The presence of a foreign accent introduces lexical integration difficulties during late semantic processing

Leah Gosselin[a,b], Clara D. Martin[b,c] , Eugenia Navarra-Barindelli[b,d], and Sendy Caffarra[b,e]

[a] *University of Ottawa, Department of Linguistics, 70 Laurier Ave. E., Ottawa, ON, Canada, K1N 6N5*

[b] *BCBL, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 20009, Donostia, Spain*

[c] *Ikerbasque, Basque Foundation for Science, Maria Diaz de Haro 3, Bilbao, 48013, Spain*

[d] *Universidad del Pais Vasco (UPV/EHU), Arriola Pasealekua 2, 20018, Donostia, Spain*

[e] *Stanford University School of Medicine, Developmental Behavioral Pediatrics, 730 Welch Road, 94304, Palo Alto CA, USA*

Corresponding author:

Leah Gosselin

University of Ottawa, Department of Linguistics, 70 Laurier Ave. E., Ottawa, ON, Canada, K1N 6N5

Email : lgoss035@uottawa.ca

The presence of a foreign accent introduces lexical integration difficulties during late semantic processing

Previous research suggests that native listeners may be more tolerant to syntactic errors when they are produced in a foreign accent. However, studies investigating this topic within the semantic domain remain conflicting. The current study examined the effects of mispronunciations leading to semantic abnormality in foreign-accented speech. While their EEG was recorded, native speakers of Spanish listened to semantically correct and incorrect sentences produced by another native speaker and a native speaker of Chinese. The anomaly in the incorrect sentences was caused by a subtle mispronunciation (typical or atypical in Chinese-accented Spanish) during a critical word production. While initial-stage semantic processing yielded no accent-specific differences, late processing revealed a persistent N400-effect in the foreign-accent but not in the native-accent. These findings suggest that foreign-accented mispronunciations are more difficult to integrate than native-accented errors, regardless of their relative typicality. The distinction between syntactic and semantic processing of foreign-accented speech is discussed.

Keywords: Error typicality, Event-related potentials, N400, non-native accent

1. Introduction

In our increasingly globalized day-and-age, native speakers must often enter into interactions with second language interlocutors, whose productions deviate from the typical speech stream (Bent, Bradlow & Smith, 2007). In fact, even highly proficient second language (L2) speakers tend to experience difficulty with the phonology of their L2 (Flege, 1995), and may produce systematic speech errors (Young-Scholten, 2011). As such, more and more, native speakers are exposed to, and must deal with, L2 speech errors. Some of these errors can lead to higher-level violations, such as syntactic or semantic violations.

The effects of speech accent on the time course of syntactic analysis have been well-studied (Xu, Abdel Rahman & Sommer, 2020; Caffarra & Martin, 2019; Grey & van Hell, 2017; Hanulíková, van Alphen, van Goch & Weber, 2012) and lead to the general conclusion that sentence processing is indeed affected by foreign accent in some regard. However, in the semantic domain, experimental results are still inconclusive. Some studies suggest that foreign accent does not modulate lexico-semantic processing (e.g., Hanulíková et al., 2012), while others imply the opposite conclusions (e.g., Goslin, Duffy & Floccia, 2012; Grey & van Hell, 2017, Romero-Rivas et al., 2015).

One way of advancing this debate is to explore an additional dimension within the range of semantic analysis, such as error typicality or expectedness. The present study compares the processing of subtle semantic violations caused by speech errors (phoneme substitutions) that are *typical* in Chinese-accented Spanish to those that are *atypical* in this accent. Unlike past research, the speech errors included in the current study are subtle mispronunciations rather than outright violations (i.e. when the expected target is substituted for an entirely different word). Subtle mispronunciations (e.g., a sentence such as "*Se escapó el pelo de tu tía*.", 'Your aunt's hair escaped.' where *perro* 'dog' is mispronounced as *pelo*, 'hair'), are more likely to occur in everyday life than outright violations (e.g., *evening* replaces *blanket*; Hanulíková et al., 2012) and are therefore more natural.

1.1. Event-related potentials (ERPs) and foreign-accented speech

A handful of studies investigating foreign-accented speech have utilized ERPs. ERPs (measured via EEG) reflect participants' real-time electrophysiological activity in response to specific language events (e.g., a word embedded in an utterance). ERPs shed light on online language processing with extremely precise temporal resolution, and thus represent a suitable measure to investigate the underlying impact of foreign-accent on speech processing. In the next sections, we will review past ERP findings relative to syntactic and semantic processing in foreign-accented speech.

The effect of foreign accent on the time course of syntactic processing

Classical studies utilizing the syntactic-violation paradigm (in which participants are exposed both to syntactically correct and incorrect sentences) persistently report P600 effects. The P600 is an ERP component which consists of a positive neural response, typically posteriorly distributed. In classical syntactic-violation paradigms, both written (e.g., Hagoort & Brown, 2000) and spoken (e.g., Hagoort & Brown, 2000; Hahne & Jescheniak, 2001; Balconi & Pozzoli, 2005) sentences containing a syntactic error generate a greater positivity than well-formed sentences around 600 ms after the target word onset; this is a classic P600 effect. Functionally, the P600 effect is most often attributed to structure reanalysis, syntactic-violation repair mechanisms (e.g., Osterhout & Holcomb, 1992; Kaan & Swaab, 2003) or to conflict monitoring (van de Meerendonk, Kolk, Vissers & Chwilla, 2010).

A number of studies have shown that syntactic errors do not exhibit the classical P600 effect when they are produced in foreign-accented speech (Xu et al., 2020; Caffarra & Martin, 2019; Hanulíková et al., 2012; Grey & van Hell, 2017). Furthermore, recent research (Caffarra & Martin, 2019) also indicates that the relative typicality of a given syntactic error plays a role in the processing of foreign-accented speech. For sentences produced in a non-native accent, atypical errors (i.e., number errors) generated a P600 effect, but this same effect was absent or reduced for typical speech errors (i.e., grammatical gender errors), particularly in the latter half of the experiment (Caffarra & Martin, 2019). This interaction was not observed for native speech. Such results suggest that listeners may adapt to language input according to their expectations about foreign-accented speech; indeed, participants who report greater exposure to the accent under study typically show smaller grammaticality-related ERP effects than accent-unfamiliar participants (Caffarra & Martin, 2019).

Taken together, the results from the literature suggest that native listeners possess a certain tolerance towards (at least some) speech errors in foreign-accented speech, as they do not attempt to repair them (reflected by the lack of a P600, or in the case of Grey & van Hell, 2017, the lack of an Nref for accent-unfamiliar participants). In any case, the largely homogenous findings indicate that the presence of a foreign accent does indeed impact native listeners’ higher-level processing mechanisms, and that this effect interacts with other variables such as error typicality, adaptation over time, and accent familiarity.

The effect of foreign accent on the time course of semantic processing

Semantic processing studies typically utilize the semantic-violation paradigm, in which participants are exposed to semantically correct and semantically incorrect sentences. Such studies generally focus on the N400 component, an increased centro-parietal negative response elicited approximately 400 ms after target onset. In native accented speech, semantic violations consistently elicit a greater N400 relative to well-formed sentences (i.e. the N400 effect; see Kutas & Federmeier, 2011 for a review). As such, the N400 effect is associated with difficulties regarding lexical-semantic integration, and semantic unexpectedness (e.g., Kutas & Hillyard, 1980; Kutas & Hillyard, 1984). For instance, low-cloze probability words (i.e. words that have a low probability of completing a sentence; “He took a sip from the *waterfall*.”), or words that render a sentence semantically anomalous (“He took a sip from the *transmitter*.”), generate larger N400s than semantically expected words (Kutas & Hillyard, 1980). In native speech, a similar effect is observed for semantic anomalies induced by more subtle errors, such as the substitution of acoustic features (e.g., tone substitution: Brown-Schmidt & Canseco-Gonzalez, 2004). However, when this classic paradigm is applied to foreign-accented speech, the typical pattern of results is blurred.

Some research has demonstrated that semantically abnormal sentences produced in foreign-accented speech continue to generate N400 effects, just like in native speech (Holt, Kung, Demuth, 2018; Hanulíková et al., 2012). Alternatively, other studies indicate that speaker accent does indeed affect semantic processing (Grey & van Hell, 2017; Romero-Rivas et al., 2015; Brunellière & Soto-Faraco, 2013; Goslin, Duffy & Floccia, 2012), although the nature of this impact is still disputed. On one hand, some studies have reported stronger or more largely distributed N400 effects in response to semantic abnormality in foreign-accented speech than in native speech (Romero-Rivas et al., 2015; see also Song & Iverson, 2018). A possible explanation for this result is that a foreign-accent is more cognitively costly during lexical access (i.e. it is more difficult to identify words in foreign-accented speech than in native speech, due to the degraded speech signal; Romero-Rivas et al., 2015). In fact, much behavioural research has demonstrated that a foreign accent may impede listeners' word identification and comprehension (e.g., Braun, Dainora & Ernestus, 2011; Floccia, Butler, Goslin & Ellis, 2009; Clarke & Garrett, 2004; van Wijngaarden, 2001).

On the other hand, ulterior research also reports *reduced* N400 effects for semantically abnormal (Xu et al., 2020 for slips of the tongue) and correct (Goslin et al., 2012) sentences produced by non-native speakers as compared to native speakers. Goslin and colleagues (2012) claim that this result is due to the fact that listeners are more reliant on top-down semantic-pragmatic cues than low-level phonological cues when they hear foreign-accented speech, as the phonological input is less reliable in non-native speech. These conclusions fall into line with the results found in the syntactic studies, that is, that listeners are more tolerant to foreign-accented errors. It is possible that listeners develop a keen ear for meaning in foreign accented speech (Cheng, 1999), and can thus overlook errors when the speaker is non-native. Minimally, foreign-accented speech errors might be processed as expected phonological "lapses", rather than semantic anomalies, resulting in reduced N400 effects.

As can be noticed—and importantly, unlike in the syntactic domain—previous studies examining the impact of foreign-accent on semantic processing produce conflicting results.

The literature suggests, all at once, that foreign-accent makes higher-level semantic processing more difficult (Romero-Rivas et al., 2015), that foreign-accent reduces the processing costs of semantic anomalies (Xu et al., 2020) or simply, that foreign-accent has no effect on semantic processing (e.g., Hanulíková et al., 2012). The variability in these findings suggests that multiple aspects of the experimental design (e.g., type of accents examined, type of violations) might account for the discordant results for semantic processing in foreign-accented speech.[1] It has been shown that the processing of foreign-accented speech is sensitive to a range of variables, such as familiarity with the accent; in contrast to the syntactic domain, accent-unfamiliar participants show *smaller* negativity effects than individuals who report more exposure to the accent under study, perhaps since their processing resources are being diverted to non-semantic levels, like pronunciation (Holt et al., 2018).

Let us note that most of the previous studies examined the effect of foreign accent by using outright semantic violations (e.g., "He spread the warm bread with *socks*."; Hanulíková et al., 2012). Results from such methodologies consistently indicate that semantic abnormality generates increased (Romero-Rivas et al., 2015; Grey and van Hell, 2017) or unchanged (Hanulíková et al., 2012; Holt et al., 2018) N400 effects relative to native speech. These findings are not particularly surprising, as outright violations are uncommon, even in L2 speech. However, when more subtle semantic errors were utilized in the experimental design, the results were reversed. For instance, Xu and colleagues (2020) incorporated slips of the tongue into their stimuli. While slips of the tongue introduce semantic oddity into their surrounding sentence, the error still closely resembled a word that *would* fit the semantic context. This type of speech error, arguably more common than outright semantic violations, generated a decreased N400 effect in foreign-accented speech compared to native speech.

