ELSEVIER

Research paper

Event related potentials at initial exposure in third language acquisition: Implications from an artificial mini-grammar study

Jorge González Alonso [a],[**], José Alemán Bañón [b], Vincent DeLuca [c], David Miller [d],
Sergio Miguel Pereira Soares [e], Eloi Puig-Mayenco [f], Sophie Slaats [g],
Jason Rothman [a],[h],[*]

[a] UiT The Arctic University of Norway, Norway
[b] Stockholm University, Sweden
[c] University of Birmingham, UK
[d] University of Illinois at Chicago, USA
[e] University of Konstanz, Germany
[f] King's College London, UK
[g] Max Planck Institute for Psycholinguistics, The Netherlands
[h] Universidad Nebrija, Spain

ARTICLE INFO

ABSTRACT

The present article examines the proposal that typology is a major factor guiding transfer selectivity in L3/Ln acquisition. We tested first exposure in L3/Ln using two artificial languages (ALs) lexically based in English and Spanish, focusing on gender agreement between determiners and nouns, and between nouns and adjectives. 50 L1 Spanish-L2 English speakers took part in the experiment. After receiving implicit training in one of the ALs (Mini-Spanish, N = 26; Mini-English, N = 24), gender violations elicited a fronto-lateral negativity in Mini-English in the earliest time window (200–500 ms), although this was not followed by any other differences in subsequent periods. This effect was highly localized, surfacing only in electrodes of the right-anterior region. In contrast, gender violations in Mini-Spanish elicited a broadly distributed positivity in the 300–600 ms time window. While we do not find typical indices of grammatical processing such as the P600 component, we believe that the between-groups differential appearance of the positivity for gender violations in the 300–600 ms time window reflects differential allocation of attentional resources as a function of the ALs' lexical similarity to English or Spanish. We take these differences in attention to be precursors of the processes involved in transfer source selection in L3/Ln.

1. Introduction

Whereas effects from specific previous language experience in adult second (L2) language acquisition can, at most, have a single

---

source (the first (L1) language), the possibilities for cross-linguistic effects are more dynamic when bilinguals of any type set out to learn/acquire yet another language. Obviously, effects can in principle: (i) come from an L1 (either in the case of simultaneous [2L1] bilinguals) or from a sequentially acquired language (i.e., L2), (ii) have distinct qualitative natures (e.g., momentary intrusions of grammar versus contributions from copies of underlying linguistic representations), (iii) be holistic (from a single language, (one of) the L1(s) or the L2) or piecemeal, property-by-property (from both the L1 or L2) and/or (iv) come all at once (likely very early on) or iteratively, as needed, over the course of L3/Ln development. This dynamic reality makes the study of cross-linguistic effects in multilingualism a challenging task. Yet it is an important one, not least due to the sheer number of people who acquire an additional language after experience with at least two others. In light of the inherent complexities, empirical methodology and timing of testing is not a one-size-fits-all issue. Rather, quite rightly, determining how we approach these problems empirically and when the best time to test is along the L3/Ln developmental continuum must be assessed against the backdrop of specifically articulated questions, situated in a context of relevant theoretical models.

An important issue that emerges from the above scenario concerns the very object of interest here. One must ask if all relevant studies are in fact talking about the same phenomena when using shared terminology: *Is one interested in any and all potential effects or, perhaps, more concerned with a specific subset of them?* This question parallels an important consideration regarding the nature of cross-linguistic influence (CLI) as a macro-construct. In the broad sense, CLI encompasses any behavioral linguistic action that can be meaningfully linked back to previous language knowledge/experience. As such, this can include behavioral reflexes that are context dependent, irregular in their appearance or unsystematic, alongside behaviors with great regularity and high systematicity. Whereas the former are more suggestive of a bleeding over of previous linguistic knowledge into Ln performance, the latter suggest a representational copy from a previous linguistic system into the developing Ln interlanguage, in other words, something that forms part of the Ln linguistic system itself.

In the context of L3/Ln, but even more generally, González Alonso and Rothman (2017) and Rothman, González Alonso, and Puig-Mayenco (2019) have argued that a distinction between at least the aforementioned two types of effects from previous language knowledge should be maintained, and that clarity of terminology in referring to them is warranted. From such a perspective, *transfer* and *CLI* are not interchangeable terms, but rather stand in a subset-superset relationship. *Transfer* refers to the nature of representational features in the grammar itself that are copied from another previously acquired grammar, whereas the more general construct of CLI would also cover all other instances of influence that do not seem to be at the level of mental grammatical representation.[1]

To the extent possible, our aim is to isolate mental representation of grammatical contributions from previously acquired grammars (transfer) and/or evidence regarding the conditioning factors for their selection from other variables affecting multilingual performances that could stem from previous linguistic knowledge (CLI proper). To do this, we believe that testing L3/Ln knowledge as early as possible in the developmental continuum affords the best chance of disentangling representational transfer and/or its precursors (evidence of what gives rise to ultimate selection) from CLI effects. While there are many theoretical and practical reasons for studying the full gamut of CLI effects across all levels of L3 development, our focus is squarely on understanding first and primarily what happens in the very beginning of L3/Ln with respect to the factors that drive representational transfer selection. Precisely because there are choices (L1 and L2) in this regard, studying how multilingual transfer selection obtains at the onset of L3/Ln acquisition provides the most accurate description of the specification of the grammatical starting point for L3 development itself (the shape of the initial L3 interlanguage representation for any given structure). Of equal importance, early testing also provides a unique window into the time course for selection itself: what evidence can we see with respect to how selection is made that also sheds light on how cognitive and linguistic factors interact in the mind more generally?

While it is not the case that all models/theories of transfer in L3 grammar are equally interested in, focused on, or limited to the initial stages, they all make direct or indirect claims about it. The available models run the entire gamut of possibilities. Some have suggested a default primacy effect for the L2 (e.g., Bardel & Falk, 2007, 2012; see Hermas, 2014, for a similar yet unformalized suggestion about the L1), whereas others claim both the L1 and L2 have, in principle, the same potential for transfer. Non-default models are differentiated on two fronts: (a) the variables they propose as deterministic for transfer selection between the L1 and L2 and/or (b) the nature and timing of transfer itself, that is, whether or not it is holistic and early on (full transfer models) or property-by-property over the course of L3/Ln development.

First among the cohort of the non-default models, the Cumulative Enhancement Model (CEM; Flynn, Foley, & Vinnitskaya, 2004) does not envisage a context of full transfer, but rather a scenario of conservative property-by-property transfer. According to the CEM, transfer is warranted under two conditions: (i) when the learner is at a point in L3 development where specification for any given particular property is ripe/needed and (ii) there is a specification in the L1 and/or L2 that facilitates the convergence on the L3 specification. The CEM precludes the possibility of transfer that would result in an "incorrect" L3 specification. The Linguistic Proximity Model (LPM; Westergaard, Mitrofanova, Mykhaylyk, & Rodina, 2017), another non-full transfer model, agrees with the CEM that transfer is property-by-property over time. However, it rejects the claim that selection is predicated on facilitation, thus leaving room for transfer that turns out to be non-facilitative. The LPM claims that underlying structural considerations (the L3 compared to what the other languages provide at the property level) is ultimately what determines transfer source. The Typological Primacy Model (TPM; Rothman, 2011, 2015; Rothman, González Alonso & Puig-Mayenco, 2019) is similar to the LPM in that it argues for underlying

---

[1] Such a distinction does not necessarily accord with all traditions studying multilingual acquisition, for example, a Dynamic Systems Theory approach, where the status of linguistic mental representations as such is contested (e.g., de Bot, Lowie, & Verspoor, 2007; Jessner, 2008), nor is it adopted equally even when the linguistic approach—for example, the generative tradition that maintains an I-language versus E-language distinction—would in principle allow for it (see Westergaard, 2019).

structure to be the motivator of transfer selection, but it differs by claiming that transfer is holistic (full transfer in the sense of Schwartz & Sprouse, 1996), obtaining from the earliest moment the parser receives enough linguistic information to determine overall structural proximity. As a result, this full transfer model advocates for an initial interlanguage grammar that is fully specified from its outset, that is, once the full copy of an underlying system has been established as the initial interlanguage of the L3.

With the above context in mind, Rothman, Alemán Bañón, and González Alonso (2015) advocated a methodological expansion in formal linguistic approaches to L3/Ln morphosyntactic studies, sitting at the cross-roads of neurolinguistics and artificial mini-grammar learning. At the time, use of electroencephalography (EEG) in formal L2 morphosyntax had been providing promising results suggesting that representational transfer effects could be captured with event-related potentials (ERPs). Work showing evidence of very early *qualitatively-similar-to-natives* processing for properties available for transfer, juxtaposed against the lack of such effects when transfer is not possible, was amassing, e.g. number versus gender processing in English learners of Spanish (Alemán Bañón, Fiorentino, & Gabriele, 2018; Bond, Gabriele, Fiorentino, & Alemán Bañón, 2011; Gabriele, Fiorentino, & Alemán Bañón, 2013). At the same time, evidence of qualitatively similar processing for gender violations after reaching advanced proficiency was showing that properties not part of the L1 could be acquired (Alemán Bañón et al., 2018). If acquired, L2 representations would in principle be available for transfer in multilingual acquisition/processing. Showing that ERPs can capture transfer effects and that learners can/do attain native-like signatures for grammatical processing meant that the precursors for applying this method were in place.

Rothman et al. (2015) argued that an ERP extension into L3 morphosyntactic transfer studies was ideal because such data register involuntary reactions at the brain level with excellent temporal resolution. They are thus not subject to metalinguistic variables in the way offline behavioral methods could be. Moreover, since ERPs provide a window of sorts into the inner working of the mind as parsing and processing occurs in real-time, EEG data could shed a unique light onto the process of transfer selection itself, especially if done very early on in the process. As such, ERPs would be as welcome an addition to the formal linguistic L3 morphosyntactic literature as they had been for the L2 literature. At the same time, Rothman and colleagues argued that employing an ERP method under the conditions of artificial mini-grammar learning would serve an additional purpose, beyond the obvious one related to capturing evidence at the very earliest stages. Artificial mini-grammar learning allows for maximal control of crucial variables, not least the quantity and the qualitative shape of language exposure. Working with semiartificial or fully artificial grammar learning is not new to non-native language research (e.g., Morgan-Short, Steinhauer, Sanz, Ullman, 2012; Rebuschat & Williams, 2012; Williams & Kuribara, 2008). In the majority of such cases, however, this paradigm is used to examine implicit versus explicit learning. Rothman et al.'s reason for using it was somewhat distinct. Using two artificial mini-grammars, as described below, offered two methodological advantages that would have been nearly impossible to achieve outside such a context. First, it allowed for control of the exact quantity and context of exposure to the L3 across all participants, ensuring everyone was at a very beginning proficiency level and received maximally comparable input. Second, it allowed for manipulation of the qualitative shape of the input the participants received, taking advantage of lexical similarities to either the L1 (Spanish) or the L2 (English) of the participant groups to make two artificial mini-grammars. As we will explain below, this was a crucial step in testing between default models and those that advocate structural similarity as the main motivator for transfer selection, in particular the TPM (the only such model existing at the time of Rothman et al.'s (2015) paper).

