Title

Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals

Authors

Bice, Kinsey
Kroll, Judith F

Publication Date

2021-05-01

DOI

10.1016/j.jneuroling.2020.100963

Published in final edited form as:
J Neurolinguistics. 2021 May ; 58: . doi:10.1016/j.jneuroling.2020.100963.

Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals

Kinsey Bice[1,2], **Judith F. Kroll**[3]

[1]Department of Psychology, Pennsylvania State University

[2]Department of Psychology, University of Washington

[3]Department of Language Science, University of California, Irvine

Abstract

Recent studies have demonstrated variation in language processing for monolingual and bilingual speakers alike, suggesting that only by considering individual differences will an accurate picture of the consequences of language experience be adequately understood. This approach can be illustrated in ERP research that has shown that sentence contexts that traditionally elicit a P600 component in response to a syntactic violation, elicit an N400 response for a subset of individuals. That result has been reported for monolingual speakers processing sentences in their L1 and also for bilinguals processing sentences in their L2. To date, no studies have compared variation in L1 and L2 ERP effects in the very same bilingual speakers. In the present paper, we do that by examining sentence processing in heritage bilinguals who acquired both languages from early childhood but for whom the L2 typically becomes the dominant language. Variation in ERPs produced by the non-dominant L1 and dominant L2 of heritage bilinguals was compared to variation found in monolingual L1 processing. The group-averaged results showed the smallest N400 and P600 responses in the native, but no longer dominant, L1 of heritage bilinguals, and largest in the monolinguals. Individual difference analyses linking ERP variation to working memory and language proficiency showed that working memory was the primary factor related to monolingual L1 processing, whereas bilinguals did not show this relationship. In contrast, proficiency was the primary factor related to ERP responses for no longer dominant L1 for bilinguals, but unrelated to monolingual L1 processing, whereas bilinguals' dominant L2 processing showed an intermediate relationship. Finally, the N400 was absent for bilinguals performing the task in the same language in which they initially learned to read, but significantly larger when bilinguals performed the task in the other language. The results support the idea that proficient bilinguals utilize the same underlying mechanisms to process both languages, although the factors that affect processing in each language may differ. More broadly, we find that bilingualism is an experience that opens the language system to perform fluidly under changing

---

Direct correspondence to: Kinsey Bice, Department of Psychology, University of Washington, 119A Guthrie Hall, Seattle, WA 98195 USA, biceroni@gmail.com.

circumstances, such as increasing proficiency. In contrast, language processing in monolinguals was primarily related to relatively stable factors (working memory).

Keywords

individual differences; ERPs; grammatical processing; bilingualism; heritage language

Research on second language (L2) processing often compares learners or bilinguals with monolingual native speakers to answer questions about constraints on acquisition, such as whether adult learners can achieve native-like sensitivity to the grammar of the L2. A variety of methodological and theoretical approaches have been taken to examine this issue. While variability is expected in the L2 as a function of proficiency and/or age of acquisition, native language (L1) processing is assumed to be relatively stable; therefore, any deviations from the monolingual norm have been considered as “non-native” L2 processing. Contrary to the assumption of L1 stability among adult native speakers, studies of individual differences in L1 processing (e.g., Kim, Oines, & Miyake, 2017; Tanner & Van Hell, 2014) have revealed variation that is strikingly similar to the variation observed in the L2. However, unlike studies that attribute L2 variability to factors such as proficiency and age of acquisition, L1 variability has been attributed to more stable factors, such as working memory capacity (Dąbrowska (2018); but see Hopp, 2014 & Hopp, 2015).

The approach in the study we report was to examine variation in each of the bilingual’s two languages and to compare that variation to monolingual native speakers. The bilinguals in this study were highly proficient Spanish-English bilinguals who were heritage speakers of Spanish. Although heritage language speakers are the most representative bilingual speakers in the US (American Academy of Arts and Sciences, 2017), they are also among the least studied. The typical experience for heritage speakers is that they acquire the native language as the home language but then, upon entering school or with increasing contact with the majority community, the L2 becomes the dominant language. Little is known about the dynamics of the two languages or the consequences of a dominance switch on processing in either language. Heritage speakers may therefore be the ideal bilinguals to provide insight into the debate regarding the contributions of age of acquisition or proficiency in constraining language processing.

The approach in the present study, using similar grammatical structures in English and Spanish, enabled us to examine individual differences within bilinguals for both English and Spanish, and between bilinguals and monolinguals for English. We further included measures of working memory and proficiency/fluency to address the contributions of cognitive and/or linguistic constraints on processing in each language. In the sections that follow, we first review previous studies on grammatical processing in heritage bilinguals and then explain the logic of how we used event related potentials (ERPs) to address the questions we have raised.

Grammatical Processing in Heritage Bilinguals

As noted above, *heritage speaker* broadly refers to an individual who acquires linguistic competency in a minority language at home in a naturalistic setting, and also acquires the majority language either simultaneously, or at a later, but still early, age (see Rothman, 2009). This pattern of language acquisition often results in high proficiency and literacy in the majority L2 while achieving variable levels of proficiency and literacy in the native L1. It is common for the L2 to become the dominant language. Heritage bilinguals as a population have complex language experience profiles that are influenced by highly variable levels of exposure, speaker diversity, literacy, and community support, all of which interact to influence the processing of the heritage language.

Several decades of research have investigated the factors that promote the acquisition or maintenance of the heritage language (e.g., Montrul, 2010; Montrul, 2012; Rivera-Mill, 2012), which aspect(s) of the heritage language differ from traditionally-defined L1 or L2 acquisition and processing (e.g., Kondo-Brown, 2005; Montrul, 2009; Rothman, 2007; Scontras, Fuchs, & Polinsky, 2015; Sorace & Serratrice, 2009), and to what extent any observed differences are due to incomplete acquisition, attrition, or contact-induced change (e.g., Domínguez, 2009; Pires & Rothman, 2009; Putnam & Sánchez, 2013). Many past studies have identified grammatical structures that differ in heritage speakers compared to age- and socioeconomic status-matched L1 speakers (Benmamoun, Montrul, & Polinsky, 2013; Flores, 2015). Yet the majority of these studies have used offline measures of grammatical competency, such as acceptability judgments, recognition tasks, production elicitation tasks, and natural corpora.

While there is no doubt that heritage speakers differ in a number of ways from monolingual speakers, it remains to be seen how *online* processing reflects those differences. Recent studies using online measures of comprehension, such as self-paced reading (Jegerski, 2018) and eye-tracking (Jegerski & Sekerina, 2020; Sekerina, Laurinavichyute, & Dragoy, 2019), have uncovered greater similarity between heritage language processing and traditional L1 processing than previously assumed (for review, see Bolger & Zapata, 2011). These similarities raise the question of whether heritage languages are truly “deficient”, or whether the observed differences in offline comprehension or production measures are a manifestation of some other surface phenomenon such as perceived competence/confidence, metalinguistic knowledge, or late-process checking. Any unifying explanations for heritage languages must have relevant evidence at its disposal from both online and offline measures before they can account for all the observed similarities and differences.

The current study is among the first to investigate variability in language processing using online neurophysiological measures of brain activity to compare heritage bilinguals in both of their languages with monolinguals. In order to be able to compare across Spanish and English processing within the heritage bilinguals, we used a grammatical structure that manifests similarly across the two languages: present tense sentences that were either grammatical or contained a subject-verb agreement violation:

English: The football game starts/*start in an hour.

Spanish: Algunos árboles viven/*vives por cienes de años.

Some trees live$_{\text{3rd person plural}}$/*live$_{\text{2nd person singular}}$ for hundreds of years. Although subject-verb agreement is a very basic feature that is present in both English and Spanish (and may therefore benefit from cross-language transfer in heritage bilinguals), its manifestation is more complex in Spanish than in English. English has stricter word order and the present indicative has only two forms: for the verb *to start*, there is the form start (I start, you start, we start, they start) and starts ((s)he starts). In contrast, Spanish has relatively freer word order yet more restrictive subject-verb agreement in the present indicative, with a 1-to-1 subject-to-verb form mapping (for the verb *vivir,* meaning *to live:* yo viv**o**, tú viv**es**, él/ella viv**e**, nosotros viv**imos**, ellos/ellas viv**en**). Past offline work using sentence judgment tasks has shown that tense and mood in heritage languages may be vulnerable features (Montrul, 2009), whereas online self-paced reading with person and number violations in the present tense found that heritage bilinguals had similar processing to traditional native speakers, with differences only appearing in reading times of the word appearing three positions after the verb (Rodríguez & Reglero, 2015). Because there have been few neurophysiological investigations of heritage bilingual language processing, it seemed important to start with a grammatical feature that may shed light on the discrepancy between online and offline measures.

Electrophysiological Measures of Language Processing

Event-related potentials (ERPs) are one type of on-line measure that have played an important role in characterizing language processing. The N400 component is a negative-going waveform that peaks around 400 ms post-stimulus and has been linked to lexico-semantic processing (Kutas & Federmeier, 2011). Its amplitude is modulated by factors that affect retrieval from long-term memory (e.g., frequency, novelty, relatedness), and is also elicited in tasks that manipulate expectancy (Boudewyn, Long, & Swaab, 2015). At a higher level, the contexts that elicit the N400 are those that manipulate conceptual and linguistic retrieval of perceptual, categorical, or event knowledge stored in long-term memory.

Another ERP component associated with grammatical processing is the P600, which is a positive-going waveform that peaks around 600 ms post-stimulus. The P600 is typically found in studies that manipulate the grammar or morphosyntax, requiring a person to detect, (re)analyze, and/or repair the sentence (Caffarra, Molinaro, Davidson, & Carreiras, 2015; Osterhout & Holcomb, 1992). However, the P600 can also be found in grammatical sentences that create syntactic ambiguities or processing difficulties, such as garden path sentences and long-distance dependencies (Kaan, Harris, Gibson, Holcomb, 2000). The P600 appears to be more sensitive to task demands, as it is attenuated for tasks that do not require a by-item acceptability or plausibility judgment (Kolk, Chwilla, van Herten, & Oor, 2003), whereas the N400 can be elicited even during REM stages of sleep (Brualla, Romero, Serrano, Valdizán, 1998). The N400 and P600 have distinct but partially overlapping scalp distributions in midline central and posterior electrodes, although the N400 tends to be more anterior/central and the P600 more posterior.

These two ERP components have been the subject of extensive research and the center of various debates on their exact interpretation, whose nuances are beyond full discussion in the current review (see Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011; Kuperberg, 2007). While the general characterization has been that the N400 is related to lexical/semantic processing and the P600 related to sentence-level processing and integration (e.g., Osterhout & Holcomb, 1992; Caffara, Molinaro, Davidson, & Carreiras, 2015), some have argued that these two components do not simply reflect the sequential application of different levels of processing (Kuperberg, 2007). Instead, they appear to function independently, yet jointly, to detect and resolve incongruities in language input that may yield the most probable interpretation (Delogu, Brouwer, & Crocker, 2019).

Many different linguistic manipulations have been documented to elicit or modulate the N400 and P600. Typically, a P600 is seen in native speakers upon encountering an ungrammatical word form. Yet an ungrammatical word is often unexpected and may render an alternative meaning of the sentence, producing an N400. Several studies have reported that ungrammatical sentences elicit the expected P600 in a majority of monolingual native speakers, but a predominant N400 effect in a subset of those speakers (Grey, Tanner, & Van Hell, 2017; Kim et al., 2017; Osterhout, 1997; Tanner & Van Hell, 2014).

One of the first studies to report variation in the presence of the P600 for syntactic anomalies was conducted by Osterhout (1997; but see Kluender & Kutas, 1993). Across two experiments, agreement violations and reduced relative clauses elicited the expected P600 in most monolingual speakers, but an N400 in a subset drawn from the same group. Osterhout discussed several potential explanations for why some individuals may have produced a syntactic N400 rather than a P600, including the idea that different pathways of language acquisition might produce biases toward semantic vs. syntactic focuses in language processing. That proposal is of interest in considering heritage speakers whose early language experience may vary, especially with respect to the amount of form- versus meaning-based exposure. In the present study we use the distribution of N400 and P600 patterns during sentence processing as a tool to identify differences in variation for heritage bilinguals in each language and relative to monolingual speakers.

Proficiency

L2 proficiency is a primary factor that has been investigated for modulating the N400 and P600 components for bilingual speakers. The N400 for vocabulary-level processing increases with proficiency (McLaughlin, Osterhout, & Kim, 2004; Pu, Holcomb, & Midgley, 2016), just as the P600 for syntax-level processing likewise increases with proficiency (e.g., McLaughlin et al., 2010)- rapidly for grammatical structures that overlap across languages but with mixed results for grammatical features in the L2 that are unique or missing in the L1 (Van Hell & Tokowicz, 2010). Even within a bilingual population processing the same type of violation, proficiency has perhaps the largest impact on the magnitude of the P600 (Ojima et al., 2005; Kasparian & Steinhauer, 2016). The presence, magnitude, and timing of the P600, in particular, has been used as a means to determine how “native-like” late bilinguals may be in processing the L2 (e.g., Hahne & Friederici, 2001).

