracting attention from the critical incongruent items, as well as from the fact that the critical items could be mapped to L1 counterparts. Words were not repeated across any of the items in each list, which were matched on log frequency, collocational strength, noun syllable length and orthographic length. Half of the nouns in each list were cognates.
To determine the collocational status of verb-noun phrases, t-scores were used as the statistical association measure (all t-scores > 4.0) (Evert, 2005; Gries, 2010). Table 3 presents two sample sets of matched items (the complete list is available in Appendix A).

Table 3
Sample Stimuli Matched Across Lists

	Language	Incongruent	Congruent	Filler
1.	Spanish (L2)	*gastar bromas*	*jugar partidos*	*marcar [un] número*
	English (L1) equivalent	*play jokes*	*play matches*	*dial [a] number*
	Literal L1 translation	'spend jokes'	'play matches'	'mark [a] number'
2.	Spanish (L2)	*pedir pizza*	*ordenar [este] caos*	*cambiar [un] cheque*
	English (L1) equivalent	*order pizza*	*order chaos*	*cash [a] check*
	Literal L1 translation	'ask for pizza'	'order chaos'	'change [a] check'

3.3.2. Materials and Procedure for the Familiarization phase

Based on the lists of collocations, stimuli were created for visual and auditory presentation of the target Spanish collocations and their idiomatic English equivalents. The Spanish stimuli were recorded by the author, who is a male native speaker of Spanish, and their English counterparts were recorded by a female native English speaker.

All learners were familiarized with the same list of materials. First, a screen showing the Spanish collocation was presented, followed by the English collocation (each lasting 2000 ms); the corresponding audio recording was played simultaneously for each screen. Then, participants were instructed to repeat twice the target Spanish collocation: they first pronounced it out loud into a microphone; next, they typed it and received immediate feedback. A second round was administered. In order to ensure that both the Spanish phrases and their meaning were learned, during this second round, participants were also prompted to remember and type the English meaning of the Spanish collocations after every seven trials. All participants reached a minimum threshold of 80% accuracy during familiarization; groups did not differ in their ability to remember the Spanish collocations (Unrelated group: 98.4%, SD: 12.4; L1-Interference group: 96%, SD: 19.5; $t(42.77) = -0.66$, $p = .51$) or their meaning (Unrelated: 83%, SD: 37.4; L1-Interference: 81%, SD: 39.5; $t(42.99) = 0.71$, $p = .48$).

3.3.3. Materials and Procedure for the Practice Phase

Practice materials. Based on the conventional phrases, two sets of materials were created, one for each of the two experimental conditions. These sets critically differed with respect to the distractor verbs presented for 'L1-L2 Congruent' and 'L1-L2 Incongruent' experimental trials, but were otherwise identical. For one group, the distractors of 'Incongruent' items were unrelated verbs; I will henceforth refer to this as the 'Unrelated' group. The other group of learners saw distractors that would be congruent with the native language ('L1-Interference' group). For the above example, "gastar bromas" ('play jokes'), the 'Unrelated' distractor was "ordenar bromas" 'order', while the 'L1-Interference' distractor was "jugar bromas" 'play' (to reiterate, "jugar" was the literal translation of the verb in the English equivalent "play jokes"). Finally, each set contained all the items in the filler items list. As indicated above, this list was practiced in identical conditions in both groups of participants. In this set, verbs from the same list were used as distractors by re-pairing them. To illustrate, for the above example "marcar un número" 'dial a number', the verb "entregar" (from "entregar una propuesta" 'submit a proposal') was used as a distractor.

Examples are shown in Table 4. Each of the forty-two collocations was repeated eight times, yielding a total of 112 trials per list, and 336 trials per participant. This is similar to previous studies that presented each collocation nine times (Pulido & Dussias, 2020; Toomer & Elgort, 2019), but allowed for items to be evenly divided into target-first and distractor-first trials.

Table 4

Experimental Design of Distractor Matching and Sample Practice Trials

1. Distractor verb matching				
Language	Incongruent item		Matched congruent item ('L1-Interference')	Matched congruent item ('Unrelated')
Spanish (L2)	***gastar*** *bromas*		***jugar*** *partidos*	***ordenar*** *[este] caos*
L1-Interf. distractor	***gastar*** ('**play** jokes')	↔ =	***jugar*** ('**play** matches')	

Unrelated distractor	***gastar*** ←→ (‘**play** jokes’) ≠	***ordenar*** (‘**order** [this] chaos’)
	2. Practice trials	
	Difficult ‘L1-Interference’ condition	*Easy ‘Unrelated’ condition*
gastar bromas Incongruent	***gastar*– *jugar*** – *bromas* ‘spend’ – ‘play’ – ‘jokes’	***gastar*– *ordenar*–** *bromas* ‘spend’ – ‘order’ – ‘jokes’
jugar partidos Congruent	***jugar* – *gastar*** – *partidos* ‘play’ – ‘spend’ – ‘matches’	*jugar* – *poner* – *partidos* ‘play’ – ‘put’ – ‘matches’
ordenar caos Congruent	*ordenar* – *pedir* – *caos* ‘order’ – ‘request’ – ‘chaos’	***ordenar* – *gastar*** – *caos* ‘order’ – ‘spend’ – ‘chaos’

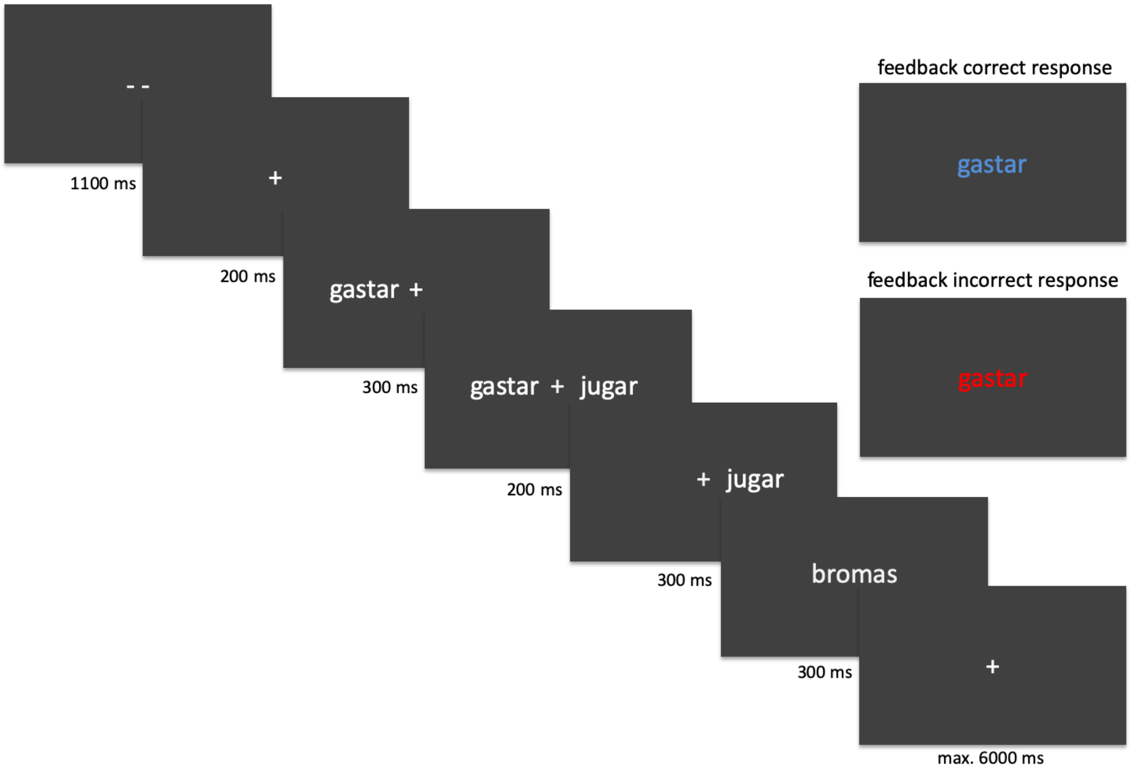
Note. Verbs that were targets for selection in each Practice trial appear underlined. Order of presentation of the target and distractor verbs for each phrase was counterbalanced. Bold-face verb pairs illustrate the same target-distractor pair used across trials within a condition.

Practice procedure. In each Practice trial, participants were presented with two verbs (target and distractor) followed by a noun. To control for the order of processing and activation of target and distractor verbs, the two verbs were displayed in sequence. First, a fixation cross was presented on the center of the screen. After 200 ms, one verb appeared to the left and one to the right of the fixation cross; each was shown for 500 ms, with a 300 ms presentation delay for the second verb. They were followed by the noun, which was displayed for 300 ms. Responses were made manually after seeing the noun, by pressing a button corresponding to location where the verb had been previously shown on the screen (left or right). Immediate feedback was provided by showing the target (in blue for correct responses, red for incorrect). Prior to beginning the experimental trials, participants completed ten practice trials so that they could get used to the task. In preparation for each trial, a “blinking” screen showing two hyphens “- -” was presented for 1100 ms. Participants were instructed to blink during this 1100 ms screen, and to avoid blinking in the 1300 ms second interval during which the words were presented. Words were displayed in font Arial 30 (trial sequencing is illustrated in Figure 1).

Fillers were practiced in a separate block that preceded the experimental block. In critical trials, learners in the “Unrelated” group saw unrelated distractors, e.g. “ordenar – gastar – broma” (correct response “gastar” ‘spend’), while learners in the “L1-Interference” group had to correctly discard the L1-equivalent verb, e.g. “jugar – gastar – broma” (i.e., discard “jugar” ‘play’ to choose “gastar” ‘spend’). The order of presentation of target and distractor was counterbalanced, such that “gastar – jugar – bromas” appeared for half of the trials, and “jugar – gastar – bromas” in the other half. Importantly, all verbs were potential candidates, and it was not until the noun was displayed that participants were able to select the appropriate verb (“gastar” in this example, based on the familiarized “gastar bromas”).