Using mini grammars in L3/Ln research is rare (e.g. Grey, Williams, & Rebuschat, 2014; Sanz, Park, & Lado, 2015) and, to date, has investigated similar questions pertaining to implicit versus explicit learning as when applied to L2. Combining ERP and a (semi) artificial grammar paradigm in non-native language research is a bit more common, although the vast majority of this regards adult L2 acquisition (see Morgan-Short, 2020 for review). To our knowledge, the few studies (e.g., Grey, Sanz, Morgan-Short, & Ullman, 2018) that have combined ERP with artificial grammar learning in L3 contexts have done so for distinct questions to those we pursue here. However, findings from Grey et al.'s (2018) study are highly relevant. They examined behavioral and neural correlates of learning an artificial language in early Mandarin–English bilinguals, compared to English monolinguals as a control. After grammar instruction, participants were tested by means of a grammaticality judgement task at low- and high proficiency while EEG was recorded. Most interesting to the present context are the low proficiency results, where bilinguals and monolinguals did not differ on behavioral measures, but showed distinct ERP patterns. Only the bilinguals showed a P600, a common ERP correlate of combinatorial processing, for sentences that were ungrammatical in the artificial grammar. Grey et al. (2018) did not align their results to the L3 literature as presented above, as they wanted to test the claim that bilinguals are "better" learners of subsequent languages (Cenoz, 2003). And so, the present study is not only the fulfillment of the discussion and methodology presented by Rothman et al. (2015), but remains the first ERP paper to weigh in on the formal linguistic L3 literature.

2. Research questions

The fundamental research questions guiding this study are:

RQ1. What linguistic and extra-linguistic variables constrain the selection of a transfer source in non-native language acquisition, when previous language experience instantiates multiple (and potentially conflicting) options?

RQ2. (How) Can neurolinguistic measures of implicit language processing, such as those provided by EEG/ERPs, help us to adjudicate between the competing predictions of several models developed on the basis of behavioral data?

In particular, we test for the first time if the sensitivity of ERPs to (potentially) different neurocognitive underpinnings of converging behavior can permit us to distinguish between performance rooted in transferred linguistic representations vs. performance based on a variety of learning/task strategies (rhyming, pattern matching, etc.).

2.1. Predictions

With these questions in mind, we first follow the predictions of Rothman et al. (2015) with respect to the three main models existing at the time (the L2 Status Factor, the CEM and the TPM) for potential ERP effects for the processing of gender violations in two artificial mini-grammars, lexically based on Spanish (Mini-Spanish) and English (Mini-English), respectively, by different groups of Spanish-English sequential bilinguals. The key feature of these semiartificial systems, as originally conceived by Rothman et al., was that *both* would display gender agreement between nouns and adjectives. Number agreement was used as a control, because both Spanish and English realize it within the NP—although only Spanish displays noun-adjective agreement. In contrast, grammatical gender applies only to Spanish. Every Spanish noun is inherently assigned masculine or feminine gender. Adjectives and determiners display a matching systematic morphophonological agreement pattern. English lacks this feature altogether.

ERP predictions for the L3/Ln transfer models were based, first, on the foundational work conducted on sentence processing in native speakers (e.g., Haagort, Brown & Groothusen, 1993; Osterhout & Holcomb, 1992), specifically on the literature exploring the processing of syntactic agreement (e.g., Osterhout & Mobley, 1995), and second, on the growing body of work examining sentence processing through electroencephalography (EEG) in non-native speakers—typically, adult second language learners. In native speakers, 25 years of research into agreement processing have yielded robust evidence of a strong association between violations of this type of morphosyntactic dependency and the appearance of the ERP component known as the P600 (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2012; Barber & Carreiras, 2005; Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Hagoort, 2003; Nevins, Dillon, Malhotra, & Phillips, 2007; Osterhout & Mobley, 1995; O'Rourke & Van Petten, 2011).

In this connection, it is worth noting that the relationship between the P600 and agreement violations in native speakers is not a one-to-one correspondence. The P600 can be elicited by non-linguistic events (e.g., disruptions in musical harmony, Patel, Gibson, Ratner, Besson, & Holcomb, 1998; visual degradation, van de Meerendonk, Chwilla, & Kolk, 2013), suggesting that it might reflect sensitivity to the structural configuration of rule-based knowledge beyond (linguistic) syntax. Also, it sometimes appears in response to linguistic anomalies that are typically subsumed under semantics (e.g., Gunter, Friederici, & Schriefers, 2000; Kuperberg, 2007; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006). Taken together, these findings led to a reinterpretation of the P600 as reflecting late integration of syntactic, semantic and thematic information (see Friederici, 2017, pp. 62–65, for an overview and discussion). Furthermore, agreement violations have been reported, in a subset of studies, to elicit other types of electrophysiological responses, such as a biphasic pattern where the P600 is preceded by a negativity in the 200–500 ms time window, often with a left-anterior distribution, known as the LAN (e.g., Friederici, Hahne, & Mecklinger, 1996; Molinaro, Barber, Caffarra, & Carreiras, 2015).[2]

Non-native speakers display more variability in their electrophysiological responses to agreement violations, which vary as a function of factors such as L2 proficiency and L1-L2 combination (see, e.g., Alemán Bañón et al., 2018). N400 responses without a subsequent positivity, which are rare in sentential contexts in the native speaker literature, can be found in low-proficiency L2 speakers for agreement violations of both shared (L1-L2) and novel (L2 only) features (e.g., Carrasco et al., 2017; Osterhout, McLaughlin, Pitkanen, Frenck-Mestre & Molinaro, 2006; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013), which tend to elicit more native-like, positive-dominant responses at higher proficiency (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2014, 2018; Foucart & Frenck-Mestre, 2012; see also Rossi, Kroll, & Dussias, 2014, where the novel feature, gender, only elicited native-like responses at very high proficiency). The rate of development to native-like processing of agreement dependencies, however, is arguably faster for non-novel features (i.e., those already present in the L1; e.g., Gabriele et al., 2013), for which P600-like responses have been reported even at low levels of proficiency (e.g., Alemán Bañón et al., 2018; Alemán Bañón, Hoffman, Covey, Rossomondo & Fiorentino, under review; Tokowicz & MacWhinney, 2005; cf. Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; Morgan-Short, Sanz, Steinhauer & Ullman, 2010, who report negative-dominant responses at this stage). Crucially for our purposes, the fact that prior experience with certain linguistic features can result in early native-like electrophysiological indices suggests that linguistic transfer can be captured and meaningfully teased apart from other potential sources of target-like behavior in grammatical agreement comprehension (for example, phonological associations between word pairs, which would yield qualitatively different ERP components, such as an N400, as shown by studies manipulating rhyming; e.g., Coch, Hart, & Mitra (2008) from the earliest stages of non-native language development.

The relevant predictions in Rothman et al. (2015) are schematized in Table 1.

Let us walk through the reasoning behind the predictions appearing in the table. The predictions of the L2 Status Factor and the CEM are the same for both participant groups (L1 Spanish learners of L2 English) regardless of which artificial mini-grammar they are exposed to, Mini-Spanish or Mini-English because they relate to aspects that are fixed across either the participant groups or the sum total of previous language experience for both. In other words, the L2 is the same for both groups, and the property of interest (grammatical gender) can also be found in only one of the previously acquired languages (Spanish; incidentally the L1) in both cases. The L2 Status Factor would predict that L2 English is the default source of transfer across the board, which, for gender, means that the

---

[2] The LAN does not appear in all relevant studies, leading some to argue that when it obtains it is more likely a by-product of averaging across within-group individual differences related to negative-dominant vs. positive-dominant responses to agreement violations (e.g., Guajardo & Wicha, 2014; Tanner, 2019; Tanner & Van Hell, 2014; see also; Grey, Tanner, & Van Hell, 2017; Osterhout, 1997; Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004; but see; Caffarra, Mendoza, & Davidson, 2019). Nevertheless, the P600 is indeed consistently found for agreement violations when group-level grand averages are considered (see Molinaro et al., 2011, for a comprehensive review).

Table 1

Original ERP predictions for the processing of gender violations in Mini-Spanish and Mini-English by L1 Spanish-L2 English speakers, derived from the L2SF, CEM and TPM (adapted from Rothman et al., 2015).

Language combination	L2 Status Factor	CEM	TPM
L1 Spanish-L2 English L3 Mini-English	No effect	(N400)–P600	No effect
L1 Spanish-L2 English L3 Mini-Spanish	No effect	(N400)–P600	(N400)–P600

L3 Mini-Spanish and Mini-English grammars should not transfer a full specification for this property. For this reason, the ERP prediction derived in Rothman et al. (2015) was that neither group should show sensitivity, as reflected in electrophysiological measures, to gender violations. The CEM makes the opposite prediction, given that transferring grammatical gender from Spanish would provide a target-like representation in the L3. Since both groups have the property in their previous linguistic experience, then, they both should show sensitivity to gender violations, reflected in an ERP pattern that is qualitatively (though almost certainly not quantitatively) similar to those elicited by grammatical gender violations in L1 (dominant) Spanish speakers (e.g., Alemán Bañón et al., 2012): a P600, potentially preceded by an N400/LAN (Molinaro, Barber, & Carreiras, 2011).

The TPM makes distinct predictions. Each group is expected to perform differently as a consequence of the artificial mini-grammar to which they are exposed, despite having the same linguistic backgrounds. The TPM claims transfer is conditioned on overall structural similarity, that is, that either the L1 or the L2 grammar is copied in full as the initial L3 interlanguage grammar based on how much structural crossover is estimated between them and the incoming L3 input as early as possible during initial exposure. This is captured by the TPM's processing hierarchy. The parser is argued to use a set of linguistic cues in a particular order to assess the relative similarity between the L3 and each of the previously acquired grammars. This cue hierarchy takes into consideration the differential salience of various elements in the linguistic input, and the relative precedence that some of these elements must logically take over others (e.g., it is simply not possible to parse functional morphology, or even basic word order, without at least some knowledge of the words involved in a given sentence). The first of these levels is the lexicon, followed by phonology/phonotactics, functional morphology and, finally, syntax (see Fig. 1).