However, an important study by Pakulak and Neville (2010) demonstrated that proficiency differences also drive variability in the P600 component for even monolingual speakers. Using comprehension and production measures of vocabulary and grammar, they demonstrated that monolingual English speakers significantly varied in their English proficiency, and that proficiency correlated with socioeconomic status. Critically, they found that the lower proficiency monolingual speakers exhibited a sustained, bilateral anterior negativity and a P600 in posterior regions that was present, but significantly smaller than the P600 found in higher proficiency monolingual speakers. In contrast, higher L1 proficiency participants produced an early and discrete anterior negativity that was left lateralized (LAN), followed by a large P600 in posterior regions that extended into some anterior medial sites; this pattern has typically been reported in native speakers and serves as the benchmark standard of comparison for many studies on L2 learners (e.g., Hahne & Friederici, 2001). These findings suggest that L2 learners may, in fact, process the L2 similarly to L1 speakers, but perhaps more similarly to the L1 speakers who are in the less proficient (or fluent) range of the average L1 fluency. Moreover, proficiency itself seems to play a critical role in both L1 and L2 processing, whereby L2 learners and monolingual speakers who are in the same range of fluency may show more similar patterns to each other than to L2 learners or monolinguals who are in different ranges of fluency.

L2 learners also sometimes produce an N400 component under circumstances in which native speakers reveal a P600. For example, Weber and Lavric (2008) tested German-English bilinguals and English monolinguals in a sentence processing task in English (both groups) and German (bilinguals only). The English monolinguals produced the expected LAN followed by a large P600 for morphosyntactic violations. The German-English bilinguals processing in German (L1) did not produce a LAN, and the P600 was present but restricted in magnitude and distribution. The bilinguals’ English (L2) processing showed an N400 followed by a P600 for the morphosyntactic violations, with the N400 similar in magnitude to the LAN in the English monolinguals, and the P600 similar in magnitude to the P600 in the German-English bilinguals’ L1. Weber and Lavric interpreted the shift from the N400 to the P600 in L2 processing as a function of proficiency (Mclaughlin et al., 2010; Tanner, Mclaughlin, Herschensohn, & Osterhout, 2013).

At lower levels of proficiency, learners often reveal an N400 for some morphosyntactic processes, but with greater exposure, proficiency, and time in the classroom, they “progress” to a more appropriate P600 response. For certain morphosyntactic violations, the presence of an N400 in L2 learners where L1 speakers exhibit a P600 has been taken as evidence that the learners have yet to progress to the next level of processing. Weber and Lavric (2008) and Osterhout and colleagues have interpreted the presence of the N400 for morphosyntactic processing along the lines of other studies that have suggested lower proficiency L2 processing may be associated with a greater emphasis on semantic and pragmatic information (Clahsen & Felser, 2006; Hopp, 2010, 2015). McLaughlin et al. (2010) also suggested that the N400 found in earlier stages of learning could be the result of violated expectations and transitional probabilities that language learners track throughout learning. In contrast, the P600 is thought to be achieved once the cognitive focus can be shifted to more automatic grammatical aspects of the language, which L1 speakers are more likely to have already automatized and emphasize during on-line processing. Additional evidence for the shift from an N400 at lower proficiency to a P600 at higher proficiency comes from L2 attrition; L2 learners who had reached a P600-stage of grammatical processing “regressed” to N400-dominance after a prolonged period of forgetting with no exposure (Osterhout, Pitkänen, McLaughlin, & Zeitlin, 2019). Yet, as reviewed above, even among relatively homogenous monolingual speakers, variation in L1 processing is observed (Tanner & Van Hell, 2014), raising the question of what these individual differences in ERPs tell us about language processing and proficiency.

What is proficiency? The terms proficiency, competency, and fluency have typically been applied to L2 learners, heritage bilinguals, and traditional native speakers, respectively. The connotations of these three terms (proficiency, competency, and fluency) overlap in their reference to linguistic ability, but differ in the extent to which they assume uniform linguistic representations and how those representations interact with processing mechanisms. That is, proficiency for L2 learners primarily speaks to the breadth and depth of one’s representations with little emphasis on how those representations are drawn upon in real-time (except see Hopp, 2014). Heritage language competency often assumes that speakers have representations that are uneven, influenced by variation in usage, or affected by interference or disuse. For traditional L1 speakers, the assumption of fluency is that the representations are fully and uniformly developed, but individual differences in processing speed, memory, or executive function contribute to variability in online processing. However, one can just as easily imagine that as an L2 learner approaches native-like processing, their ability may be constrained by online processing mechanisms more so than representational deficiencies, moving them from the “proficiency” domain into the “fluency” domain. That is, while these terms are generally applied to a population, there instead appears to be an overarching effect of linguistic ability, and perhaps usage-based factors such as language dominance, that modulates the extent to which a person is primarily affected by representations, usage, or cognitive resources. Although there is no account of proficiency or fluency that is yet adequate, the goal of the current study in measuring linguistic ability (hereafter “proficiency”) was to capture aspects of each of these concepts (representation, usage, and on-line processing). It is also important to note that low L1 fluency does not fall within the same range as low L2 proficiency or heritage competency; lower L1 fluency is often in the upper range of L2 proficiency. In the current study, we exploited the naturally occurring variation that is found within the L1 of heritage bilinguals to consider the role of proficiency in accounting for variation in L1 and L2 processing.

Working Memory

In the L2, working memory is one of the factors that has been identified as important for language learning aptitude (Miyake & Friedman, 1998) and for impacting both L2 processing and proficiency outcomes (Linck, Osthus, Koeth, & Bunting, 2014). In the L1, differences in fluency are often attributed to differences in working memory (e.g., Daneman, 1991). Working memory has also been investigated in conjunction with language comprehension from the late 1980’s and has been central to debates regarding language processing (Daneman & Merikle, 1996; King & Just, 1991). Individuals with higher working memory capacity have been shown to be better able to integrate semantic, syntactic, and pragmatic information on-line relative to their lower span counterparts (Just &

Carpenter, 1992). They also appear able to entertain a larger number of alternative interpretations at points of ambiguity in a sentence while simultaneously maintaining sensitivity to the probability constraints of those interpretations (Pearlmutter & MacDonald, 1995). ERP research on language comprehension has revealed a relationship between working memory and the types of ERP responses that are elicited across individuals, most commonly showing that low span participants show an N400 during L1 grammatical processing where individuals with higher working memory produce a P600 (Bornkessel, Fiebach, & Friederici, 2004; Nakano, Saron, & Swaab, 2010).

Kim et al. (2017) used working memory measures to predict whether monolinguals would be more likely to produce an N400 or P600 in sentence processing. Using two measures each of verbal working memory, non-verbal working memory, and language experience/knowledge, they found that only verbal working memory measures were related to the predominant brain response (N400 vs. P600) during sentence processing. In line with past research, participants with higher verbal working memory showed a strong P600 effect and those with lower verbal working memory were more likely to reveal an N400. The finding that verbal working memory, but not non-verbal working memory, was critical in identifying patterns of sentence processing suggests that language-specific aspects of working memory best account for individual differences.

Current Study

While on-line ERP measures of language processing have not yet been documented in heritage speakers, perhaps the most relevant predictions can be drawn from looking at a series of studies conducted on L1 attrition (Kasparian & Steinhauer, 2016; Kasparian & Steinhauer, 2017; Kasparian, Vespignani, & Steinhauer, 2017). L1 attrition refers to the phenomenon of losing access and features of one’s L1 upon prolonged immersion or immigration to an L2-majority context. In these studies, L1 attrition was reported in native Italian speakers who had immigrated as adults to an English-dominant region of Canada and reported having limited use in Italian. Kasparian and colleagues compared Italian attriters and monolingual Italian controls living in Italy and observed a number of meaningful differences in their neural and behavioral responses. In one study, they probed the processing of lexico-semantic access among the attriters by swapping the grammatical gender of a minimal pair word in a sentence (e.g., *il cappello*, meaning the hat, changed to *la cappella*, meaning the chapel). Attriters and controls both produced an N400 and P600, and each component was larger in individuals with higher Italian proficiency (Kasparian & Steinhauer, 2016). Another clever study manipulated word order, such that all sentences were grammatically correct but only differed in whether the word order was canonical or not in English (attriter’s L2). Monolingual controls produced an N400 and late P600 complex, whereas attriters did not exhibit an N400 and instead produced an earlier, stronger, and more broadly-distributed P600 (Kasparian & Steinhauer, 2017). Although the attriters as a group did not demonstrate an N400, correlations uncovered that larger N400 effects (more like the controls) were related to lower English proficiency and a shorter length of residence in Canada.

Finally, and perhaps most relevantly, Kasparian, Vespignani, & Steinhauer (2017) manipulated subject-verb and subject-modifier agreement. Focusing on the subject-verb agreement similar to the violations used in the current study, they found that the attriters produced a larger and more broadly distributed early negativity compared to the controls who showed a smaller and left-lateralized effect. Both groups produced a similar P600 effect, which was positively related to proficiency. But only the controls maintained the P600 effect into later time windows. They attributed some of the differences in these studies to the freer word order and occurrence of post-verbal subjects in Italian compared to English. For attriters relying more heavily on English word order cues, this led to the detection of a grammatical "violation" (a large P600) in grammatically correct sentences with non-canonical English word order (Kasparian & Steinhauer, 2017) and the larger and more robust early negativity upon encountering a subject-verb agreement, because they are less likely than controls to wait for further input that may otherwise match the verb. Overall, these studies provide some initial background into understanding how variation in L1 experience may manifest or impact L1 processing.

The goal of the present study was to better understand the source of variability in sentence processing. Comparisons across the native and non-native languages have typically involved different groups of speakers. Here we focus on two comparisons, one within group and the other across groups. We ask whether the patterns of N400 vs. P600 patterns of grammatical processing are similar or different for a bilingual's two languages by comparing the two languages within the same bilingual speakers. We then compare those patterns to the variability observed in monolingual speakers in their native language. Where patterns diverge, we ask whether the observed differences can be accounted for by proficiency and/or working memory.

As noted earlier, research on heritage bilinguals has not previously incorporated electrophysiological measures of brain activity to illuminate online language processing, leaving unanswered questions about whether and how heritage grammars differ. The purpose of the current study was twofold; (a) to examine variability in language processing within and across languages, and whether the source of observed variation is similar or distinct, and (b) to report some of the first ERP results on heritage speakers processing a grammatical structure known to be vulnerable. Doing so will allow us to determine the range of variability in each language and population (heritage bilinguals and monolinguals), where those ranges overlap (or do not overlap), and how proficiency and working memory contribute to the patterns of variability. None of the past studies has compared these effects across languages but within the same bilingual.

The heritage speakers in the present study acquired two languages (L1 Spanish and L2 English) early in life and were highly proficient and relatively balanced across their two languages. Because they had been formally educated in English and lived in a predominantly English-speaking context for most of their lives, the majority were English-dominant. Therefore, Spanish is their native L1 and English is their dominant L2. Monolingual English speakers were also included. By using grammatical forms and violations that overlapped between languages (subject-verb agreement), we were able to directly compare variation in grammatical processing between the L1 and L2 *within* the heritage bilinguals, as well as to compare variation in English, the L1 of the monolinguals and the dominant language of the bilinguals, *between* speakers.

Method

Participants

A total of 67 participants (45 females) were tested at Pennsylvania State University (PSU; n = 23) and the University of California-Riverside (UCR; n = 44)[1]. Data from 13 subjects were excluded for a number of reasons (failure to complete all required tasks: 11, missing EEG data: 1, monolingual with early L2 experience: 1). Therefore, data from 29 functionally monolingual (PSU: 17, UCR: 12) and 25 Spanish-English heritage bilinguals (all UCR) were included in analyses. Inclusion criteria for monolinguals were to be native speakers of English, between the ages of 18–35, have normal or corrected-to-normal vision and not be colorblind, and no history of concussion, epilepsy, neurological, or speech disorders. Monolinguals must not have taken more than 2 years of foreign language classes in high school, and must have self-rated their proficiency in any other languages as 4 or less on a scale from 1 (no knowledge) to 10 (native-like fluency). Heritage bilinguals must have reported learning Spanish as a native language in the home setting and must have learned English early (either in the home or at the onset of school), and they must have been able to read in Spanish to complete the requirements of the tasks in the study. Table 1 reports the descriptive characteristics of the participants.

Materials

Language History Questionnaire.—The language history questionnaire (LHQ) was designed to assess subjective language proficiency, exposure, and use, and was modified from the LEAP-Q (Marian, Blumfeld, & Kaushanskaya, 2007) to include questions regarding the mother's education (as a proxy for socioeconomic status), whether any immediate family members were left-handed, in which language(s) the participant initially learned to read, and contained a shortened version of the Edinburgh handedness questionnaire to obtain a more continuous measure of handedness (Oldfield, 1971; Veale, 2014).