As such, the conflicting results with regard to semantic processing of foreign-accented speech may be at least partially caused by variations in the stimuli. It appears that subtle semantic errors are more likely to be overlooked by listeners than outright violations in non-native speech, especially when it comes to understanding the meaning of the speech signal. Nevertheless, the specific features of semantic analysis that are sensitive to speech accent have not yet been defined. This study represents a step forward in the investigation of this topic: error typicality will be systematically manipulated. In particular, we compare subtle semantic violations caused by speech errors that are either *typical* or *atypical* in Chinese-accented Spanish.

1.2 Present study

The present study investigates the impact of foreign-accent on native listeners' semantic processing, as well as the role of error typicality in foreign-accented speech. As such, native speakers of Spanish listened to semantically correct and semantically incorrect sentences produced by a native Spanish speaker and a Chinese-accented speaker (who was highly proficient in Spanish but still possessed a clear foreign-accent). Instead of outright violations, the anomaly in the semantically incorrect sentences was caused by a phoneme substitution that is either typical in Chinese-accented Spanish (e.g., /r-l/ substitution, as in *ca**r**a* 'face' → *ca**l**a,* 'cove'), or atypical in Chinese-accented Spanish (e.g., /n-s/ substitution, as in *pa**n**a*, 'corduroy' → *pa**s**a,* 'raisin'). The typicality of the /r-l/ substitution is based on the fact that the phonological inventory of Chinese does not contain Spanish rhotics (trill or the tap). Consequently, Chinese speakers of Spanish (even at high levels of Spanish proficiency) have been known to substitute the Spanish alveolar /r-ɾ/ with the lateral /l/ sound (Ortí Mateu, 1996), which can sometimes lead to a different word in Spanish; native Chinese speakers of Spanish may in fact "never [reach] a fully native-like pronunciation" of the Spanish trill or tap (Chen, 2007: 135). This pattern of error is stereotypically known by Spanish native speakers (Pichardo, 1953). Crucially, the subtle /r-l/ and /n-s/ phoneme substitutions in the current study change the critical word's meaning and make the subsequent sentence semantically anomalous. While participants listened to the stimuli, ERPs time-locked to the target words were obtained.

Our predictions are focused on the N400 time window as past research has shown that semantic violations due to mispronunciations lead to N400 effects similar to those reported in classical violation paradigms (i.e. where the congruent target word is replaced with an incongruent one; Brown-Schmidt & Canseco-Gonzalez, 2004).

For the native accent, we expect that both types of mispronunciations will elicit similar N400 effects compared to semantically well-formed sentences (i.e., classic semantic-violation response; McCallum, Farmer & Pocock, 1984). We do not expect any effect of typicality as /r-l/ and /n-s/ errors are equally unlikely in native-accented speech.

For the foreign accent, we expect to see an effect of error typicality (Caffarra & Martin, 2019). We anticipate that semantic violations produced by *atypical* mispronunciations in Chinese-accented Spanish (/n-s/ phoneme substitutions) will elicit N400 effects that are similar (Hanulíková et al., 2012) or greater (Romero-Rivas et al., 2015) than those observed in the native accent, since these types of errors are unexpected. This prediction is also motivated by the wealth of past behavioural research which indicates that the presence of a foreign accent may impede listeners' word identification and comprehension (e.g., Braun, Dainora & Ernestus, 2011; Floccia, Butler, Goslin & Ellis, 2009; Clarke & Garrett, 2004; van Wijngaarden, 2001). Contrastively, semantic violations produced by *typical* mispronunciations in Chinese-accented Spanish (/r-l/ phoneme substitutions) should elicit a reduced N400 effect as compared to native accent. This hypothesis is bolstered by past results indicating that native listeners are tolerant towards frequently-heard mispronunciations (Caffarra & Martin, 2019; Xu et al., 2020). As such, typical /r-l/ substitutions may be overlooked semantically (eliciting mitigated integration difficulties), as they are processed as simple phonological "lapses". The reduction in the N400 effect may be particularly robust in the later parts of the experiment, since the listener's reduced sensitivity for typical foreign-accented errors might become more evident as the experiment progresses (similar results to Caffarra & Martin, 2019).

2. Methods

2.1 Participants

The project was approved by Research Ethics Board of the University of Ottawa (Ethics file #S-05-19-3940), as well as the Ethics Committee from the Basque Center on Cognition, Brain and Language (study #140519DS), and is in conformity with the World Medical Association Declaration of Helsinki. Forty native speakers of Spanish in the Basque Country, Spain, gave informed consent for their participation and were provided monetary compensation for their time (10 € per hour). Spanish was their current dominant language, though all possessed high proficiency in Basque and English. Three participants were subsequently excluded, as more than 30% of their trials were rejected during data pre-processing. One additional participant was excluded due to being left-handed. The remaining 36 right-handed participants (11 men, 25 women) ranged from 19 to 39 years of age ($\overline{x}$ = 25.8) and had no history of head trauma or language disorders. They reported no background, schooling, or knowledge of Chinese, and none had visited China for a prolonged period of time (i.e., over two weeks). Their language background details are presented in Table 1, below. Assuming an effect size of 0.25, a post-hoc power analysis (G*power, 3.1; Heinrich Heine University Düsseldorf, Germany) confirms that a sample of 36 participants ensures a high statistical power; the probability for a type I error is below .0001%.

<Insert Table 1 about here>

2.2 Materials

Target words were composed of 32 Spanish minimal pairs involving /l/ and /r, ɾ/ phonemes (e.g., *cala* ‘cove’, and *cara* ‘face’) and 32 minimal pairs involving /s/ and /n/ phonemes (e.g., *pasa* ‘raisin’, and *pana* ‘corduroy’). Minimal pairs could be either singular or plural words. Each set was included in the stimuli list twice, due to the limited number of minimal pairs available in Spanish. As such, the complete target word list was composed of 128 non-unique pairs.

Semantically correct sentences were constructed involving each target word. None of the target words had high predictability (i.e. the sentence context was low-constraining). On average, the target word was the fifth word in each sentence ($\bar{x}$ position = 5.08 for /r-l/-target-words and $\bar{x}$ = 5.20 for /n-s/-target-words). In total, 256 sentences (128 including /r-l/-target-words, and 128 including /n-s/-target-words) were constructed by a native speaker of Spanish and later reviewed by three other L1-Spanish speakers on two occasions (six informants total) to ensure syntactic and semantic correctness. Subsequently, in /r/-target-words (e.g., *cara*), the tap or trill was substituted by the lateral phoneme /l/, so as to create a condition of sentences involving errors stereotypically produced by native Chinese speakers (e.g., *cara* became *cala*). Critically, this substitution changed the meaning of the word, thereby rendering the corresponding surrounding sentence semantically anomalous. Similarly, in /n/-target-words (e.g., *pana*), the nasal phoneme /n/ was substituted by the fricative /s/, thus creating a condition of sentences involving an atypical error in Chinese-accented Spanish (e.g., *pana* became *pasa*). Once again, as a consequence of this substitution, the surrounding sentence became semantically anomalous. Atypical errors occurred on the /n-s/ pair, as these phonemes match the place of articulation of the /r-l/ typically-substituted pair. Examples from each condition are illustrated in Table 2. As can be seen, the pre-target word was identical in both the correct and the incorrect versions of each target-word sentence pairs. In addition, targets occurred at the same word-position in corresponding sentence pairs (e.g., the third and fourth word, in the examples in Table 2), and were never sentence-final.

<insert Table 2 about here>

The critical phoneme (/l/ or /s/) occurred only once per target word. Since lexical mispronunciations involved the /r/ and the /n/ phonemes, these sounds did not occur anywhere else in the stimuli (including in the pre- and post-target sentence context). The reason for this design was twofold. Firstly, it maximized the foreign speaker's pronunciation reliability, since she did not constantly switch from a trill to a liquid pronunciation of /r/. Second, this design ensured that participants perceived the speech deviations as mispronunciations instead of simple lexical errors, as they did not hear the speaker pronounce the /r/ or /n/ phonemes correctly at any point throughout the experiment. Only /l/ and /s/-target-words were heard by the participants.

The stimuli were recorded by one native female Chinese speaker, highly proficient in Spanish, and one native female Spanish speaker from the Basque Country (similar to other previous studies; Grey and van Hell, 2017; Hanulikova et al. 2012). Their language background details are included in Table 3. Note that Chinese was chosen as the foreign accent in the present study, as Chinese-accented speakers of Spanish produce the phonological substitutions that allow us to answer the research questions. Furthermore, native Spanish speakers have stereotypical knowledge of these mispronunciations (Pichardo, 1953). Both speakers produced the semantically correct and incorrect versions of each sentence (256 sentences each). To reduce the risk of order effects, the sentences were alternated by condition during the recording sessions. However, correct and incorrect versions of each target word were produced consecutively, in order to ensure that identical target words possessed similar acoustic properties. All auditory clips were trimmed to the precise onset and offset of each sentence, and scaled to an intensity of 75 dB. The native speaker listened to each unique sentence recorded by the Chinese speaker before producing it herself, thereby matching the speed and prosody of the foreign-accented speech.

<insert Table 3 about here>

Twenty-one native Spanish individuals (none of whom participated in the EEG experiment) evaluated the quality of the auditory recordings during a pre-test survey (randomized and counterbalanced across lists). Two participants were excluded, as they did not originate from the Basque Country. As a result, the responses from 19 individuals were considered (10 women, 9 men, $\bar{x}$ age = 25.7 years). These participants were asked to rate the strength of speakers' accent for each individual sentence on a scale of 1 to 5 (1 corresponding to 'very weak' and 5 corresponding to 'very strong'), to identify sentences with incorrectly pronounced target words, and to transcribe the penultimate word from every recording. Based on their responses, four items per conditions (32 sentences in total) were eliminated from the pool, due to low intelligibility and error identification rate.

A by-item ANOVA including the experimental factors of Accent (native, foreign), Error typicality (typical-/l/, atypical-/s/) and Correctness (correct, incorrect) was conducted on the responses to the pre-test survey for the remaining 224 sentences (the final item list). Crucially, a main effect of Accent confirmed that the respondents rated the native Chinese speaker as having a significantly stronger accent in Spanish ($\bar{x}$ rating = 3.16, SD = .28) compared to the native speaker ($\bar{x}$ = 1.57, SD = .21; F (1,447) = 5071.40, $p < .0001$), regardless of the correctness of the sentence. This demonstrates that the presence of Chinese speaker's accent was not solely based on mispronunciations: even when she did not consciously mispronounce a word (semantically correct sentences), participants still rated her accent as significantly stronger than that of the native Spanish speaker (p_{tukey}<.001).

Nevertheless, both the native speaker and the foreign-accented speaker were similarly intelligible, as pre-test respondents were able to transcribe the penultimate word in the individual auditory sentence accurately ($\bar{x}$ accuracy = 95.0% for foreign speech; $\bar{x}$ = 96.5% for native speech; F (1,447) = 2.774, p = .10). Respondents also accurately identified sentences with speech errors in both accents, with an advantage for native speech ($\bar{x}$ = 85.0%) compared to foreign speech ($\bar{x}$ = 80.2 %; F (1,440) = 8.10, p = .005). This difference falls in line with the presupposed hypothesis that errors in foreign-accented speech are more likely to be overlooked than errors produced by native speakers (similar error identification results are also reported in Caffarra & Martin, 2019; Grey & van Hell, 2017; Hanulikova et al. 2012).[2] A main effect of Correctness revealed that sentences containing an incorrectly pronounced target word received significantly higher accent ratings (F (1,440) = 39.47, p > .001), regardless of accent. This confirms that there is likely a connection between speaker accent and error detection: listeners associate speech errors with highly accented speech. There was no main effect of Error typicality (/l/-errors vs. /s/-errors) for accent strength, error identification, or intelligibility (all Fs < 1.6, all ps > .20). Finally, this analysis confirmed that there were no interactions between the main experimental conditions in regard to accent strength, intelligibility, or error identification (all Fs < 2.5, all ps > .11). For these reasons, any results emerging from the ERP experiment can safely be attributed to the processing of foreign-accented speech, and not to a confounding interaction between the variables within the study.