Figure 1

Sample Sequence of the Practice Procedure



Note. The figure illustrates the sequencing in the Practice procedure for a sample trial. Following a response (or the absence thereof after 6000 ms), participants were presented with one of the two feedback screens (blue for correct, red for incorrect/no response).

3.3.4. Materials and Procedure for the Testing phase

Testing materials. Three types of tests were created for the study, serving three different purposes: (1) a Baseline multiple-choice pre-test, which was administered to confirm that participants were not able to recognize the target collocations even when the correct response was presented to them; (2) a contextualized Gap-fill post-test, which gauged the ability to use the target verbs in context, but in the absence of the L1 equivalents; (3) a Translation post-test, which assessed the ability to recall the collocation in connection with its meaning in the native language.

Test administration and scoring. The multiple-choice Baseline test was administered in the first session using Qualtrics (Qualtrics, Provo, UT). In addition to selecting responses, participants provided confidence ratings for their responses, using a scale from 1 to 5 (1 = no knowledge; 5 = certainty in the response). No feedback was provided.
In the second session, following the Practice, first the Gap-fill test was administered and then the Translation test. This order was used so that the L1-equivalents would not be shown again until the end of the session. Both tests were completed on a computer by typing the responses.
Feedback was provided only after responses in the Translation test, so that erroneous attributions of meanings could be corrected at the end of the session. The Gap-fill and Translation tests were

completed again in the same order when participants returned for the next session. The delay between the immediate and the delayed tests averaged 5.61 days (SD: 0.21). This time lapse did not significantly differ across the Unrelated (mean: 5.75, SD: 1.52) and L1-Interference (mean: 5.48, SD: 0.95) groups ($t(30.99) = -0.69$, $p = .49$).

The gap-fill test was considered a form recall test, in which the nouns in the collocations learned were embedded in a sentence. Participants were asked to use the corresponding verbs that were learned and practiced to complete the gap in each sentence. This test allowed to assess form retrieval in an L2-only context in which recollection of the meaning of the collocation was not strictly necessary. For Translation, only the L1 English collocation was presented; this was considered the more stringent of the two post-tests, as it required recalling both meaning and form.

Prior familiarity was assessed based on the accuracy scores and confidence ratings of the Baseline test were used. As described above, participants provided ratings for any correct response in an incongruent trial, using a scale from 1 to 5 (1 = no knowledge; 5 = certainty in the response). Ratings of 1 indicated that the correct option had been selected by chance and were recoded as inaccurate. Participants who indicated previous knowledge of more than one collocation (rating = 5) or substantial familiarity with more than two items (rating ≥ 4) for any of the incongruent lists were excluded from the study. For the remaining subjects, weighted pre-test familiarity scores for correct responses were calculated based on confidence ratings (ratings of 5, 4, 3, and 2 received weights of 0.5, 0.4, 0.3 and 0.2, respectively). Baseline familiarity scores (Unrelated mean: 0.04, SD: 0.1; L1-Interference mean: 0.04, SD: 0.11) revealed no significant differences across groups ($t(1296.1) = 0.88$, $p = .38$). Typed responses in the Translation and Gap-fill post-tests were coded for accuracy. Misspellings were not penalized as long as no more than two phonemes were incorrect, and the response could be interpreted unambiguously.

3.4. EEG recording and analysis

The continuous electroencephalogram (EEG) was recorded from 32 electrodes mounted in an elastic cap (EasyCap; Brain Products, GmbH) and an ActiChamp amplifier (Brain Products, GmbH) with a 24-bit analog to digital conversion (online sampling rate: 500 Hz, 0.1–100 Hz band-pass filter). Electrode impedances were kept below 5 KΩ. During recording, electrodes were referenced to the right mastoid. Grounding electrodes were mounted on the forehead and beneath the right eye. Datasets were filtered online with a 25 Hz low pass and 0.1 Hz high pass noncausal IIR Butterworth digital filter. Segments with excessive muscular artifacts on the continuous data were manually rejected. Independent component analysis (ICA) was performed to correct remaining ocular artifacts (Jung et al., 2000); the number of components rejected averaged 2.23 for the Unrelated group and 2.09 for the L1-Interference group (range: 1-4). Then all epochs with activity exceeding ±100 µV at any electrode site were automatically removed using a peak-to-peak moving window (2.29% of data in Unrelated group; 1.77% in L1-Interference group). Epochs ranging from -200 to 1000 ms after onset of the noun were extracted. Baseline correction was done relative to pre-stimulus activity. Inaccurate trials were excluded from the analysis (Unrelated: 2.31%; L1-Interference: 4.74%). Data from all remaining participants (Unrelated N = 22; L1-Interference N = 23) contained at least 30 valid trials per condition.

Based on the predictions, the analysis of the EEG data focused on the canonical N400 window (300–500 ms post-stimulus presentation), and the RFE window (600 and 800 ms) identified in research using a similar paradigm (Pulido & Dussias, 2019). The behavioral data revealed that average RTs for button presses across conditions ranged from 687 to 759 ms, coinciding with the expected 600-800 time-window.

Repeated measures ANOVAs were conducted, with mean amplitudes as dependent variables, and Group, Congruency and Order as independent variables. Midline ANOVAs included Frontality as a predictor (frontal, central, parietal) and were performed on Fz, Cz and Pz. Lateralized ANOVAs with Frontality and Hemisphere (left, right) predictors were conducted on F3, F7, F8, FC1, FC2, FC5, FC6, C3, C4, CP1, CP2, CP5, CP6, P3, P4, P7, P8. Greenhouse–Geisser-corrected values are reported where appropriate. To better characterize the topography of significant effects in each group, pairwise t-tests were performed on each scalp region (right frontal: F4, F8, FC2, FC6; left frontal: F3, F7, FC1, FC5; right posterior: CP2, CP6, P4, P8; left posterior: CP1, CP5, P3, P7). Results of post-hoc paired pairwise t-tests are reported with FDR-corrected *p* values.

Individual ERP measures for the N400 Congruency effect were calculated as the difference waves of average amplitudes (Incongruent - Congruent) at Cz, which showed significant effects for both groups. Behavioral scores for the congruency-based cost were calculated as the difference between individual averages, taking the accuracy in congruent trials as the individual baseline (proportion Congruent – proportion Incongruent). For ease of interpretability, the resulting score was subtracted from 1, with a value of 1 indicating no difference in accuracy (1 – 0 = 1).

3.5. Behavioral analysis

The accuracy results from tests were analyzed using mixed-effects logistic regression with the lme4 package (Bates, Maechler, Bolker & Walker, 2015) in R version 3.5.2 (R Core Team, 2018). Given that each participant produced fourteen data points per type of collocation in each test, and because the same participants responded to identical questions, a combined analysis was conducted for both tests, while including Test type as a fixed effect. The fixed effects considered in the analysis included Type of collocation (Congruent, Incongruent), Group (Unrelated or L1-Interference distractors), Test (Translation, Gap-fill) and Session (2: immediate tests, 3: delayed tests). Additionally, the contribution of the individual N400 difference means and its interaction with the factors above were also considered. No substantial collinearity was found among the variables considered.

Initial maximal random effects structure included all the factors above (Barr, Levy, Scheepers & Tily, 2013). Following initial attempts which led to convergence issues, the random terms were simplified; the reduced structure included by-subject and by-item random intercepts, as well as random by-subject slopes for Type of Collocation and random by-item slopes for Group and N400 means. Model selection was conducted in a forward step procedure, by adding each predictor and their interactions one by one for model comparison; predictors were kept if the model fit was significantly improved (likelihood ratio test, $p < .05$). The reference levels were set to L1-Interference for Group, Congruent for Type of collocations, Gap-fill for Test type and Session 2. All continuous variables were centered (Baayen et al., 2008). Parameter-specific *p*–values were estimated using the lmerTest package in R (Kuznetsova, Brockhoff & Christensen, 2017). The results and the model output are reported below. Section 4 reports on the results of the ERP analysis from the Practice procedure; and on exploratory correlations between ERPs recorded during Practice and subsequent recall during testing. Section 5 reports on the results of linear mixed-effects models that investigated learning outcomes in recall post-tests.

4. Results

4.1. ERP Results from Practice Phase

Learners completed the Practice phase with high accuracy in selection of the target verbs in both groups (‘Unrelated’ mean: 93.32%, SD: 24.87; ‘L1-Interference’ mean: 91.78%, SD: 27.48; $t(83.85) = -1.35$, $p = .18$). Response times averaged 729 ms (SD: 583) in the Unrelated group and 732 ms (SD: 597) in the L1-Interference group ($t(10166) = 0.32$, $p = .75$). This section presents the analysis of the EEG data, acquired during the experimental block of the Practice procedure, which contained the trials of congruent and incongruent collocation lists that were subject to the experimental manipulation. The analysis of the 300-500 ms and 600-800 ms time windows allows to address the first research question, which asked what ERP components are elicited by the easy Unrelated condition and the difficult L1-Interference condition.

4.1.1. 300-500 ms time window (N400)

Recall that for the 300-500 ms time-window, it was hypothesized that conditions of desirable difficulty would result in significant modulations of the N400 component: a smaller N400 was predicted for L1-congruent collocations at baseline (Unrelated condition), but not if the native language was inhibited (L1-Interference condition); in the latter case, a larger N400 was predicted for L1-congruent items. The midline ANOVA revealed a main effect of Frontality ($F(1.33, 57.04) = 15.71$, $p < .0001$, $\eta^2 = .05$) and a critical Group x Congruency interaction ($F(1, 43) = 15.62$, $p < .001$, $\eta^2 =.01$), as well as a Congruency x Frontality interaction ($F(1.44, 61.92) = 4.82$, $p < .05$, $\eta^2 = .0007$). Follow-up comparisons confirmed this interaction was due to divergent effects of Congruency in each group. In the Unrelated group, incongruent trials elicited a significantly greater negativity at Cz, maximal at Pz. In contrast, in the L1-Interference group *congruent* trials elicited a significantly more negative peak, which was maximal at Fz and also significant at Cz and Pz. This is in line with the hypothesis that conditions of desirable difficulty would result in inhibition of the L1 information, reversing the relative cost of processing incongruent collocations.