In our experiment, the artificial mini-grammars of Mini-Spanish and Mini-English are, in fact, identical in morphosyntactic structure (see the Methods section for details): They have the exact same word order and the exact same morphosyntactic properties of gender and number agreement with the exact same morphophonological exponents. However, their lexical bases are radically dissimilar, as they are entirely based on Spanish and English, respectively. From the perspective of the TPM's processing hierarchy, then, being exposed to Mini-Spanish vs. Mini-English represents a completely different task. The very high percentage of lexical overlap between the L3 and one of the previous languages only (the L1 in the case of Mini-Spanish, the L2 for Mini-English) should bias the selection of a transfer source in favor of the lexically closest language (Spanish or English), irrespective of the presence or absence of grammatical gender in that language—and, crucially, despite it featuring in both artificial mini-grammars. In short, L1 Spanish L2 English speakers exposed to Mini-Spanish should, after exposure training, transfer Spanish as the initial L3 interlanguage, and English should be transferred for those speakers exposed to Mini-English. In terms of ERP responses, Rothman et al. (2015) suggested this should translate into a sensitivity to gender violations in the form of P600-type components (potentially as a biphasic response, preceded by an N400) *only* for the Mini-Spanish group, with L3 Mini-English learners showing no sensitivity to these violations.

3. Method

3.1. Artificial mini-grammars

Two artificial mini-grammars (henceforth ALs [artificial languages], for convention) were created, based on the English (Mini-English) and Spanish (Mini-Spanish) lexicons, respectively (e.g., Grey et al., 2014; Rebuschat & Williams, 2012; Williams & Kuribara, 2008). Each contained 12 nouns (six masculine, six feminine), 12 adjectives, one article (four allomorphs: masculine singular, feminine singular, masculine plural, feminine plural), a copula (two inflected forms: singular and plural), one coordinative conjunction, one adverb and two locatives. Adjectives carried number (singular/plural) and gender (masculine/feminine) morphology in both languages. These morphemes, as well as the article, were common to both ALs, and thus phonotactically plausible in both English and Spanish. Gender and number information was marked on both the article and the adjective, but not on the noun. Gender assignment
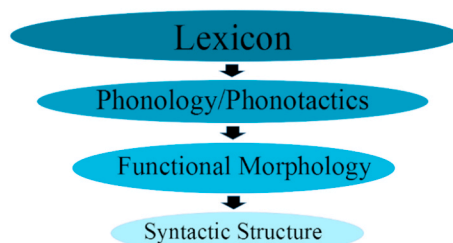
Fig. 1. The TPM's cue hierarchy.

was inherent to each noun, congruent with the gender of the equivalent nouns in Spanish. These were selected so that half of the nouns in each gender display transparent assignment in Spanish (feminine nouns ending in *-a* and masculine nouns ending in *-o*), and half have nontransparent assignment (ending in *-e* or a consonant). Mini-Spanish nouns were stripped of *-a* and *-o* endings, and any original diacritics or exclusive characters (accents, *ñ*) were removed or substituted by approximate characters that were nonexclusive (i.e., shared by the Spanish and English alphabets; e.g., *ñ > n*). Table 2 contains a full list of stimuli in both ALs.

To facilitate implicit training through adjectival contrasts (see below), adjectives consisted of six pairs of opposing qualities (e.g., *old/new*, *expensive/cheap*). The adverb and conjunction were introduced in order to construct spillover regions after the critical violations in the ERP experiment. The locatives were used in filler sentences, to reduce the predictability of an adjective continuation after the copula.

Number morphology consisted of an *-r* affix to mark plural adjectives and articles, akin to Spanish and English *-s*. Gender morphology was built on a consonantal contrast, *(−)j-* (feminine) vs. *(−)z-* (masculine), carefully selected to avoid consonants with strong associations to functional morphology in either source language (e.g., *n-*as a recurrent segment in negation-related morphology). The use of a consonantal instead of a vowel contrast was necessary to achieve this same objective (all vowels have associations to one or several functional morphemes in Spanish) as well as to reduce the opportunity for potential misleading rhyming strategies with the *-e* endings of some of the nouns. Allomorphs of the article and the inflectional affix on the adjective had the following structure: In the adjective, the distinctive consonant (*j* or *z*) was invariably followed by the letter *-e*. In the adjective, the epenthetic vowel *-e-* preceded the distinctive consonant, which was invariably followed by the letter *-u*. The different vocalic endings on the article and the adjective were designed to avoid the development of rhyming strategies in the learning of the ALs' gender morphology.

To isolate the lexical similarity factor in accordance with the implicational hierarchy of the TPM, all typological differences between Spanish and English that may have been potentially inherited by Mini-Spanish and Mini-English were methodologically neutralized. First, our study employed visual (as opposed to auditory) training and presentation of stimuli, in order to remove all possible phonological bias. Second, gender and number morphology were common to both ALs and compliant with each of the source languages' phonotactics. Finally, the experiment examined morphosyntactic violations on the marking of predicative adjectives following a copulative verb, a structure in which English and Spanish word order coincides (as opposed to the pre-/postnominal contrast displayed by most attributive adjectives in these two languages). These manipulations ensured that the only difference between Mini-Spanish and Mini-English lay on their lexical base, while both languages instantiated gender agreement.

3.2. AL training

A pre-training phase familiarized participants with the nouns, by presenting a randomized list of picture-phrase combinations containing two instances of each noun in simple article-noun phrases (e.g., *je window*). This phase thus introduced both the lexical material and the nouns' lexical gender.

The critical items in the ensuing training session exploited adjectival contrasts to provide implicit exposure to number and gender morphology through meaningful sentences in the AL without metalinguistic explanations. These sentences were presented as

Table 2
Full AL design and vocabulary items, and example sentence (feminine singular).

MINI-SPANISH				MINI-ENGLISH			
Nouns				**Nouns**			
Feminine		Masculine		Feminine		Masculine	
mochil, taz, ventan, pared, llave, calle.		*cuchil, gor, roper, camion, reloj, lapiz*		*bag, cup, window, wall, key, street*		*knife, hat, closet, truck, watch, pencil*	
Adjectives				**Adjectives**			
amarill-, roj-, pequen-, grand-, nov-, vej-, suci-, limpi-, barat-, car-, cort-, larg-				*yellow-, red-, small-, big-, new-, old-, dirty-, clean-, cheap-, expensiv-, short-, long-*			
Inflectional affixes				**Inflectional affixes**			
Feminine		Masculine		Feminine		Masculine	
Singular	Plural	Singular	Plural	Singular	Plural	Singular	Plural
-eju	*-ejur*	*-ezu*	*-ezur*	*-eju*	*-ejur*	*-ezu*	*-ezur*
Article				**Article**			
Feminine		Masculine		Feminine		Masculine	
Singular	Plural	Singular	Plural	Singular	Plural	Singular	Plural
je	*jer*	*ze*	*zer*	*je*	*Jer*	*ze*	*zer*
Copula				**Copula**			
Singular		Plural		Singular		Plural	
es		*son*		*Is*		*are*	
Conjunction				**Conjunction**			
Y				*And*			
Adverb				**Adverb**			
Tambien				*Too*			
Locatives				**Locatives**			
arriba, abajo				*above, below*			
Example sentence				**Example sentence**			
Je mochil es barategu.				*Je bag is cheapegu.*			
"The bag is cheap."				"The bag is cheap."			

descriptions of pictures showing the object and property denoted by the noun and the adjectives, respectively (see Fig. 2 for an example). The training session lasted approximately 45 min.

In order to maintain an implicit focus on the morphology, none of the critical items in the training session contained either number or gender contrasts, relying instead exclusively on the semantic differences encoded in the adjectives. That is, an image depicting a masculine or feminine plural object was always contrasted to an image depicting the same object in the plural—and likewise with singular items. To encourage mapping of the prenominal element to the intended functional categories (i.e., singular and plural articles instead of numeral determiners), half of the images depicting plural nouns contained two objects, and the other half contained three. The training contained a total of 144 critical items, resulting from crossing the 12 nouns and 12 adjectives. Two different lists were created to counterbalance the presence of a given noun-adjective combination in singular or plural form. The training stimuli were completed with 144 fillers—for a total of 288 items—using the locative prepositions (e.g., *The bag is below the hat*).

At the end of the training, the participant's command of the AL was tested through a multiple-choice sentence-picture matching task. Upon presentation of a single picture (e.g., three cheap bags), the participant was asked to select the sentence that best described the image among five alternatives: correct (e.g., Mini-Spanish *Jer mochil son baratejur*), a sentence containing a gender violation (*Jer mochil son baratezur*), a sentence containing a number violation (*Jer mochil son barateju*), a sentence containing a double violation (*Jer mochil son baratezu*) or a sentence containing only a semantic violation (that is, the opposite adjective with the right morphology, e.g., #*Jer mochil son carejur*). The test consisted of a total of 36 items, in which knowledge of all possible gender and number morphology was assessed.

3.3. Participants

A group of 60 Spanish native speakers (52 female), L2 speakers of English, were recruited as participants from three different areas of South-East England (Southampton, London, Reading) and tested at the University of Reading. Completion of a short questionnaire ensured that they met the study's language background requirements (Spanish as a mother tongue, English as a late-acquired L2). The age range was 18–51 (mean = 28.5; $SD = 8.61$). Their average proficiency in English, as measured by the *Oxford Quick Placement Test* (Oxford University Press et al., 2001) was 40.1 for the Mini English group ($SD = 11.08$; 95% CI = 4.63) and 35.7 ($SD = 8.32$; 95% CI = 3.74) for the Mini Spanish group, thus placing them between lower intermediate and advanced at the individual level. Importantly, the groups did not differ in their English proficiency ($t(43) = 1.60$; $p = 0.12$), and all participants were proficient enough to have acquired English number agreement, the only L2 grammatical property involved in the design. Average age of acquisition of English was 11.08 (range: 0–26; $SD = 7.92$). Participants were randomly assigned to one of the two AL groups, which meant that they completed the training and ERP experiment (if eligible, see below) exclusively in that language (i.e., they are never exposed to the other AL). Therefore, all participants had Spanish as their L1, English as their L2, and either Mini-English or Mini-Spanish as their L3. All participants were neurotypical adults with normal or corrected-to-normal vision, and were right-handed according to the Edinburgh Inventory for the Assessment of Handedness (Oldfield, 1971). Participants were compensated for their time.

