Cross-linguistic Influence in Bilingual Processing: An ERP Study

Gita Martohardjono, Ian Phillips, Christen N. Madsen II, and Richard G. Schwartz

1. Introduction

The study of heritage languages and their speakers has been the topic of investigation in at least two subfields of linguistics: Second Language Acquisition (SLA) and Bilingualism, and Sociolinguistics. Research in SLA/Bilingualism has largely taken the approach of experimental studies and group comparisons in an effort to delineate similarities and differences between heritage and native speakers. The ultimate goal is to characterize the cognitive processes that the heritage speaker has undergone which would in turn explain his/her use of the language. Typically, heritage speakers, sometimes classified along proficiency measures, are given various language tasks, usually involving acceptability or truth value judgments, sentence manipulations, and other explicit measures. Results are then compared to those taken from non-heritage bilingual groups, who are considered “native speakers” of the examined language variety, or sometimes even from monolingual speakers of the language. For many structures tested, the results usually point to group differences especially at lower levels of proficiency and the general conclusion is that heritage speakers, because of extended contact with their now-dominant language, have either undergone attrition or “incomplete acquisition” (e.g., Polinsky 2011; Montrul 2002).

Sociolinguistic studies take a different approach, and indeed have a different goal from that of SLA/Bilingualism studies. Heritage languages are defined as those who stand in a minority position to a dominant, societal language, but are regarded as separate and valid systems in and of themselves, rather than deficit versions of the variety spoken in the home country. Heritage speakers are defined by characteristics of language use and provenance. For example, in a study on heritage speakers of Cantonese, Italian, Russian, and...
Ukrainian in Toronto, Nagy (2015) includes any speaker who is exposed to the heritage language naturalistically since birth, regardless of eventual dominance. Heritage languages are investigated for evidence of language change from the variety spoken in the homeland to the variety spoken under conditions of language contact. The aim is to find innovations in the speech of heritage speakers, and the data are usually drawn from corpora of conversational speech. Nagy observed that experimental and sociolinguistic approaches stand in contrast to each other, often producing contrasting results, and that few or no attempts have been made to bridge the two approaches. The present study attempts to do just that.

Heritage language studies can and should also be related to a relatively new subfield of research in bilingualism that inspects cross-linguistic influence, not of the first language (L1) on the second language (L2), as is the norm in SLA studies of transfer from the first language, but of the reverse case where a later-learned language affects the first-learned language both in L2 learners (e.g., Cook 2003), and in fluent bilinguals (e.g., Fernández 2003; Dussias and Sagarra 2007; Tsimpli et al. 2004). Finally, heritage languages have rarely been studied from the perspective of online processing, nor has there been a systematic comparison of results from implicit tasks, which tap unconscious language processes, and explicit tasks, which require conscious linguistic judgment and manipulation on the part of the subject.

In an attempt to address the above issues, we designed a study that includes both heritage speakers of Spanish, who were exposed to a minority home language (Spanish) since birth, who are in a language contact situation (English in New York City), and who are now dominant in English; and late bilinguals, who grew up speaking Spanish in a Spanish-speaking country and who immigrated to the U.S. and started using English on a daily basis later in life (age 18 and above). We further combined experimental approaches, both implicit (event-related potentials (ERP)) and explicit (acceptability judgments), with a sociolinguistic questionnaire probing proficiency, literacy, use, and exposure to both Spanish and English.

Our aims were two-fold: First, to see whether heritage speakers differ qualitatively from non-heritage speakers in the way they process the first-learned language, Spanish. And second, to identify those speaker variables derived from the questionnaire, that are most predictive of divergence, as measured by ERP components (implicit) and acceptability judgments (explicit).