The omnibus lateralized ANOVA also revealed a main effect of Frontality ($F(1.30, 55.88) = 16.0$, $p < .0001$, $\eta^2 = .05$) and the same crucial two-way interaction for Group x Congruency ($F(1, 43) = 16.11$, $p < .001$, $\eta^2 = .001$), as well as additional interactions for Order x Hemisphere ($F(1, 43) = 4.86$, $p < .03$, $\eta^2 = .0007$), Congruency x Frontality ($F(1.34, 57.52) = 5.73$, $p < .01$, $\eta^2 = .0008$), Hemisphere x Frontality ($F(1.70, 73.21) = 6.59$, $p < .01$, $\eta^2 = .002$), and a three-way Order x Hemisphere x Frontality interaction ($F(1.99, 85.91) = 2.06$, $p < .0007$, $\eta^2 = .001$). There were marginally significant interactions of Group x Order x Frontality ($F(2, 85) = 2.63$, $p = .08$, $\eta^2 = .0001$) and Congruency x Hemisphere x Frontality ($F(2, 85) = 2.85$, $p = .06$, $\eta^2 = .0001$). Given these two- and three-way interactions, the effects of Order and Congruency in each group were further investigated.

Follow-up comparisons on these effects again confirmed that the group-by-congruency interaction was due to significantly more negative peaks for incongruent trials in the Unrelated group, and a reversal of this pattern in the L1-Interference group, with significantly more negative waves for incongruent trials. Further, the effects of Congruency and Order were greater over left posterior electrodes in the Unrelated group, where all comparisons were significant with the exception of those between incongruent distractor-last and congruent target-last trials ($p =$

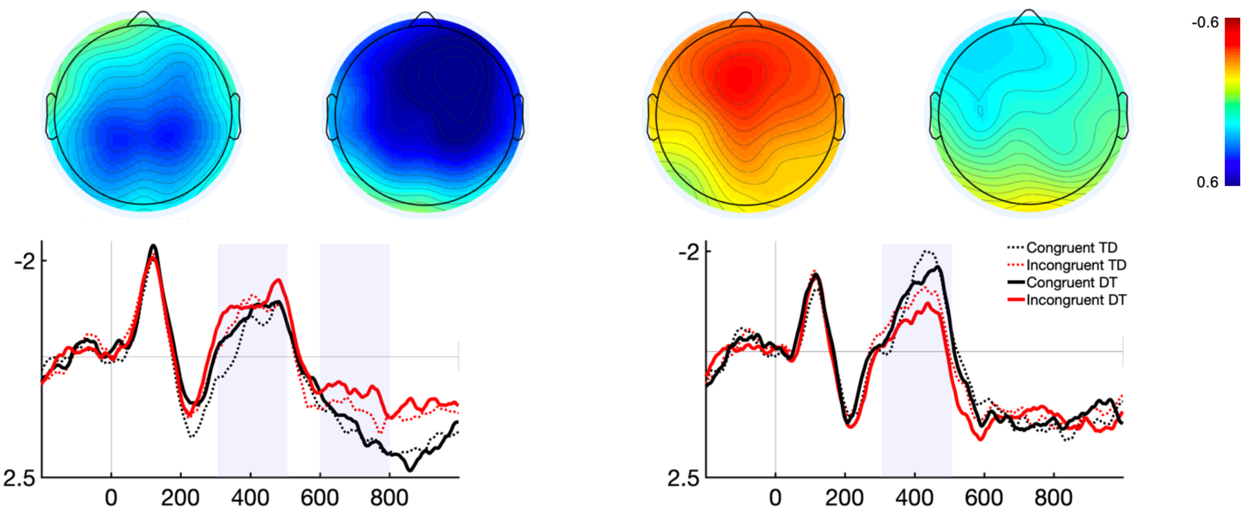
.07). In contrast, in the L1-Interference group Congruency produced significant differences in all electrode regions, with the exception of right posterior electrodes, where congruent distractor-last and incongruent distractor-last trials were not significantly different (p = 0.38). Further comparisons revealed that significant differences emerged between congruent distractor-last trials and all other trials in right posterior electrodes in the Unrelated group, while no effects were found in right frontal electrodes. In the L1-Interference group, over left posterior electrodes, only differences between incongruent target-last trials and all other trials emerged. Significant Order effects emerged for congruent trials in right posterior and left frontal regions, as well as a marginal effect (p = .05) in incongruent trials in left posterior electrodes.
The scalp plots in Figure 2 show the distribution of the effects of Congruency and Order in each group within 300-500 ms and 600-800 ms, as well as the grand averages for the Unrelated (A) and L1-Interference (B) groups, respectively.

4.1.2. 600-800 ms time-window

As discussed, ERP modulations were predicted based on increased monitoring during response selection. The omnibus midline ANOVA revealed significant effects of Congruency (F(1, 43) = 12.83, p < .001, η^2 = .01) and Frontality (F(1.72, 73.83) = 6.49, p < .01, η^2 = .02). Importantly, there was a significant interaction of Group x Congruency (F(1, 43) = 5.86, p < .02, η^2 = .005), as well as a borderline significant interaction between Congruency and Frontality (F(1.10, 47.19) = 3.58, p = .05, η^2 = .0004).
Similarly, the lateralized ANOVA showed main effects of Congruency (F(1, 43) = 13.05, p < .001, η^2 = .01) and Frontality (F(1.59, 60.80) = 4.24, p < .01, η^2 = .01), and a significant two-way interaction of Group x Congruency (F(1, 43) = 5.84, p < .05, η^2 = .05). Additionally, there were significant interactions of Congruency x Frontality (F(1.50, 64.35) = 3.69, p < .05, η^2 = .0004) and Hemisphere x Frontality (F(1.63, 70.06) = 11.79, p < .001, η^2 = .003), as well as a highly significant three-way interaction of Order x Hemisphere x Frontality (F(1.88, 80.64) = 109.81, p < .0001, η^2 = .004). Pairwise comparisons indicated that there was no significant Congruency effect for the L1-Interference group, whereas Congruency had a significant effect for the Unrelated group. The Congruency effect was maximal at Fz over the midline for the Unrelated group; over the four regions of electrodes, the effect of Congruency was maximal at right posterior electrodes, and significant in all other regions with the exception of a lack of significant differences between congruent target-last trials and incongruent distractor-last trials over right parietal electrodes. A significant effect of Order also emerged across groups in left frontal and maximally in right parietal electrodes.

Figure 2
ERP Results from the Practice in Each Group

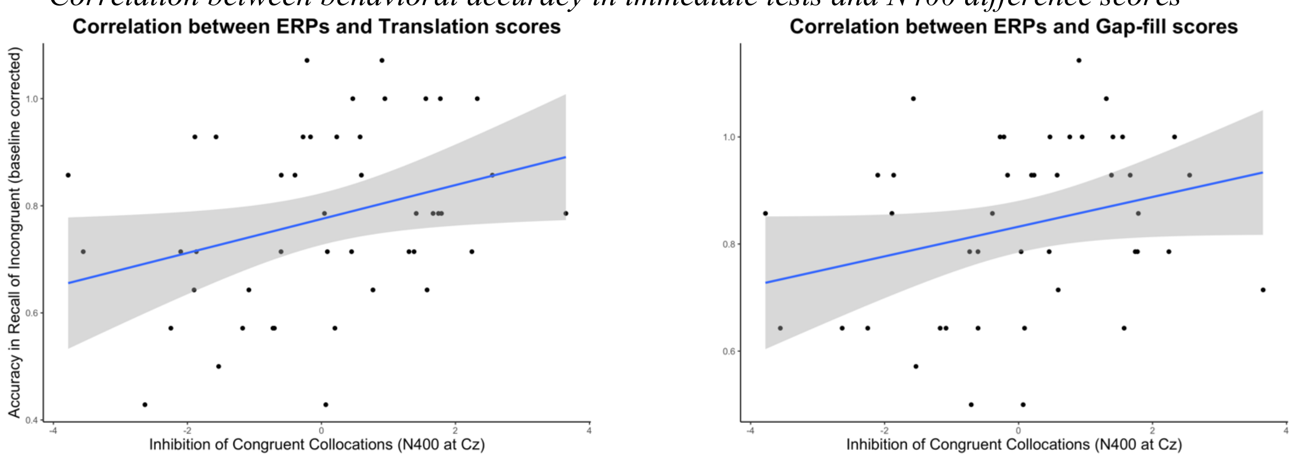
Unrelated group L1-Interference group
300-500 ms 600-800 ms 300-500 ms 600-800 ms



Note. Unrelated group (left panel), and L1-Interference group (right panel). ERPs are time-locked to the presentation of the noun. Scalp plots show voltage (µV) differences between congruent and incongruent trials. The two larger wave form plots show overlaid ERP modulations by congruency and by order of the distractor (TD: target, distractor; DT: distractor, target) at representative electrode Cz (additional electrodes in Appendix B). Shaded boxes indicate significant differences elicited by congruency.

Figure 3

Correlation between behavioral accuracy in immediate tests and N400 difference scores



Note. The horizontal axis presents the index of inhibition of native-language competitors, based on the difference in N400 amplitude between L1-congruent and incongruent trials.