3.4. Procedure

Upon arrival, participants were informed of the general procedure of the study, provided written consent, and completed a short questionnaire that gathered demographic and language background data. Next, they were assigned to an AL group and seated comfortably in a quiet room, in front of a 32-inch monitor. After being exposed to the pre-training and training phases described above, the participants were tested through the sentence-picture matching task. If they scored above 80% in this task (equivalent to 29 correct responses out of 36 total), they moved on to the ERP experiment. Otherwise, they were re-exposed to the training. Failure to reach the 80% accuracy threshold on a second take of the test resulted in the participant's removal from the study. This resulted in the exclusion of less than 10% of the original participants (N = 4).

After the ERP experiment was completed, participants performed a gender assignment task (GAT) in Spanish that contained the 12 nouns which served as the lexical base in Mini-Spanish (and were translated as the 12 critical nouns in Mini-English) together with 108 fillers, for a total of 120 items. This was done to ensure that gender assignment in our participants' lexical representations of these nouns matched the intended values. In the task, each noun was presented in isolation on the center of the screen, and participants were asked to respond, by pressing one of two buttons on the mouse, matching the noun to the appropriate determiners. A short set of six trials familiarized participants with the procedure before the critical and filler items were presented in a single block of trials. Accuracy and response times (RTs) were collected for each response, although there was no timeout in the GAT and participants were encouraged to favor accuracy over speed. As expected (given that these were all native speakers of Peninsular Spanish), results showed full alignment across the board with the expected gender assignment for the critical nouns.

3.4.1. ERP experiment

In the ERP experiment, participants read sentences on a computer screen while electrophysiological activity was recorded at the scalp. Sentences were presented word by word, employing the rapid serial visual presentation (RSVP) method. Each word appeared on screen for 450 ms, with an inter-stimulus interval of 250 ms. After the last word of the sentence, a prompt with a happy and a sad face on each side of the screen indicated that a grammaticality judgement was required. Using their right hand index and middle fingers, participants responded by pressing the left or right button on a mouse. While response times were recorded as an additional behavioral measure, there was no time limit on the grammaticality judgement.

The ERP experiment contained the same 144 critical sentences as the training, resulting from crossing the 12 nouns and 12

Jer mochil son carejur   *Jer mochil son baratejur*

Fig. 2. Example of an item in the Mini-Spanish training session: Item showing N-ADJ agreement between a FEM-PL noun and the predicative adjective.

adjectives. Three versions of each sentence were created: grammatical, gender violation and number violation. No double violations (i. e., gender *and* number) were ever present in the stimuli. Table 3 contains a sample of the conditions in the Mini-English and Mini-Spanish experiments, up to the critical word.

Violations were always on the adjective, which represented the critical word in our analyses. All critical sentences contained a final region consisting of the conjunction (*and/y*) plus another determiner phrase (DP) of the same gender and number as the sentence's initial DP and the adverb *too/tambien* (e.g., *je bag is cheapegu and je cup too*; Mini-English, grammatical) (Alemán Bañón et al., 2012, 2014, 2018). This extra material was introduced to prevent any potentially confounding wrap-up effects from interfering with ERP responses elicited by the critical word.

Each participant saw a total of 240 sentences: 96 critical sentences (48 gender violation, 48 grammatical) plus 144 fillers (48 number violations and 96 grammatical fillers). The grammatical fillers followed the same structure as those in the training session (DP + copula + locative + DP), with a final region akin to that of critical sentences (e.g., *Zer truck are below zer closet and zer watch too*). Different experimental lists counterbalanced the conditions in which the different noun-adjective combinations appeared, as well as the number of times in which each initial DP was used in the singular or plural form. The 240 trials were divided into six blocks of 40. Participants were encouraged to rest their eyes and take as long as they needed before continuing with the experiment.

3.4.2. EEG recording and analysis

The data from six participants was lost due to technical problems. Together with those excluded for not having reached the accuracy threshold in the sentence-picture matching task, this left a total of 50 participants: 24 in the Mini English group, 26 in the Mini Spanish group.

The EEG signal was continuously recorded from the scalp by means of 64 active electrodes (ActiCap, Brain Products, Inc.) fitted in an elastic cap in a 10–20 system. AFz served as the ground electrode. EEG recording was referenced online to electrode FCz, and re-referenced offline to the average mastoids (TP9/10). The fronto-parietal electrodes FP1 and FP2, located above the eyebrows, were used to monitor eye blinks. Impedance was kept below 10 kΩs for all electrodes. The recordings were amplified by a BrainAmp MR Plus amplifier (Brain Products, Inc.) with a bandpass filter of 0.01–200 Hz, and digitized continuously at a sampling rate of 1 kHz.

Preprocessing of the EEG data was performed on Brain Vision Analyzer 2.0 (Brain Products, Inc.). We used all trials, regardless of accuracy in the GJT (e.g., Mickan & Lemhöfer, 2020; Tanner et al., 2013; VanRullen, 2011). First, data were filtered with a band-pass filter of 0.1–30 Hz. The continuous EEG was then segmented into 1500 ms epochs with reference to the critical word (adjective). Epochs contained a 300 ms pre-stimulus baseline and ended 1200 ms after stimulus onset. Trials were manually inspected for artifacts (drifts, excessive muscle artifact, blinks, blocking, etc.). Rejecting trials with artifacts resulted in the exclusion of 43.3% trials in the Mini-English data, and 24.7% trials in the Mini-Spanish data.[3] The average number of trials kept did not differ by condition within each AL (Mini English, Number: 26, Gender: 27, Grammatical: 29; Mini-Spanish, all conditions: 36; all *ps* > .1). The remaining epochs were baseline-corrected relative to the 300 ms pre-stimulus baseline and averaged by condition for each participant.

Based on the previous literature on the processing of agreement features in nonnative speakers, ERPs were identified by measuring mean amplitudes in three time windows of interest: 200–500 ms, roughly corresponding to the typical time window for components such as the N400 or the LAN (e.g., Kutas & Hillyard, 1980; Molinaro et al., 2015), and 300–600 ms and 400–900 ms, where positive components such as the P600 tend to occur (e.g., Alemán Bañón et al., 2012). Analyses were conducted on a subset of the electrodes based on nine regions of interest (ROIs; left anterior: F1/3/5, FC1/3/5; right anterior: F2/4/6, FC2/4/6; left medial: C1/3/5, CP1/3/5; right medial: C2/4/6, CP2/4/6; left posterior: P1/3/5/7, PO3/7; right posterior: P2/4/6/8, PO4/8; midline anterior: FCz, Fz; midline medial: Cz, CPz; midline posterior: Pz, POz) (e.g., Alemán Bañón, Miller, & Rothman, 2017; Alemán Bañón & Rothman, 2016, 2019).

---

[3] Using a higher number of items per condition (48) than is usual in this type of studies (e.g., in their review of agreement processing studies, Molinaro et al., 2011, report use of as low as 30 items per condition and recommend using at least 40), allowed us to preserve an appropriate number of trials despite the relatively high rates of trial rejection.

Table 3
Sample stimuli for the experimental conditions in the Mini-English and Mini-Spanish ERP experiments. Examples do not include the final region.

		Condition		
		Grammatical	Number violation	Gender violation
AL	**Mini-English**	je bag is cheapegu	*je bag is cheapegu<u>r</u>	je bag is cheape<u>z</u>u
	Mini-Spanish	je mochil es barategu	*je mochil es barategu<u>r</u>	*je mochil es barate<u>z</u>u
		"The bag is cheap"		

Table 4
Mean accuracy (in percentage) and RTs (in milliseconds), with standard deviations and 95% confidence intervals, for each condition in the two ALs.

		Condition	Accuracy (%)			Response times (ms)		
			Mean	SD	CI	Mean	SD	CI
AL	**Mini English**	Grammatical	82.5	38	2.15	1657	2236	127
		Gender violation	68.8	46.3	2.62	1728	2821	160
	Mini Spanish	Grammatical	79.1	40.6	2.22	1636	2139	117
		Gender violation	68.2	46.5	2.54	1592	2148	117

The mean amplitudes for electrodes in these regions were entered as dependent variables into factorial repeated-measures ANOVAs for each time window of interest. Separate analyses were carried out for lateral and midline electrodes (e.g. Alemán Bañón & Rothman, 2016; Barber & Carreiras, 2005; Gillon Dowens, Vergara, Barber, & Carreiras, 2010). ANOVAs for the lateral electrodes included Condition (grammatical, gender violation), Hemisphere (left, right) and Caudality (anterior, medial, posterior) as predictors, whereas the models for the midline included only Condition and Caudality. In those cases in which the assumption of sphericity was violated (as determined by Mauchly's test) for a given main effect or interaction, we report adjusted degrees of freedom and *p*-values following from corrections based on Huynh-Feldt epsilon estimates. While this correction is less conservative than Greenhouse-Geisser estimates, it is particularly appropriate for data of this kind, where the nature of the study (*ab initio* learners, artificial mini-grammars) is likely to yield relatively large between- and within-subject variability. Follow-up analyses on any significant interactions were conducted by means of one-way repeated-measures ANOVAs on the appropriate subsets of the data.

All analyses of both the behavioural and the ERP data were conducted in the statistical software *R* (R Core Team, 2019), using the packages *lme4* (Bates, Maechler, Bolker & Walker, 2015; for linear mixed models), *multcomp* (Hothorn, Bretz, & Westfall, 2008; for post hoc comparisons of effects in the linear mixed models), and *ez* (Lawrence, 2016; for repeated-measures ANOVAs).

4. Results

4.1. Behavioural data

Accuracy and response time (RT) data were collected for the participant's responses to each grammaticality judgement. Table 4 provides a summary of descriptive statistics. After removing the data from filler sentences, the raw accuracy scores and RTs were submitted to statistical analysis.

We fit a generalized linear mixed model (binomial family) on the accuracy data, with condition (grammatical, gender) and AL as fixed factors (main effects and interaction), and the maximal random structure justified by design (Barr, Levy, Scheepers, & Tily, 2013). This included random intercepts for subjects and items (in our case, the critical adjective) as well as random slopes for condition within subjects and within items. There were no main effects or interactions of AL (all $ps > .4$). Overall, responses to gender violations were only numerically less accurate ($\beta = 0.71$; $z = 1.93$; $p = .053$).

For the RT data, we fit generalized linear mixed models with an inverse Gaussian family. We began by fitting a maximal model that included condition and AL as fixed factors (main effects and interaction), as well as random intercepts for subjects and items and random slopes for condition within both. We proceeded to decorrelate random slopes and intercepts and to remove random slopes accounting for the least amount of variance until a model achieved convergence. This final model included random intercepts for subjects and items. There were no main effects or interactions of either AL or condition (all $ps > .4$). In sum, while the behavioral data show certain numerical trends, with overall higher accuracy in grammatical sentences across both ALs, there were no significant differences in accuracy or RTs as a function of condition in either AL group.