2. Experiment 1: Event-Related Potentials (ERP)

We administered an ERP experiment with auditorily presented stimuli to address the following question: Do heritage speakers and late bilinguals exhibit the same or different processing patterns when hearing syntactically anomalous Spanish sentences? The use of ERP with auditorily presented stimuli is especially well suited for studying processing in the first-learned language of heritage speakers, as subjects are not required to provide any conscious response
to the stimuli, which avoids possible confounds related to literacy and education. Heritage speakers typically do not receive formal schooling in the heritage language and often lack age-appropriate literacy skills in the heritage language (Beaudrie and Fairclough 2012; City of New York Department of Education 2011; Menken and Kleyn 2010; Valdés et al. 2006), which often results in poor performance on experimental tasks that require metalinguistic judgments about acceptability or grammaticality (Klein and Martohardjono 2009). ERP measures voltage changes at the scalp associated with the synchronous firing of brain cell populations in response to a stimulus and the ERP technique has been used successfully to study language processing in monolingual (Friederici 2002; Kutas and Hillyard 1980; Osterhout 1994), bilingual (Kotz 2009; Moreno, Rodríguez-Fornells, and Laine 2008), and learner populations (Osterhout et al. 2008). The ERP literature shows that recognition of semantic and lexical anomaly is typically indexed by an N400 component—a negative-going deflection largest over centro-parietal regions of the scalp that peaks around 400ms following the anomaly (Kutas and Hillyard 1980; Swaab et al. 2012)—while syntactic anomaly is typically indexed by a P600 component—an increase in positivity over posterior sites that peaks around 600ms following the anomaly (Friederici 1995; Osterhout and Holcomb 1992).

Given these well-establish ERP components, we asked three questions: (1) Are ERP components that indicate recognition of syntactic anomaly weaker or absent for heritage speakers compared to late bilinguals when hearing Spanish sentences that exhibit grammatical patterns that are anomalous in Spanish but acceptable in parallel English sentences? (2) Are ERP components that indicate recognition of syntactic anomaly weaker or absent for heritage speakers when hearing Spanish sentences that exhibit grammatical patterns that are anomalous in both Spanish and English? (3) For all bilingual subjects, are these ERP components modulated by linguistic and/or demographic factors such as generation, language use, or socioeconomic status (SES)? We addressed these questions by collecting ERP data for heritage speakers and late bilinguals while they listened to anomalous Spanish sentences (described in Stimuli) and analyzing the results with linear mixed modeling, using responses collected in the linguistic background questionnaire (described in Methods) as individual-level predictor variables in the model.

2.1. Subjects

N=38 Spanish-English bilingual adults from the New York City area (N=23 female) completed Experiment 1 and were paid for their participation. All subjects were right handed based on the Edinburgh Handedness Inventory (Oldfield 1971), had normal or corrected-to-normal vision, and had no history of neurological disorder. Subjects were categorized as heritage speakers (U.S.-born, second immigrant generation adults) or late bilinguals (Latin American-born, first immigrant generation adults) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, and Polinsky 2013). This
information was collected via a 38-item linguistic background questionnaire that was administered during the recruitment process. The questionnaire included standard items pertaining to language history (Li, Sepanski, and Zhao 2006) and additional items pertaining to demographics (education, SES), language ability (self-reported literacy and speaking abilities in English and Spanish), current language use (language choice with different interlocutors and contexts), and language exposure (language used during childhood and adolescence, amount of exposure to English over the lifetime). Late bilinguals (N=20): were born and raised in a Spanish-speaking Latin American country; came to the U.S. between ages 18–40 (M=25.56, SD=4.77); had been living in the U.S. no more than 20 years at the time of testing (M=5.84, SD=6.06); and were between 23–50 years old at the time of testing (M=31.7, SD=3.70). Heritage speakers (N=18): were born in the continental U.S. or were born in the Spanish-speaking Caribbean and brought to the U.S. before age eight (M=1.52, SD=0.96) (Benmamoun, Montrul, and Polinsky 2010); were born no more than 20 years after the parents arrived in the U.S. (M=6.81, SD=6.82); were 18–50 years old at the time of testing (M=26.61, SD=7.30); and were raised speaking primarily Spanish until at least age 10 by caregivers who met the late bilingual criteria described above.

To ensure adequate levels of Spanish proficiency for our heritage speakers, who are typically English dominant, we tested their auditory comprehension of complex sentences using a computerized picture-pointing task: the RISLUS Multilingual Syntax Test (RMST) (Klein and Martohardjono 2009). All heritage speakers performed at ceiling on this test, indicating fluency in Spanish.

2.2. Stimuli

To test the processing of Spanish sentences exhibiting grammatical patterns that are anomalous in both Spanish and English, subjects listened to N=15 Wh-questions containing a complex noun-phrase violation like (1a), and their grammatical counterparts like (1b). Sentences like (1a) are ungrammatical due to the presence of the head noun in the complex noun phrase (cf. 1b), and this property is common to both Spanish and English.