Author Recognition Task.—The Author Recognition Task (ART) is used to measure print exposure, known to be related to reading experience. It contained a mixed list of 50 English author names, 50 English non-authors, 50 Spanish author names, and 50 Spanish non-authors. Non-author names were either fabricated names or were real people who were not authors. The English author names were previously validated in other studies of print exposure (Acheson, Wells, & MacDonald, 2008; Moore & Gordon, 2014). The Spanish author names were drawn from popular Spanish books, and efforts were made to constrain the list to casual reading literature and avoid author names from literature that is typically

---

[1] In addition to the analyses reported in the Results, a full set of analyses were run to compare any differences between the monolinguals collected from each location (PSU, UCR). No differences in the EEG data were found based upon location. Another full set of analyses was conducted comparing only the monolinguals collected at UCR with the bilinguals, which produced very similar patterns of results with statistical differences only appearing due to reduced power. Importantly, the similar EEG data suggest that differences equipment did not produce systematic variance or patterns of results.

presented in Spanish literature classes, since the purpose was to measure the amount of print exposure in leisure reading. The Spanish ART was independently validated post-hoc in a pilot study at the University of California-Riverside, using the subject pool to examine print exposure through various questionnaires and other print exposure measures in Spanish-speaking students. An exploratory factor analysis of the Spanish ART with the other questionnaires revealed that the ART loaded onto factors with other variables that measured leisure reading, such as the number of hours spent reading for fun in Spanish, reading on the internet in Spanish, sending emails in Spanish, and online chatting in Spanish. The English and Spanish names were mixed into one questionnaire, such that every participant (including monolinguals) was asked to indicate whether they recognized the English and Spanish author names. The ART score in each language was the total number of correctly-identified authors for books in each language minus the total number of false alarms.

Grammatical Processing Task: English.—The English grammatical processing task included 100 sentences, and each sentence had a grammatical and ungrammatical version. The ungrammatical sentences contained subject-verb agreement violations, similar to those used in Tanner and Van Hell (2014). Sentences contained a lexical verb (not auxiliary verb) in the present tense that varied in sentence position, such that it never occurred earlier than 3 words into the sentence to allow for baselining, and was never the last word in the sentence to avoid sentence wrap-up effects. In order to make a grammatical sentence ungrammatical, the lexical verb was modified to disagree in conjugation with the subject of the sentence (e.g., *No natural lakes <u>exist</u> in Maryland* became *No natural lakes <u>exists</u> in Maryland*). The sentences were split into two lists, which were used to counterbalance which grammatical sentences and ungrammatical sentences a given participant read following a Latin square design. Additionally, lists were matched on the number of words in the sentence, the position of the critical word, and the frequency, length, number of orthographic and phonological neighbors, and average lexical decision and word naming reaction times based on the norms in the English Lexicon Project (Balota et al., 2007), of the preceding word (for baselining purposes) and the critical word (in grammatical form).

During the task, each sentence was preceded by a “Ready?” screen, which remained on the screen until the participant pressed the space bar. Each word was presented one at a time in the center of the screen for 350 ms with 100 ms inter-stimulus interval. The last word of each sentence contained punctuation. After the last word of each sentence, a screen asking “Good/Bad?” was displayed, at which point the participant indicated with a button press whether they judged the sentence to be well-formed, grammatical, and made sense, or if they detected an error in the sentence.

Grammatical Processing Task: Spanish.—The Spanish grammatical processing task also had 100 sentences, with similar constraints to the sentences in English. They were not translations of the English sentences. All verbs were conjugated in present tense and never occurred earlier than the third position in the sentence or as the last word in the sentence. Given the richer morphological system in Spanish and the more flexible word order, sentences were constructed to present the subject before the verb. The ungrammatical versions of the sentences were balanced in which conjugation was used to render the sentence ungrammatical. For example, grammatical sentences conjugated in the 3rd person singular were ungrammatically conjugated to the 3rd person plural in 18 sentences (e.g., *Este sobre contiene información muy importante*, meaning "this envelope contains very important information", became *Este sobre contienen información muy importante*), and likewise the grammatical sentences conjugated in the 3rd person plural were ungrammatically conjugated to the 3rd person singular in 18 sentences (e.g., *Las mascotas consuelan a los niños*, meaning "pets comfort children", became *Las mascotas consuela a los niños*). Each sentence was checked by two native speakers of Spanish from different regions (Spain and Paraguay) to remove colloquialisms and verify the grammaticality (and ungrammaticality). Like the English sentences, lists were counterbalanced in a Latin square design and were additionally matched on the number of words in the sentence, the position of the critical word, and the frequency, length, number of orthographic and phonological neighbors, based on the norms in the Clearpond Database (Marian, Bartolotti, Chabal, & Shook, 2012), of the preceding word (for baselining purposes) and the critical word (in grammatical form).

The Spanish task was exactly the same as the English task, with the exception that the instructions, ready screen, and response screen were in Spanish rather than English and the sentences were different. The presentation rate and ISI remained the same.

Operation Span.—The Operation Span task (O-Span) was adapted from Turner and Engle (1989) to assess working memory. The processing component of the task required participants to solve arithmetic problems within 3 s and indicate with a button press whether the provided solution was correct or incorrect. The storage component of the task required participants to hold in memory a list of English words, presented one at a time interleaved with the arithmetic problems that ranged from a set size of two words and increased to a maximum set size of six words. A given trial was counted as correct if the participant provided an accurate response to the arithmetic problem and successfully recalled the following word in the recall portion at the end of the set, for a maximum score of 60.

Verbal Fluency.—The verbal fluency tasks measured semantic fluency in four categories per language (List 1 categories: animals, family members, vegetables, school supplies; List 2 categories: body parts, professions, fruits, colors). Participants were given 30 seconds to name as many exemplars of the given category as they could. Lists were counterbalanced across participants and languages. There were no significant differences in number of items produced per list, in English (difference between lists: $t(51) = 0.77$, $p = .45$) or in Spanish (difference between lists: $t(23) = 0.58$, $p = .57$).

Procedure

Participants returned to complete the various tasks for this study over the course of four sessions as part of a larger study. The questionnaires and grammatical processing tasks were conducted during the first session and the operation span task and verbal fluency tasks were in the fourth session (time between sessions: M = 6.78, SD = 3.30). Additional tasks were conducted as part of a larger study, which are not reported in full here.

Upon arrival to the first session, participants provided informed consent and completed the language history questionnaire, author recognition task, and handedness questionnaire. Next, the EEG cap was placed on their head and the procedures for preparing for EEG collection were conducted. Once impedances were at an acceptable level, participants performed the English grammatical processing task, which lasted approximately 30 minutes. Bilingual participants remained to complete the Spanish grammatical processing task, which lasted approximately another 30 minutes.

In the fourth session, participants completed the operation span task followed by the verbal fluency task, in English first and then in Spanish for the bilinguals.

EEG Acquisition and Processing

Pennsylvania State University.—EEGs at PSU were acquired from 30 Ag/AgCl scalp electrodes placed in accordance with the 10–20 system, 4 electro-oculogram (EOG) electrodes to measure vertical and horizontal eye movements, and one on-line reference electrode placed on the right mastoid with simultaneous recording from another electrode placed on the left mastoid. Impedances were kept below 5 kΩ. The signal was amplified using a Neuroscan SynAmps2 amplifier with a 24-bit analog to digital conversion (Compumedics NeuroScan, Inc., El Paso, TX) at a 500 Hz sampling rate and filtered with an online high-pass filter of .01 Hz.

University of California-Riverside.—EEGs at UCR were acquired from 32 Ag/AgCl scalp electrodes placed in accordance with the 10–20 system, 4 electro-oculogram (EOG) electrodes to measure vertical and horizontal eye movements, and one on-line reference electrode placed on the right mastoid with simultaneous recording from another electrode placed on the left mastoid. Impedances were kept below 10 kΩ. The signal was amplified using a Brain Vision actiCHamp amplifier with a 24-bit analog to digital conversion (Brain Products, München, Germany) at a 500 Hz sampling rate and filtered with an online high-pass filter of .01 Hz.

All data were pre-processed offline using Brain Vision Analyzer 2 (Brain Products, München, Germany). Electrodes were re-referenced offline to the average of both mastoids and filtered using a 0.1–30 Hz IIR Butterworth filter. The data collected from UCR also had a 60 Hz notch filter applied to the EOG channels. An independent components analysis (ICA) was used to remove the components capturing eye movements (blinks, horizontal eye movements). In the uncommon case that a single component could not be determined to capture the eye movements, then no components were removed and instead the normal artifact rejection steps were used. A whole-head artifact rejection moving window was applied to the all electrodes of interest in the continuous EEG data with parameters adjusted to capture each participant’s artifacts, but the default settings were +/− 150 μv within 200 ms, or any single step >50 μv. A second pass of artifact rejection was meant to remove any trials in which the participant was blinking or moving their eyes during the exact moment a critical word in the sentence was presented, using a moving window to reject any trials in which the EOG electrodes deviated by +/− 200 μv within 150 ms between −100 to 100 ms surrounding the presentation of the critical word (baseline or target word). Target and baseline words were extracted from the continuous data in epochs that began 200 ms before the presentation of the word and extended until 1000 ms after the presentation. ERPs were baseline-corrected across conditions from −200 to 0 ms, and then averaged by condition. On average, each participant's final data set contained 45 trials with correctly answered grammatical targets in English (*SD* = 4.13 trials) and 42 trials with correctly answered grammatical targets in Spanish (*SD* = 5.05 trials), and an average of 42 trials with correctly answered ungrammatical targets in English (*SD* = 5.49 trials) and 34 trials with correctly answered ungrammatical targets in Spanish (*SD* = 10.46 trials).

To reconcile the discrepancies between the electrode arrays for each system, a subset of 22 electrodes were used and mapped into electrode regions. The electrode regions were defined as the conjunction of two factors used in subsequent analyses: anteriority and laterality. The anteriority factor had 3 levels: anterior, central, and posterior electrodes. Across both systems, anterior electrodes included F3, Fz, and F4, common central electrodes were C3, Cz, and C4, and all posterior electrodes were in common: P3, P7, O1, Pz, Oz, P4, P8, and O2. On the Neuroscan system, anterior electrodes additionally consisted of electrodes FC3, FT7, FC4, and FT8, and central electrodes included CP3, TP7, CP4, and TP8. For the Brain Product system, FC1, FC5, FC2, and FC6 were also defined as anterior electrodes, and central electrodes additionally included CP1, CP5, CP2, and CP6. A similar process for laterality was conducted. Across both systems, left electrodes included F3, C3, P3, P7, and O1 and right electrodes were the right-hemisphere homologues. All midline electrodes were shared across systems (Fz, Cz, Pz, and Oz). The Neuroscan additionally included left electrodes: FC3, FT7, CP3, and TP7, with the right-hemisphere homologues. The Brain Products system further included in the left electrodes: FC1, FC5, CP1, and CP5, with the right-hemisphere homologues.

Results and Discussion

Approach

As one of the first studies investigating online sentence processing in heritage bilinguals using ERPs, we first report data on the magnitude and distribution of N400 and P600 effects across groups and languages. Subsequently, we consider whether and which factors drive individual differences in each component, using measures of working memory, proficiency, and reading experience (heritage bilinguals only). Importantly, many would not expect to find an N400 component for the ungrammatical sentences presented in the current study, particularly for traditional native speakers. Past work that has investigated the individual differences in ERP responses has not identified what factors drive the presence or magnitude of the N400, specifically, in these kind of sentence contexts.

In the first set of analyses with the goal of comparing the magnitude and distribution of the components across groups and languages, we used multi-level modeling in R (R Core Team, 2019; Bates, Maechler, Bolker, & Walker, 2015) with fixed effects of condition (grammatical, ungrammatical), anteriority (anterior electrodes, central electrodes, posterior electrodes), and laterality (left, midline, right), with random intercepts for subjects and random slopes by condition. Separate models were run for the ERP mean amplitudes of the N400 and P600 component. Additionally, separate models were run to compare the English responses, between monolinguals and heritage speakers, and to compare the responses in heritage speakers, between English and Spanish. These comparisons (Group: monolingual vs. heritage speaker in English; Language: English vs. Spanish in heritage speakers) were added as fixed factors to the models. All interactions were included between the fixed effects. In addition to the model coefficients and intercepts, an ANOVA with type II sum of squares was conducted on each model using the *car* package (Fox & Weisberg, 2019) to determine main effects and interactions, which were then interpreted in conjunction with the model coefficients. In all models, we report only the effects that include condition or group/language.

In the second set of analyses with the goal of investigating which factor(s) drive individual differences in responses, we also used multi-level models. The outcome variable was the effect magnitude (ungrammatical – grammatical mean amplitude) for the N400 and P600 for each electrode; therefore, more negative N400 values indicate larger canonical N400 effects, and more positive P600 values indicate larger canonical P600 effects. The base model included fixed effects of anteriority, laterality, and group/language, with full interactions, and random intercepts for subjects. Consecutive model comparisons were conducted using AIC values to compare the addition of each factor (O-Span scores, proficiency scores) first as a main effect, then as an interaction with group/language, and finally with full interaction terms in the model. For the best models, we report effects that include the factor of interest (O-Span, proficiency) only, because the distribution of the effects was already reported in the first set of findings.