4.2. ERP data

We report here only main effects or interactions with the Condition factor (that is, we leave out main effects or interactions which involve exclusively Caudality and Hemisphere). The full output of the omnibus ANOVA for each time window in each AL can be found in Appendix A (Tables A1 to A12).

4.2.1. Mini-English
Topographical maps and grand averaged ERP waveforms for the Mini-English group can be seen in Figs. 3 and 4.
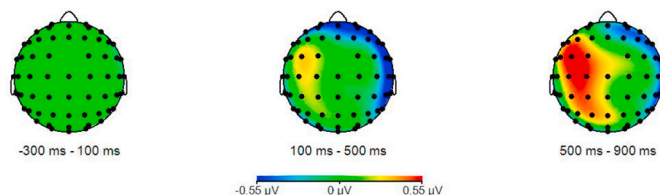
Fig. 3. Topographical distribution maps of the difference wave for the gender-grammatical condition contrast in the Mini-English experiment.
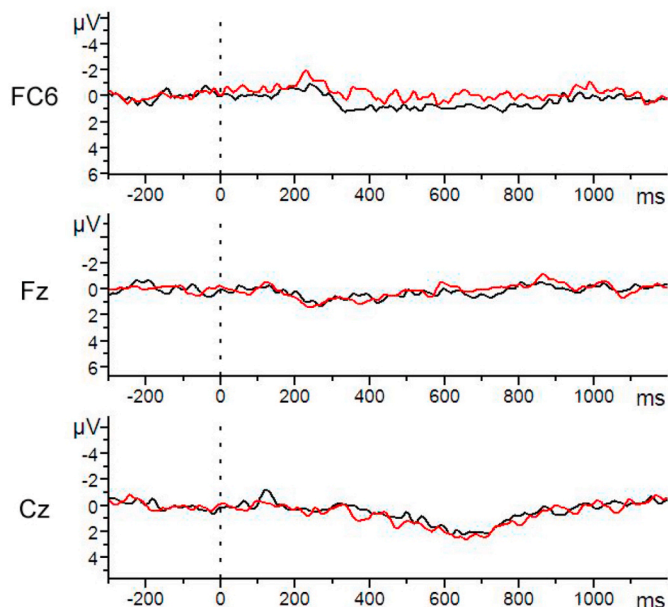

Fig. 4. Grand averaged ERP waveforms for the grammatical (black) and gender (red) conditions in Mini-English at electrodes, FC6 (right anterior), Fz (midline anterior) and Cz (midline medial). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.2.1.1. 200–500 ms. In the lateral electrodes, the omnibus ANOVA did not reveal a main effect of Condition ($F(1, 23) = 0.02, p = .88$) for gender violations in Mini-English in the 200–500 ms time window. The analysis yielded a significant three-way interaction between Condition, Hemisphere and Caudality ($F(1.41, 32.48) = 3.85, p < .05, \eta^2 = 0.002$). Follow-up analyses revealed that the effect of Condition was significant at right-anterior electrode sites only ($F(1, 286) = 5.07, p < .05$; left-anterior: $F(1, 286) = 1.89, p = .17$; left-medial: $F(1, 286) = 2.00, p = 0.16$; right-medial: $F(1, 286) = 0.95, p = .33$; left-posterior: $F(1, 286) = 0.05, p = .83$; right-posterior: $F(1, 286) = 0.01, p = .92$), indicating more negative voltages in response to gender violations as compared to grammatical sentences.

The analysis of the midline electrodes did not reveal a significant main effect of Condition ($F(1, 23) = 0.004, p = .95$) nor any significant interactions with this factor.

4.2.1.2. 300–600 ms. No significant main effects of Condition were found in the lateral ($F(1, 23) = 0.001, p = .98$) or midline electrode sites ($F(1, 23) = 24.35, p = .1$) in the 300–600 ms time window. Likewise, there were no significant interactions of Condition with either of the topographical factors.

4.2.1.3. 400–900 ms. As in the 300–600 ms time window, there were no main effects (lateral: $F(1, 23) = 0.06; p = .82$; midline: $F(1, 23) = 0.13; p = .72$) or interactions with Condition, indicating that gender violations in Mini-English did not elicit a different response from grammatical sentences in the 400–900 ms time window.

4.2.2. Mini-Spanish

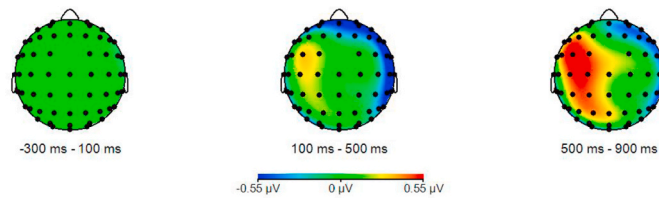Topographical maps and grand averaged ERP waveforms for the Mini-Spanish group can be seen in Figs. 5 and 6.

Fig. 5. Topographical distribution maps of the difference wave for the gender-grammatical condition contrast in the Mini-Spanish experiment.
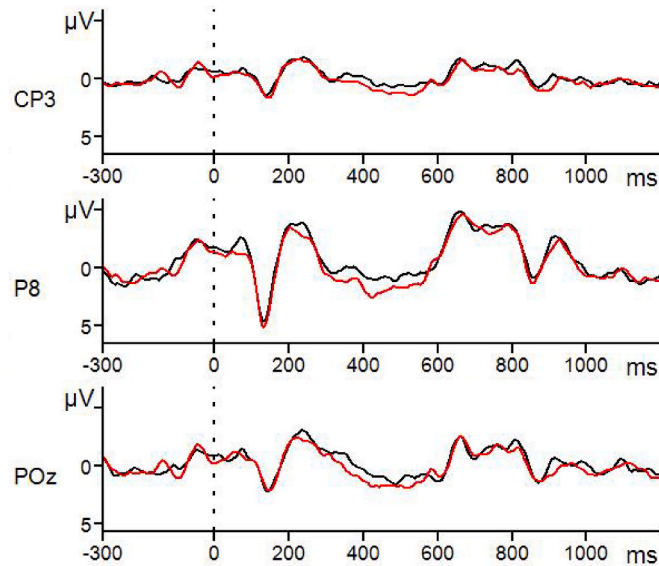

Fig. 6. Grand averaged ERP waveforms for the grammatical (black) and gender (red) conditions in Mini-Spanish at electrodes CP3 (left medial), P8 (right posterior) and POz (midline posterior). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.2.2.1. 200–500 ms. Figure 5 and 6. In lateral electrodes, the omnibus ANOVA performed on the data corresponding to the 200–500 ms time window in the Mini-Spanish experiment returned a marginally significant main effect of Condition ($F(1, 25) = 4.00, p = .057, \eta^2 = 0.01$), indicating that scalp voltages were more positive in the gender violation condition as compared to grammatical sentences. This effect appeared to be broadly distributed, as all two-way and three-way interactions with the topographical factors were nonsignificant (Condition by Hemisphere: $F(1, 25) = 0.76, p = .39$; Condition by Caudality: $F(1.39, 34.85) = 1.67, p = .21$; Condition by Hemisphere by Caudality: $F(1.49, 37.3) = 0.41, p = .61$).

This marginally significant main effect of Condition was also found in the analysis of the midline electrodes ($F(1, 25) = 3.96, p = .058, \eta^2 = 0.01$). The two-way interaction between Condition and Caudality was not significant here either ($F(2, 50) = 1.67, p = .2$), indicating that the increase in positive voltage in gender violations as compared to grammatical sentences was distributed evenly along the midline.

4.2.2.2. 300–600 ms. Gender violations in Mini-Spanish elicited significantly more positive voltages over all electrode regions in the 300–600 ms time window, as indicated by the main effect of Condition ($F(1, 25) = 4.61, p < .05, \eta^2 = 0.01$), which was not qualified by any interactions with Hemisphere ($F(1, 25) = 1.40, p = .25$) or Caudality ($F(1.57, 39.4) = 1.97, p = .16$).

The analysis of the midline electrodes yielded similar results, with a significant main effect of Condition ($F(1, 25) = 4.25, p < .05, \eta^2 = 0.01$) which did not interact with Caudality ($F(2, 50) = 1.71, p .19$), indicating that gender violations elicited more positive responses across the midline, as compared to grammatical sentences.

4.2.2.3. 400–900 ms. In lateral electrodes, no significant main effect of condition was returned by the omnibus ANOVA performed on the data from the 400–900 ms time window ($F(1, 25) = 1.16, p = .29$). Likewise, there were no interactions with the Hemisphere or Caudality factors.

No significant main effect of Condition ($F(1, 25) = 0.72, p = .4$) nor an interaction with Caudality were found in the analysis of the midline electrodes.

4.2.3. Summary of ERP results

Gender violations elicited a fronto-lateral negativity in Mini-English in the earliest time window (200–500 ms), although this effect was not followed by any other differences in subsequent periods. This effect was highly localized, surfacing only in electrodes of the right-anterior region. In contrast, gender violations in Mini-Spanish elicited a broadly distributed positivity that starts trending in the 200–500 ms time window, emerges significantly between 300 and 600 ms, and dissipates in the later period.

5. Discussion and conclusion

The results described above present an interesting scenario; one that does not conform to any expectations in light of current theories, yet captures a difference between how the AL groups perform. It is thus important to first acknowledge that these results are unexpected and discuss how they deviate from the predictions above, and then proceed to engage with the data themselves in order to interpret what it is they tell us about the earliest stages of L3 processing. Thus, related to our original research questions, we note that the data are rather limited with respect to adjudication between the L3/*Ln* models we originally sought to test. Nevertheless, they offer rich insights as it pertains to the perennial question of what is happening at the very beginning of L3 exposure, thus potentially addressing what factors feed into eventual transfer selection itself, the process as opposed to its outcome.

5.1. Revisiting the original predictions

Put simply, the ERP predictions derived above from the TPM, the CEM, and the L2SF were not substantiated by the present data. Recall that, as an extension of their different claims regarding the (under)specification of initial L3 interlanguage representations, we stated that (i) the CEM would predict both groups to show sensitivity to gender violations that is qualitatively similar to that of native speakers (a P600 or biphasic N400–P600 response), because both would transfer the linguistic representation for gender in Spanish; (ii) the L2SF would predict that both groups should lack this sensitivity and therefore show no effects in response to gender violations, because they would always transfer from their L2 (English in both cases); and (iii) the TPM would predict only those participants exposed to Mini-Spanish to show sensitivity (in the form of a P600 or a biphasic N400–P600 response) to gender violations, since the model predicts that English should be transferred as the initial interlanguage grammar for L3 Mini-English—thus lacking this property—and, conversely, a copy of the Spanish grammar should become the initial L3 interlanguage grammar for Mini-Spanish—thus instantiating gender agreement (see Table 1 for ERP predictions).