4.2. Correlations between Practice phase ERP data and performance in recall post-tests

Exploratory tests were performed to investigate the potential correlation between the neurophysiological responses elicited during Practice and accuracy in the recall tests that were subsequently completed. This allowed to address the second research question, and to test the prediction that ERPs indexing inhibition (i.e., an N400 for L1-congruent items) should be correlated with performance in post-tests. Individual ERP indices were calculated for each participant based on the mean amplitude from each collocation type (Incongruent – Congruent) in the two time windows analyzed (300-500 ms and 600-800 ms, respectively). Positive values in this difference score indicate a more positive N400 in processing the noun of Incongruent collocations relative to Congruent trials, i.e., inhibition of Congruent collocations. Differences in the N400 time window were greatest at Fz in the L1-Interference group, and maximal at Pz for the Unrelated group; values were calculated for Cz, which showed significant differences in both groups. For the 600-800 ms time window, differences were computed for the mean of the right frontal region. Recall that scores for each participant were also calculated as the difference between collocation type. Positive values indicate greater ability to recall incongruent collocations (see additional details in the *Methods* section). Difference scores for ERPs and behavior were calculated for all participants and correlations were performed on the whole dataset.

For the Translation test, a significant positive correlation was found between difference scores in the N400 time window and accuracy in recall (r = .31, p < .05). A marginally significant correlation was found for the Gap-fill test (r = .28, p = .07). The results are illustrated in Figure 3. No significant correlations were found between differences in the 600-800 ms time window of ERPs and performance in post-tests. Correlations between the N400 effect and one-week delayed post-tests were not significant for any of the time windows.

Finally, to further investigate the source of individual variability in inhibition of congruent items, as indexed by differences in the N400, exploratory correlations were conducted between the N400 index and the individual measures reported in Table 1 (including Flanker effect, PSTM, AoA, weekly L2 exposure, L1 and L2 verbal fluency, and baseline L2 vocabulary knowledge). Of particular interest was the possibility that cognitive control ability, captured by Flanker effect scores, might be predictive of L1 inhibition. No significant correlations were found between the N400 index and any individual behavioral predictors (all p values > .25).

4.3. Interim discussion (Practice phase-ERP results)

A main goal, formulated in the first research question, was to investigate the neurophysiological signatures associated with different retrieval Practice conditions, by analyzing ERPs that were time-locked to the presentation of nouns, as well as their association with recall in two subsequent post-tests. The results showed that the two experimental groups displayed distinct neurophysiological signatures during completion of the Practice. First, divergent patterns were found for the N400 window across groups. Secondly, significant by-congruency differences emerged after 600 ms only in the Unrelated group. The results for each time window are discussed in turn.

4.3.1. Differences within 300-500 ms

Recall that for the 300-500 ms time-window, it was hypothesized that conditions of desirable difficulty would result in significant modulations of the N400 component. It was hypothesized that if the L1 is inhibited in conditions of desirable difficulty (L1 Regulation Hypothesis), this would reverse the relative cost of processing L1-congruent and incongruent trials.

The analysis of the N400 window allowed to examine differences in the processing of the noun resulting from the distractor verbs that preceded it, while keeping target verbs constant. As in previous experiments that capitalized on collocational priming, it was predicted that a known association between the verb and the noun would result in priming during processing of the noun. The effects of priming were expected to be greater when such an association was strong (i.e., in L1-L2 congruent items) in the form of an attenuated N400. For new incongruent collocations, more successful learning should therefore lead to a smaller difference in the N400 of incongruent relative to congruent collocations. At the same time, it was hypothesized that if inhibition of L1-L2 congruent collocations played a role in learning, a critical reversal in processing costs would be observed: congruent trials would reveal inhibition, potentially observable as an enhanced N400 relative to incongruent trials. In line with these predictions, a significant N400 in the Unrelated group revealed greater facilitation in processing the nouns of congruent vis-à-vis incongruent collocations. Significant differences were also found in the L1-Interference group but, critically, these showed a relative cost for congruent collocations, with a less negative wave for incongruent collocations. The results provide compelling evidence that practice in selecting the correct verb resulted in inhibition of L1 congruent verbs, when these were plausible alternatives based on the native language (L1-Interference group).

Additionally, inhibition was expected to play a critical role in the ability to retrieve the verbs of incongruent collocations beyond the Practice task. More specifically, it was hypothesized that evidence of inhibition in ERPs during Practice might predict performance in subsequent recall tests. The exploratory correlations between individual ERPs and test scores were also considered important given that individual variability was expected within- and across-groups. Specifically, previous behavioral evidence also suggests that L1 inhibition was present in more successful learners, even across different practice conditions (Pulido & Dussias, 2020). One aspect to note is that the N400 for incongruent collocations (Unrelated group) had a centro-parietal distribution, whereas the reversed effect (greater N400 for congruent items) in the L1-Interference group was maximal at frontal electrodes. There is an unsettled debate as to whether the classic centro-parietal N400 and its frontal counterpart (FN400) are associated with differences in functionality, with some researchers arguing that they are equivalent (e.g., Voss & Federmeier, 2012) while others defending that FN400 modulations are attributable to familiarity (e.g., Woodruff, Hayama & Rugg, 2010; Yu & Rugg, 2010). Since our materials cannot directly speak to this issue, it is beyond the scope of this study to adjudicate between the two views. Nevertheless, given these differences, voltage at the representative electrode Cz was employed for all subjects for consistency and replicability, as it showed significant differences for both groups.

The results of correlations showed that, indeed, the individual EEG activity during the Practice task was correlated with performance in recall tests conducted later. The association reached significance in the Translation test, in which the ability to resist interference from the L1 was most critical, and trended toward significance in the Gap-fill test, in which native language equivalents were not directly presented. Altogether, this pattern of results provides compelling evidence that Practice conditions that aim to induce L1-interference successfully produce inhibition of L1 competitors, and that this is directly associated with higher rates of recall of learned items. Finally, while the degree of effectiveness of a particular learning condition may vary across individuals, exploratory correlations with the behavioral measures available did not identify a specific source of variability; this question should be explored in future studies by considering predictors that may include additional measures of inhibition and cognitive control, as well as measures of attention or working memory.

4.3.2. Differences within 600-800 ms

ERPs were expected to be modulated based on increased monitoring for response selection. As expected, for learners in the Unrelated group, modulations in the N400 were followed by later differences in the 600-800 ms window, with a greater negativity in incongruent collocations that was maximal at right frontal electrodes. That is, the negativity was associated with those trials for which uncertainty was higher and for which confidence in recognition would have been lower. This right frontal effect (RFE) has indeed been described in memory studies as an index of generic memory monitoring processes (Hayama, Johnson & Rugg, 2008; Leynes & Kakadia, 2013), and as different from the posterior late positive component (LPC) associated with recollection. The RFE found in this Practice task is consistent with the results from Pulido and Dussias (2019), where a similar paradigm was employed, and in which an RFE was also elicited for all trials requiring enhanced monitoring in response selection. Interestingly, no effect was observed for the L1-Intereference group, suggesting that different mechanisms were deployed to avoid incorrect selection of the distractor. Because learners in the L1-Interference group showed a “flipped N400” (i.e., more negative for congruent than for incongruent collocations), this would suggest that the inhibition of distractors acted as a mechanism to reduce interference. That is, inhibition acted in lieu of monitoring processes indexed by the ERPs in the Unrelated group. I therefore propose that the different neurophysiological signatures indicate that different mechanisms were engaged in each group. While Unrelated learners required enhanced monitoring during response selection, learners who practiced selection in conditions of L1 interference exerted inhibition to regulate the influence of the native language. If this interpretation is correct, one would predict that the amplitudes of the RFE should be negatively correlated with the N400 inhibition wave at the individual level. A post-hoc correlation analysis was performed to test this hypothesis, which confirmed that both components were indeed negatively correlated at the individual level ($r = -.35$, $p < .05$). An important aspect to note is that, while both components index mechanisms that would have provided effective pathways to perform the task, only the N400 differences associated with inhibition correlated with higher recall ability during subsequent testing.

Finally, Order produced localized differences limited to right parietal electrodes in the 600-800 ms time-window. In both groups of learners, a more positive-going wave was elicited for distractor-last trials. Because distractor-last trials require retrieving a previously presented verb, one might speculate that the sustained positivity may be related to a higher working memory load. This is unlikely, however, because the effect does not share the scalp topography of components associated with variable working memory conditions. Order did not produce any notable differences across groups. Given that the motivation of counterbalancing the order of target and distractor in our design was simply to render the position of the target unpredictable, this is an aspect that may be addressed in future research.