The same by-item ANOVA was applied to the durational information of the final item pool. This analysis confirmed that there were no differences in regards to the total duration of the pre-target (native $\bar{x}$ = 222 ms, SD = 100 ms and foreign $\bar{x}$ = 217 ms, SD = 106 ms; F (1,440) = .308, p = .58), or the duration of the target word (native $\bar{x}$ = 396 ms, SD = 103 ms and foreign $\bar{x}$ = 411 ms, SD = 97 ms; F (1, 440) = 2.893, p = .09) between the native speaker and the foreign-accented speaker. A main effect of Phoneme also indicated that target words involving the /s/ phoneme were significantly longer than /l/-target-words (F (1,440) = 72.767, p < .001; see Table 4 below), for both native and foreign-accented speech. This is not unexpected, as phoneme identity can play a role in segment duration in Spanish (e.g., Córdoba et al., 2002). There were no durational differences between correctly pronounced target words ($\bar{x}$ = 401 ms, SD = 102 ms) and incorrectly pronounced target words ($\bar{x}$ = 406 ms, SD = 99 ms; F (1,440) = 0.373, p = .54). Furthermore, there were no interactions between Accent, Error typicality and Pronunciation correctness in regard to the total duration of each sentence, nor the length of the pre-target or the target words (all Fs < 1, all ps > .40).

The lexical properties of the target words included in the final item pool are listed in Table 4. Independent-sample t-tests confirmed that /l/-target-words and /s/-target-words did not differ in frequency, number of syllables, phonological neighbors and familiarity (all ps > .12), suggesting that any results found in regards to error typicality cannot be attributed to the lexical properties of the target words.

<insert Table 4 about here>

Finally, in order to ensure that target words in the final item pool did not differ in cloze probability, all semantically normal sentences (the ‘correctly pronounced’ condition) were trimmed immediately before the target word. Thirteen native Spanish informants (none of whom participated in the EEG experiment or the pre-test survey; 8 men, 5 women, $\bar{x}$ age = 32.5 years) were asked to indicate the next expected word in the sentence. While the sentences are low-constraining contexts (target words generally appear early in each sentence), cloze probability (CP) did not differ according to the critical phoneme involved in the target word (t (62) = .741, p = .46), suggesting that both /l/-target-words (CP = 0.08) and /s/-target-words (CP = 0.05) were equally unexpected in their semantically normal sentence contexts.

All 448 recordings from the final item pool were counterbalanced across two lists of 224 trials. Of these trials, half included correctly pronounced target words (and the sentences were therefore semantically normal; 50% typical /l/ errors and 50% atypical /s/ errors). The other half of the sentences included incorrectly pronounced target words (and were therefore semantically anomalous; 50% typical /l/ errors and 50% atypical /s/ errors). Participants heard every sentence only once, produced either by the foreign-accented speaker or the native speaker. Each list (224 items; half produced by the native speaker; 28 items per condition) was pseudo-randomized into three separate orders. Participants heard no more than 5 consecutive trials involving the same critical phoneme, no more than 5 consecutive trials recorded by the same speaker, and no more than 5 consecutive trials involving an incorrect pronunciation. Furthermore, as target words appeared more than once in each order (e.g., the correct and incorrect sentence-pairs for "*cala*"), pseudo-randomization ensured that trials involving the same target word were separated minimally by twenty trials.

Lastly, as adaptation over time may play a role in the processing of recurring errors in foreign-accented speech (see Caffarra & Martin, 2019), experimental conditions were distributed equally across each quartile of every experimental list and order. This is particularly important for future analyses involving experimental blocks (see Section 3.3).

2.3 Procedure

The experiment was run through Presentations® software. During the study, participants sat in front of a desktop computer in a sound-attenuated room. They began by listening to short auditory introductions from both speakers, stating their name and their region of origin. This was done in order to ensure that participants were forming appropriate expectations according to their preconceived notions of Chinese-accented speech. The experiment was split into four blocks of 56 trials (following four practice trials). A single trial is depicted in Figure 1.

Firstly, the symbol *.* appeared on the computer screen. Participants were instructed to blink during this time and to minimize any ocular movements throughout the rest of the trial. Following a 300 ms blank-screen interval, the auditory sentence ($\bar{x}$ duration=3157 ms, SD=882 ms) began to play through external speakers while a central fixation cross was displayed on the monitor. In order to ensure active attention on the participants' behalf, approximately one-fifth of auditory sentences were followed by a yes/no comprehension question (48 questions; equally distributed across conditions). Participants were instructed to press one of two keys to indicate their response (50% of questions required a "yes" answer). For incorrectly pronounced trials (e.g., *Me cosió la pasa de la chaqueta,* 'He sewed the raisin to the jacket.'), the comprehension question never referred to the target word. In consequence, each question could be answered accurately, regardless of whether participants assimilated mispronunciations to their intended meaning or not. Three-hundred milliseconds after the end of the auditory stimuli (or following the participant's button press, if the trial had a comprehension question), the next trial began. The experiment lasted approximately 30 minutes. All participants completed the study in its entirety.

<insert Figure 1 about here>

After the experiment, participants filled out a language background questionnaire and an additional offline questionnaire upon which they were asked about their familiarity with Chinese-accented Spanish in general, their impressions about the occurrence of /r/ to /l/ and /n/ to /s/ mispronunciations in Chinese-accented Spanish, as well as their feelings of trust or reliability towards the speakers. They were also asked to rate their perceived cognitive effort in processing the stimuli during the experiment (i.e. "From 1 to 10, how easy was it to identify word-by-word what the Chinese/Spanish native speaker said during the experiment?"). Lastly, the participants were asked to provide a definition for every target word included in the experiment in order to ensure that the critical words used in the item pool were known by the native speakers.

2.4 EEG recording and analysis

EEG signal was monitored via an elastic cap fitted with 27 electrodes, positioned according to the 10-20 system (EasyCap, Brain Products; Herrsching, Germany). In addition, six external electrodes were applied, respectively, to both ocular canthi, above and below the right eye, and on the left and right mastoids. Impedances were kept below 10 kΩ for facial electrodes, and below 5 kΩ for all other electrodes. Each EEG chamber was equipped with a BrainAmp DC amplifier (Brain Products, GmbH) and data was recorded at a sampling rate of 500 Hz. The left mastoid served as the initial online reference. The data was subsequently re-referenced offline to the average of both mastoids. A high-pass filter of 0.01 Hz and a low-pass filter of 30 Hz (24 dB/oct) were applied to the data offline.

EEG peaks exceeding +/-100 μV were rejected, once a visual inspection confirmed that these effects were truly artifacts. Blinks and horizontal eye movements were corrected via the Independent Components Analysis (ICA) in the Brain Analyzer software. ICA decomposes the EEG signal into distinct components. This allows the researcher to isolate the components that account for the highest variance in the eye electrodes in order to subtract them from the original data. As a result of these methods, 7.6% of trials were discarded, with no difference across conditions ($F (1,7) = .850$, $p = .520$). ERPs were time-locked to the onset of the critical word. A 200 ms pre-stimulus corrected baseline was established, extending to 1000 ms following the onset of the critical word.

The ERPs obtained from the nine posterior-most electrodes (CP1, CP2, P7, P3, Pz, P4, P8, O1, O2) were included in the main analyses, as N400 and P600 effects related to foreign-accented speech processing are posteriorly distributed (e.g., Caffarra & Martin, 2019;

Hanulíková et al., 2012; Holt et al., 2018; Romero-Rivas et al., 2015).[3] Using version 0.11.1 of the JASP software (University of Amsterdam, Netherlands), a repeated-measures ANOVA was conducted on these ERPs, averaged across trials from a given condition for each participant. In order to answer our research questions, we aimed to compare correct (i.e. *cala* correctly pronounced as *cala*, and therefore in a semantically normal context) and incorrect conditions (i.e. *cala* as a mispronunciation of *cara*, and therefore in a semantical abnormal context) according to accent and error typicality. Recall that there are no correct pronunciations of /r/ or /n/ target words in our stimuli, in order to maintain the speakers' pronunciation reliability throughout the experiment (i.e., this creates an ecologically valid situation where the Chinese speaker is consistently mispronouncing /r/ or /n/). The ANOVA thus included the experimental factors of Accent (native, foreign), critical Phoneme (typical /l/, atypical /s/), and Correctness (semantically correct, incorrect). The topographic factor of Laterality, which contained three levels (left: P7, P3, O1; medial: CP1, CP2, Pz; right: P4, P8, O2) was also included (see Figure 2), since previous studies have observed distribution-related differences according to accent (Romero-Rivas et al., 2015). These ANOVAs were conducted on two separate time-windows: 350-600 ms (similar to Goslin et al., 2012; Romero-Rivas et al., 2015; Brunellière & Soto-Faraco, 2013), and 600-900 ms (similar to Romero-Rivas et al., 2015; Caffarra & Martin, 2019). The later time-window was examined as foreign-accented-related negativities may extend beyond the typical N400 window (see Caffarra & Martin, 2019; Grey & van Hell, 2017).

<insert Figure 2 about here>

Only interactions involving an experimental variable are reported. In the event of violations of the sphericity assumption, the Greenhouse-Geisser procedure was implemented. For post-hoc tests, p-values are corrected using the Holm-Bonferroni method, in order to avoid familywise error rates (see Eichstaedt, Kovatch & Maroof, 2013). Using the same design as the ANOVAs noted above, a Bayesian analysis of effects across matched models was performed in JASP when null results were obtained, in order to verify their reliability. The Bayes inclusion factor (BF_{incl}), which compares a given effect or interaction to matched models stripped of this effect, is reported.

Finally, we examined the relationship between the elicited ERPs and the participant responses to the offline questionnaire, as well as their LexTALE proficiency score (Izura, Cuetos & Brysbaert, 2014). A Pearson correlation matrix was conducted between the ERP effects and the participant's self-reported experience with Chinese accented-speech, their proportion of foreign friends, their tendency to let pass or correct an error in speech, and their feelings of reliability, trust and perceptibility towards the speakers in the study. The values for ERP effects included in those analyses were obtained by subtracting the average response to semantically correct sentences (across the three parietal ROIs, *n* = 9 electrodes) from the responses elicited by semantically incorrect sentences, for both accents independently. In the case of significant correlations for both accents, the cocor package in R (Diedenhofen & Musch, 2015) was used with a confidence level of .95, adopting a one-tailed Pearson's *r* and Filon's *z*, in order to verify whether the independent correlations differed significantly from one another.

3. Results

3.1 Comprehension questions and offline questionnaire

Participants responded to the comprehension questions with an overall average accuracy of 95.2% (range: 88-100%). There were no differences across conditions (F (1,7) = 1.381, p = .233). A minimum accuracy was calculated by subtracting 2.5 standard deviations to the participants' average accuracy. Based on this value, no outliers were found, and all participants were thus deemed to have carefully attended to the stimuli.

In the post-test questionnaire, 15 participants noted that they came into contact with Chinese-accented Spanish, although this exposure was rather minimal ($\bar{x}$ = 1.5 hours per week); the remainder of the participants reported no notable exposure to this variant of foreign-accented speech. On average, participants noted that 8.9% of their friends were foreign (range: 0-70%).