A quick glance at the results shows that the data did not conform to the scenarios described in (i)-(iii). The early, broadly distributed positivity in the Mini-Spanish group does not resemble the P600 that should be associated with the transfer predictions of the CEM or the TPM, and neither does it validate the no-effects prediction of the L2SF.[4] At best, the effects of the Mini-Spanish group could be consistent with Steinhauer, White and Drury's (2009) claim that non-native learners will initially show a broad P600 (across the scalp), which then becomes focalized. We will return to this below. Similarly, the early right anterior negativity displayed by the Mini-English group was not anticipated by any of the models. Importantly, we know of no current theory, including those not addressed in detail here (e.g., the LPM), that would have predicted the current pattern of results. It is important to keep in mind, however, that all predictions were predicated on the assumption that testing would have happened after *transfer* would have taken place. Thus, the onus was on examining the *outcome* and using it for evidential adjudication. However, if indeed the current methodology did not allow sufficient time (from initial exposure to testing) to get at the outcome stage, as we will argue occurred below, then the data are not only still valid but potentially crucial, albeit at a different level: examining the process of how selection is conditioned in the first place.

5.2. ERP indices of pre-transfer processing?

Although other mini grammar studies have claimed to show evidence of transfer after minimal exposure, this has been the case in L2 studies where there is no competition between more than one system. In L3 acquisition, by comparison, it proved in our case to be problematic to assume that transfer selection among choices would take place in a similar time course. No transfer model offers a rubric of timing for selection—even the most articulated current models of L3/*Ln* acquisition offer only indirect estimations of the timing of transfer. The TPM's cue hierarchy, for example, provides an implicit rubric of timing that is ultimately dependent on overall L1/L2 to L3 similarity (typology, in a broad sense): moving through the levels of the hierarchy requires increasing exposure/time, which means that transfer will inevitably be delayed in language triads where the L3 bears no immediately apparent resemblance to the L1 or the L2 in, at least, vocabulary and/or segmental and suprasegmental phonology.

It follows from the above reasoning that the inverse relation should also hold: (full) transfer should be accelerated in language triads where the L3 has a strong similarity at the lexical level with either the L1 or the L2. This was in fact the rationale for our design, where the ALs hide identical underlying grammars under a heavily Spanish- or English-biased lexicon. Given this similarity, which is only disrupted by the introduction of novel agreement morphology, we *should* have been able to tap into an early L3 interlanguage grammar after relatively brief AL exposure. However, there was no guarantee that our training phase, which consisted of about 45 min

---

[4] Even under an L2SF account where the Declarative Procedural (DP) model is argued to underlie the basis of L2 default selection (see Bardel & Falk, 2012), predictions are not supported. Under such an account, which assumes that L2/*Ln* grammatical representation is subserved by declarative memory, one might predict N400s for gender errors (as the DP model takes modulations of this component to reflect the involvement of the declarative memory system; Ullman, 2016), but this is not seen in either group.

under implicit learning conditions, would not fall short of the actual amount of exposure required to trigger any kind of representational transfer from a previous language. Given this possibility, we must consider an additional scenario with a corresponding research question, extending the original scope of Rothman et al. (2015): *What should we expect to observe in terms of brain evidence, according to the different models of transfer in L3/Ln acquisition, if transfer has not (yet, or at all) taken place? Do they differ at all?* We think so.

Framing the above in the context of the current study is not an easy task, as none of the transfer models has explicitly dealt with these questions in the published record. Nonetheless, there are certain predictions that must ultimately follow from the different conceptual frameworks of these theories. In particular, default models of L3/*Ln* transfer, for which order of acquisition (i.e., L1 vs. L2) is the most deterministic factor in transfer source selectivity, have grounded this position on the basis of processing arguments related to a relative prominence of the L1 or L2 as a first-pass filter to the incoming L3 input. The L2 Status Factor, in particular (see Bardel & Falk, 2012), assumes Paradis' (2009) Declarative/Procedural (DP) model, where the L1 grammar is argued to be subserved by procedural memory while all other linguistic material, including L1/*Ln* lexicon and L2/*Ln* grammars are subserved by declarative memory systems. In other words, the L2 Status Factor attributes the L2's purportedly dominant influence on L3 grammar building to the fact that any subsequent grammar to those acquired natively is dealt with by the same neurocognitive subcomponents, in all cases distinct from the L1. Since the L2 should be guiding all handling of the L3 input, it seems reasonable to assume that the L2SF would expect electrophysiological activity reflecting an English bias in both groups.

The TPM, on the other hand, does not envision a predetermined, input-independent privileged role for either previous language upon initial exposure to an L3/*Ln*. On the contrary, the mechanics of the cue hierarchy and the principled approach to a best-guess estimation of overall structural similarity between the L3 and previously acquired languages have been explicitly argued to capture a "sink or swim" reaction to the task of parsing the incoming L3 input (see Chapter 4 in Rothman et al., 2019, for a comprehensive discussion). In this sense, the specific linguistic properties of the input are crucial: any property of the input that allows the processor to adjudicate between potentially conflicting parsing strategies should be capitalized on. Lexical similarity is likely to be a powerful trigger—and thus stands at the top of the TPM processing hierarchy. In our current design, this should translate into the English-like vocabulary of Mini-English biasing the parser in favor of processing strategies that focus resources on those aspects of sentences that are critical to the processing of English, and the Spanish-like vocabulary of Mini-Spanish biasing the allocation of processing resources in ways that are necessary when parsing Spanish sentences. While there is little theoretical apparatus to base conjectures on for other non-default models, such as the CEM or the LPM, we assume here that a similar logic would apply to them in these very early stages.

We believe that the current results are more in line with the latter type of accounts. The first, and perhaps most obvious, reason is that default models based on order of acquisition, irrespective of whether they envision a privileged role of the L1 or the L2, seemingly preclude the possibility of different behavior across our two groups, which are identical populations exposed to different L3 ALs. The second is that these divergent responses to gender agreement violations in Mini-Spanish and Mini-English are expressed through differences in ERP patterns that suggest our participant groups' attentional resources were differentially allocated, and this affected their perception of the critical manipulation in our experiment. In particular, we submit that the early positivity in the Mini-Spanish group's response to gender violations is reminiscent of the P300 (or P3), a well-known ERP component that is thought to reflect the recruitment of attentional and memory resources (e.g., Polich, 2012). The P300 is a (family of) positive going wave(s), elicited by both auditory and visual stimuli that are motivationally significant (Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). Its latency and amplitude vary greatly as a function of different manipulations in stimulus frequency and task-relevance (e.g., Luck, 2014; Polich, 2012; Woodman, 2010), and different subcomponents have different topographical maxima (frontal and parietal being the most common; Luck, 2014). While first observed within the oddball paradigm (Squires, Squires, & Hillyard, 1975), where a series of frequent stimuli is occasionally interrupted by an infrequent deviant stimulus (see, e.g., Donchin, 1979, 1981, for early reviews), it was quickly established that the P300 could not be indexing an estimation of absolute frequency, as it was elicited by task-relevant stimuli even when the frequency distributions of different stimulus categories were balanced in a given experiment (e.g., Duncan-Johnson & Donchin, 1977). In our study, the ratio of deviant (violations) to frequent stimuli (grammatical sentences) can be considered as low as 1:4 or as high as 2:3, depending on what is counted as a distinct stimulus category—i.e., depending on whether number violations, fillers here, are included in the deviant stimulus count. In any case, and since participants were explicitly asked to evaluate the grammaticality of these sentences, violations can be considered task-relevant stimuli. It is worth noting that our effect is broadly distributed and thus does not seem to present either a frontal or a parietal maximum, but it is difficult to ascertain with the present data whether this could be related to an averaging effect of different overlapping subcomponents—whereby the nature of our study (artificial mini-grammars, first exposure) may have increased the weight of anterior effects, typically linked to novelty (e.g., Squires et al., 1975; Verleger, Jaskowski, & Wauschkuhn, 1994), counterbalancing the parietal bias present in most P300 studies employing a standard oddball paradigm.

As pointed out by Sassenhagen et al. (2014:30), the fact that the P300 is sensitive to deviations that are not infrequent suggests that the component reflects subjective rather than inherent salience. In fact, some studies have found that unattended stimuli do not elicit a P300 (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Spencer, Dien, & Donchin, 2001)—not even the frontally maximal P3a subcomponent associated with task-nonrelevant stimuli ("ignore" stimuli) in the seminal paper by Squires et al. (1975). Since our Mini-Spanish participants show this positivity and our Mini-English participants do not, it is reasonable to posit that only the former are attending to the relevant properties of the deviant stimuli—i.e., the gender violations—and that they do so under the influence of a larger focus on word-final gender morphology that a Spanish bias should induce.

While our focus here has been on gender, precisely because this is the grammatical feature not shared across Spanish and English, if our hypothesis is on the right track converging evidence from the number condition is relevant. While both English and Spanish have syntactic number, morphological number agreement on the adjective—herein we targeted both gender and number in the EEG

recording for both mini-grammars on the adjective—only pertains to Spanish. And so, it is not clear what the effect of having number as a syntactic feature alone, as is the case in English, would translate to under our current hypothesis; however, the insight is that it would not necessarily matter because there is no word final number morphology on English adjectives. For the case of the Spanish mini-grammar group we would expect a similar P3 effect as was shown for gender. Indeed, visual inspection of the data from the Mini-Spanish group shows a broadly distributed early positivity (300–500 ms), which is followed by a later more focused positivity over typical P600 sites, although occurring later than the typical P600 time window (starting at 700 ms). This contrasts with the mini-English group, which shows a late negativity (500–900 ms) through canonical N400 sites, though a positivity emerges around 300–500 ms in frontal sites—dissipating before 700 ms.