(1) a. *Qué vecino contó Juan ||el chisme que robó el carro anoche? 
What neighbor told Juan the rumor that robbed the car last night 
‘What neighbor did Juan say the rumor that robbed the car last night?’

b. Qué vecino contó Juan ∅ ||que robó el carro anoche? 
What neighbor told Juan that robbed the car last night 
‘What neighbor did Juan say robbed the car last night?’

---

1 We define the Spanish-speaking Caribbean to include Cuba, Puerto Rico, and the Dominican Republic. N=4 heritage speaker subjects were raised by caregivers from Puerto Rico and N=14 were raised by caregivers from the Dominican Republic.

2 In Ex. 1–2, vertical lines “||” indicate the comparison point for the ERP analysis.
To test the processing of Spanish sentences exhibiting grammatical patterns that are anomalous in Spanish but not in English, subjects listened to N=20 Wh-questions with a missing complementizer *que* like (2a), and their grammatical counterparts like (2b). In sentences like (2a), the absence of the complementizer *que* introducing the subordinate clause is ungrammatical. This is in contrast to parallel English sentences that are rendered ungrammatical by the presence of the complementizer *that*.

(2) a. *Qué hermana confesó Inés ∅ ||había comdío la tarta?*  
    ‘What sister did Inés confess had eaten the cake?’

b. Qué hermana confesó Inés ||*que* había comdío la tarta?  
    ‘What sister did Inés confess that had eaten the cake?’

The stimuli were recorded by a female native speaker of Spanish and were presented as natural running speech in two-sentence trials. Each trial consisted of a declarative sentence that provided a context followed by a *Wh*-question related to the context. Half of the trials contained an ungrammatical *Wh*-question, like (1a) and (2a); and half contained a grammatical *Wh*-question, like (1b) and (2b), which served as controls and provided baseline processing measures. These trials were interspersed with an additional N=270 filler items presented in the same trial format, and were distributed evenly over five blocks and pseudorandomized so that the same condition never appeared in consecutive trials. Each block lasted approximately 15 minutes with short breaks in between.

2.3. Methods

Subjects completed Experiment 1 seated in a comfortable chair, 70cm from a computer monitor and external speakers in a shielded IAC booth. At the start of the experiment, subjects listened to recorded instructions in English and Spanish and completed a practice session. Trials were presented auditorily and following 40% of the trials, a visual cue appeared that prompted subjects to answer a comprehension question about the sentences in the trial. The experiment resumed after subjects answered the question. Subjects were not asked to make any metalinguistic judgments during the ERP experiment.

Continuous EEG data were recorded during the experiment for 32 Ag/AgCl sintered electrodes mounted in a QuikCap, positioned according to the 10-20 International Electrode System (Jasper 1958)\(^3\) and referenced online to the nose. EEG data were amplified with a Neuroscan SynAmps\(^2\) amplifier and recorded

---

\(^3\) Data was collected for: FP1, FP2, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T7, C3, CZ, C4, T8, TP7, CP3, CPZ, CP4, TP8, P7, P3, PZ, P4, P8, O1, OZ, O2.
between DC–100Hz with a 1000Hz sampling rate and electrode impedances kept below 15kΩ. After recording, EEG data were pre-processed with Neuroscan Edit software, which involved: interpolating blocked or disconnected electrodes; applying a digital 0.1–30Hz (FIR) bandpass filter; correcting for ocular artifacts (i.e., blinks); segmenting the EEG data into 1000ms epochs starting at the onset of the words marked with “||” in (1–2); baseline correcting epochs with a 200ms pre-stimulus interval; and rejecting epochs that contained artifacts exceeding +/- 70µV between 0–900ms. Four subjects (N=3 heritage speakers, N=1 late bilingual) who had greater than 40% data loss following this procedure were removed from further analysis.

2.4. Results

Pre-processed data for individual epochs were exported to R where they were downsampled to 250Hz and grand mean centered before analysis. A point-by-point subtraction wave was calculated for the centered data by subject and by item by subtracting amplitude for the grammatical control items (1b) and (2b) from the corresponding ungrammatical items (1a) and (2a). To isolate ERP components consistent with an N400 or P600, a spatial Principal Component Analysis (PCA) with a promax rotation was performed on the subtraction waves for all subjects during the 200–400ms (for N400) and 400–800ms (for P600) time window. For missing items, the subject-condition average was used. Following this procedure, subtraction waves were scaled by the weightings for the rotated component that was spatially consistent with either an N400 or P600, i.e., the rotated components with the highest weightings over the centro-parietal region (for N400) or posterior region (for P600). The results of this procedure are given below, by condition.