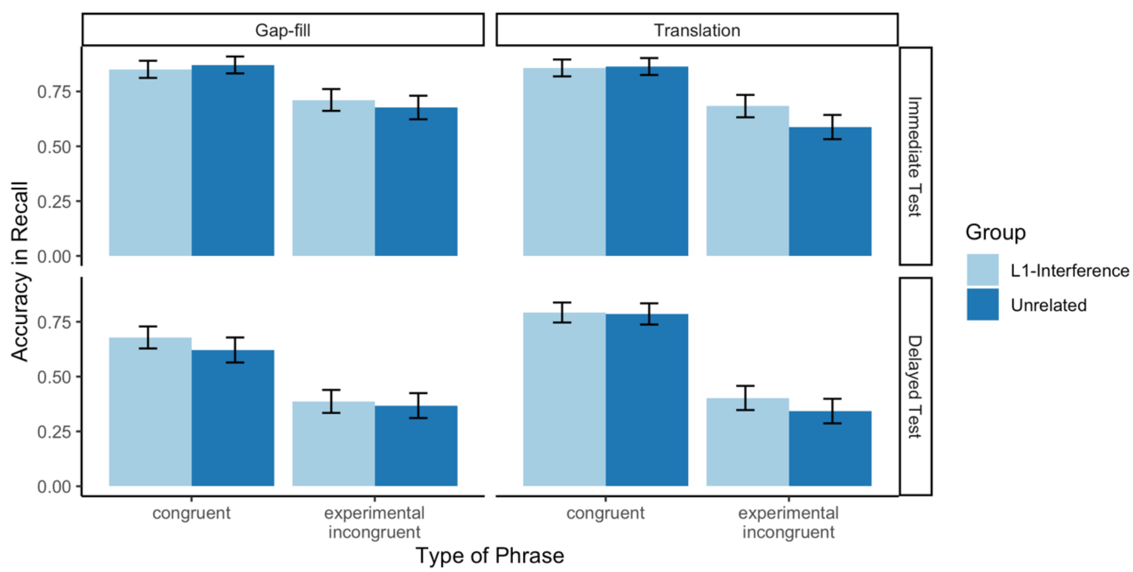
5. Tests results (learning outcomes)

The results from the gap-fill and the translation tests completed after the practice revealed the expected main effect of Congruency such that, relative to congruent collocations, accuracy in recall was overall significantly lower for incongruent phrases (β: -1.88, SE: 0.44, $p < .001$). A main effect of Session indicated that recall was significantly lower in the one-week delayed tests (β: -0.76, SE: 0.18, $p < .001$). Session interacted with Test type, revealing that recall was worse for the gap-fill test, for which the specific Spanish verb (that is, the specific action expressed)

had to be remembered (β: -1.03, SE: 0.25, $p < .001$). Additionally, a three-way interaction of Congruency x Session 3 x Test indicated higher accuracy in the gap-fill test (β: 0.72, SE: 0.32, $p < .05$), with a greater loss of accuracy for translation of incongruent items (Figure 4).
No significant main effect of Group or interactions were found. However, a critical significant interaction of the individual N400 by-condition difference and Session emerged, indicating that higher individual N400 means (indicating L1 inhibition) were associated with significantly higher recall in the immediate tests completed after the Practice (β: 0.19, SE: 0.08, $p < .05$).
Figure 4 shows the results of the Translation and the Gap-fill tests for Session 2 (Immediate tests) and Session 3 (Delayed tests). The model output is presented below in Table 5.

Figure 4

Results of Accuracy by Group in Responses to Gap-fill and Translation Tests



Note. The figure shows the results of tests completed immediately after the practice (upper row), and after a one-week delay (lower row).

Error bars represent 95% CIs.

Table 5

Output of Mixed-Effects Analysis on Post-Test Recall

Variable	Estimate	Std. Error	*z* value	*p* value
(Intercept)	2.70	0.37	7.29	< .001***
Incongruent	-1.88	0.44	-4.31	< .001***
Session	-0.76	0.18	-4.14	< .001***
Test type (Gap-fill)	0.01	0.19	0.04	.97
N400 difference mean	0.25	0.19	1.31	.19
Incongruent x Session 3	-0.75	0.23	-3.25	< .001***
Incongruent x Gap-fill test	0.33	0.24	1.41	.16
Session 3 x Gap-fill test	-1.03	0.25	-4.10	< .001***

Session 3 x N400 difference mean	-0.19	0.08	-2.42	.02*
Incongruent x Session 3 x Gap-fill test	-0.72	0.32	2.26	.02*

Note. *$p < .05$. ** $p < .01$. ***$p < .001$.

5.1. Interim discussion (tests results)

The results of the behavioral analysis revealed the expected congruency effect in the main effect and interactions for Type of collocation. The data showed a numeric difference between incongruent collocations across the two experimental groups, with higher scores for the L1-Interference group. However, despite a numeric trend, there was no main effect of experimental Group nor did it significantly interact with any of the variables considered. Therefore, while the results align with the expected pattern, the fact that the differences observed did not reach statistical significance in the mixed-effects model gives support to the hypothesis that manual responses (e.g., button presses) engaged memory and control systems to a lesser extent than practice requiring overt vocal responses in a previous study (Pulido & Dussias, 2020).

Notably, while the effect of Group was non-significant, individual ERPs did significantly correlate with recall rates in the immediate tests. Two aspects are worth noting in this regard. First, the individual ERP index provided a continuous measure which, in regression models, are not hindered by power loss in the way that categorical variables are. Relatedly, the ERP measure captured individual variability, which also carries theoretical import. Because individual variability is expected, the ERP data are additional evidence of a direct association between cognitive processes measured at the individual level during Practice and performance in testing.

Finally, the third research question was concerned with the relationship between ERPs that may index regulation of the L1, based on the predictions above, and different behavioral measures of assessment, which require regulating the L1 to a greater (Translation) or lesser extent (Gap-fill). While the results of the mixed-effects regression revealed no main effect of Test, an interaction with session revealed that recall was lower for the delayed Gap-fill test. This is not surprising, given the fact that selecting the correct verb in this test required remembering both the target meaning and form, whereas in the Translation test the exact target equivalent was provided. That is, it is quite plausible that learners experienced some difficulty remembering the particular action to be expressed in the verb, from among potential completions.

6. General Discussion

There is currently limited understanding of what neural pathways are activated as learning unfolds, and which patterns of activation lead to successful learning. By measuring EEG and behavior, the goal of this study was to investigate the cognitive mechanisms engaged by different practice conditions in real time, and to examine the association between brain potentials and subsequent behavioral learning outcomes.

The results revealed that different conditions of retrieval practice, which differed solely in their difficulty level based on the distractors included, elicited different patterns of neurophysiological activity in each group of participants. Grounded on previous research on bilingualism, the two training conditions were devised to explore the engagement of cross-language regulation, believed to be a learning-supporting mechanism in adult second language learning (Bogulski et al., 2019). To recapitulate, in an easier 'Unrelated' condition, a group of adult language learners

(L1 English—L2 Spanish) practiced selecting among two verbs, consisting of the target verb and an unrelated distractor. For another group of learners assigned to the difficult 'L1-Interference' condition, the distractors of critical items were acceptable in English but inadequate in Spanish, e.g., *gastar* ('spend') – *jugar* ('play') – *broma* ('joke) (target: *gastar* 'spend'). Thus, the level of difficulty was based on the need to suppress information from the native language. Because the distractors presented to the L1-Interference group were acceptable for native English speakers, they had to be more strongly inhibited (i.e., the verb 'play' will become inhibited). If inhibiting the native language is an important aspect of second language learning, ERPs were expected to show (a) an effect of Practice on inhibition of native-like verbs and (b) an association with performance in learning tests.

6.1. Triggers of neurocognitive pathways during learning

Differences in the components elicited at the group level revealed that practice conditions had an impact on the way selection of correct responses was approached during the task. While participants in the Unrelated group presented a clear RFE associated with monitoring response selection, those in the L1-Interference group showed no such effect, but a reversed N400. A subtle manipulation (presenting different distractors to each group) appeared to have deeply influenced participants' strategy and the cognitive mechanisms engaged to complete the task. A negative correlation between the RFE and the reversed N400 confirmed the trend in participants to show either one component or the other at the individual level. This indicates that individuals relied predominantly on one of two strategies: either monitoring closely for the response to be selected, or inhibiting the strongly competing distractors to reduce interference during response selection. This finding illustrates how purportedly small changes in the practice conditions of adult learners may result in different cognitive paths being taken to accomplish the same task, each with a different level of effectiveness and unequal consequences on measurable learning outcomes.

Thus, qualitative differences in the type of conflict induced resulted in alternative pathways for control. The pattern in the results is congruent with the well-established notion that conditions of high conflict recruit enhanced executive control (e.g., Bialystok et al., 2005; Larson, Clayson, & Clawson, 2014; West, Jakubek, Wymbs, Perry, & Moore, 2005), including conditions where conflict stems from competing linguistic representations (in syntactic representations Hsu & Novick, 2016; Navarro-Torres, Garcia, Chidambaram, & Kroll, 2019; Novick, Hussey, Teubner-Rhodes, Harbison, & Bunting, 2013; individual words, Coderre, Van Heuven, & Conklin, 2013; Kroll, Bobb, Misra, & Guo, 2008; multiword units, Pulido & Dussias, 2019). In both the easy ('Unrelated') and the difficult ('L1-Interference') conditions, the status of cross-linguistically incongruent trials triggered some form of control, although the critical difference was in how control was exerted. In the Unrelated group, the RFE indexed increased monitoring during response selection; but the consequences of this effect were transient and ceased after practice, with no consequences on recall performance. Importantly, in this condition the level of conflict between target and non-target verbs was kept the same across trials through unrelated distractors, regardless of cross-linguistic differences in phrases. However, the non-arbitrary association between target and distractor in the L1-Interference condition triggered targeted inhibition of the non-target verb, following the logic that greater interference requires stronger inhibition.

Crucially, the degree of inhibition indexed by the reversed N400 effect was correlated with behavioral measures of learning, while the RFE was not. The ERP data were compared with the results of recall post-tests, to examine the hypothesis that L1 inhibition is associated with L2 attainment. The results confirmed that the relative magnitude of N400 components at the individual level was a significant predictor of performance in the post-test completed immediately after training. That is, a greater cost in ERPs in processing *congruent* collocations (i.e., literal translations of the native language) was associated with more accurate recall of incongruent L2 collocations in testing. Therefore, the results of the experiment identified inhibition, as indexed by the reversed N400 effect, as a neurophysiological correlate associated with the practice condition that posed a desirable difficulty for learning.