Based on a scale of 1 to 10, participants reported similar feelings of reliability ($\bar{x}$ foreign-accent=7.69; $\bar{x}$ native accent=7.61) and trust ($\bar{x}$ foreign-accent = 6.25; $\bar{x}$ native accent = 5.72) for both accents (ps > .35), but significantly higher perceived cognitive effort for the foreign-accented speaker ($\bar{x}$ = 6.42) than the native-accented speaker ($\bar{x}$ = 7.75; p = .002; see similar results in Caffarra & Martin, 2019). When asked to rate how likely they were to “let pass” (as opposed to correct) a speech error produced in a foreign-accent, most participants responded in the affirmative ($\bar{x}$ rating = 7.75, range: 3-10). In order to determine whether participants’ expectations to Chinese-accented Spanish were as foreseen, they were asked to rate how often they heard native Chinese speakers produce various errors in Spanish (1=never, 10=always). The /r/ to /l/ errors (i.e., the typical errors in this study) were rated as significantly more frequent ($\bar{x}$ = 7.91) than the /n/ to /s/ errors ($\bar{x}$ = 3.91; t(34) = 9.432, p <.001), as anticipated (and regardless of past exposure to Chinese-accented Spanish). The high rating of /r/ to /l/ errors also support the idea that the productions from the foreign-accented speaker were not simply caricaturist realizations of this type of speech.

Finally, participants were asked to define all critical words presented as stimuli after the experiment. On average, the participants provided accurate definitions for over 90% of target words (50.44 out of 56, range = 41-56), indicating that the critical words were well-known by the participants. There were no differences in accuracy between critical /l/-words and critical /s/-words (F (1,54) = .236, p = .629).

3.2 Main ERP results

First time-window (350-600 ms)

Upon visual inspection of the ERP waves (see Figures 3, 4), it appears as though incorrectly pronounced target words (semantically incorrect sentences) elicited a more negative response than correctly pronounced target words (semantically correct sentences). Statistical analyses indeed confirm the presence of a main effect of Correctness ($F (1,35) = 9.048$, $p = .005$). The factor of Correctness also interacted with the topographic factor of Laterality ($F (1,35) = 4.332$, $p = .017$). Post-hoc comparisons revealed that incorrect sentences elicited significantly more negative waves than correct sentences on medial electrodes ($t (35) = 3.624$, $p_{holm} = .008$) and right electrodes ($t (35) = 2.784$, $p_{holm} = .049$), but not on left sites ($t (35) = 2.296$, $p_{holm} = .135$).

A main effect of Phoneme was also observed ($F (1,35) = 9.124$, $p = .005$), such that target words with the /l/ phoneme elicited more negative responses than target words with the /s/ phoneme. This effect interacted with Laterality ($F (2,70) = 3.633$, $p = .032$): the difference between /l/- and /s/-words was most robust on medial electrodes ($t (35) = 3.643$, $p_{holm} = .006$), compared to left ($t (35) = 2.436$, $p_{holm} = .096$) and right ($t (35) = 2.626$, $p_{holm} = .073$) sites. Lastly, an Accent * Phoneme interaction was observed ($F (2,70) = 4.871$, $p = .034$). Post-hoc comparisons revealed that /l/- and /s/-words elicited significantly different waves in the foreign accent ($t (35) = 3.722$, $p_{holm} = .002$), but not in the native accent ($t (35) = 0.773$, $p_{holm} = .885$).

No other effects were significant (all ps $\geq .159$), including the interactions of interest, notably Accent * Correctness ($F (2,70) = 0.621$, $p = .436$) and Accent * Phoneme * Correctness ($F (1,35) = 0.375$, $p = .544$), as well as the same interactions involving the topographic factor (Fs < 1.89, ps $\geq$. 159). The Bayes inclusion factor for these potential effects (all $BF_{incl} < .245$) indicated that the data was more likely to occur under the null hypothesis, providing substantial evidence for the absence of these interactions. Note that the Bayesian inclusion factors also matched the results from the ANOVA for the significant main effects of Correctness ($BF_{incl} > 5.4E+6$) and Phoneme ($BF_{incl} > 1.40E+4$), indicating the data is substantially more likely under the alternate hypothesis than the null hypothesis.

Second time-window (600-900 ms)

In the second time-window, the main effect of Correctness persisted (F (1,35) = 4.195, p = .048), with incorrect sentences eliciting significantly more negative responses than correct sentences (see Figures 3, 4). An interaction between Accent and Correctness was also observed (F (1,35) = 6.721, p = .014). This interaction was carried by the significant difference between correct and incorrect sentences in the foreign accent (t (35) = 3.269, p_{holm} = .010), but not in the native accent (t (35) = 0.243, p_{holm} = 1.000). A Bayesian analysis of these effects confirms that there is substantial evidence for the presence of this interaction (BF_{incl} = 217.923). No other main effects or interactions were significant (all ps > .20), including the Accent * Phoneme * Correctness interaction (F (1,35) = .707, p = .406; BF_{incl} = .349), and its counterpart involving the topographic factor of laterality (F (2,70) = .679, p = .510; BF_{incl} = .097).

<insert Figure 3 about here>

<insert Figure 4 about here>

Summary

From 350-600 ms, critical words containing the /l/ phoneme elicited more negative responses than /s/-words in foreign-accented speech. Semantically incorrect sentences elicited an N400 component, which was not modulated by the speakers' accent or the typicality of the errors. However, from 600-900 ms, this increased negativity persisted in foreign-accented trials, but not in native-accented trials, indicating that the late phase of the N400 effect is indeed modulated by speakers' accent, but not by error typicality.

3.3 Blocked ERP results

As an Accent * Correctness interaction was observed in the second time-window of the main analysis, we opted to investigate the nature of this effect over time. The data was segmented into two halves (blocks 1 and 2; blocks 3 and 4) in order to assess possible adaptation to the speech errors throughout the experiment. There were no differences between the number of trials included between both Blocks (F (1,1) = .562, p = .459) and no interaction between Block and conditions (F (1,7) = 2.211, p = .10). An ANOVA (2 X 2 X 2 X 2 design) with the same experimental factors was conducted on these segments, with the added factor of Experiment Block (first or last). The topographic factor of latitude was discarded, in order to ensure adequate statistical power (power = 99.9%).

First time-window (350-600 ms)

As in the main analysis for the first time window, the current analysis (see Figure 5) yielded a main effect of Phoneme (F (1,35) = 9.124, p = .005), a main effect of Correctness (F (1,35) = 9.048, p = .005) and an interaction between Accent and Phoneme (F (1,35) = 4.871, p = .034); no main effect of Block was present (F (1,35) = 2.278, p = .140, BF_{incl} = .225). An interaction between Block and Correctness was observed (F (1,35) = 4.708, p = .037), but not between Block and Phoneme (F (1,35) = .599, p = .444, BF_{incl} = .273). A post-hoc test for the Block by Correctness interaction revealed that correct and incorrect sentences elicited significantly different results in the second block (t (35) = 3.709, p_{holm} = .003), but not in the first block (t (35) = 1.211, p_{holm} = .461). A significant three-way interaction between Block, Accent, and Phoneme was observed (F (1,35) = 5.595, p = .024). However, post-hoc tests revealed that there were no significant differences between the two blocks for any of the conditions (all $p_{holm} > .44$). A low Bayes factor of inclusion ($BF_{incl} = .945$) indeed indicates that this interaction, while significant, is rather unreliable.

There was also no interaction between Block and Accent ($F(1,35) = 0.271$, $p = .606$), Block, Accent and Correctness ($F(1,35) = .067$, $p = .798$), nor Block, Accent, Correctness and Phoneme ($F(1,35) = .031$, $p = .862$; all $BF_{incl} < .161$).

<insert Figure 5 about here>

Second time-window (600-900 ms)

In the second time window, the analysis once again revealed a main effect of Correctness ($F(1,35) = 4.195$, $p = .048$), and a significant interaction between Accent and Correctness ($F(1,35) = 6.721$, $p = 0.014$), but no Accent * Phoneme * Correctness interaction ($F(1,35) = .707$, $p = .406$). These effects were not modulated by Block ($Fs < 1.798$, $ps > .189$) and the factor of Block did not yield a significant main effect ($F(1,35) = 1.230$, $p = .275$). The Bayes inclusion factor for both the Block * Accent * Correctness interaction ($BF_{incl} = .210$) and the Block * Accent * Phoneme * Correctness interaction ($BF_{incl} = .274$) provide substantial evidence against that presence of these effects.

A marginal interaction between Block and Accent was observed ($F(1,35) = 3.919$, $p = .056$). However, post-hoc tests indicated that the ERPs elicited during the first and second blocks did not differ significantly in the foreign accent ($t(35) = 0.879$, $p_{holm} = 1.000$), nor in the native accent ($t(35) = 2.243$, $p_{holm} = .170$). Indeed, a low Bayes factor of inclusion for this effect ($BF_{incl} = 0.714$) indicated that there was little compelling evidence supporting the reliability of this interaction.

Summary

The segmented analysis confirmed the effects observed in the main analysis. Furthermore, in the first time-window, the factor Block modulated the experimental factor of Correctness, with a larger N400 effect in the second block than the first. The factor Block did not impact the elicited ERP responses in the second time-window.

3.4 Relation between ERP effects and questionnaire responses

Correlations

In both time-windows, the ERP effect in both the foreign accent and in the native accent were not correlated with the participants' LexTALE results, their self-reported experience with Chinese accented-speech, their proportion of foreign friends, their tendency to let pass or correct an error, nor their self-reported feelings of reliability or trust towards the speakers (all r <. 28, ps > .10). The ratings obtained for the perceptibility ("From 1 to 10, how easy was it to identify word-by-word what the Chinese/Spanish native speaker said during the experiment?") of the foreign-accented speaker were positively correlated with the ERP effect in the foreign-accent (first time-window: r (34) = .374, p = .025; second time-window: r (34) = .424, p = .010). Contrastively, the perceptibility ratings for the native-accented speaker were not correlated to the ERP effect elicited in native accent trials (first time-window: r (34) = .132, p = .442; second time-window: r (34) = .009, p = .957). The cocor package revealed that the correlations in each accent were not significantly different from one another in the first time-window (z = 1.148, p = 0.126), but did differ significantly in the second time-window (z = 1.935, p = 0.027; i.e. the same time-window that yielded significant differences in the ANOVA analysis). These effects are represented graphically in Figure 6.

<insert Figure 6 about here>

Post-hoc experience-grouped analysis

As past research has yielded ERP differences with regard to accent familiarity (Grey & van Hell, 2017) and only 15 participants in the current study reported regular contact with Chinese-accented Spanish, we also opted to conduct an additional 2 x 2 x 2 ANOVA with the added factor of Familiarity (accent-familiar participants: n=15; $\bar{x}$ exposure = 1.5 hours per week; accent-unfamiliar participants: n=21, exposure = 0 hours per week). The other factors included Accent (foreign, native) and Correctness (correct, incorrect); the data was collapsed across critical phonemes and parietal ROIs. In the first time-window, Familiarity was not significant, and did not interact with any other experimental factor ($F < 1.4$, $p > .23$, $BF_{incl} < .39$). The same null results were observed in the second time-window ($F < 2.3$, $p \geq .14$, $BF_{incl} < 1$). Altogether, these results suggest that, when treated categorically, Familiarity did not impact the results.