ERP Responses Across Groups and Languages

English and Spanish in Heritage Speakers—The mean amplitudes of the N400 component (300–500 ms) in heritage bilinguals were modeled with a fixed effect of language (English, Spanish) in addition to the other base fixed effects (condition: grammatical vs. ungrammatical; anteriority: anterior, central, posterior; laterality: left, midline, right) and random slopes for condition by subjects. The results of the model showed no main effect of condition ($\chi^2(1) = 1.46$, $p = .23$), but a significant interaction between condition and language ($\chi^2(1) = 15.48$, $p < .01$). The interaction reflects the finding that the N400 magnitude was larger in English than in Spanish. In English, the heritage bilinguals had a marginal main effect of condition ($\chi^2(1) = 2.88$, $p = .09$) and a significant condition x anteriority interaction ($\chi^2(2) = 10.33$, $p < .01$) showing that the effect was strongest over anterior and central electrodes but was not present over posterior electrodes. In Spanish, there was no effect or interaction involving condition (all ps > .2).

The P600 mean amplitudes were also modeled within the heritage speakers across languages. The effect of condition ($\chi^2(1) = 7.22$, $p < .01$) was qualified by a higher-order language x condition interaction ($\chi^2(1) = 16.74$, $p < .01$). Follow-up models for each language separately showed that the P600 magnitude was larger in English, in which the effect of condition was significant ($\chi^2(1) = 9.55$, $p < .01$), than it was in Spanish, in which the effect of condition was marginally significant ($\chi^2(1) = 3.46$, $p = .06$). The full model also uncovered a significant condition x anteriority interaction ($\chi^2(2) = 112.82$, $p < .01$). The P600 effect was present over central and posterior electrodes, but not over anterior electrodes.

Monolinguals and Heritage Speakers in English—The N400 mean amplitudes were modeled in English with a fixed effect of group (heritage bilinguals, monolinguals) instead of language as in the previous section. The results of the model revealed a significant effect of condition ($\chi^2(1) = 6.05$, $p = .01$), such that ungrammatical sentences were more negative than grammatical sentences. A significant condition x anteriority interaction ($\chi^2(2) = 8.63$, p = .01) revealed the distribution of the N400 effect. Follow-up analyses uncovered that the effect was present over anterior and central electrodes but not over posterior electrodes. There were no condition x group interactions, suggesting that the magnitude and distribution of the N400 in English did not significantly differ between monolinguals and heritage speakers.

The mean amplitudes of the P600 component (600–900 ms) were also modeled. The effect of condition ($\chi^2(1) = 43.61$, p < .01) was qualified by two-way interactions between condition x anteriority ($\chi^2(2) = 113.83$, p < .01) and condition x laterality ($\chi^2(2) = 8.45$, p = .01), and a significant higher-order interaction between group, condition, and anteriority ($\chi^2(2) = 8.02$, p = .02). Follow-up analyses within each group and for each electrode region (anterior, central, posterior) revealed that there was not a significant difference in the magnitude of the P600 between the monolinguals and heritage speakers over the posterior or central electrodes, but for the monolinguals only, the effect extended into anterior electrodes.

Individual Differences in ERP Responses

Working Memory

English and Spanish in Heritage Speakers: The base model comparison for the N400 and P600 models included interactions between group/language, anteriority, and laterality; therefore, for a factor such as working memory to improve the fit of the model, it must explain variance above and beyond the group and distributional effects. The model of the N400 magnitude (ungrammatical – grammatical) with O-Span scores as an added main effect or interaction with language showed no main effect of O-Span scores ($\chi^2(1) = 0.25$, p > .60), and no significant language x O-Span interaction ($\chi^2(1) = 2.41$, p > .10). No models that included O-Span scores significantly improved the fit over the base model (all ps > .2).

The P600 magnitudes were similarly modeled, which showed no main effect of O-Span scores ($\chi^2(1) = 0.81$, p > .10), but the O-Span x language interaction was significant ($\chi^2(1) = 8.21$, p < .01) and significantly improved the fit of the model ($\chi^2(2) = 8.98$, p = .01). While O-Span scores did not significantly modulate the P600 component for heritage speakers in either language, the slope was nevertheless significantly steeper in English than in Spanish. No higher-order distributional interactions were observed in the model with full interaction terms.

Monolinguals and Heritage Speakers in English: To compare English between the monolinguals and heritage bilinguals, models of the N400 magnitude were similarly fit by consecutively adding O-Span scores and interaction terms and comparing models with the base model. The model of the N400 magnitude (ungrammatical – grammatical) with O-Span scores as an added main effect showed a marginal effect of working memory ($\chi^2(1) = 3.69$, p = .05). The effect showed that individuals in both groups who had higher working memory capacity also had larger N400 effects. In contrast, individuals with lower working memory scores actually demonstrated an early positivity, likely showing an early P600 effect being captured in the N400 time window.

Higher-order models were fitted to determine whether working memory differentially affected N400 magnitudes across groups, or if the working memory scores only modulated N400 magnitudes over certain electrode regions. While adding the group interaction did not significantly improve model fit ($\chi^2(2) = 4.30$, $p = .12$), the full interaction terms did improve model fit over the base model ($\chi^2(18) = 37.65$, $p < .01$). The full model including interactions revealed an interaction between O-Span scores and anteriority ($\chi^2(2) = 9.81$, $p < .01$), qualified by higher-order interactions between O-Span scores x group x anteriority ($\chi^2(2) = 9.69$, $p < .01$) and laterality ($\chi^2(2) = 7.77$, $p = .02$). Follow-up analyses and visualization of the effects showed that for monolinguals, the effect was largely uniform across electrode sites, but O-Span scores had little relation to the N400 magnitude in heritage bilinguals except over left-hemisphere and posterior electrodes.

The model of the P600 magnitude also showed a main effect of working memory ($\chi^2(1) = 8.75$, $p < .01$). The effect of working memory on the P600 magnitude was such that individuals with higher working memory had a smaller P600 magnitude. Higher-order models that included the group interaction did not improve fit ($\chi^2(1) = 0.80$, $p > .30$), but the full interaction terms significantly improved the fit of the model over the model with O-Span scores as a main effect ($\chi^2(17) = 33.33$, $p = .01$). The higher-order model uncovered a significant O-Span x anteriority interaction ($\chi^2(2) = 8.25$, $p = .02$), an O-Span x laterality interaction ($\chi^2(2) = 12.90$, $p < .01$), and a marginal group x O-Span x anteriority interaction ($\chi^2(2) = 5.3$, $p = .07$). Follow-up models were fit for each group separately to examine the interaction. The follow-up model of the monolinguals in English showed the same effects (main effect of O-Span scores: $\chi^2(1) = 12.67$, $p < .01$; O-Span x anteriority: $\chi^2(2) = 12.52$, $p < .01$; O-Span x laterality: $\chi^2(2) = 12.25$, $p < .01$). Monolinguals with low working memory capacity had significantly larger P600 magnitudes particularly over central and anterior midline electrodes. The follow-up model of the heritage speakers in English showed no main effect of O-Span scores ($\chi^2(1) = 1.31$, $p = .25$), and no higher-order interactions.

Proficiency

Calculating Proficiency: In order to assess proficiency, we created a composite proficiency measure as reported in previous studies of language processing (McMurray, Samelson, Lee, & Bruce Tomblin, 2010; Pivneva, Palmer, & Titone, 2012). We included measures of production fluency (verbal fluency), comprehension (d` scores on grammaticality judgment task), self-rated proficiency averaged for speaking, understanding, and reading, and age of acquisition. Scores on each measure were pooled across languages and then z-scored (i.e., the Spanish verbal fluency scores were z-scored relative to the English verbal fluency scores). The composite proficiency measure was calculated by combining the z-scored measures: z(Verbal Fluency) + z(d′ values) + z(Self-Rated Proficiency) - z(Age of Acquisition). The resulting proficiency measure provided a proficiency score separately for each participant in each language (i.e., a separate proficiency score for heritage bilinguals in English and Spanish, and monolinguals in English), relative to other participants and languages, such that proficiency could be directly compared across languages. Higher values indicate higher proficiency, and 0 indicates average proficiency across the languages and participants. Overall, the composite score was normally distributed ($W = .98$, $p = .15$), and the heritage bilinguals had significantly higher proficiency scores in English ($M = 0.93$, $SD = 1.27$) than in Spanish ($M = -0.93$, $SD = 1.74$; $t(24) = 5.23$, $p < .01$), as would be expected since English was the dominant language (i.e. higher proficiency score in English than in Spanish) for all but 2 of the heritage bilinguals, who had also self-reported being Spanish-dominant.

English and Spanish in Heritage Speakers: The same model fitting process was applied to the proficiency measure as was conducted for the working memory analyses. Proficiency scores were significantly related to the N400 magnitude for heritage bilinguals ($\chi^2(1) = 10.84$, $p < .01$), such that individuals with lower proficiency had larger N400 magnitudes. The higher-order model showed a significant proficiency x language interaction ($\chi^2(1) = 52.89$, $p < .01$) and significantly improved model fit ($\chi^2(1) = 51.09$, $p < .01$). The proficiency x language interaction reflected the finding that N400 magnitudes were more closely related to proficiency scores in Spanish than in English for the heritage speakers. The higher-order model with full interaction terms did not significantly improve fit over the model with the language interaction ($\chi^2(16) = 12.57$, $p > .70$).

For heritage bilinguals, the P600 magnitudes were significantly related to proficiency scores ($\chi^2(1) = 86.95$, $p < .01$), such that heritage bilinguals with higher proficiency had larger P600 magnitudes. The interaction with language was significant ($\chi^2(1) = 39.76$, $p < .01$), showing that the relationship between P600 magnitudes and proficiency scores was stronger in Spanish than in English for the heritage bilinguals. Finally, the model including full interactions had significantly better fit than the lower-order model with the language interaction term ($\chi^2(16) = 41.78$, $p < .01$). A significant proficiency x anteriority interaction ($\chi^2(2) = 31.11$, $p < .01$) uncovered that the relation between P600 magnitudes and proficiency scores was strongest over the posterior electrodes.

Monolinguals and Heritage Speakers in English: The magnitude of the N400 component in English was unaffected by proficiency scores ($\chi^2(1) = 0.14$, $p > .70$), which did not interact with group ($\chi^2(1) = 81.48$, $p > .20$). No models that included proficiency scores significantly improved the fit over the base model (all ps > .3).

The magnitude of the P600 in English was marginally affected by proficiency scores ($\chi^2(1) = 3.50$, $p = .06$), such that higher proficiency scores were related to larger P600 magnitudes. Adding the group interaction term significantly improved the model fit ($\chi^2(1) = 5.03$, $p = .02$), given that the effect of proficiency scores on P600 magnitudes was significantly stronger for heritage speakers in English than for monolinguals. Finally, the higher-order model including full interaction terms further improved model fit ($\chi^2(16) = 36.84$, $p < .01$), due to the proficiency x anteriority interaction ($\chi^2(2) = 25.00$, $p < .01$). The interaction showed that the proficiency had a stronger effect on posterior P600 magnitudes than over anterior electrodes.

Reading Experience—Literacy and reading experience are among a number of factors known to exhibit large variability in heritage populations (e.g., Bayram et al., 2019; Kondo-Brown, 2005; Kupisch & Rothman, 2018; Montrul, 2011). The population of heritage bilinguals in the current study all must have been able to read at some level in Spanish to be included, due to the fast-paced reading skills required to complete the grammaticality judgment task in Spanish. Nevertheless, the modifications made to the language history questionnaire included an additional question asking participants in which language they first learned to read, which we used to split the bilinguals based on early reading experience and examine the N400 and P600 magnitudes.

Among the heritage bilinguals, 11 indicated that they initially learned to read in English, 9 reported first learning to read in Spanish, and 5 reported learning to read in both languages. Given the small number of those who learned to read in both languages ($n = 5$), we excluded those bilinguals and examined the bilinguals who learned to read in one language (English or Spanish) more closely. For each group of bilinguals, we fit multi-level models on the magnitude of the N400 and the P600 (separately), with fixed effects of anteriority, laterality, language, and their full interactions, along with random slopes for each subject. Type II sum of squares ANOVAs were conducted on the model terms to determine the main effects and interactions.

For bilinguals who learned to read in English, there was a main effect of language on the N400 magnitude ($\chi^2(1) = 42.62$, $p < .01$), qualified by higher-order interactions between language x anteriority ($\chi^2(2) = 15.10$, $p < .01$) and language x laterality ($\chi^2(2) = 10.79$, $p < .01$). Follow-up models for each language uncovered that the magnitude of the N400 was larger in Spanish than in English, and the difference between the two languages was particularly pronounced over central and anterior midline electrodes. For the P600 magnitude, there was once again a main effect of language ($\chi^2(1) = 94.80$, $p < .01$) and a language x laterality interaction ($\chi^2(2) = 6.69$, $p = .04$). Follow-up models uncovered that the P600 magnitude was significantly larger in English than in Spanish, particularly over midline posterior electrodes. Figure 6 shows the ERP waveforms for bilinguals who learned to read in English while processing sentences in English and Spanish.