These results are in line with accounts indicating that the ability to regulate the L1 is critical for successful L2 usage (Bogulski et al., 2019; Levy et al., 2007). More specifically, the data give support to the L1 Regulation hypothesis (Boguslki et al., 2019), which proposes that (a) bilinguals' regular use of a second language confers experience and greater skill in suppressing the L1; and that (b) greater ability to regulate L1 activation facilitates the learning of a new language. It is important to note that regulation of the native language is not predicted to occur indiscriminately but, rather, it is expected to emerge at loci of conflict between the novel language and the L1. This is precisely what was indicated by the results reported here, where inhibition was specifically found for L1 information that was in conflict with the newly learned L2 multiword units. Going one step further, the present study created the specific conditions to favor development of L1 regulation; the results of a "flipped" N400 effect in the present study indicated that conditions of desirable difficulty were particularly effective in engaging L1 inhibition. Importantly, these effects were found in the more successful learners, indicating a direct connection between L1 regulation in ERPs and behavioral outcomes. Future research will be needed to further expand these findings in important ways, e.g., by clarifying how L1 regulation may emerge even in conditions that do not specifically create "desirable difficulties"; and by investigating whether the development of L1 regulatory skills afford benefits in subsequent language learning.

6.2. On the scope of practice and inhibition effects

As indicated, inhibition of distractors in the difficult condition predicted accuracy in recall tests, indicating that the ability to reduce L1 interference improved L2 performance. Recall that, in the current experiment, the effect was present in immediate tests but not after one week. This replicates the behavioral results in a previous study by Pulido & Dussias (2020), who also found that the effect of learned inhibition from one training session disappeared after one week; however, the effect was persistent after additional practice. An important question is then how L1 interference reduced during recall tests, improving accuracy. One possibility is that the inhibition of competitors during immediate tests was a carryover effect following the practice (i.e., inhibition occurs during practice and rapidly decays). In other words, a plausible explanation is that the correlation between ERPs and immediate test performance might be explained as residual inhibition from the practice, which would be short-lived. Alternatively, differences might be based on a learned inhibitory association. A second account is therefore that practice might produce a more durable effect through an inhibitory 'marker' in specific items. In this second explanation, the reason why effects were significant for the immediate tests only is because one single training session was not enough to produce long-lasting effects, and more practice would be needed. Only this second explanation is supported by previous results of Pulido and Dussias which revealed that, after further practice in two additional training sessions, significant by-group effects of training persisted after a one-week delay, and were marginally significant even one month later.[4]

If the interpretation that the L1-Interference practice trained targeted inhibition is correct, the results fit well with models within the framework of the conflict-monitoring hypothesis (Botvinick et al., 2001) and, particularly, conflict-modulated Hebbian learning models such as Egner's proposal (2014). This framework considers multiple levels of control that are encoded into an event file, including bottom-up associations as well as perceptual information (Hommel 1998), e.g., top-down control states. In this approach, a given top-down control state may be encoded as part of the event, allowing to capitalize on a learned inhibitory response. While the data do not warrant more than speculation, it is plausible that experience in language regulation affects the cognitive system not only through momentous ad-hoc recruitment of reactive control (Green, 1998; Green & Abutalebi, 2013), but perhaps also through long-term associations that may include markers for inhibitory responses, e.g., for specific words and linguistic structures.

Additionally, an important gap in the previous literature on inhibition during lexical selection is that the evidence has come from experiments at the individual word level. The results reported here add to our current understanding of cross-language competition and selection by expanding the findings to selection of L2 multiword units. Specifically, the experimental manipulations in previous paradigms typically required controlling interference between languages using single words (Kroll, Bobb, Misra, & Guo, 2008; Rodriguez-Fornells et al., 2005; Rossi, Newman, Kroll, & Diaz, 2018), and often by establishing a clear-cut dichotomy between the two languages (e.g., based on language switching, Meuter & Allport, 1999; Hernandez, Martinez, & Kohnert, 2000; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; Price, Green, & Von Studnitz, 1999). In actual language usage, however, speakers not only need to avoid using the wrong language, and they must also avoid transferring information from one language onto the other beyond the word level, to avoid producing infelicitous utterances such as "spend a joke" (instead of "play a joke"), "roll a movie" (for "shoot a movie"), etc. To do so, lexical selection must not only be guided by single word-level properties, but it must operate at a more complex level, involving connections between words. By exploring lexical selection at the phrase level, the results here provide preliminary evidence that inhibition during selection is also triggered through intra-lexical connections in L2 phrase selection, affecting associations that stem from experience of the L1. While, certainly, not every aspect of learning a second language triggers conflict (i.e., L1-L2 congruent multiword units arguably require little, if any, actual learning; e.g., Carrol & Conklin, 2017), it is precisely L1-L2 incongruent information that is problematic to learn (Peters, 2016; Pulido & Dussias, 2020) and to deploy during second language use (Nesselhauf, 2003).

6.3. Indirect evidence for the effect of modality in response selection

A secondary goal of the study (Research Question 4) was to test whether the group differences produced by a comparable manipulation in Pulido and Dussias (2020), in which response selection was made vocally, would be replicated through manual response selection. The results revealed that the experimental between-group manipulation was still effective in engaging inhibitory mechanisms as shown by the ERP data. However, at the group level, there was a lack of significant differences in post-test performance. That is, the by-group comparison of behavioral recall alone would have suggested that the manipulation was not effective when responses are made manually. Importantly, this possibility was ruled out by significant between-group differences in ERPs.

The differences between previous research and the results reported here, however, lend support to the idea that inhibition training is most effective when exerted directly on oral production, but less so when suppression of competitors is more indirect and mediated by manual responses. As indicated in our predictions, research has suggested that engaging the production system may play an important role in learning (i.e., the "generation" or "production effect") (Forrin & MacLeod, 2017; Ozubko & MacLeod, 2010; Zormpa, Brehm, Hoedemaker & Meyer, 2018). On the one hand, overt responses feed into the auditory loop and have greater impact when produced by oneself rather than by another person (Forrin & MacLeod, 2017). On the other hand, overt speech engages the production system, encompassing several levels of retrieval, including gesture and motor planning. The available evidence suggests that production may play an important role in learning by enforcing retrieval of form-meaning associations at multiple levels of representation, with enduring effects even after several days (Ozubko, Hourihan & MacLeod, 2012). Given this, the inclusion of vocal responses in previous research may have provided the optimal practice conditions to test the effectiveness of the manipulation. Unlike vocal gestures, button presses are not directly linked to linguistic representations, but to somewhat arbitrary task-related mappings, i.e., in this case target responses were identified according to their position on the screen (left or right). However, in real-world language learning, adults are regularly asked to select a correct choice by making a response through the click of a computer mouse or by circling the correct option on a piece of paper. While manual responses are certainly an ecologically valid form of response selection, the evidence on the production effect suggests that such non-verbal responses may be less effective.

The question of whether non-vocal responses are equally effective as a mode of retrieval may be particularly relevant at this time, given the fact that several factors (e.g., the shift towards remote instruction) have contributed to diminishing the number of opportunities for overt vocal responses in practice. Therefore, the question of whether modality of lexical selection affects learning is one of both theoretical and practical import. However, the comparison with previous data is an indirect one, and this idea will need to be investigated systematically in future studies.

6.4. Implications for adult language learning, individual differences and personalized learning

The approach undertaken in this study is but one way in which the pathways to approach one task might diverge based on subtle differences in practice conditions. The present data explored the effect of training context on individual neurophysiological responses. An important question for future work is to elucidate what individual traits condition ERP responses even within the same training conditions. In particular, future studies may explore the potential role of individual cognitive differences, and their interaction with practice conditions. For example, research on working memory has revealed important differences in how baseline individual traits interact with training conditions. In some cases, the training condition that is optimal to one subgroup of individuals is disadvantageous for others (Li, Ellis, & Zhu, 2019; Perrachione, Lee, Ha, & Wong, 2011). To illustrate, individuals with better perceptual abilities show enhanced learning of L2 phonological contrasts in high phonetic variability training conditions. However, the same high-variability training impairs learning in individuals with lower perceptual abilities (Perrachione et al., 2011). Similarly, higher working memory (WM) has been linked to better L2 learning (Kormos & Sáfár, 2008; Martin & Ellis, 2012; Sunderman & Kroll, 2009). At the same time, those with higher WM require greater amounts of input to learn information-integration category structures relative to those with lower WM (DeCaro, Thomas & Beilock, 2008). These and similar findings represent gradual steps toward optimization of training, and underscore the importance of identifying the cognitive processes underlying individual performance. Future studies that embrace the multifaceted nature of learning will need to consider the complexity within the content being learned, the effectiveness of training approaches and individual traits. Such approaches bear the promise of expanding our understanding of variability in cognition and behavior.

7. Conclusion

Understanding the neurobiological basis of adult language learning remains a challenge across disciplines concerned with language, memory and cognition. This study provided novel insights into the connection between individual variability in brain responses and successful second language learning. Relative to previous studies that examined ERPs after training, the data presented here provide novel evidence from ongoing training, and show how training conditions may modulate brain responses as practice unfolds. The findings showed that the level of difficulty of practice can be manipulated to engage specific neurological pathways. Importantly, even when practice performance looks similar on the surface, the individual brain responses predict rates of learning success measured in subsequent tests. Furthermore, by building on recent behavioral findings (Pulido & Dussias, 2020), the data provide direct evidence that more successful second language learning is achieved through inhibition of interference from the native language. By drawing a connection between the brain mechanisms and the ensuing learning outcomes, this approach advances our understanding of the cognitive processes engaged under enabling training conditions.

Acknowledgments

This work was supported by NSF Doctoral Dissertation grant BCS-1844188 and a *Language Learning* Doctoral Dissertation Award. Thanks to Katrina Connell, Mary Henderson, Anna Northcott, Zeal Patel, Enissa Ramos and Laura Rodrigo for help with data collection; and to Matt Carlson, Judith Kroll, Rena Torres Cacoullos, and Janet van Hell for their input. I am also grateful to Malte Nieuwland and two anonymous reviewers for their helpful comments. Special thanks to Paola E. Dussias for comments, insight and advice at different stages of the study. Any errors are my own.