To explore the effect of speaker variables (collected via the linguistic background questionnaire) on ERP component amplitude, scaled amplitude data was analyzed with linear mixed-effects models with random intercepts for subject and item across groups. Before modeling, questionnaire responses were analyzed and collinearity was addressed by combining related and correlated items to derive theoretically- and sociolinguistically-motivated composite variables. For example, a number of items probed language preferences with different forms of media (listening to music/radio, watching television, reading novels) and we grouped these items into one composite variable, language used for media. We ultimately derived 13 variables of language use, language exposure, and demographic variables, including a variable for heritage speaker (second generation bilingual) vs. late bilingual (first generation bilingual). The distribution and correlations of these final variables were examined to ensure collinearity was avoided for the mixed-effects modeling.

A backwards elimination procedure was used for all predictor variables, removing in a step-wise fashion the variable with the highest p-value. To determine if a given variable should have been retained, models with and without the variable were compared to check if the fit of the model without the
variable was significantly degraded. The process was completed upon reaching a model for which no other variables could be removed without significantly degrading the fit. The final models contained only significant predictor variables of the dependent variables: N400 scaled amplitude or P600 scaled amplitude.

![Figure 1: Grand mean amplitude by grammaticality for the complex noun phrase condition (top, electrode P3 shown) and comp-trace condition (bottom, electrode CZ shown).](image)

**2.4.1. Complex Noun Phrase**

For the complex noun phrase condition, an ERP component consistent with a P600 was found in the 400–800ms time window (Fig. 1). Mean scaled subtraction amplitude for complex noun phrase items in the 400–800ms window was positive across groups (M=0.03, SD=0.24) indicating that ungrammatical items (1a) elicited a greater positive-going deflection in the time window compared to control items (1b). There was considerable overlap in component amplitude between late bilingual subjects (M=0.03, SD=0.25, Range: -0.40–0.60) and heritage speaker subjects (M=0.03, SD=0.23, Range: -0.35–0.41).

To analyze the effect of speaker variables on this component, mean scaled subtraction amplitude was modeled following the procedure outlined above. The final model, shown in Table 1, contained no significant predictor variables of P600 amplitude for complex noun phrase items, $\chi^2(1)=1.01, p=.32$. No variables
of language use or language exposure significantly improved the fit of the model, and generation was also not a significant predictor of P600 amplitude.

Table 1: Best fitting model for P600 scaled subtraction amplitude for complex noun phrase sentences.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.04</td>
<td>0.05</td>
<td>--</td>
<td>0.77</td>
<td>--</td>
</tr>
</tbody>
</table>

† p<.1 *p<.05 ** p<.01 ***p<.001

2.4.2. Comp-trace

For the comp-trace condition, an ERP component consistent with an N400 was found in the 200–400ms time window (Fig. 1). Mean scaled subtraction amplitude for comp-trace items in the 200–400ms time window was negative across groups (M=-0.13, SD=0.78) indicating that ungrammatical items (2a) elicited a greater negative-going deflection in the time window compared to control items (2b). Again, there was considerable overlap in component amplitude between late bilingual subjects (M=-0.21, SD=0.70, Range: -1.08–0.16) and heritage speaker subjects (M=-0.04, SD=0.86, Range: -0.33–0.27).

These results were analyzed following the same procedure described above. The final model, shown in Table 2, which contained only significant predictor variables of N400 amplitude for comp-trace sentences included variables of language use, language exposure, and SES, \( \chi^2(1)=5.78, p<.01 \). There were significant main effects of English use in the home, social settings, and work, proportion of life in an English dominant environment, and SES. The mean scaled subtraction amplitude for 200–400ms decreased as subjects’ English use increased (\( \beta=0.43, SE(\beta)=0.15, t(32.44)=2.86, p=.007 \), as subjects had been in an English-dominant environment longer (\( \beta=0.63, SE(\beta)=0.13, t(32.34)=5.05, p<.001 \), and as subjects’ SES increased (\( \beta=0.01, SE(\beta)=0.003, t(32.25)=2.55, p=.016 \)). Again, generation was not included in the final model.