For bilinguals who learned to read in Spanish, there was also a main effect of language on the N400 magnitude ($\chi^2(1) = 69.78$, $p < .01$), qualified by a language x anteriority interaction ($\chi^2(2) = 13.66$, $p < .01$). In contrast to those who learned to read in English, heritage bilinguals who learned to read in Spanish had a significantly larger N400 magnitude in English than in Spanish, particularly over midline/right anterior electrodes. For the P600 magnitude, there was a marginally significant effect of language ($\chi^2(1) = 3.11$, $p = .08$), showing a marginally larger P600 magnitude in English than in Spanish. Figure 7 illustrates the ERP waveforms for bilinguals who learned to read in Spanish while performing the sentence reading task in Spanish and in English.

General Discussion

The primary goal of the present study was to investigate the role of individual differences in brain responses during L1 and L2 sentence processing within the same individuals. As one of the first studies to examine ERP responses in heritage bilinguals, we reported group-level brain responses during language processing in both of the bilinguals' two languages, in comparison to traditional monolingual speakers. The group ERP comparisons generally found weaker ERP effects (both the N400 and P600) in heritage bilinguals than monolinguals, with a graded effect of dominance (i.e., heritage bilinguals had larger effects in their more dominant L2). Unlike most ERP research on language processing that attempts to minimize variability, the current study aimed to capture a larger range of variability to better understand what individual differences drive the observed patterns; therefore, one benefit of studying heritage bilinguals was the significant natural variability in both the L1 and L2 of the population. The individual difference analyses uncovered that language processing in bilinguals became more similar to monolingual language processing as proficiency in each language increased; in contrast, working memory was the primary factor driving variability in monolingual language processing. Early reading experience in the bilinguals also had a significant impact on their brain responses, especially the N400 effect.

Proficiency and Working Memory

Overall, the pattern of results, with a number of qualifications, confirms the assumptions of previous research on L1 and L2 processing with respect to which variables have the largest influence on language processing. The results of the current study support previous findings showing that proficiency is the primary source of variability for processing the less proficient language. However, these results extend past work by demonstrating that proficiency modulated brain responses in both of the bilinguals' languages. Language dominance had a higher-order effect, such that the bilinguals' less dominant L1 was more affected by proficiency than their more dominant L2. Such a finding may simply be attributed to proficiency itself; the more dominant language is also higher proficiency. However, such an explanation would not account for why the *relation* between proficiency and brain responses became weaker with increasing proficiency. Language dominance may instead have a higher-order influence, perhaps related to exposure and usage. For these heritage bilinguals, the dominant language is most often the majority or community language to which they have more exposure and use more frequently. The effect of language dominance may reflect that when a language is used less frequently, one's proficiency has a stronger influence on how it is processed than when it is used less frequently. However, given that our proficiency measure contained aspects of representation (d'), usage (AoA), and processing (verbal fluency), we cannot strictly disentangle this account of language dominance with our measures. Nevertheless, these findings are an important step in understanding the interplay between proficiency, dominance, and age of acquisition, and suggest avenues for future studies.

For monolinguals, brain responses were primarily related to working memory. The pattern of results was a striking reversal of the proficiency results, with the largest effect for monolinguals and the smallest effect for bilinguals in their less dominant L1. Monolinguals with higher working memory produced larger N400 and smaller P600 components. This could be due to the anti-correlation between the two components, showing spillover from a larger effect in the N400 time window into the P600 time window and/or the tendency for

N400-dominant individuals to not produce a P600 and vice versa (e.g., Kim et al., 2017; Tanner & Van Hell, 2014).

The working memory results reported here differ from past research on individual differences in ERPs (Kim, Oines, & Miyake, 2017; Tanner & Van Hell, 2014). Kim et al. (2017) used multiple measures of working memory and showed that higher-span individuals produced a larger P600, whereas Tanner & Van Hell (2014) used the same working memory measure (O-Span) and grammatical violation (subject-verb agreement) as in the current study and found no relation between working memory and brain responses. One difference between the current study and past studies that may account for the discrepancy in results could be related to the fact that our grammatical processing task did not include filler sentences. The lack of fillers may have changed the strategic nature of language processing for some individuals more than others, resulting in the observed patterns of brain responses.

An interesting and important distinction between working memory and proficiency is the trait-like vs. state-like behavior. Working memory is considered to be a relatively stable characteristic of an individual, and any individual only has one working memory. Proficiency, in contrast, is measured separately for each language and can increase or decrease with relative independence. Because of this, individual differences in language processing in each language have separate proficiency scores, but draw upon the same pool of working memory scores. Some important questions for future work is to consider what this distinction means for predictions about the stability of language processing (is it state-like or trait-like?), to what degree language processing is influenced by experience (captured in its influence on proficiency), and how working memory and proficiency are (or are not) related in different populations.

Heritage Language Processing

As one of the first published studies on electrophysiological measures of language processing in heritage bilinguals, the results of the current study advance our understanding of what occurs in heritage language processing. At the group level, the population of heritage bilinguals had a reduced N400 and P600 in their less dominant L1 than in their more dominant L2. Compared to monolingual speakers of English, the heritage bilinguals' brain responses in English were not statistically significantly different, but models of the two groups separately indicated a marginal N400 effect in the bilinguals where there was a significant effect in monolinguals, and a P600 in the monolinguals that extended to more anterior electrodes. Therefore, at the group level, there appear to be significant differences in heritage language processing, between both languages, as well as compared to monolingual language processing.

However, as expected, proficiency played a key role in influencing brain responses. The results depicted in Figure 5 showing the relation between proficiency and brain responses highlight another important point- namely, the range of overlap. The fitted regression line for bilinguals' brain responses in Spanish (and the shaded region indicating the confidence interval) overlaps with the line for monolinguals' brain responses in English at approximately where proficiency is equal to 0 (i.e., average proficiency across groups and languages in our sample). That would suggest that when speakers of either group have roughly equivalent proficiency in their L1, their brain responses are similar. The range of overlaps for bilinguals in English is on average at a higher proficiency level, possibly hinting at one consequence of later acquisition that remains after accounting for proficiency.

While not a direct comparison, these results differ in a number of ways from the ERP research conducted on L1 attriters. Kasparian, Vespignani, and Steinhauer (2017) investigated native Italian speakers who immigrated to an English-dominant country and reported losing access/proficiency in Italian. Their brain responses to Italian to subject-verb violations showed a larger N400 than the monolingual control group and a P600 of similar magnitude that was positively related to proficiency in both groups. This contrasts with the significantly smaller N400 found at the group level for heritage speakers of Spanish reported here. While both studies showed a relation between the P600 magnitude and proficiency measures in the groups of interest (attriters, heritage speakers), the two studies differed in whether that relation also applied to the control group (monolinguals). With respect to the differences found in the N400, Kasparian et al. argue that the smaller effect found in monolingual controls was likely due to a wait-and-see response. Given that Italian (and Spanish) have freer word order than English, they attributed the smaller negativity in monolingual controls as their adjusted expectations that a post-verbal noun could come after a verb that fails to agree with its preceding noun, whereas attriters who have adjusted their expectations to strict English word order no longer consider the possibility of a post-verbal noun, producing a more robust early negativity. Another important difference between the studies was that unlike the heritage bilinguals in the current study, the attriters had been dominant in Italian for most of their lives before moving to the English-speaking context. While the specific findings of these studies differ in a number of important ways that future research will need to consider, the combined findings make a strong case for the role of proficiency over and above age of acquisition by showing that language processing is dynamic in responding to experiential and contextual factors, and may even change quantitatively and qualitatively from its initial configuration.

The results of the reading experience analyses were particularly striking. The bilinguals' initial experience with learning to read continued to impact language processing in the now-adult heritage bilinguals. When performing the grammaticality judgment task in the same language as they initially learned to read, bilinguals produced no N400 component (see Figure 8 for heatmap of ERP effects as a function of whether bilinguals were performing the task in the first-literate vs. second-literate language). But the N400 component was significantly larger when the same bilinguals performed the task in the other language. For both groups, the P600 was larger in English than in Spanish, likely due to the many years spent reading English in educational settings. Just as Kasparian et al. (2017) found that the magnitude of the late P600 in response to violations on a modifier was related to English exposure, the P600 results also suggest greater sensitivity to cumulative experience. These results were surprising, in part due to the fact that there was low power in each group to detect such differences (n = 11 and n = 9), and especially given that all the bilinguals were literate in both languages regardless of which language they initially learned to read as children. As one might expect, the effect of early reading experience was the largest for bilinguals reading sentences in Spanish who had initially learned to read in English, who produced the largest N400 and smallest P600. The divergence in how the N400 was affected by initial reading experience vs. how the P600 was affected by continued/developed reading experience provides an avenue for future research to examine the development and flexibility of these language processing components.

These findings in heritage bilinguals have important implications for understanding the role of literacy and education in language processing. Indeed, these results are supported by past findings from a large European project on simultaneous child bilingual development (Kupisch & Rothman, 2018). In a longitudinal study of language development comparing well-matched French and Italian HSs, the French HSs maintained more monolingual-like linguistic features. One of the arguments put forth to explain the stronger similarity in French HSs was that unlike the Italian HSs who primarily attended school in German, the French HSs attended French schools throughout their childhood. Their schooling provided more comprehensive experience with formal French registers and exposed them to extensive French literature. In the current study, we only examined a self-reported measure of the language of initial literacy, and the results showing a large N400 in the second-literate language that was absent in the first-literate language. The modulation of the N400 as a function of early literacy experience did not make one language more monolingual-like than the other, given that even the monolinguals, as a whole, exhibited an N400. However, these findings do support the claim that early literacy experience has long-lasting effects on language processing, especially among heritage bilinguals.

Limitations and Future Directions

The results reported here were part of a larger, intensive 5-session language training study that constrained our ability to collected larger sample sizes for each of our groups and our ability to employ multiple measures of working memory as others have done previously (Kim, Oines, & Miyake, 2017). Such time constraints limited our ability to collect an independent sample of monolinguals in California (to exclude the monolinguals collected in a different context and using different EEG equipment) and our ability to collect additional bilinguals who uniformly had English dominance. However, we did have multiple converging measures of proficiency, and the within-subjects nature of the comparisons within the heritage bilinguals leveraged additional power. The issue of including heritage bilinguals who were dominant in Spanish (n = 2) was partly addressed in the individual difference analyses with proficiency. Future studies should replicate and extend the current findings with larger sample sizes and by including bilinguals with other language profiles.

Another possible limitation is that there was not a group of monolingual Spanish speakers; because the Spanish materials were not direct translations of the English sentences, it is possible that the differences between Spanish and English processing in heritage bilinguals were due to intrinsic properties of the materials. While extensive efforts to match the materials on a number of dimensions was made and the materials were checked with two native Spanish speakers from different dialectal regions, it is quite possible that the languages themselves may lead to differences in processing subject-verb agreement. One reason why there could be language-level differences is due to the relatively free word order in Spanish, such that in regular speech, the subject can (and often does) appear after the verb, which could lead to delayed responses while Spanish speakers await additional information. The sentences were constructed to clearly indicate a pre-verbal subject in the current study, but a lifetime of adjusting expectations may have resulted in language-level differences in ERP responses.

Although a standardized proficiency test in each language may not have been as directly comparable across languages, including them would have enabled a more precise and rigorous definition of proficiency. In contrast, we employed the same measures (d', verbal fluency, self-rated proficiency, age of acquisition) across languages in determining proficiency, which allowed for a direct comparison across languages, but lacked the standardization and breadth of measures captured in standardized proficiency tests. As the focus on variation in language processing and in bilingualism moves forward, it will be important to grapple more explicitly with what is meant and what is measured by proficiency.

Conclusion

As research on bilingualism continues to delve deeper into understanding variation in language experiences and how they shape language and cognitive processing, it will be important to identify and utilize methods and analyses that can capture that variation. Although traditionally ERP analyses were not developed for investigating individual differences, the current study was one among many that illustrated the sensitivity of these online electrophysiological methods for harnessing, explaining, and understanding variation in language processing. The grand-averaged waveforms reported here stand in contrast to the working memory and proficiency results showing the full range of brain responses in each group and language, including where they overlap in range. Combined with powerful multi-level modeling statistical methods, subtle effects were detected that would otherwise be buried in inter- and intra-individual variability.

To conclude, the present study found that individual differences in language processing are produced by the interplay between proficiency and working memory across languages, and are further modulated by higher-order language dominance. The results speak to the growing literature on heritage grammars (e.g., Polinsky & Scontras, 2020), how variation in language experience impacts the effects of bilingualism (for review, see Zirnstein, Bice, & Kroll, 2019), and new approaches that seek to harness individual differences and variation rather than average over them (e.g., Pliatsikas, DeLuca, Voits, 2019). The nature of bilingual experiences is inherently variable, yet only by seeking to understand the variability and its source(s) can future research begin to garner a deeper understanding of how language experiences translate into differences in performance.

Acknowledgments

This research was supported by the William Orr Dingwall Neurolinguistics Fellowship and Penn State College of Liberal Arts RGSO Award to K. Bice, NSF Grant BCS-1551892 to J. F. Kroll and K. Bice, NIH Grant HD082796 and NSF Grants BCS-1535124 and OISE-1545900 to J. F. Kroll. The authors would like to thank Dr. Eleonora Rossi for her feedback on a previous version of the manuscript.