4. Discussion

The current study examined whether native listeners are more tolerant to typical speech errors when they are produced in foreign-accented (vs. native-accented) speech. To do so, speakers of Spanish listened to sentences recorded by a native Spanish speaker and a Chinese-accented speaker of Spanish. Some of these stimuli contained typical or atypical phonological mispronunciations, which led to sentential semantic abnormality. The findings provided no evidence for native listeners' tolerance to foreign-accented errors. Indeed, mispronunciations produced in Chinese-accented Spanish elicited longer-lasting negativities relative to native-accented trials (rather than mitigated effects, as would be expected in the case of reduced error sensitivity).

In the earlier time-window (350-600 ms), semantically incorrect sentences were indexed by a predominantly central-posterior increased negativity. Crucially, this N400 effect was not modulated by the speaker's accent, nor by the typicality of the mispronunciations which yielded the semantic abnormality. This means that participants experienced lexical-semantic integration difficulties for mispronunciations, regardless of whether they were produced in a native or a foreign accent, and irrespective of their assumed frequency/typicality. The findings obtained in the first time-window are thus similar to the overall results observed by Hanulíková et al. (2012) and Holt, et al. (2018). Indeed, both of these studies found congruous N400 effects for foreign-accented and native-accented speech in response to outright semantic violations (though see modulating effects of familiarity in Holt et al., 2018). Although no tolerance was observed towards foreign-accented mispronunciations, foreign- and native-accented trials displayed similar electrophysiological correlates in regard to semantic analysis. The null effect observed in the first time-window of the current study supports this notion: during the initial stages of processing, the ERP signature attributed to semantic processing was homogenous across both accents.

This first time-window additionally yielded a main effect of phoneme, such that critical /l/-words elicited significantly more negative responses than critical words containing the /s/ phoneme. Research indeed indicates that, within this time frame, lateral phonemes elicit more negative ERP responses than fricatives over parietal sites (Kovács, Winkler & Vicsi, 2017). An interaction between accent and phoneme was also observed, revealing that the effect was particularly robust in the foreign accent. This suggests that the accent was clearly detected from the early stages of processing, and that it may have introduced more variable phonetic features into the speech input. As such, it appears that the presence of a foreign accent may influence aspects of lexical analysis, presumably those related to phonology, minimally at the initial stages of semantic analysis.

In the later time-window (600-900 ms), a sustained N400 effect was observed, but only for foreign-accented sentences. This effect was not modulated by the typicality of the phonological mispronunciations (i.e. typical /l/ errors compared to atypical /s/ errors). As such, while mispronunciations initially yielded comparable heightened processing costs in both accents, the processing cost persisted uniquely for foreign-accented speech. This finding is comparable to previous studies which reported stronger (Grey & van Hell, 2017, Song & Iverson, 2018) or more largely distributed (Romero-Rivas et al., 2015) negativities in response to semantic abnormality in foreign-accented speech than in native speech. However, it contrasts Xu et al.'s (2020) study: for semantic abnormality induced by slips of the tongue, these authors found an N400 for native-accented trials but *not* for foreign-accented trials.

In this case, the long lasting N400 effect elicited by foreign accented speech (but not by native-accented speech) suggests that lexical access is more cognitively effortful for listeners who are attending to non-native input. In other words, the observed processing cost may be due to the fact that it is more difficult to identify words in real-time in foreign-accented speech, potentially due to the degraded speech signal (see Romero-Rivas et al., 2015 for a similar interpretation). Indeed, a wealth of behavioural literature illustrates that a foreign accent may impede listeners' word identification and comprehension (e.g., Braun et al., 2011; Floccia et al., 2009; Clarke & Garrett, 2004; van Wijngaarden, 2001). Relatedly, electrophysiological studies indicate that foreign-accented speech leads to a reduction in ERP components attributed to phonological processing (see Crowley & Colrain, 2004; Newman & Connolly, 2009), such as the P2 (Romero-Rivas et al., 2015) or the Phonological Mapping Negativity response (PMN; Goslin et al., 2012). A reduction in these components suggests that "the extraction of phonetic/acoustic information" is more difficult in foreign-accented speech compared to native speech (Romero-Rivas et al., 2015, p.7). Such a complication may very well hinder lexical access, which is indexed within the N400 (e.g., Lau, Almeida, Hines & Poeppel, 2009). This explains why a latent increased negativity was observed uniquely for foreign-accented speech. Going further, the present study indicates that, even though native listeners can successfully understand foreign-accented speech (as indicated by similar transcription accuracies for both speakers during the norming procedure), the time course of their online semantic analysis is very much accent-dependent, with a persistent processing difficulty for foreign-accented, but not native-accented speech.

Let us note that the difficulty regarding the deviated foreign-accented speech signal may be (at least partially) circumvented. Indeed, Goslin et al. (2012) suggest that top-down information can be particularly useful for listeners who are attending to non-native input, as the phonological information is unstable. Indeed, studies testing native speakers attending to speech in degraded listening environments (i.e. speech-in-noise) demonstrate that listeners actually restrict their lexical expectations regarding upcoming input when listening conditions are not ideal, perhaps as a result of cognitive overload (Strauß, Kotz & Obleser, 2013). Importantly, top-down processing mechanisms and reliable contextual cues were not available to the participants in the current study, since sentences were semantically low-constraining. As such, the observed lexical processing cost for foreign-accented mispronunciations may be due to the absence of two crucial features: stable phonological information and reliable contextual cues. Perhaps the absence of such features also characterizes similar real-life speech situations, particularly those involving individuals with contrasting accents who are unknown to each other.

This line of reasoning may also begin to explain why the main result of the current study contrasts Xu et al.'s (2020) finding with respect to slips of the tongue. Even though our stimuli are more similar to slips of the tongue than outright violations (i.e., both slips of the tongue and mispronunciation errors still result in an output that resembles the intended target), Xu et al. (2020) report no negativity for foreign-accented slips, while we detected an increased late negativity for foreign-accented speech errors. Once again, perhaps this divergence is the consequence of contextual cues, which were arguably more present in the

Xu et al. study than in the current experiment. Firstly, these authors presented their stimuli alongside photos of Asian and Caucasian faces (in accordance with the accent-condition of the trial), supplying the participants with precursory phonological information. Secondly, though the authors do not present cloze probability scores for targets in their stimuli, slips of the tongue appear to have occurred later in the sentence than in the current study; thus, Xu et al.'s participants may have had narrowed semantic expectancies, facilitating lexical processing. Relatedly, the slips of tongue in Xu et al. (2020) did not give rise to other real words, but rather pseudo-words or "illegal constitutes" (p.3). Undoubtedly, this design produced less lexical competition than our own (wherein mispronunciations produced other real Spanish words), resulting in the greater possibility for participants to assimilate the error (i.e. they could have been judged as phonological lapses rather than semantic anomalies). Altogether, these factors may have led to the absence of an N400 in the Xu et al. study, but not our own.

Going back to the latent processing cost observed for foreign-accented speech in the present study, a second possibility that might account for this finding is the participants' perceived cognitive effort towards foreign-accented speech. This is conceivable, as the current study revealed a correlation between the offline perceptibility ratings and the effect of semantic correctness: individuals who rated the foreign-accent speech as more effortful to perceive (i.e. those that gave it the lowest perceptibility rating) displayed the largest N400 effects. Crucially, the same correlation was not observed for native-accented speech, even though the native speaker also produced semantically anomalous sentences. This suggests that the impact of perceived cognitive effort is unique to foreign-accented speech, perhaps as negative biases already exist towards this type of input (e.g., Gluszek & Dovidio, 2010). Recent behavioural research indeed indicates that perceptions about foreign-accented speech can affect its processing. For instance, Vaughn (2019) demonstrated that participants who are instructed to transcribe L1 speech perform more accurately than those who are told that the same speaker is non-native, even if the fluency of the L2 speaker is emphasized.

In the current study, it is also feasible that this perceived cognitive effort was impacted by the participants' lack of veritable exposure to the accent in question. Previous studies have illustrated that familiarity with a given accent may indeed modulate its processing. In the syntactic domain, less familiar individuals show the greatest grammaticality-related ERP effects (e.g., Caffarra & Martin, 2019; Grey & van Hell, 2017; see Porretta, Tremblay & Bolger, 2017 for comparable results regarding the PMN); in the semantic domain, less familiar individuals show the smallest negative ERP effects (Holt et al., 2018). In the current paper, we did not observe familiarity-related differences. However, we note that even the 'accent-familiar' participants self-reported relatively infrequent exposure to Chinese-accented speech. Future studies are needed in order to examine whether late effects of cognitive effort are modulated by foreign accent familiarity.

To recapitulate, native listeners were not tolerant to subtle speech errors when they were produced with a foreign accent, even though mispronunciations (particularly /r-l/ errors in Chinese-accented Spanish) are more common in foreign-accented speech than in native speech (see Ortí Mateu, 1996; Chen, 2007). In theory, participants' real-world knowledge about the foreign accent in question could have allowed them to unconsciously 'correct' the speech error, and therefore, to process semantically abnormal sentences as correct even when they contained a mispronunciation (perhaps similarly to the phoneme restoration effect; Samuel, 1996). Evidence for this 'self-correction' was indeed demonstrated in similar studies examining the processing of foreign-accented syntactic errors (e.g., Caffarra & Martin, 2019; Grey & van Hell, 2017; Hanulíková et al., 2012; Xu et al., 2020). From our data, it is difficult to draw conclusions about the origin of this potential tolerance to foreign-accented speech (e.g., ease of semantic integration or lexical access, assimilation of the phonological error, etc.), as we did not observe this effect; we leave this empirical question for future research.

The question may be asked as to why the findings relative to foreign-accented semantic processing clearly differ from those reported by similar research in the syntactic domain. In these studies, grammatical gender errors elicited reduced P600 responses when they were produced in a foreign accent, but not when they were produced in the native accent (Caffarra & Martin, 2019, Hanulíková et al., 2012; Xu et al., 2020), suggesting that listeners possessed a relaxed sensitivity towards non-native errors. On the other hand, semantic processing of foreign-accented speech is marked by an intolerance (i.e. a processing cost) on the native listener's behalf. A possible explanation for the distinction between the results found in the syntactic and semantic domains derives from the nature of processing that both types of errors evoke.

Lexical-semantic N400 effect have been attributed to automatic processing mechanisms (e.g., Kiefer, 2002; Liu, Wu, Meng & Dang, 2013; Wang, Huang & Mao, 2009; though see Kutas & Federmeier, 2011 for a review of competing positions), while syntactically-driven ERP positivities index more controlled processes (Hahne & Friederici, 1999). Importantly, frequency distributions have been found to impact controlled processes, like the P600, but *not* automatic processes: P600 effects tend to be absent for highly frequent violations (Hahne & Friederici, 1999). This notion is important, as speech errors possess distinct distributional frequencies according to accent. Notably, they occur at higher rates in foreign-accented speech than in native-accented speech. This may explain why a tolerance is observed in response to foreign-accented syntactic errors but not towards semantic errors. In short, syntactic errors elicit controlled P600 responses, and are therefore sensitive to the higher distributional frequency of errors in foreign-accented speech. Contrastively, no tolerance is found for foreign-accented semantic violations, as such errors elicit automatic

N400 responses that are not impacted by this same frequency distribution. An identical argumentation can also be applied to the factor of error typicality, which modulates syntactically-driven P600 effects, but not the more automatic N400 effects in the semantic domain, as in the current study. Indeed, Caffarra & Martin's results (2019) support this notion: these authors observed reduced P600 responses for typical (but not atypical) syntactic errors throughout their experiment, but no modulation of the N400 component.