References

Abutalebi, J., Annoni, J. M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., Lazeyras, F., Cappa, S., & Khateb, A. (2008). Language control and lexical competition in bilinguals: An event-related fMRI study. *Cerebral Cortex, 18*(7), 1496–1505. https://doi.org/10.1093/cercor/bhm182

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242-275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods, 50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering Can Cause Forgetting: Retrieval Dynamics in Long-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

Arnon, I., & Christiansen, M. H. (2017). The Role of Multiword Building Blocks in Explaining L1–L2 Differences. *Topics in Cognitive Science, 9*(3), 621–636. https://doi.org/10.1111/tops.12271

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language, 83,* 62-78. https://doi.org/10.1016/j.jml.2015.04.004

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning and Memory, 22*(11), 544–556. https://doi.org/10.1101/lm.037986.114

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24*(1), 40–49. https://doi.org/10.1016/j.neuroimage.2004.09.044

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties y vocabulary learning. *The American Journal of Pyschology, 128*(2), 241–252.

Bogulski, C. A., Bice, K., & Kroll, J. F. (2019). Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. *Bilingualism: Language and Cognition, 22*(5), 1052-1067. https://doi.org/10.1017/S1366728918000858

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652. doi:10.1037/0033-295X.108.3.624

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology, 109*(4), 383–396. https://doi.org/10.1016/j.jecp.2011.03.012

Calabria, M., Costa, A., Green, D. W., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences, 1426*(1), 221–235. https://doi.org/10.1111/nyas.13879

Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea "has legs." *Bilingualism: Language and Cognition*, *20*(2), 299–317. https://doi.org/10.1017/S1366728915000103

Coderre, E. L., Van Heuven, W. J. B., & Conklin, K. (2013). The timing and magnitude of Stroop interference and facilitation in monolinguals and bilinguals. *Bilingualism: Language and Cognition, 16*(2), 420–441. https://doi.org/10.1017/S1366728912000405

Davies, M. (2016). *Corpus del Español: Two billion words, 21 countries.* http://www.corpusdelespanol.org/web-dial/.

De Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning, 50*(1), 1–56.

Degani, T., & Tokowicz, N. (2010a). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition, 13*(3), 299–314.

Degani, T., & Tokowicz, N. (2010b). Semantic ambiguity within and across languages: An integrative review. *Quarterly Journal of Experimental Psychology, 63*(7), 1266–1303. https://doi.org/10.1080/17470210903377372

Egner, T. (2014). Creatures of habit (and control): A multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology, 5*, 1247. doi:10.3389/fpsyg.2014.01247

Evert, S. (2005). *The Statistics of Word Cooccurrences Word Pairs and Collocations*. Unpublished dissertation.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English, *Applied Psycholinguistics*, *24*(3), 453-489.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP "old/new" effects: Memory strength and decisional factor(s). *Neuropsychologia*, *40*(13), 2288–2304. https://doi.org/10.1016/S0028-3932(02)00113-6

Forrin, N. D., & Macleod, C. M. (2017). This time it's personal: the memory benefit of hearing oneself. *Memory, 26*(4), https://doi.org/10.1080/09658211.2017.1383434

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking? *Journal of Memory and Language, 66,* 226-248. https://doi.org/10.1016/j.jml.2011.07.007

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism: Language and Cognition, 14*(3), 379–399. https://doi.org/10.1017/S136672891000012X

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can Bilinguals See It Coming? Word Anticipation in L2 Sentence Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469.

Grady, C. L., Luk, G., Craik, F. I. M., & Bialystok, E. (2014). Brain network activity in monolingual and bilingual older adults, *Neuropsychologia*, *66*, 170-181. https://doi.org/10.1016/j.neuropsychologia.2014.10.042

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Gries, S. T. (2010). Useful statistics for corpus linguistics. *A mosaic of corpus linguistics: Selected approaches*, *66*, 269-291.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801-811. https://doi.org/10.1098/rstb.2007.2089

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

Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old / new ERP effect and post-retrieval monitoring: Specific or non-specific? *Neuropsychologia,* 46, 1211–1223. https://doi.org/10.1016/j.neuropsychologia.2007.11.021

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition, 5*(1-2), 183–216. doi:10.1080/713756773

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hsu, N. S., & Novick, J. M. (2016). Dynamic Engagement of Cognitive Control Modulates Recovery from Misinterpretation During Real-Time Language Processing. *Psychological science*, *27*(4), 572-582. https://doi.org/10.1177/0956797615625223

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology, 111*(10), 1745–1758. https://doi.org/10.1016/S1388-2457(00)00386-2

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*(3), 416–430. https://doi.org/10.1016/j.actpsy.2008.02.001

Kutas, M., & Federmeier, K. D. (2010). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual review of psychology*, *62*, 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science, 207*(4427), 203–205. http://www.jstor.org/stable/1683915.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163. https://doi.org/10.1038/307161a0

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology, 93*(3), 283–297. https://doi.org/10.1016/j.ijpsycho.2014.06.007

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting Your Native Language. *Psychological Science, 18*(1), 29–34. https://doi.org/10.1111/j.1467-9280.2007.01844.x

Leynes, P. A., & Kakadia, B. (2013). Variations in retrieval monitoring during action memory judgments : Evidence from event-related potentials (ERPs). *International Journal of Psychophysiology, 87*(2), 189–199. https://doi.org/10.1016/j.ijpsycho.2013.01.004

Li, S., Ellis, R., & Zhu, Y. (2019). The associations between cognitive ability and L2 development under five different instructional conditions. *Applied Psycholinguistics, 40*, 693–722. https://doi.org/10.1017/S0142716418000796

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*(1), 29–41. https://doi.org/10.1016/j.cognition.2009.08.014

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-57. https://doi.org/10.1044/1092-4388(2007/067)

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition 3434*(34), 379–413. https://doi.org/10.1017/S0272263112000125

Meara, P. (1996). The dimensions of lexical competence. *Performance and competence in second language acquisition, 35*, 33-55.

Mendez, M. F. (2019). Bilingualism and Dementia: Cognitive Reserve to Linguistic Competency. *Journal of Alzheimer's Disease, 71*(2), 377–388. https://doi.org/10.3233/JAD-190397

Mendez, M. F. M. (1999). Bilingualism and Dementia: Cognitive Reserve to Linguistic Competency. *Journal of Alzheimer's Disease*, *71*(2), 377-388. https://doi.org/10.3233/JAD-190397

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

Morgan-Short, K., Steinhauer, K., Sanz, C., Ullman, M. T., Steinhauer, K., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24(*4), 933–947. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Navarro-Torres, C. A., Garcia, D. L., Chidambaram, V., & Kroll, J. F. (2019). Cognitive Control Facilitates Attentional Disengagement during Second Language Comprehension. *Brain sciences*, *9*(5), 95. https://doi.org/10.3390/brainsci9050095

Nesselhauf, N. (2003). The use of collocations by advanced learners of English. *Applied Linguistics, 24*(2), 223–242. Retrieved from http://eprints.ibu.edu.ba/78/

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., ... & Matthew Husband, E. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2013). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience, 29*(2), 186–217. https://doi.org/10.1080/01690965.2012.758297

Ozubko, J. D., Hourihan, K. L., & Macleod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*(7), 717-727. https://doi.org/10.1080/09658211.2012.699070

Ozubko, J. D., & Macleod, C. M. (2010). The Production Effect in Memory: Evidence That Distinctiveness Underlies the Benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1543. https://doi.org/10.1037/a0020604

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design*. The Journal of the Acoustical Society of America, 130,* 461. https://doi.org/10.1121/1.3593366

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113-138. https://doi.org/10.1177/1362168814568131

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143*(2), 644–667. https://doi.org/10.1037/a0033194

Pulido, M. F., & Dussias, P. E. (2019). The Neural Correlates of Conflict Detection and Resolution During Multiword Lexical Selection: Evidence from Bilinguals and Monolinguals. *Brain Sciences, 9*(5), 110. https://doi.org/10.3390/brainsci9050110

Pulido, M. F., & Dussias, P. E. (2020). Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism, 23*(3), 652–667. https://doi.org/10.1017/S1366728919000622

Rodriguez-Fornells, A., van der Lugt, Rotte, M., Britti, B., Heinze, H-J., Münte, T. F (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience, 17*, 422–433. https://doi.org/10.1162/0898929053279559

Rodriguez-Fornells, A., Rotte, M., Helnze, H. J., Nösseit, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature, 415*(6875), 1026–1029. https://doi.org/10.1038/4151026a

Rossi, E., Newman, S., Kroll, J. F., & Diaz, M. T. (2018). Neural signatures of inhibitory control in bilingual spoken production. *Cortex, 108*, 50–66. https://doi.org/10.1016/j.cortex.2018.07.009

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences*, *11*(6), 251-257. https://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., Herron, J., & Morcom, A. (2002). Electrophysiological studies of retrieval processing. *Neuropsychology of memory*, *3*, 154-165

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46*(2), 419–440. https://doi.org/10.1006/jmla.2001.2813

Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning, 63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139-152. https://doi.org/10.1080/09571730802389975

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics, 30*(1), 79–99. https://doi.org/10.1017/S0142716408090048

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition, 17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Tokowicz, N., & MacWhinney, B. (2005). Implicit and Explicit Measures of Sensitivity to Violations in Second Language Grammar. *Studies in Second Language Acquisition, 27*(2), 173–204. https://doi.org/10.1017/S0272263105050102

Toomer, M., & Elgort, I. (2019). The Development of Implicit and Explicit Knowledge of Collocations: A Conceptual Replication and Extension of Sonbul and Schmitt (2013). *Language Learning, 69*(2), 405–439. https://doi.org/10.1111/lang.12335

Van Berkum, J. J. A, Hagoort, P. & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671, https://doi: 10.1162/089892999563724.