Before delving deeper into the nature of the positive component in the Mini-Spanish group, we should devote some attention to the negativity found in the Mini-English results. Indeed, besides the absence of the positivity that we interpret as a P300 in the other group, responses to gender violations in the Mini-English group seem to be characterized by a very localized right anterior negativity in the 200–500 ms time window. There are no precedents, to the best of our knowledge, of a comparable component in language-related studies. Music processing research has identified a replicable ERP known as the ERAN (early right anterior negativity), reliably elicited by structural violations such as disruptions of musical scale (e.g., Bharucha & Krumhansl, 1983; Koelsch, Schmidt, & Kansok, 2002; Leino, Brattico, Tervaniemi, & Vuust, 2007; see Koelsch, 2009, for review), normally in conjunction with later components. The ERAN was only reported in one study employing concurrent musical and linguistic stimuli, and there the presence of syntactic violations in conjunction with harmonic expectancy violations seemed to *reduce* the amplitude of the component (Steinbeis & Koelsch, 2007). In linguistic research, Hahne and colleagues have observed right anterior negativities in nonnative speakers (e.g., Hahne, 2001; Hahne & Friederici, 2001), but these obtain (i) in auditory experiments; (ii) in later time windows (e.g., following an N400), and (iii) in response to distinctly *semantic* violations. It might be tempting to think that, in comparison with the potentially attention-related positive component shown by the Mini-Spanish group, the early right anterior negativity of the Mini-English participants reflects an automatic detection of the gender violation that is otherwise unattended. However, even in studies that report an early negativity in response to unattended structural violations, this is either markedly left-dominant (Batterink & Neville, 2013) or broadly distributed (Hasting & Kotz, 2008), in which case it has a significantly earlier onset (100 ms vs. 200 ms) and shorter latency (200 ms vs. ~350 ms) than the early right anterior negativity present in our data. We are, in short, unable to find a functionally credible account of the Mini-English negativity (if it is indeed a real effect), although we believe that there are sufficient reasons to dismiss the possibility that it indexes any attentional or anomaly-detection processes. Crucially, however, it is of significant value to point out that whatever underlies the difference between the groups, their brainwave patterns are distinct despite being equal in all other relevant ways concerning language experience and other profile factors. Thus, this effect must have something to do with the discrepancy in the stimuli they were given, which we controlled to be at the lexical level in relation to one or the other language (Spanish or English) they have matched knowledge on. This means that the language to which one is exposed in L3/Ln acquisition matters for determining how one initially parses/processes it above and beyond order of acquisition of previous languages.

If the positive component in the Mini-Spanish group is indeed an instance of the P300, it seems meaningful to ask whether this could be a precursor of more identifiable language-related responses at later stages—i.e., after longer exposure to the L3 ALs. We have no verifiable way to know, but there are reasons to highlight this as a valuable (and viable) inquiry for future research. A healthy amount of studies have suggested that the P600 may in fact be a member of the P3 family (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2015; Coulson, King, & Kutas, 1998; Sassenhagen & Bornkessel-Schlesewsky, 2015; Sassenhagen et al., 2014; cf.; Osterhout, McKinnon, Bersick, & Corey, 1996; Osterhout & Hagoort, 1999), given similarities such as the fact that the P600 is also absent or reduced for unattended relevant stimuli (e.g., Batterink & Neville, 2013; Hahne & Friederici, 2002; Hanulíkova, van Alphen, van Goch, & Weber, 2012) that it also peaks around centro-parietal sites, and that it seems to be response-time aligned (Sassenhagen & Bornkessel-Schelewsky, 2015; Sassenhagen et al., 2014). In this view, the later onset of the P600 relative to a standard P300 would be explained by the higher complexity of linguistic material as compared to the visual stimuli typically found to elicit a P300. Indeed, increasing the complexity of stimuli in a P300 experiment reliably delays the onset of this waveform (see Luck, 2014, pp. 98–101, for review).

What, then, could we expect of this component in terms of its developmental sequence, if it is indeed tied to the allocation of attentional resources that may over time result in the detection of syntactic anomalies? Studies examining the processing of non-novel agreement features in low-proficiency L2 learners have reported incipient P600-like effects, which already have the expected onset and scalp distribution but are much reduced in amplitude (Gabriele et al., 2013, under review). However, even the low-proficiency learners in Gabriele et al. (2013) had significantly more exposure to the target language than our participants. It is, therefore, possible that our data capture a much earlier stage where linguistic representations are underspecified, yet processing of the L3/Ln input is already guided by selective focus on those elements of the linguistic material that are most relevant to constructing reliable syntactic parses in one of the previous languages—e.g., affixes in Spanish, which typically contain agreement morphology necessary to process different kinds of long-distance dependencies. To the extent that this is not grammatical processing per se yet and, as such, evaluations of deviancy in the stimuli may not involve higher-order (linguistic) processing, an earlier onset of this positive component is not unexpected.

In such a short training exposure, it does make sense that we would tap processes of economizing for preparing an individual for the task of multilingual acquisition. And so, what we believe we are seeing is what the mind does in the most initial sense to get into the real process of subsequent linguistic acquisition, even before transfer takes place, by first finding the simplest ways to justify transfer selection among choices while still performing the task at hand. The simplest way to detect differences between grammatical and ungrammatical sentences would be to focus on the morphology and really only the morphology. In this sense, the participants in this task do not seem to be processing entire phrases *per se*, but rather some are detecting patterns that happen to be linguistic ones at a

much smaller level. Crucially, however, only one of the two groups is doing this, indicating that something more general about the input one but not the other group receives determines whether smaller-unit pattern matching is activated to try to parse the incoming input stream. Given our manipulation of the mini-grammars, lexical similarity stands out as the only obvious candidate for the difference noted, which is compatible with models of transfer that focus on structural similarity as the main conditioning factor (such as the TPM and the LPM).

The above discussion highlights the importance of future research into the earliest stages of L3/Ln processing. Future work with longer training sessions, we believe, would determine the amount of exposure needed for one to have a better chance at capturing an initial true L3 grammar, thus allowing one to adjudicate between competing models of transfer source selection as was our original goal. Such studies have the potential to uncover the very dynamic nature of the processes that lead to early transfer in the first place, which are themselves connected to important questions regarding the implementation of cognitive economy in language. In fact, longitudinal studies making use of multiple testing and mid-term consolidation have the best chance to capture transitional stages that illuminate not only the mechanisms that establish initial specifications in the new (L3) grammar, but also potential missing links between the neural correlates of sensory and higher-order processing within a single developmental sequence. If it turns out that the P300 here reflects more focused attention at a stage before representational transfer has taken place—the selection phase—because the target language of exposure is lexically based on Spanish, then we would predict that in a design where there is a consolidation phase, re-exposure in an additional (or more) phase(s) and subsequent rerunning of the EEG/ERP we would see a P600 emerge. Such a design is not uncommon in the implicit vs. explicit non-native learning literature that employs similar designs for distinct questions (e.g., Morgan-short, Finger, Grey, & Ullman, 2012). We would hypothesize that under such a design the Rothman et al. (2015) predictions would be in the best position to be tested.

Acknowledgements

Appendix A

Model outputs for omnibus factorial repeated-measures ANOVAs in the different time windows of the Mini-English and Mini-Spanish experiments. For those main effects or interactions where the sphericity assumption was violated, we report adjusted degrees of freedom and p-values (based on Huynh-Feldt corrections).

Table A1
Model output for omnibus ANOVA in Mini-English, 200–500 ms: lateral electrodes

Effect	DFn	DFd	F	p	η^2
Condition	1	23	0.024	0.879	0.000
Hemisphere	1	23	18.386	0.000	0.074
Caudality	1.15	26.36	0.697	0.430	0.010
Condition*Hemisphere	1	23	3.387	0.079	0.003
Condition*Caudality	1.31	30.08	0.118	0.800	0.000
Hemisphere*Caudality	1.64	37.72	5.528	0.012	0.010
Condition*Hemisphere*Caudality	1.41	32.48	3.851	0.045	0.002

Table A2
Model output for omnibus ANOVA in Mini-English, 200–500 ms: midline electrodes

Effect	DFn	DFd	F	P	η^2
Condition	1	23	0.004	0.953	0.000
Caudality	1.36	31.19	0.429	0.577	0.009
Condition*Caudality	2	46	0.243	0.785	0.001

Table A3
Model output for omnibus ANOVA in Mini-English, 300–600 ms: lateral electrodes

Effect	DFn	DFd	F	p	η^2
Condition	1	23	0.001	0.977	0.000
Hemisphere	1	23	18.573	0.000	0.073
Caudality	1.18	27.14	4.834	0.031	0.069
Condition*Hemisphere	1	23	3.175	0.088	0.004
Condition*Caudality	1.60	36.89	0.184	0.785	0.000
Hemisphere*Caudality	1.48	34.09	6.270	0.009	0.014
Condition*Hemisphere*Caudality	1.24	28.43	3.071	0.083	0.003

Table A4
Model output for omnibus ANOVA in Mini-English, 300–600 ms: midline electrodes

Effect	DFn	DFd	F	p	η²
Condition	1	23	24.351	0.100	0.000
Caudality	1.27	29.30	253.616	0.037	0.093
Condition*Caudality	2	46	46.202	0.515	0.002

Table A5
Model output for omnibus ANOVA in Mini-English, 400–900 ms: lateral electrodes

Effect	DFn	DFd	F	p	η²
Condition	1	23	0.055	0.817	0.000
Hemisphere	1	23	8.635	0.007	0.030
Caudality	1.42	32.71	7.161	0.006	0.075
Condition*Hemisphere	1	23	2.274	0.145	0.002
Condition*Caudality	1.30	29.85	0.070	0.854	0.000
Hemisphere*Caudality	1.36	31.28	9.799	0.002	0.019
Condition*Hemisphere*Caudality	1.26	28.93	1.997	0.166	0.002

Table A6
Model output for omnibus ANOVA in Mini-English, 400–900 ms: midline electrodes

Effect	DFn	DFd	F	P	η²
Condition	1	23	0.132	0.720	0.000
Caudality	1.34	30.73	9.575	0.002	0.190
Condition*Caudality	1.58	36.25	0.696	0.473	0.003

Table A7
Model output for omnibus ANOVA in Mini-Spanish, 200–500 ms: lateral electrodes

Effect	DFn	DFd	F	p	η²
Condition	1	25	3.996	0.057	0.011
Hemisphere	1	25	17.024	0.000	0.043
Caudality	1.30	32.45	0.756	0.424	0.006
Condition*Hemisphere	1	25	0.757	0.393	0.000
Condition*Caudality	1.39	34.85	1.668	0.207	0.002
Hemisphere*Caudality	1.33	33.25	4.796	0.026	0.007
Condition*Hemisphere*Caudality	1.49	37.3	0.412	0.606	0.000

Table A8
Model output for omnibus ANOVA in Mini-Spanish, 200–500 ms: midline electrodes

Effect	DFn	DFd	F	P	η²
Condition	1	25	3.961	0.058	0.012
Caudality	1.51	37.85	0.272	0.701	0.005
Condition*Caudality	2	50	1.668	0.199	0.003

Table A9
Model output for omnibus ANOVA in Mini-Spanish, 300–600 ms: lateral electrodes

Effect	DFn	DFd	F	p	η²
Condition	1	25	4.611	0.042	0.013
Hemisphere	1	25	22.654	0.000	0.060
Caudality	1.36	34	3.446	0.060	0.030
Condition*Hemisphere	1	25	1.403	0.247	0.000
Condition*Caudality	1.57	39.35	1.973	0.161	0.001
Hemisphere*Caudality	1.31	32.65	8.564	0.003	0.013
Condition*Hemisphere*Caudality	1.58	39.55	0.286	0.701	0.000

Table A10
Model output for omnibus ANOVA in Mini-Spanish, 300–600 ms: midline electrodes

Effect	DFn	DFd	F	P	η²
Condition	1	25	4.254	0.050	0.010
Caudality	1.48	36.9	2.617	0.101	0.046
Condition*Caudality	2	50	1.709	0.191	0.003

Table A11

Model output for omnibus ANOVA in Mini-Spanish, 400–900 ms: lateral electrodes

Effect	DFn	DFd	F	p	η^2
Condition	1	25	1.164	0.291	0.009
Hemisphere	1	25	27.732	0.000	0.089
Caudality	1.37	34.25	2.809	0.092	0.025
Condition*Hemisphere	1	25	2.034	0.166	0.001
Condition*Caudality	1.55	38.8	1.957	0.163	0.002
Hemisphere*Caudality	1.44	35.95	18.375	0.000	0.024
Condition*Hemisphere*Caudality	1.53	38.3	0.177	0.780	0.000

Table A12

Model output for omnibus ANOVA in Mini-Spanish, 400–900 ms: midline electrodes

Effect	DFn	DFd	F	P	η^2
Condition	1	25	0.723	0.403	0.004
Caudality	1.34	33.55	4.203	0.037	0.073
Condition*Caudality	2	50	1.288	0.285	0.006

References

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2012). The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research, 1456*, 49–63.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Language Research, 30*(3), 275–306.