Regarding the main analysis, one final matter to address is the possible effect of semantic constraint on the results of the current study. As was noted, the stimuli included in the experiment were semantically low-constraining (similar to Caffarra & Martin, 2019); it is possible that the low-predictability of the sentences may have impacted the observed findings. Up to this point, research examining semantic processing in foreign-accented speech appear to have primarily analyzed predictable target words embedded into constraining sentences (e.g., Hanulíková et al., 2012). In native speech, it has been shown that semantically abnormal sentences elicit N400 responses even in the case of low predictability (Thornhill & Van Petten, 2012). Recent evidence indeed suggests that this component reflects both prediction and integration (see Nieuwland et al., 2020). As such, while we theorize that the long-lasting negativity observed for foreign-accented speech reflects semantic integration difficulties, at least one past study has demonstrated that anticipatory processes may be accent-dependent (Romero-Rivas, Martin & Costa, 2016). More research is thus required in order to disentangle the effects of predictability versus integration in non-native input.

Finally, in regard to the processing of the speech errors across the time course of the experiment, participants showed 'reverse adaptation' to the mispronunciations in both accents. That is, the difference between correct and incorrect sentences was more robust in the second block than the first (regardless of accent), suggesting that the subtle phonological variations (leading to semantic abnormality) were more strongly perceived in the latter half of the study; in turn, this also suggests that the processing costs attributed to incorrect sentences may have increased throughout the study. These results are somewhat at odds with those observed by Romero-Rivas et al., (2015), who reported a reduction in amplitude of the N400 response in the second half of their experiment for foreign-accented correct sentences (see also Caffarra & Martin, 2019 for similar results with semantically abnormal sentences). However, the findings are comparable to those observed by Xu et al. (2020): in response to grammatical gender errors, these authors noted a P600 effect only in the last half of the study. These contradictory results may be partially due to other confounding factors, such as familiarity, which has been found to modulate foreign-accented speech processing (e.g., Holt et al., 2018). While adaptation to foreign-accented speech is detected at the behavioural level (e.g., Clarke & Garrett, 2004), support for this phenomenon remains somewhat conflicting within the ERP literature. Comparable studies using other online techniques, such as eye-tracking, find little evidence for adaptation to non-native speech (Trude, Tremblay & Brown-Schmidt, 2013). These results fall into line with the findings from the current study, as participants consistently elicited N400 responses throughout the time course of the experiment.

To conclude, past research examining the processing of foreign-accented syntactic errors has largely demonstrated that listeners are tolerant to such speech errors, particularly when they are frequent in non-native speech. However, similar studies examining outright semantic errors provide conflicting results. The current project thus examined more subtle semantic anomalies produced in foreign-accented speech (i.e. typical and atypical single-phoneme mispronunciations leading to semantic incoherencies). The results indicate that native listeners' semantic processing is accent-dependent: in early stages of processing, the presence of a foreign-accent may influence aspects of lexical analysis (presumably related to phonology), while in later stages, foreign-accent introduces a persistent difficulty with lexical access. This means that Spanish native listeners do not show a relaxed sensitivity to subtle speech errors, even those that are frequent, and therefore expected, in Chinese-accented Spanish. Nevertheless, the late semantic processing cost does not appear to prevent native speakers from reaching a high comprehension performance. The current study demonstrates that a clear delineation exists between syntactic and semantic processing of foreign-accented speech, perhaps due to the differing automaticity of these two types of underlying processing mechanisms.

Notes:

[1] For instance, semantic violations in past studies differed with respect to the position of the critical word in each sentence (fixed vs. variable; close to the sentence-onset vs. close to the end of the sentence). The absence of a common standard for EEG analyses should also be mentioned, as differences regarding time-windows, the number and distribution of electrodes chosen, as well as other criteria, may impact results.

[2] Note that intelligibility, comprehension and error identification are not necessarily interrelated (see Munro & Derwing, 1995; Derwing & Munro, 1997). As such, it is not contradictory to report that respondents found the speakers similarly intelligible, but were still less successful at detecting errors in foreign-accented speech as compared to native speech.

[3] See supplementary materials for an analysis including all 27 electrodes and the additional topographic factor of Longitude (frontal, central, parietal). Additional figures involving the

topographic factor of Latitude, as well as individual conditions for the Block analysis, are included in the supplementary material.

Acknowledgements: This international collaboration was made possible by a Globalink Research Award Abroad (MITACS), as well as a CGS-M scholarship (SSHRC) conferred to the first author. The research was supported by the Spanish Ministry of Economy and Competitiveness (SEV-2015-490 to the BCBL; PSI2017-82941-P to C.D.M.), the Basque government (PIBA18_29 to C.D.M.), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 837228 to S.C. and No 819093 to C. D. M.). The authors are grateful for the help of all research assistants at the BCBL who aided with data collection and the two anonymous Reviewers whose comments greatly improved the manuscript.

Disclosure of interest: The authors declare no competing interests.

Data availability statement: A full stimuli list, as well as the offline questionnaire and the additional analysis and figures can be found online at: https://github.com/leahgosselin/Pelo-Perro. All data are available upon request.

References

Balconi, M., & Pozzoli, U. (2005). Comprehending semantic and grammatical violations in Italian. N400 and P600 comparison with visual and auditory stimuli. *Journal of Psycholinguistic Research*, *34*(1), 71–98. https://doi.org/10.1007/s10936-005-3633-6

Bent, T., Bradlow, A. R., & Smith, B. (2007). Phonemic errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In M. J. Munro, & O. S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 331–347). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.28ben

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down online speech comprehension. *Language and Cognitive Processes*, *26*(3), 350–375. https://doi.org/10.1080/01690965.2010.492641

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research*, *33*(2), 103–135. https://doi.org/10.1023/B:JOPR.0000017223.98667.10

Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, *125*(1), 82–93. https://doi.org/10.1016/j.bandl.2013.01.007

Caffarra, S., & Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error typicality in foreign accented speech perception. *Cortex*, *116*, 308–320. https://doi.org/10.1016/j.cortex.2018.03.007

Chen, Y. (2007). *A comparison of Spanish produced by Chinese L2 learners and native speakers—an acoustic phonetics approach*. [Doctoral dissertation, University of Illinois at Urbana-Champaign]. http://hdl.handle.net/2142/86153

Cheng, L. R. L. (1999). Moving beyond accent: Social and cultural realities of living with many tongues. *Topics in Language Disorders*, *19*(4), 1–10. https://doi.org/10.1097/00011363-199908000-00004

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Córdoba, R., Montero, J. M., Gutiérrez, J. M., Vallejo, J. A., Enriquez, E., & Pardo, J. M. (2002). Selection of the most significant parameters for duration modelling in a Spanish text-to-speech system using neural networks. *Computer Speech & Language*, *16*(2), 183–203. https://doi.org/10.1006/csla.2002.0190

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1–16. https://www.jstor.org/stable/44488664

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLOS ONE*, *10*(6), e0131499. https://doi.org/10.1371/journal.pone.0131499

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, *45*, 1246–1258. https://doi.org/10.3758/s13428-013-0326-1

Eichstaedt, K. E., Kovatch, K., & Maroof, D. A. (2013). A less conservative method to adjust for familywise error rate in neuropsychological research: The Holm's sequential Bonferroni procedure. *NeuroRehabilitation*, *32*(3), 693–696. https://doi.org/10.3233/NRE-130893

Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research*, *38*(4), 379–412. https://doi.org/10.1007/s10936-008-9097-8

Gluszek, A., & Dovidio, J. F. (2010). Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology*, *29*(2), 224–234. https://doi.org/10.1177/0261927X09359590

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108. https://doi.org/10.1016/j.jneuroling.2016.12.001

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531–1549. https://doi.org/10.1016/S0028-3932(00)00053-1

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. https://doi.org/10.1162/089892999563328

Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*(2), 199–212. https://doi.org/10.1016/S0926-6410(00)00071-9

Hanulíková, A., Van Alphen, P. M., Van Goch, M. 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. https://doi.org/10.1162/jocn_a_00103

Holt, R., Kung, C., & Demuth, K. (2018). Listener characteristics modulate the semantic processing of native vs. foreign-accented speech. *PLOS ONE*, *13*(12), e0207452. https://doi.org/10.1371/journal.pone.0207452

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49–66. http://hdl.handle.net/1854/LU-5774107

Kaan, E., & Swaab, T. Y. (2003). Repair, Revision, and Complexity in Syntactic Analysis: An Electrophysiological Differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98–110. https://doi.org/10.1162/089892903321107855

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: Further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*(1), 27–39. https://doi.org/10.1016/S0926-6410(01)00085-4

Kovács, A., Winkler, I., & Vicsi, K. (2017). EEG correlates of speech: examination of event related potentials elicited by phoneme classes. *Proceedings of the 8th IEEE*

International Conference on Cognitive Infocommunications (CogInfoCom), 000115-000120. Debrecen, Hungary, September 11-14.. https://doi.org/10.1109/CogInfoCom.2017.8268226

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

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

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and Language*, *111*(3), 161–172. https://doi.org/10.1016/j.bandl.2009.08.007

Liu, B., Wu, G., Meng, X., & Dang, J. (2013). Correlation between prime duration and semantic priming effect: Evidence from N400 effect. *Neuroscience*, *238*, 319–326. https://doi.org/10.1016/j.neuroscience.2013.02.010

McCallum, W. C., Farmer, S. F., & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and*

Clinical Neurophysiology, *59*(6), 477–488. https://doi.org/10.1016/0168-5597(84)90006-6

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–121. https://doi.org/10.1016/j.biopsycho.2008.04.008

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Husband, E. M., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … & Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Ortí Mateu, R. (1996). Comparación fonética, diagnóstico y tratamiento de las dificultades de los estudiantes chinos para aprender español [Phonetic comparison, diagnosis and treatment of the difficulties of Chinese students to learn Spanish]. In P. Guerrero Ruiz, & A. López Valero (Eds.), *Aspectos de didáctica de la lengua y la literatura: actas del*

III Congreso Internacional de la Sociedad española de didáctica de la lengua y la literatura (pp. 319–334). Servicio de Publicaciones.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Pichardo, E. (1953). *Pichardo Novísimo o Diccionario Provincial casi Razonado de Vozes y Frases Cubanas* [Almost Reasoned Provincial Dictionary of Cuban Voices and Phrases]. (E. Rodríguez Herrera, Ed.). La Habana: Editorial Selecta.

Porretta, V., Tremblay, A., & Bolger, P. (2017). Got experience? PMN amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics*, *44*, 54–67. https://doi.org/10.1016/j.jneuroling.2017.03.002

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience*, *9*, 167. https://doi.org/10.3389/fnhum.2015.00167

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255. https://doi.org/10.1016/j.neuropsychologia.2016.03.022

Samuel, A. (1996). Phoneme restoration. *Language and Cognitive Processes, 11*(6), 647–654. https://doi.org/10.1080/016909696387051

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*, 163–170. https://doi.org/10.1016/j.cognition.2018.06.001

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395. https://doi.org/10.1162/jocn_a_00389

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. https://doi.org/10.1016/j.ijpsycho.2011.12.007

Trude, A. M., Tremblay, A., & Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *Journal of Memory and Language*, *69*(3), 349–367. https://doi.org/10.1016/j.jml.2013.05.002

van de Meerendonk, N., Kolk, H. H., Vissers, C. T. W., & Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82. https://doi.org/10.1162/jocn.2008.21170

van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, *35*(1-2), 103–113. https://doi.org/10.1016/S0167-6393(00)00098-4

Vaughn, C. R. (2019). Expectations about the source of a speaker's accent affect accent adaptation. *The Journal of the Acoustical Society of America*, *145*(5), 3218–3232. https://doi.org/10.1121/1.5108831

Wang, Q., Huang, H., & Mao, L. (2009). N400 repetition effect in unidentifiable Chinese characters: evidence for automatic process. *NeuroReport*, *20*(7), 723–728. https://doi.org/10.1097/WNR.0b013e32832ad310

Xu, J., Abdel Rahman, R., & Sommer, W. (2020). Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials. *Language, Cognition and Neuroscience*, *35*(6), 752-765. https://doi.org/10.1080/23273798.2018.1562558

Young-Scholten, M. (2011). *The acquisition of prosodic structure in a second language* (Vol. 304). Walter de Gruyter. https://doi.org/10.1515/9783111715254

Table 1. Language background details for participant sample (*n*=36). Proficiency measure is reflective of the LexTALE test result (Izura, Cuetos & Brysbaert, 2014). Standard deviations are indicated in parentheses.