References

Abutalebi J, Della Rosa PA, Green DW, Hernandez M, Scifo P, Keim R, … Costa A (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. Cerebral Cortex, 22(9), 2076–86. 10.1093/cercor/bhr287 [PubMed: 22038906]

Acheson DJ, Wells JB, & MacDonald MC (2008). New and updated tests of print exposure and reading abilities in college students. Behavior Research Methods, 40(1), 278–89. 10.3758/BRM.40.1.278 [PubMed: 18411551]

Ameel E, Malt BC, Storms G, & Van Assche F (2009). Semantic convergence in the bilingual lexicon. Journal of Memory and Language, 60(2), 270–290. 10.1016/j.jml.2008.10.001

American Academy of Arts and Sciences (2017). America's Languages: Investing in Language Education for the 21st Century.

Ashby FG, & Maddox WT (2005). Human Category Learning. Annual Review of Psychology, 56(1), 149–178. 10.1146/annurev.psych.56.091103.070217

Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, Loftis B, … Treiman R (2007). The English Lexicon Project. Behavior Research Methods, 39(3), 445–59. [PubMed: 17958156]

Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. doi:10.18637/jss.v067.i01.

Bayram F, Rothman J, Iverson M, Kupisch T, Miller D, Puig-Mayenco E, & Westergaard M (2017). Differences in use without deficiencies in competence: passives in the Turkish and German of Turkish heritage speakers in Germany. International Journal of Bilingual Education and Bilingualism, 22(8), 919–939.

Benmamoun E, Montrul S, & Polinsky M (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. Theoretical linguistics, 39(3–4), 129–181.

Bice K, & Kroll JF (2015). Native language change during early stages of second language learning. NeuroReport, 26(6), 966–971. 10.1097/WNR.0000000000000453 [PubMed: 26351964]

Bice K, & Kroll JF (in prep). Consequences of bilingual language experience for the trajectory and outcome of new language learning.

Bolger PA, & Zapata GC (2011). Psycholinguistic approaches to language processing in heritage speakers. Heritage Language Journal, 8(1), 1–29.

Bornkessel ID, Fiebach CJ, & Friederici AD (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. Cognitive Brain Research, 21(1), 11–21. 10.1016/j.cogbrainres.2004.05.007 [PubMed: 15325409]

Boudewyn MA, Long DL, & Swaab TY (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. Cognitive, Affective, & Behavioral Neuroscience, 15(3), 607–624. 10.3758/s13415-015-0340-0

Brouwer H, Crocker MW, Venhuizen NJ, & Hoeks JC (2017). A neurocomputational model of the N400 and the P600 in language processing. Cognitive Science, 41, 1318–1352. [PubMed: 28000963]

Brualla J, Romero MF, Serrano M, & Valdizán JR (1998). Auditory event-related potentials to semantic priming during sleep. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 108(3), 283–290.

Caffarra S, Molinaro N, Davidson D, & Carreiras M (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. Neuroscience and Biobehavioral Reviews, 51, 1–17. 10.1016/j.neubiorev.2015.01.010 [PubMed: 25528684]

Chandrasekaran B, Yi HG, & Maddox WT (2014). Dual-learning systems during speech category learning. Psychonomic Bulletin & Review, 21(2), 488–495. 10.3758/s13423-013-0501-5 [PubMed: 24002965]

Chang CB (2013). A novelty effect in phonetic drift of the native language. Journal of Phonetics, 41(6), 520–533. 10.1016/j.wocn.2013.09.006

Clahsen H, & Felser C (2006). Grammatical processing in language learners. Applied Psycholinguistics, 27(1), 3–42. 10.1017/S0142716406060024

Dąbrowska E (2018). Experience, aptitude and individual differences in native language ultimate attainment. Cognition, 178, 222–235. [PubMed: 29886057]

Daneman M (1991). Working memory as a predictor of verbal fluency. Journal of Psycholinguistic research, 20(6), 445–464.

Daneman M, & Merikle PM (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin & Review, 3(4), 422–433. 10.3758/BF03214546 [PubMed: 24213976]

Delogu F, Brouwer H, & Crocker MW (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. Brain and Cognition, 135, 103569. [PubMed: 31202158]

Domínguez L (2009). Charting the route of bilingual development: Contributions from heritage speakers' early acquisition. International Journal of Bilingualism, 13(2), 271–287.

Dussias PE, & Sagarra N (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. Bilingualism: Language and Cognition, 10(1), 101–116. 10.1017/S1366728906002847

Flores CMM (2015). Understanding heritage language acquisition. Some contributions from the research on heritage speakers of European Portuguese. Lingua, 164, 251–265.

Fox J, & Weisberg S (2019). An {R} Companion to Applied Regression, Third Edition Thousand Oaks CA: Sage URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frenzel S, Schlesewsky M, & Bornkessel-Schlesewsky I (2011). Conflicts in language processing: A new perspective on the N400–P600 distinction. Neuropsychologia, 49(3), 574–579. [PubMed: 21145903]

Friederici AD (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. Brain and Language, 50(3), 259–281. 10.1006/brln.1995.1048 [PubMed: 7583190]

Green DW., & Abutalebi J. (2013). Language control in bilinguals: The adaptive control hypothesis. Journal of Cognitive Psychology, 25(5), 1–16. 10.1080/20445911.2013.796377

Grey S, Tanner D, & Van Hell JG (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. Brain Research, 1669, 27–43. 10.1016/j.brainres.2017.05.024 [PubMed: 28554807]

Gross M, & Kaushanskaya M (2018). Contributions of nonlinguistic task-shifting to language control in bilingual children. Bilingualism: Language and Cognition, 21(1), 181–194. 10.1017/S1366728916001097

Hahne A, & Friederici AD (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. Bilingualism: Language and Cognition, 4(2), 123–141. 10.1017/S1366728901000232

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

Hopp H (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. Language Acquisition, 21(3), 250–278. 10.1080/10489223.2014.892943

Hopp H (2015). Individual differences in the L2 processing of object-subject ambiguities. Applied Psycholinguistics, 36(2), 129–73.

Hsu NS, Jaeggi SM, & Novick JM (2017). A common neural hub resolves syntactic and non-syntactic conflict through cooperation with task-specific networks. Brain and Language, 166, 63–77. [PubMed: 28110105]

Jegerski J (2018). Sentence Processing in Spanish as a Heritage Language: A Self‐Paced Reading Study of Relative Clause Attachment. Language Learning, 68(3), 598–634.

Jegerski J, & Sekerina IA (2020). The processing of input with differential object marking by heritage Spanish speakers. Bilingualism: Language and Cognition, 23(2), 274–282.

Just MA, & Carpenter PA (1992). A capacity theory of comprehension: individual differences in working memory. Psychological Review, 99(1), 122. [PubMed: 1546114]

Kaan E, Harris A, Gibson E, & Holcomb P (2000). The P600 as an index of syntactic integration difficulty. Language and cognitive processes, 15(2), 159–201.

Kasparian K, & Steinhauer K (2016). Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second language acquisition. Neuropsychologia, 93, 200–217. [PubMed: 27751710]

Kasparian K, & Steinhauer K (2017). When the second language takes the lead: Neurocognitive processing changes in the first language of adult attriters. Frontiers in Psychology, 8, 1–22. 10.3389/fpsyg.2017.00389 [PubMed: 28197108]

Kasparian K, Vespignani F, & Steinhauer K (2017). First language attrition induces changes in online morphosyntactic processing and re‐analysis: an ERP study of number agreement in complex Italian sentences. Cognitive Science, 41(7), 1760–1803. [PubMed: 27868225]

Kim AE, Oines L, Miyake A, (2017). Individual Differences in Verbal Working Memory Underlie a Tradeoff Between Semantic and Structural Processing Difficulty During Language Comprehension: An ERP Investigation. Journal of Experimental Psychology: Learning, Memory and Cognition, 44(3), 406–420.

King J, & Just MA (1991). Individual Differences in Syntactic Processing: The Role of Working Memory. Journal of Memory and Language, 30(5), 580–602.

Kluender R, & Kutas M (1993). Subjacency as a processing phenomenon. Language and Cognitive Processes, 8(4), 573–633.

Kolk HH, Chwilla DJ, Van Herten M, & Oor PJ (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. Brain and Language, 85(1), 1–36. [PubMed: 12681346]

Kondo–Brown K (2005). Differences in language skills: Heritage language learner subgroups and foreign language learners. The Modern Language Journal, 89(4), 563–581.

Kotz SA (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. Brain and Language, 109(2–3), 68–74. 10.1016/j.bandl.2008.06.002 [PubMed: 18657314]

Kroll JF, Dussias PE, Bice K, & Perrotti L (2015). Bilingualism, Mind, and Brain. Annual Review of Linguistics, 1(1), 377–394. 10.1146/annurev-linguist-030514-124937

Kuperberg GR (2007). Neural mechanisms of language comprehension: Challenges to syntax. Brain research, 1146, 23–49. [PubMed: 17400197]

Kupisch T, & Rothman J (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. International Journal of Bilingualism, 22(5), 564–582.

Kutas M, & Federmeier KD (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annual Review of Psychology, 62(August), 621–47. 10.1146/annurev.psych.093008.131123

Lev-Ari S, & Peperkamp S (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. Journal of Phonetics, 41(5), 320–331.

Linck JA, Osthus P, Koeth JT, & Bunting MF (2014). Working memory and second language comprehension and production: a meta-analysis. Psychonomic Bulletin & Review, 21(4), 861–83. 10.3758/s13423-013-0565-2 [PubMed: 24366687]

Marian V, Bartolotti J, Chabal S, & Shook A (2012). Clearpond: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. PLoS ONE, 7(8). 10.1371/journal.pone.0043230

Marian V, Blumfeld H, & Kaushanskaya M (2007). Bilinguals and Multilinguals. Hearing Research, 50(August), 940–967. 10.1044/1092-4388(2007/067)

McKeever WF, Seitz KS, Hoff AL, Marino MF, & Diehl JA (1983). Interacting sex and familial sinistrality characteristics influence both language lateralization and spatial ability in right handers. Neuropsychologia, 21(6), 661–668. 10.1016/0028-3932(83)90064-7 [PubMed: 6664485]

McLaughlin J, Osterhout L, & Kim A (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. Nature Neuroscience, 7(7), 703–704. [PubMed: 15195094]

McLaughlin J, Tanner D, Frenck-Mestre C, Valentine G, & Osterhout L (2010). Brain Potentials Reveal Discrete Stages of L2 Grammatical Learning. Language Learning, 60(s2), 123–150.

McMurray B, Samelson VM, Lee SH, & Bruce Tomblin J (2010). Individual differences in online spoken word recognition: Implications for SLI. Cognitive Psychology, 60(1), 1–39. 10.1016/j.cogpsych.2009.06.003 [PubMed: 19836014]

Montrul S (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. International Journal of bilingualism, 13(2), 239–269.

Montrul S (2010). Current issues in heritage language acquisition. Annual Review of Applied Linguistics, 30, 3–23.

Montrul S (2011). Morphological errors in Spanish second language learners and heritage speakers. Studies in Second Language Acquisition, 33(2), 163–192.

Montrul SA (2012). Is the heritage language like a second language?. Eurosla Yearbook, 12(1), 1–29.

Moore M, & Gordon PC (2014). Reading ability and print exposure: item response theory analysis of the author recognition test. Behavior Research Methods, 47(4), 1095–1109. 10.3758/s13428-014-0534-3

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

Nakano H, Saron C, & Swaab TY (2010). Speech and Span: Working Memory Capacity Impacts the Use of Animacy but Not of World Knowledge during Spoken Sentence Comprehension. Journal of Cognitive Neuroscience, 22(12), 2886–2898. 10.1162/jocn.2009.21400 [PubMed: 19929760]

Ojima S, Nakata H, & Kakigi R (2005). An ERP study of second language learning after childhood: Effects of proficiency. Journal of Cognitive Neuroscience, 17(8), 1212–1228. [PubMed: 16197679]

Oldfield RC (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9(1), 97–113. 10.1016/0028-3932(71)90067-4 [PubMed: 5146491]

Osterhout L (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. Brain and Language, 59(3), 494–522. 10.1006/brln.1997.1793 [PubMed: 9299074]

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

Osterhout L, Pitkänen I, McLaughlin J, & Zeitlin M (2019). Event-related Potentials as Metrics of Foreign Language Learning and Loss. In The Oxford handbook of language attrition.

Pakulak E, & Neville HJ (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. Journal of Cognitive Neuroscience, 22(12), 2728–44. 10.1162/jocn.2009.21393 [PubMed: 19925188]

Pearlmutter NJ, & MacDonald MC (1995). Individual Differences and Probabilistic Constraints in Syntactic Ambiguity Resolution. Journal of Memory and Language, 34(4), 521–542.

Pires A, & Rothman J (2009). Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. International Journal of Bilingualism, 13(2), 211–238.