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2018). Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish. *PloS One, 13*(7), Article e0200791.

Alemán Bañón, J., Miller, D., & Rothman, J. (2017). Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(10), 1509–1536.

Alemán Bañón, J., & Rothman, J. (2016). The role of morphological markedness in the processing of number and gender agreement in Spanish: An event-related potential investigation. *Language, Cognition and Neuroscience, 31*(10), 1273–1298.

Alemán Bañón, J., & Rothman, J. (2019). Being a participant matters: Event-related potentials show that markedness modulates person agreement in Spanish. *Frontiers in Psychology, 10*, 746.

Barber, H. A., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*(1), 137–153.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of germanic syntax. *Second Language Research, 23*(4), 459–484.

Bardel, C., & Falk, Y. (2012). Behind the L2 status factor: A neurolinguistic framework for L3 research. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). Amsterdam: John Benjamins.

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Batterink, L., & Neville, H. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience, 33*(19), 8528–8533.

Bharucha, J., & Krumhansl, C. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*(1), 63–102.

Bond, K., Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In J. Herschensohn, & D. Tanner (Eds.), *Proceedings of the 11th generative approaches to second language acquisition* (pp. 17–29). Sommerville, MA: Cascadilla Press.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S., & Rauschecker, J. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends Cogn. Sci., 19*(3), 142–150.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic view as a complementary perspective. *Bilingualism: Language and Cognition, 10*(1), 51–55.

Caffarra, S., Mendoza, M., & Davidson, D. (2019). Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain and Language, 191*, 9–16.

Carrasco-Ortiz, H., Velázquez Herrera, A., Jackson-Maldonado, D., Avecilla Ramírez, G. N., Silva Pereyra, J., & Wicha, N. Y. Y. (2017). The role of language similarity in processing second language morphosyntax: Evidence from ERPs. *International Journal of Psychophysiology, 117*, 91110.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism, 7*(1), 71–87.

Coch, D., Hart, T., & Mitra, P. (2008). Three kinds of rhymes: An ERP study. *Brain and Language, 104*(3), 230–243.

Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language & Cognitive Processes, 12*(2158).

Donchin, E. (1979). Event-related brain potentials: A tool in the study of human information processing. In H. Begleiter (Ed.), *Evoked brain potentials and behavior. The downstate series of research in psychiatry and psychology* (pp. 13–88). Boston, MA: Springer.

Donchin, E. (1981). Surprise!...Surprise? *Psychophysiology, 18*(5), 493–513.

Duncan-Johnson, C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Phsychophysiology, 14*(5), 456–467.

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

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*(1), 226–248.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica, 128*(3), 528–536.

Friederici, A. D. (2017). Neurobiology of syntax as the core of human language. *Biolinguistics, 11*, 325–337.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related potential effects. *Learning, Memory and Cognition, 22*(5), 1219–1248.

Gabriele, A., Alemán Bañón, J., Hoffman, L., Covey, L., Rossomondo, A., & Fiorentino, R. (2005). *Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials* (under review).

Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2013). Examining second language development using event-related potentials: A cross-sectional study on the processing of gender and number agreement. *Linguistic Approaches to Bilingualism, 3*(2), 213–232.

Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870–1887.

González Alonso, J., & Rothman, J. (2017). Coming of age in L3 initial stages transfer models: Deriving developmental predictions and looking towards the future. *International Journal of Bilingualism, 21*(6), 683–697.

Gouvea, A., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language & Cognitive Processes, 25*(2), 149–188.

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

Grey, S., Tanner, D., & Van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Research, 1669*, 27–43.

Grey, S., Williams, J., & Rebuschat, P. (2014). Incidental exposure and L3 Learning of morphosyntax. *Studies in Second Language Acquisition, 36*(4), 611–645.

Guajardo, L., & Wicha, N. (2014). Morphosyntax can modulate the N400 component: Event related potentials to gender-marked post-nominal adjectives. *NeuroImage, 91*, 262–272.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*(4), 556–568.

Hagoort, P. (2003). Violations, interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic. *Journal of Cognitive Neuroscience, 15*, 883–899.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Pocesses, 8*(4), 439–483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*, 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potential. *Bilingualism: Language and Cognition, 4*(2), 123–141.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Brain Research. Cognitive Brain Research, 12*(3), 339–356.

Hanulíkova, A., van Alphen, P., van Goch, M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878–887.

Hasting, A., & Kotz, S. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience, 20*(7), 1207.

Hermas, A. (2014). Multilingual transfer: L1 morphosyntax in L3 English. *International Journal of Language Studies, 8*(2), 1–24.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363.

Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *The Modern Language Journal, 92*(2), 270–283.

Koelsch, S. (2009). A neuroscientific perspective on music therapy. *Annals of the New York Academy of Sciences, 1169*, 374–384.

Koelsch, S., Schmidt, B., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology, 39*(5), 657–663.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language & Cognitive Processes, 21*(5), 489–530.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Lawrence, M. (2016). Ez: Easy analysis and visualization of factorial experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez.

Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (2007). Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research, 1142*, 169–177.

Luck, S. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: The MIT Press.

van de Meerendonk, N., Chwilla, D., & Kolk, H. (2013). States of indecision in the brain: ERP reflections of syntactic agreement violations versus visual degradation. *Neuropsychologia*, 1393–1396.

Mickan, A., & Lemhöfer, K. (2020). Tracking syntactic conflict between languages over the course of L2 acquisition: A cross-sectional event-related potential study. *Journal of Cognitive Neuroscience, 32*(5), 822–846.

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN). *Cortex, 66*, 156–159.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*, 908–930.

Morgan-Short, K. (2020). Insights into the neural mechanisms of becoming bilingual: A brief synthesis of second language research with artificial systems. *Bilingualism: Language and Cognition, 23*(1), 87–91.

Morgan-short, K., Finger, I., Grey, S., & Ullman, M. T. (2012). Second language processing shows increased native-like neural responses after months of no exposure. *PloS One, 7*(3), Article e32974.

Morgan-Short, K., Sanz, C., & Ullman, M. T. (2010). second language acquisition of gender agreement in explicit and implicit training Conditions : An event-related potential study. *Language Learning, 60*(1), 154–193.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature number and feature type in processing Hindi verb agreement. *Brain Research, 1167*, 81–94.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510–532.

Oldfield, R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 46*, 143–168.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language & Cognitive Processes, 14*, 1–14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8*(6), 507–526.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentence in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras, & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56*(1), 199–230.

Osterhout, L., & Mobley, L. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.

Oxford University Press, & Cambridge University & Association of Language Testers in Europe. (2001). *Quick placement test: Paper and pen test*. Oxford: Oxford University Press.

O'Rourke, P., & Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the eventrelated brain potentials. *Brain Research, 1392*, 62–79.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam, NL: John Benjamins.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*, 717–733.

Polich, J. (2012). Neuropsychology of P300. In E. Kappenman, & S. Luck (Eds.), *The Oxford handbook of event-related potential components*. Oxford, UK: Oxford University Press.

Rebuschat, P., & Williams, J. (2012). Implicit and explicit knowledge in second language acquisition. *Applied PsychoLinguistics, 33*(4), 829–856.

Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia, 62*, 11–25.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The Typological Primacy Model. *Second Language Research, 27*, 107–127.

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition, 18*(2), 179–190.

Rothman, J., Alemán Bañón, J., & González Alonso, J. (2015). Neurolinguistic measures of typological effects in multilingual transfer: Introducing an ERP methodology. *Frontiers in Psychology, 6*, 1087.

Rothman, J., González Alonso, J., & Puig-Mayenco, E. (2019). *Third language acquisition and linguistic transfer*. Cambridge, UK: Cambridge University Press.

Sanz, C., Park, H. I., & Lado, B. (2015). A functional approach to cross-linguistic influence in ab initio L3 acquisition. *Bilingualism: Language and Cognition, 18*(2), 236–251.

Sassenhagen, J., & Bornkessel-Schlesewsky, I. (2015). The P600 as a correlate of ventral attention network reorientation. *Cortex, 66*, A3–A20.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language, 137*, 29–39.

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research, 12*(1), 40–72.

Spencer, K., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology, 38*(2), 343–358.

Squires, N., Squires, K., & Hillyard, S. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*(4), 387–401.

Steinbeis, N., & Koelsch, S. (2007). Shared neural resources between musich and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169–1178.

Steinhauer, K., White, E., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13–41.

Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex, 111*, 210–237.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition, 16*, 376–382.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*, 289–201.

Tokowicz, N, & MacWhinney, B (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*(2), 173–204.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 953–968). San Diego, CA: Elsevier.

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology, 2*, 365.

Verleger, R., Jaskowski, P., & Wauschkuhn, B. (1994). Suspense and surprise: On the relationship between expectancies and P3. *Psychophysiology, 31*(4), 359–369.

Westergaard, M. (2019). *Microvariation in multilingual situations: The importance of property-by-property acquisition*. Second Language Research.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism, 21*(6), 666–682.

Williams, J., & Kuribara, C. (2008). Comparing nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua, 118*(4), 522–553.

Woodman, G. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics, 72*(8), 2031–2046.