	Proficiency (%)	**Amount of exposure (%)**	**Amount of speaking (%)**
Spanish (L1)	95.6 (4.4)	65.4 (15.4)	67.4 (18.0)
Basque (L2)	81.3 (10.2)	26.9 (14.3)	25.4 (16.0)
English (L3 or L4)	65.3 (8.7)	6.9 (7.2)	6.0 (6.5)

Table 2. Example of experimental materials. Target nouns are underlined. (Note: the meaning of the first Spanish example is roughly "*His/her face/cove* was happy the whole time.*")

Phoneme	Correctness	Sentence	Translation
/l/	Correct	*Visité la **cala** que me dijiste.*	(I) Visited the cove that you told me about.
	Incorrect (typical error)	*Tuvo la **cala** feliz todo el tiempo.*	(He/She) Got the cove* happy the whole time. ***Intended***: (He/She) Got the face ('*cara*') happy the whole time.
/s/	Correct	*Masticó mucho la **pasa** y luego la escupió.*	(He/She) Chewed the raisin and then spit it out.
	Incorrect (atypical error)	*Me cosió la **pasa** de la chaqueta.*	(He/She) Sewed the raisin* of the jacket. ***Intended:*** (He/She) Sewed the corduroy ('*pana*') of the jacket.

Table 3. Language background characteristics of both recorded speakers.

Speaker	**Foreign speaker**	**Native speaker**
Age (years)	21	22
Chinese AoA (years)	0	--
Spanish AoA (years)	8	0
Daily usage of Chinese (%)	50	0
Daily usage of Spanish (%)	50	70*

*Note: the native speaker also reports usage of Basque and English (25% and 5% respectively).

Table 4. Lexical properties of target nouns (extracted from the EsPal database, Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013).

	/l/			**/s/**			**t-tests**	
Lexical Property	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*	*t*	*p*
Log frequency	1.16	0.67	0.14–2.50	1.23	0.90	0.05–3.36	0.36	.72
N˚ syllables	1.93	0.38	1–3	2.07	0.47	1–3	1.26	.21
Phono. neighbor	43.22	17.68	9–83	35.07	20.82	2–85	1.56	.12
Familiarity*	5.50	1.08	3.00–6.78	5.46	1.02	4.05–6.76	0.10	.92
Duration (ms)	366.0	78.27	207–577	441.2	105.75	219–754	8.55	<.001

*Note: the measure of familiarity was available for approximately 62% of target words. Given the high matchedness of these properties, it is unlikely that the remaining target words introduce any significant differences between the conditions.

Figure captions

Figure 1. Experimental design; a trial which includes a comprehension question is depicted (Translation: 'Did he eat something?').

Figure 2. Electrode distribution, including topographical factor of Laterality. Electrodes used in the analysis are indicated in the display.

Figure 3. Average ERP responses observed across all conditions for both accents, collapsed across the three ROIs. Shaded areas represent the standard error across participants for each condition. The time-windows with the significant Accent * Correctness effect (600-900 ms) are highlighted in yellow.

Figure 4. *Top*: Difference waves for each accent. Shaded areas represent the standard error across participants for each condition. The time-window with the significant Accent * Correctness effect (600-900 ms) is highlighted in yellow. *Bottom*: Individual variation in the N400 effect in both time windows. The topography of each effect is also depicted. *Note*: For all figures, the voltage value for N400 effect was obtained by subtracting the correct from the incorrect conditions, collapsed across phoneme types (/l/ or /s/), as the factor of error typicality did not interact with the Accent * Correctness interaction.

Figure 5. Difference waves for each accent in both blocks. The voltage value for N400 effect was obtained by subtracting the correct from the incorrect conditions, collapsed across phoneme types (/l/ or /s/). The time-window with significant effect of Correctness by Block (350-600 ms) is highlighted in yellow. *Note*: For all figures, the ERP response corresponds to

an average across the three ROIs, as the topographic factor of Latitude was not included in the Block analysis. Shaded areas represent the standard error across participants for each condition.

Figure 6. Correlation between participants' N400 effect size and their perceptibility ratings for both the foreign-accented and native-accented speaker. The voltage value for the N400 effect was obtained by subtracting the correct from the incorrect conditions, collapsed across phoneme types (/l/ or /s/) and ROIs.

Figure 1. Experimental design; a trial which includes a comprehension question is depicted (Translation: ‘Did he eat something?’).

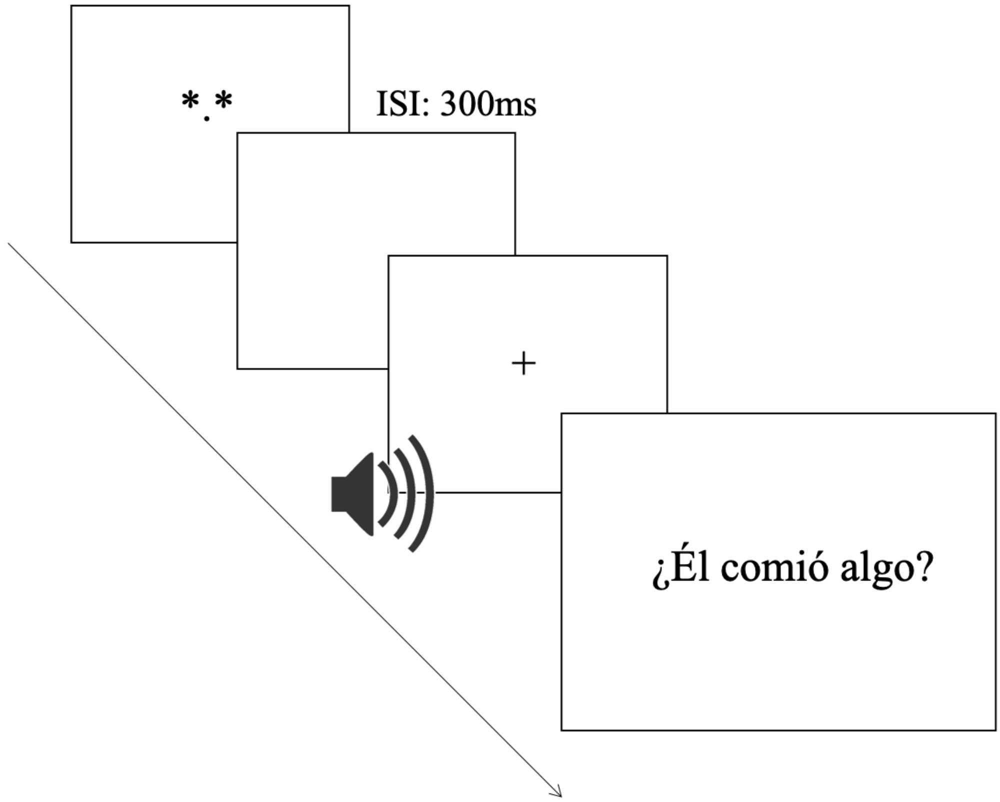

Figure 2. Electrode distribution, including topographical factor of Laterality. Electrodes used in the analysis are indicated in the display.

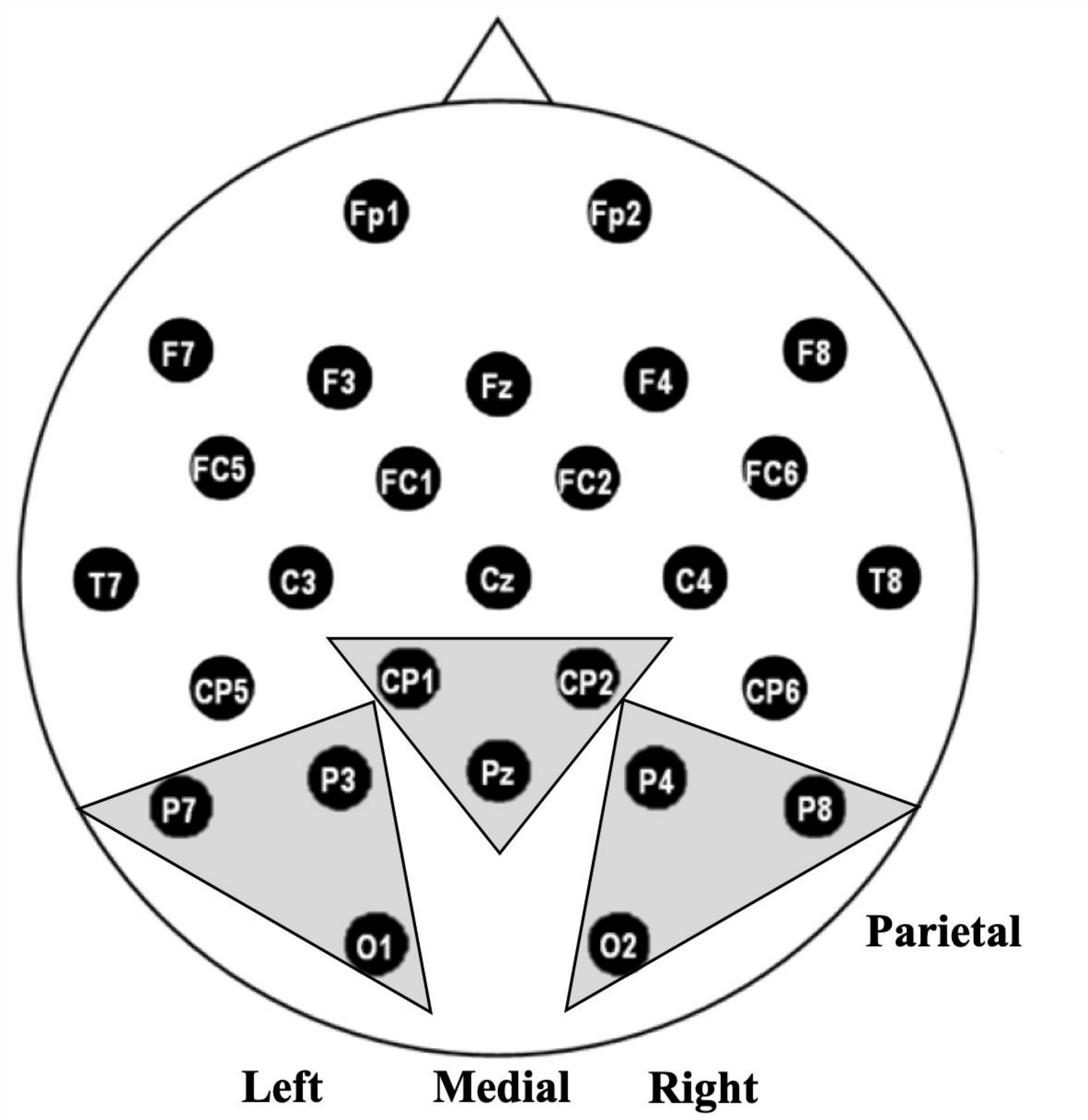

Figure 3. Average ERP responses observed across all conditions for both accents, collapsed across the three ROIs. Shaded areas represent the standard error across participants for each condition. The time-windows with the significant Accent * Correctness effect (600-900 ms) are highlighted in yellow. The topography of each effect (Incorrect – Correct trials) is also depicted.