Pivneva I, Palmer C, & Titone D (2012). Inhibitory control and L2 proficiency modulate bilingual language production: evidence from spontaneous monologue and dialogue speech. Frontiers in Psychology, 3(March), 57 10.3389/fpsyg.2012.00057 [PubMed: 22438846]

Pliatsikas C, DeLuca V, & Voits T (2019). The many shades of bilingualism: language experiences modulate adaptations in brain structure. Language Learning.

Polinsky M (2011). Reanalysis in adult heritage language: New evidence in support of attrition. Studies in second language acquisition, 33(2), 305–328.

Polinsky M, & Scontras G (2020). Understanding heritage languages. Bilingualism: Language and Cognition, 23(1), 4–20.

Pu H, Holcomb PJ, & Midgley KJ (2016). Neural changes underlying early stages of L2 vocabulary acquisition. Journal of Neurolinguistics, 40, 55–65. [PubMed: 28983152]

Putnam MT, & Sánchez L (2013). What's so incomplete about incomplete acquisition?: A prolegomenon to modeling heritage language grammars. Linguistic Approaches to Bilingualism, 3(4), 478–508.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL https://www.R-project.org/.

Rivera-Mills SV (2012). Spanish heritage language maintenance. Spanish as a heritage language in the United States: The state of the field, 21–42.

Rodríguez E, & Reglero L (2015). Heritage and L2 processing of person and number features: Evidence from Spanish subject-verb agreement. EuroAmerican Journal of Applied Linguistics and Languages, 2, 11–30.

Rothman J (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. International Journal of Bilingualism, 11(4), 359–389.

Rothman J (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. International Journal of Bilingualism, 13(2), 155–163.

Scontras G, Fuchs Z, & Polinsky M (2015). Heritage language and linguistic theory. Frontiers in Psychology, 6, 1545. [PubMed: 26500595]

Sekerina IA, Laurinavichyute AK, & Dragoy O (2019). What eye movements can and cannot tell us about Wh-movement and Scrambling In Grammatical Approaches to Language Processing (pp. 147–165). Springer, Cham.

Sorace A, & Serratrice L (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. International Journal of Bilingualism, 13(2), 195–210.

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

Tanner D, McLaughlin J, Herschensohn J, & Osterhout L (2012). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. Bilingualism: Language and Cognition, 55(2), 367–382. 10.1017/S1366728912000302

Tanner D, & Van Hell JG (2014). ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia, 56(April), 289–301. 10.1016/j.neuropsychologia.2014.02.002 [PubMed: 24530237]

Townsend DJ, Carrithers C, & Bever TG (2001). Familial Handedness and Access to Words, Meaning, and Syntax during Sentence Comprehension. Brain and Language, 78(3), 308–331. 10.1006/brln.2001.2469 [PubMed: 11703060]

Turner ML, & Engle RW (1989). Is working memory capacity task dependent? Journal of Memory and Language, 28(2), 127–154.

Ullman MT (2001). The declarative/procedural model of lexicon and grammar. Journal of Psycholinguistic Research, 30(1), 37–69. [PubMed: 11291183]

Van Hell JG, & Tokowicz N (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. Second Language Research, 26(1), 43–74.

Veale JF (2014). Edinburgh Handedness Inventory – Short Form: A revised version based on confirmatory factor analysis. Laterality: Asymmetries of Body, Brain and Cognition, 19(2), 164–177. 10.1080/1357650X.2013.783045

Wampler EK (2017). Individual Differences in Grammatical Error Processing. University of Washington.

Weber K, & Lavric A (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. Psychophysiology, 45(6), 920–925. 10.1111/j.1469-8986.2008.00691.x [PubMed: 18778320]

White EJ, Genesee F, & Steinhauer K (2012). Brain responses before and after intensive second language learning: proficiency based changes and first language background effects in adult learners. PloS One, 7(12), e52318 10.1371/journal.pone.0052318 [PubMed: 23300641]

Zirnstein M, Bice K, & Kroll JF (2019). Variation in language experience shapes the consequences of bilingualism In Sekerina IA, Spradlin L, & Valian VV (Eds), Bilingualism and Executive Function: An Interdisciplinary Approach (pp. 35–48). Amsterdam, NL: John Benjamins Publishers.

Zirnstein M, Hell J. G. Van, & Kroll JF (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. Cognition, 176(February), 87–106. 10.1016/j.cognition.2018.03.001 [PubMed: 29549762]

Highlights:

1. We assessed individual differences in ERP responses to grammatical violations
2. Heritage bilinguals had smaller ERP effects in their less-dominant L1
3. Working memory modulated ERP responses in monolinguals
4. Proficiency modulated ERP responses in both bilinguals' languages
5. Early reading experience affected the bilinguals' N400

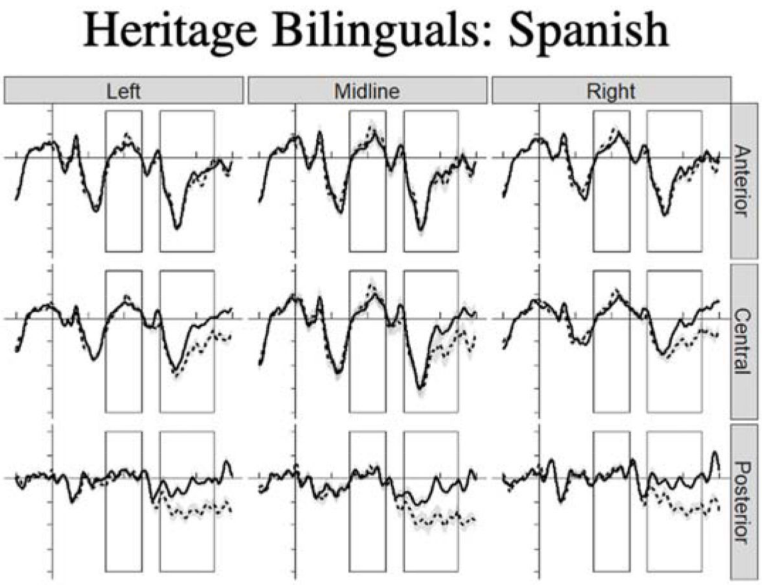


Figure 1.

Note: ERP waveforms recorded from heritage bilinguals reading grammatical sentences (solid line) and ungrammatical sentences (dashed line) in Spanish. Negative is plotted up. Waveforms were averaged across electrodes present in each electrode region (anteriority and laterality). Shaded region surrounding lines represents one standard error at each time point. Boxes highlight the time windows of interest: 300–500 ms (N400 time window) and 600–900 ms (P600 time window). Dashes on the x-axis appear every 200 ms; dashes on the y-axis mark every 2 μV.

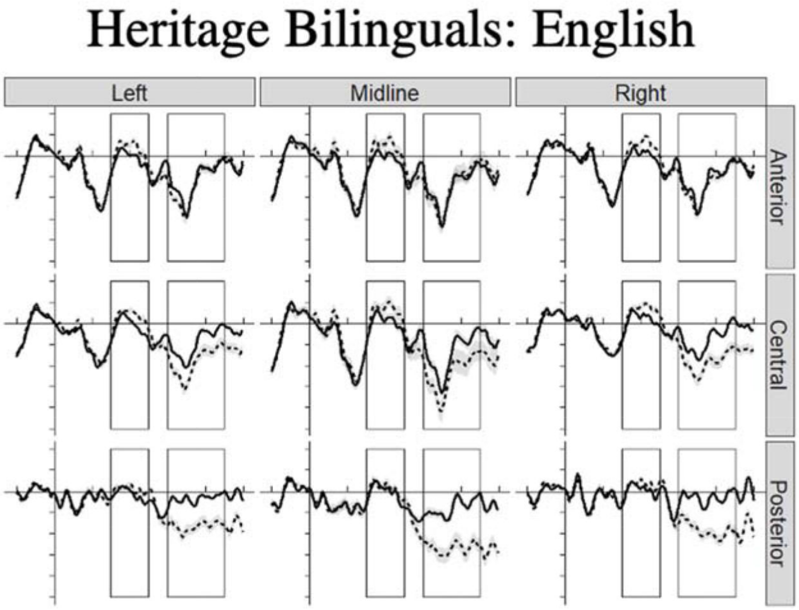


Figure 2.

Note: ERP waveforms recorded from heritage bilinguals reading grammatical sentences (solid line) and ungrammatical sentences (dashed line) in English. Negative is plotted up. Waveforms were averaged across electrodes present in each electrode region (anteriority and laterality). Shaded region surrounding lines represents one standard error at each time point. Boxes highlight the time windows of interest: 300–500 ms (N400 time window) and 600–900 ms (P600 time window). Dashes on the x-axis appear every 200 ms; dashes on the y-axis mark every 2 μV.

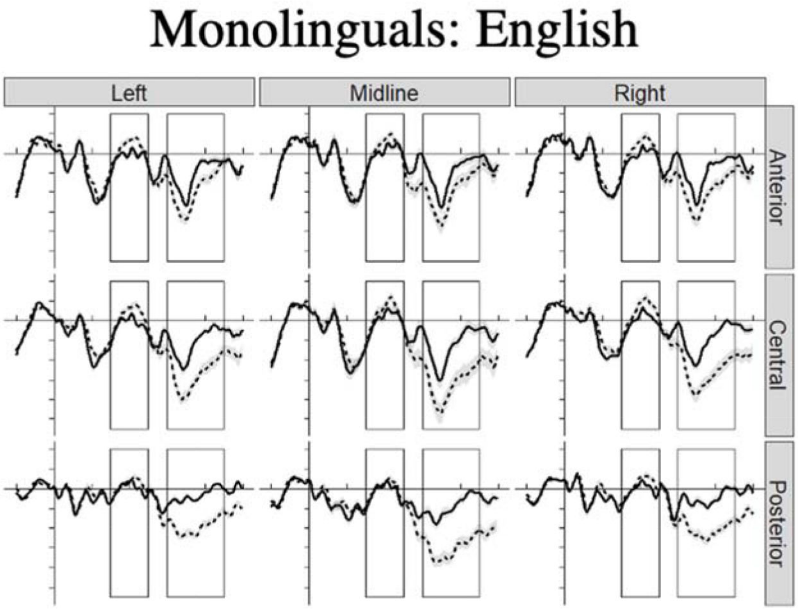


Figure 3.

Note: ERP waveforms recorded from monolingual English speakers reading grammatical sentences (solid line) and ungrammatical sentences (dashed line) in English. Negative is plotted up. Waveforms were averaged across electrodes present in each electrode region (anteriority and laterality). Shaded region surrounding lines represents one standard error at each time point. Boxes highlight the time windows of interest: 300–500 ms (N400 time window) and 600–900 ms (P600 time window). Dashes on the x-axis appear every 200 ms; dashes on the y-axis mark every 2 μV.

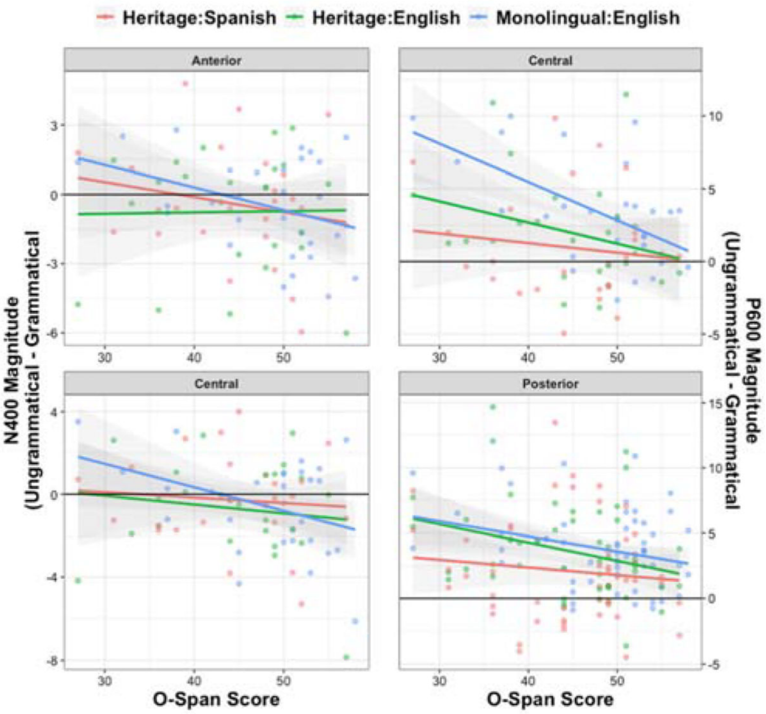


Figure 4.

Effect of working memory on brain responses for each group/language

Note: The N400 results (left) are shown for anterior and central midline electrodes, where the largest N400 effects were found. More negative values indicate larger canonical N400 effects. The P600 results (right) are shown for central and posterior midline electrodes, where the largest P600 effects were found. More positive values indicate larger canonical P600 effects. Black line at 0 indicates where no effect would be present for each component. Shaded region around best-fit regression lines indicate the 95% confidence interval.