West, R., Jakubek, K., Wymbs, N., Perry, M., & Moore, K. (2005). Neural correlates of conflict processing. *Experimental Brain Research, 167*(1), 38–48. https://doi.org/10.1007/s00221-005-2366-y

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of l2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2018). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 340-352. https://doi.org/10.1080/09658211.2018.1510966

---

[1] While an N400 reflecting lexical access would be consistent with the priming effect described, either lexical access or integration (or both) might explain the presence of the component in the paradigm employed here; adjudicating between the two goes beyond the scope of the present paper and no claims are made regarding the specific underlying process.

[2] An a priori power analysis was performed on G*Power (Faul, Erdfelder, Buchner & Lang, 2009) based on data from the group of L2 speakers in Pulido and Dussias (2020), which used a similar verb selection paradigm with a plausible/implausible distractor manipulation (F(u)=0.50). The analysis indicated that 21 participants/group would yield a power of .90, $\alpha = .05$.

[3] For example, "dial" could only be literally translated in Spanish as *discard*, but this verb is dated and is also a regionalism (i.e., non-existent in several Spanish-speaking countries). For the beginner-intermediate learners tested here, identifying a literal translation of L2 verbs would be virtually impossible.

[4] The suggestion that additional practice was necessary is also congruent with other studies on learning of multiword units, which consisted of one or more training sessions (Sonbul & Schmitt, 2013; Toomer & Elgort, 2019).

Appendix A

Table A1

List of stimuli

	Experimental Incongruent collocations			Congruent Collocations			Filler collocations		
	verb	(det.)	noun	verb	(det.)	noun	verb	(det.)	noun
1	*pedir 'ask for'*		*pizza*	*ordenar 'order'*	[*este*]	*caos*	*marcar 'mark'*	[*un*]	*número*
	order		pizza	order	[this]	chaos	dial	[a]	número
2	*dirigir 'direct'*	[*un*]	*negocio*	*correr 'run'*		*millas*	*hacer 'make'*	[*las*]	*maletas*
	run	[a]	business	run		miles	pack	[the]	bags
3	*rodar 'roll'*		*escenas*	*disparar 'shoot'*	[*una*]	*pistola*	*controlar 'control'*	[*el*]	*estrés*
	shoot		*scenes*	shoot	[a]	gun	handle	[the]	stress
4	*blanquear 'whiten'*		*dinero*	*lavar 'launder'*		*ropa*	*volar 'fly'*		*puentes*
	launder		money	launder		clothes	blow up		*bridges*
5	*despertar 'awaken'*		*dudas*	*levantar 'raise'*	[*la*]	*cabeza*	*navegar 'sail'*	[*la*]	*web*
	raise		doubts	raise	[one's]	head	surf	[the]	web
6	*poner 'put'*		*atención*	*pagar 'pay'*	[*el*]	*coste*	*reconocer 'recognize'*		*fallos*
	pay		attention	pay	[the]	cost	acknowledge		mistakes
7	*dar 'give'*		*paseos*	*tomar 'take'*		*aviones*	*presenciar 'be present'*		*muertes*
	take		walks	take		planes	witness		deaths
8	*perder 'lose'*	[*un*]	*tren*	*extrañar 'miss'*	[*sus*]	*besos*	*entregar 'deliver'*		*propuestas*
	miss	[a]	train	miss	[her]	kisses	submit		proposals

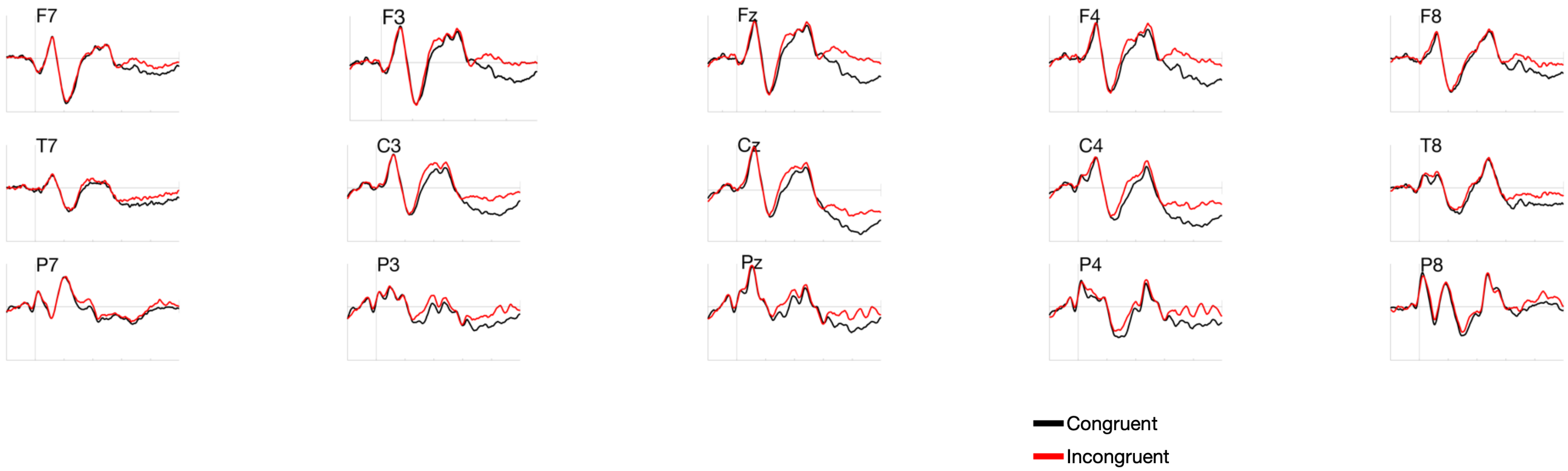
9	*ganar 'win'*		*tiempo*	*comprar 'buy'*	*bebidas*	*programar 'program'*	*citas*
	buy		time	buy	drinks	schedule	appointments
10	*gastar 'spend'*		*bromas*	*jugar 'play'*	*partidos*	*echar 'throw'*	*agua*
	play		jokes	play	matches	pour	water
11	*abrir 'open'*	[*el*]	camino	*liderar 'lead'*	*equipos*	*cambiar 'change'*	*cheques*
	lead	[the]	way	lead	teams	cash	checks
12	*montar 'assemble'*		*fiestas*	*tirar 'throw'*	*pelotas*	*subir 'take up'*	*documentos*
	throw		parties	throw	balls	upload	documents
13	*sacar* 'extract'		*fuerzas*	*dibujar 'draw'*	*líneas*	*publicar 'publish'*	*vídeos*
	draw		strength	draw	lines	post	videos
14	*revelar 'reveal'*		*fotos*	*desarrollar 'develop'*	*diabetes*	*ajustar 'adjust'*	*cuentas*
	develop		photos	develop	diabetes	settle	accounts

Note. The table presents the collocations (i.e., conventional verb-noun phrases) that were studied and practiced by learners. Below each Spanish collocation, its idiomatic English translation is provided. The literal English translations for the verbs Spanish collocations are provided to the right of each verb.

Appendix B

Figure B1

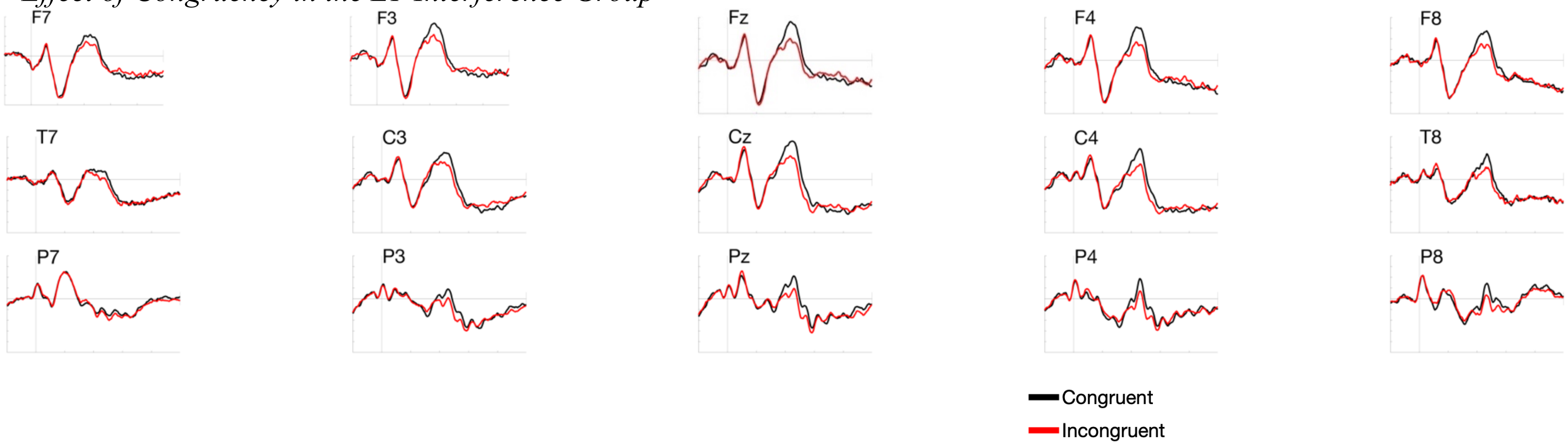
Effect of Congruency in the Unrelated Group



Note. Grand average ERPs time-locked to presentation of the noun (from -200 to 1000 ms) showing the effect of Congruency for the Unrelated group. Negativity is plotted up.

Figure B2

Effect of Congruency in the L1-Interference Group



Note. Grand average ERPs time-locked to presentation of the noun (from -200 to 1000 ms) showing the effect of Congruency for the L1-Interference group. Negativity is plotted up.