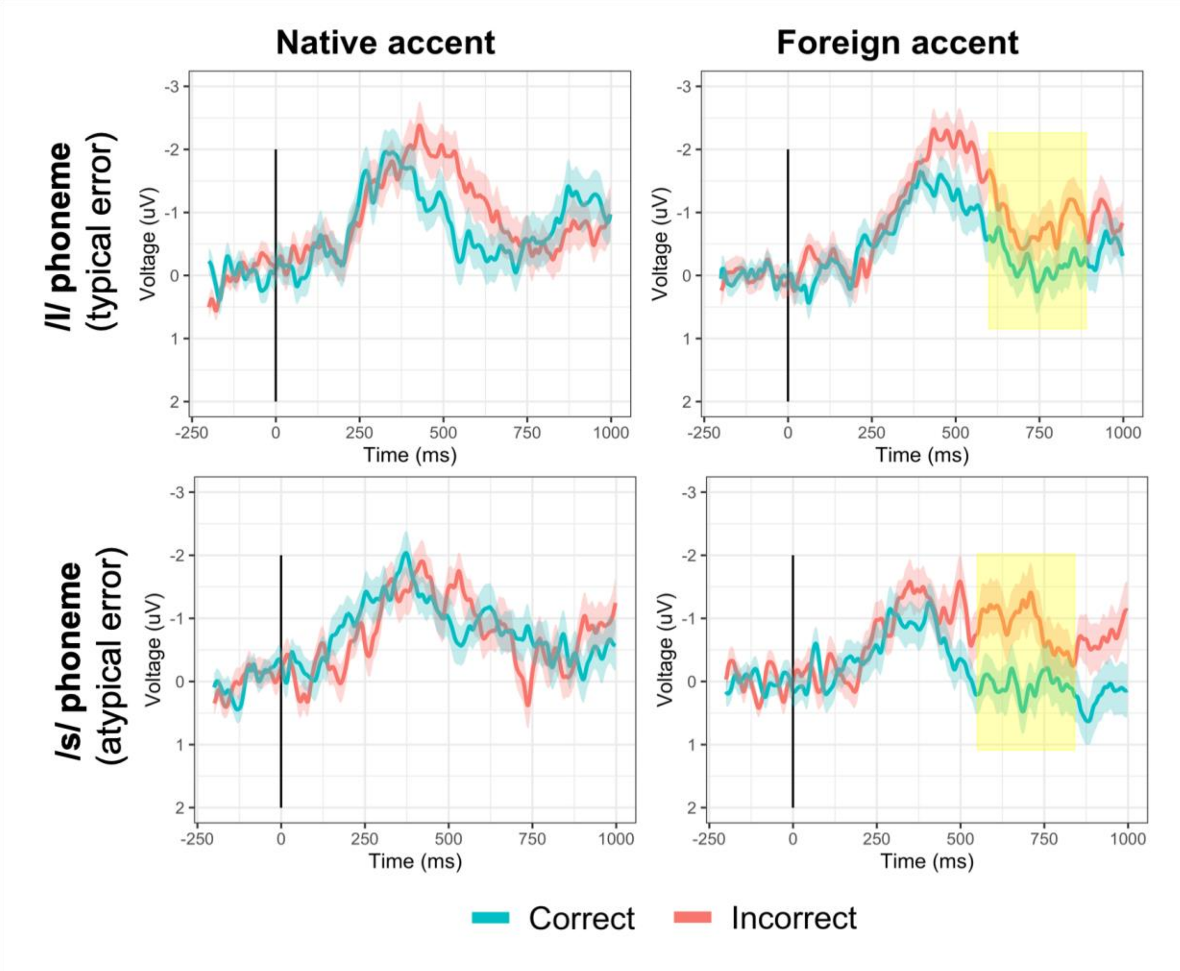

Figure 4. *Left*: Difference waves for each accent. Shaded areas represent the standard error across participants for each condition. The time-window with the significant Accent * Correctness effect (600-900 ms) is highlighted in yellow. *Right*: Individual variation in the N400 effect in both time windows. *Note*: For all figures, the voltage value for N400 effect was obtained by subtracting the correct from the incorrect conditions, collapsed across phoneme types (/l/ or /s/), as the factor of error typicality did not interact with the Accent * Correctness interaction.

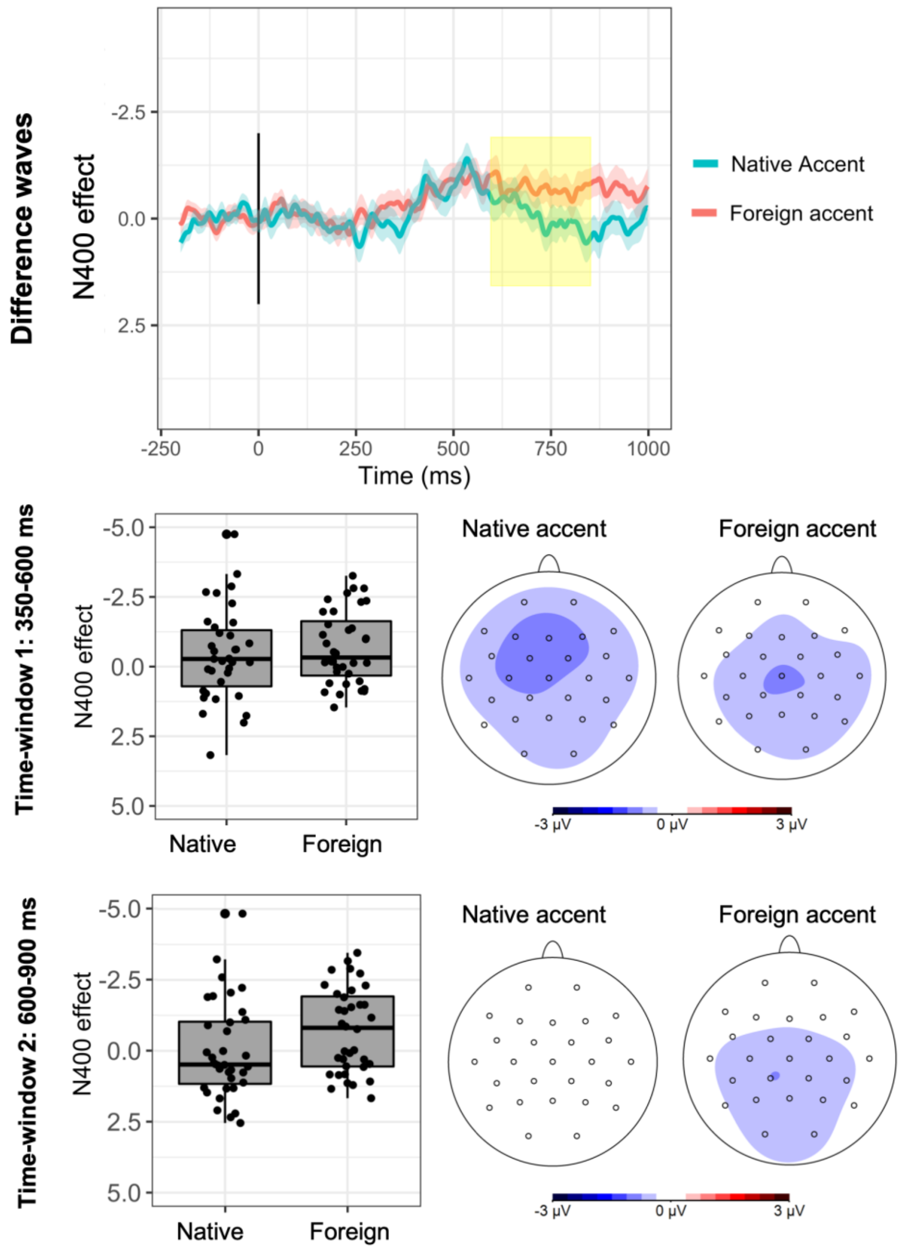

Figure 5. Difference waves for each accent in both blocks. The voltage value for N400 effect was obtained by subtracting the correct from the incorrect conditions, collapsed across phoneme types (/l/ or /s/). The time-window with significant effect of Correctness by Block (350-600 ms) is highlighted in yellow. *Note*: For all figures, the ERP response corresponds to an average across the three ROIs, as the topographic factor of Latitude was not included in the Block analysis. Shaded areas represent the standard error across participants for each condition.

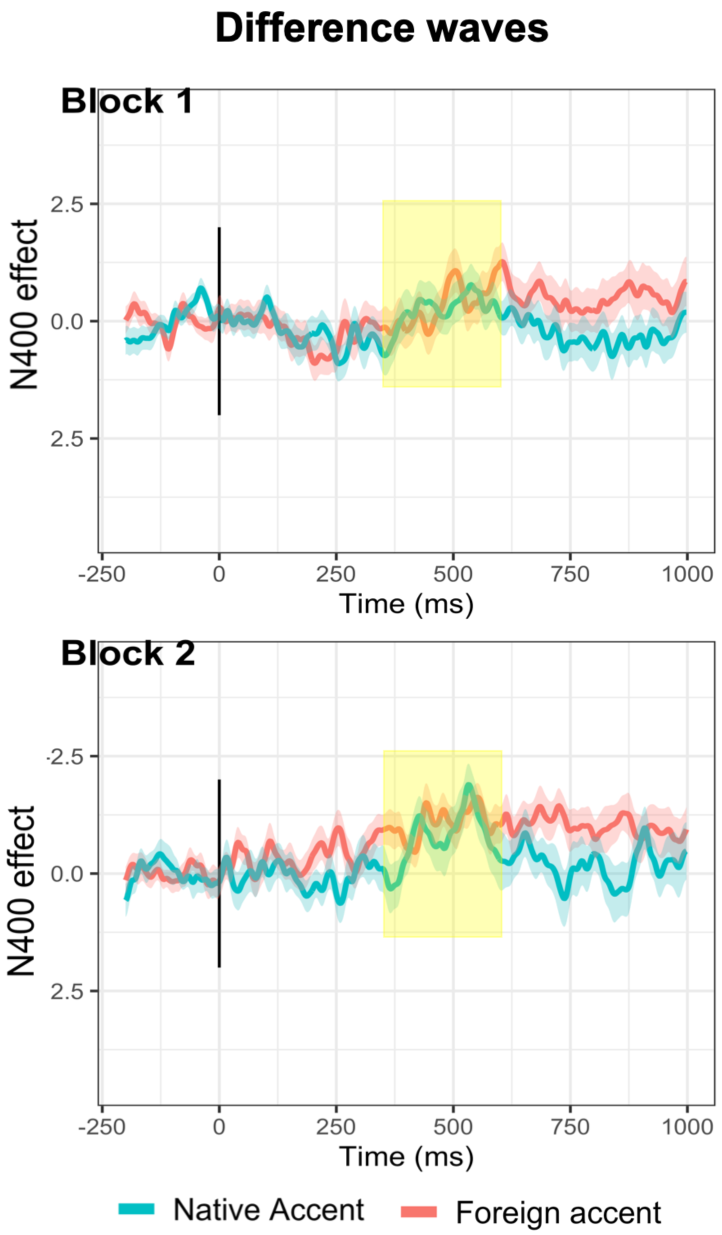

Figure 6. Correlation between participants' N400 effect size and their perceptibility ratings for both the foreign-accented and native-accented speaker. The voltage value for the N400 effect was obtained by subtracting the correct from the incorrect conditions, collapsed across phoneme types (/l/ or /s/) and ROIs.