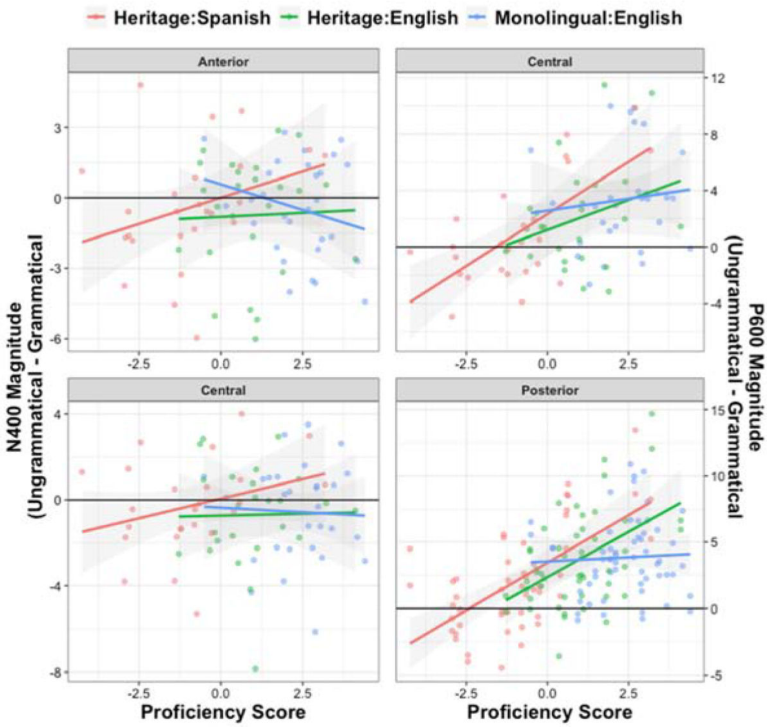
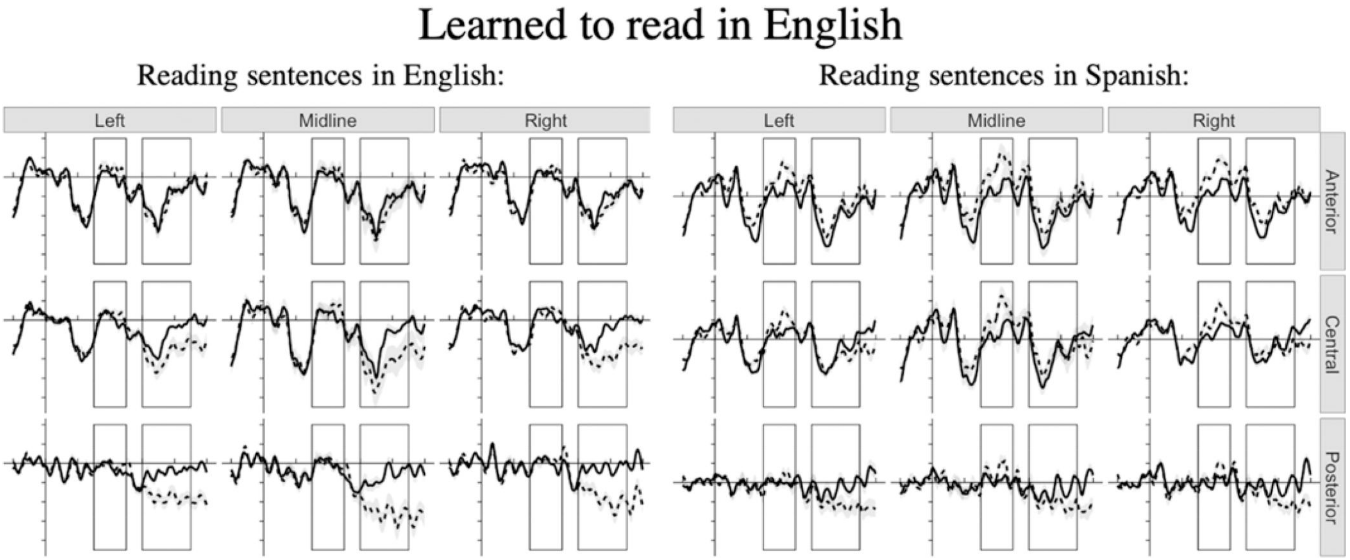


Figure 5.

Effect of proficiency on brain responses for each group/language

Note: The N400 results (left) are shown for anterior and central midline electrodes, where the largest N400 effects were found. More negative values indicate larger canonical N400 effects. The P600 results (right) are shown for central and posterior midline electrodes, where the largest P600 effects were found. More positive values indicate larger canonical P600 effects. Black line at 0 indicates where no effect would be present for each component. Shaded region around best-fit regression lines indicate the 95% confidence interval.



Learned to read in English

Reading sentences in English:

Reading sentences in Spanish:

Figure 6.

ERP waveforms recorded from heritage bilinguals who initially learned to read in English

Note: ERP waveforms recorded while reading grammatical sentences (solid line) and ungrammatical sentences (dashed line) in English (left) and in Spanish (right). Negative is plotted up. Waveforms were averaged across electrodes present in each electrode region (anteriority and laterality). Shaded region surrounding lines represents one standard error at each time point. Boxes highlight the time windows of interest: 300–500 ms (N400 time window) and 600–900 ms (P600 time window). Dashes on the x-axis appear every 200 ms; dashes on the y-axis mark every 2 μV.

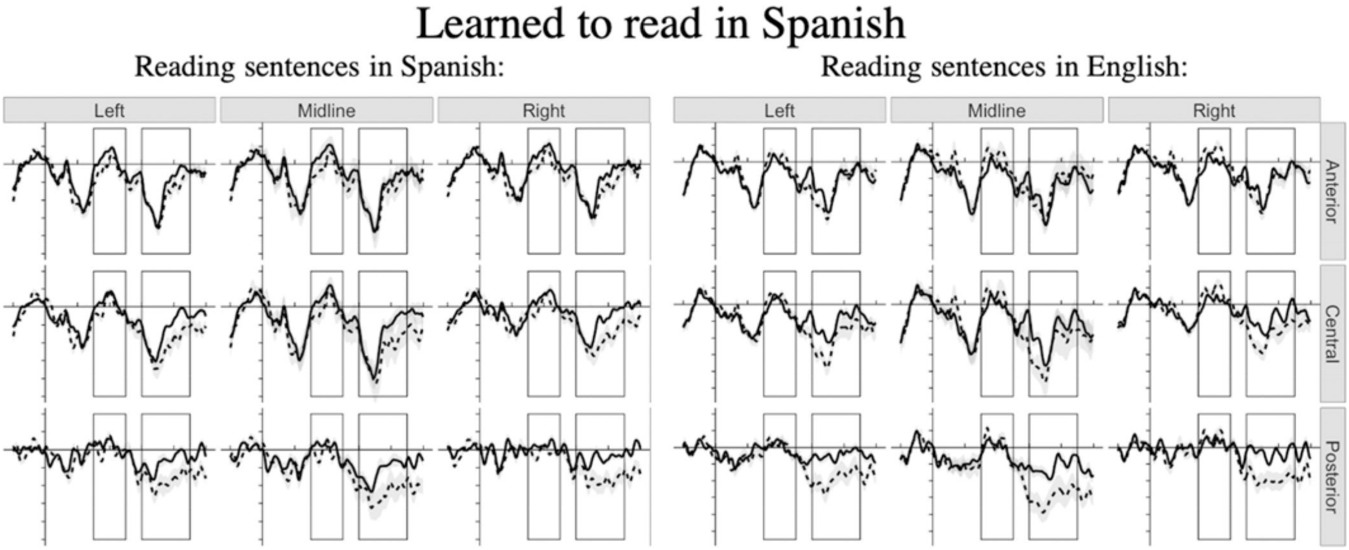


Figure 7. ERP waveforms recorded from heritage bilinguals who initially learned to read in Spanish

Note: ERP waveforms recorded while reading grammatical sentences (solid line) and ungrammatical sentences (dashed line) in Spanish (left) and in English (right). Negative is plotted up. Waveforms were averaged across electrodes present in each electrode region (anteriority and laterality). Shaded region surrounding lines represents one standard error at each time point. Boxes highlight the time windows of interest: 300–500 ms (N400 time window) and 600–900 ms (P600 time window). Dashes on the x-axis appear every 200 ms; dashes on the y-axis mark every 2 μV.

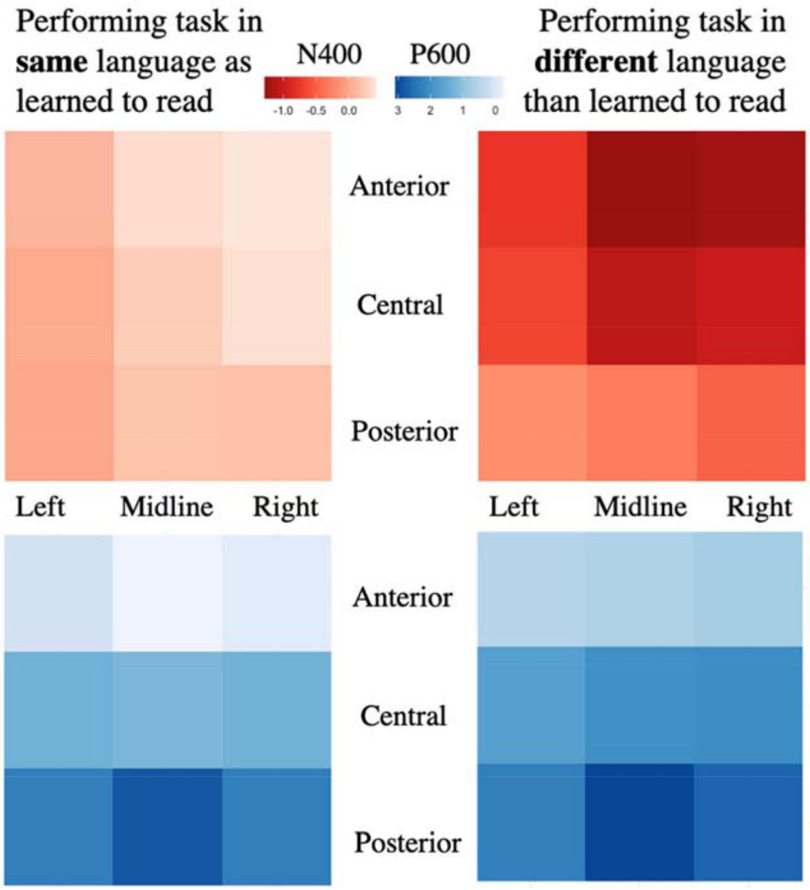


Figure 8.

Summary of the effect of early reading experience on ERP responses

Note: Heatmap of the distribution of group-averaged effects of early reading experience on the N400 (top row) and P600 (bottom row), for bilinguals performing the grammaticality judgment task in the same language as they initially learned how to read (e.g., bilinguals who learned to read in Spanish, performing the task in Spanish), and bilinguals performing the task in the other language than that in which they initially learned how to read (e.g., bilinguals who learned to read in Spanish, performing the task in English).

Table 1

Participant and demographic characteristics

	Heritage bilinguals: Spanish	Heritage bilinguals: English	Monolinguals
N (# Female)	25 (21)		29 (20)
Age (SD)	19.6 (1.85)		22.21 (4.54)
O-Span (SD; range)	44.64 (7.76; 27–57)		47.90 (7.67; 27–58)
Age of acquisition (SD; range)	1.52 (1.16; 0–4)	3.54 (1.74; 0–6)	1.00 (1.23; 0–5)
Self-rated proficiency (SD; range)	8.55 (1.21; 5.83–10)	9.73 (0.43; 8.5–10)	9.67 (0.63; 7.67–10)
D′(SD; range)	1.87 (0.94; 0.12–3.5)	2.55 (0.65; 1.33–3.8)	3.21 (0.62; 2.02–4.11)
Verbal fluency (SD; range)	9.43 (2.05; 6.75–14.75)	12.96 (2.23; 8.25–18.25	13.70 (2.61; 7.25–18.25)

Note: Self-rated proficiency was based on a scale from 1 (basic greetings) to 10 (native fluency), averaged across ratings for speaking, listening, and understanding. Ages are reported in years. D′ scores were calculated based on performance on the grammaticality judgment task in each language. Verbal fluency scores were averaged across four categories for each participant.

Table 2

Summary of group-averaged ERP effects.

	Heritage Speakers: Spanish		**Heritage Speakers: English**		**Monolinguals: English**
N400 Effect:	No effect	<	Anterior• Central•	≈	Anterior* Central*
P600 Effect:	Central• Posterior•	<	Central* Posterior*	≈	Anterior* Central* Posterior*

Note: Indicates electrode any electrode region(s) in which a significant (*) or marginally significant (•) effect was found, for each ERP component (N400, P600) and each group/language. Comparators between columns indicate whether the effect was significantly different (<) or not (≈) for each comparison (heritage speakers in Spanish vs. English, heritage speakers vs. monolinguals in English).

Table 3

Summary of analyses relating working memory and proficiency with ERP responses for each group/language

	Heritage Speakers: Spanish	**Heritage Speakers: English**	**Monolinguals: English**
Working memory	No effect on N400 or P600	No effect on N400 or P600	Higher span -> larger N400, smaller P600
Proficiency	Higher proficiency -> smaller N400, larger P600	No effect on N400 Higher proficiency -> larger P600	No effect on N400 or P600