Table 2: Best fitting model for N400 scaled subtraction amplitude for comp-trace sentences.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>df</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.64</td>
<td>0.35</td>
<td>33.07</td>
<td>-4.67</td>
<td>&lt;.001   ***</td>
</tr>
<tr>
<td>Use - English</td>
<td>0.43</td>
<td>0.15</td>
<td>32.44</td>
<td>2.86</td>
<td>.007    **</td>
</tr>
<tr>
<td>SES</td>
<td>0.01</td>
<td>0.00</td>
<td>32.25</td>
<td>2.55</td>
<td>.016    *</td>
</tr>
<tr>
<td>Age – English</td>
<td>0.63</td>
<td>0.13</td>
<td>32.34</td>
<td>5.05</td>
<td>&lt;.001   ***</td>
</tr>
</tbody>
</table>

† p<.1 *p<.05 ** p<.01 ***p<.001
3. Experiment 2: Acceptability Judgments

To examine the relationship between processing and acceptability judgments for our stimuli and to compare our results to the heritage speaker literature, the same subjects from Experiment 1 returned to the lab 10–14 days later to complete an acceptability judgment task.  

3.1. Methods

For Experiment 2, subjects were asked to provide acceptability judgments for a subset of the items presented in the ERP task, using a five-point Likert scale with the ends and midpoints labeled in Spanish “natural”, “possibly natural”, and “not natural”. Stimuli were presented in trials following the format used in Experiment 1, but subjects were not required to answer comprehension questions. Following each trial, a Likert scale appeared on the screen prompting subjects to make a judgment about the Wh-question they just heard. Subjects were instructed to judge the Wh-question as quickly as possible, using a serial response (SR) box to indicate their judgment as soon as they were confident. The next trial began 1000ms following the button press. The experimental trials were combined with N=97 filler items, pseudorandomized, and distributed evenly over two blocks as in Experiment 1. Each block lasted approximately 20 minutes with a short break in between. Subjects completed Experiment 2 seated in a comfortable chair, 70cm from a computer monitor and external speakers in the same booth used in Experiment 1.

3.2. Results

Results for Experiment 2 are shown by group in Fig. 2 (complex noun phrase) and Fig. 3 (comp-trace), below. Results were analyzed two ways: first, by group means and second, by the expectedness of ratings. For both conditions, ratings of 4 or 5 (natural) for ungrammatical items were unexpected. Prior to the analysis of unexpected ratings, natural ratings (1 or 2) were binned as 1 and unnatural ratings (4 or 5) were binned as 5 in order to control for subjects who avoided the edges of the Likert-scale (primarily heritage speakers).

3.2.1. Complex Noun Phrase

Mean ratings for complex noun phrase items are shown in Fig. 2. Across groups, subjects rated ungrammatical items as less natural (M=4.20, SD=1.45) than grammatical items (M=2.20, SD=1.65). In exploring the ratings by group, late bilingual subjects (M=2.14, SD=1.62) and heritage speaker subjects (M=2.29, SD=1.68) rated grammatical items similarly natural; however, heritage speaker subjects (M=3.62, SD=1.69) rated ungrammatical items as more natural

---

4 One heritage speaker subject did not complete Experiment 2.
than late bilingual subjects (M=4.63, SD=1.05) although both groups rated the ungrammatical items as unnatural.

Exploring only ungrammatical items, we looked at unexpected ratings. Out of a total of N=310 ratings for ungrammatical items, there were 50 unexpected ratings (16.13%). The majority of these unexpected ratings (N=37; 74.00%) came from the heritage speaker group, and nearly half (46%; n=23) came from just four heritage speaker subjects, out of the 18 heritage speakers that were tested. The number of ungrammatical items that were rated as grammatical was not sufficient to model these unexpected responses with a logistic mixed-effects model using the procedure outlined above. For this reason, we were unable to model the unexpected responses in a similar way to the ERP analysis.

**Figure 2:** Mean acceptability ratings for complex noun phrase items by group (1, 2) and across groups (1 & 2).

### 3.2.2. Comp-trace

Mean ratings for comp-trace items are shown in Fig. 3. Across groups, subjects rated both ungrammatical (M=2.03, SD=1.52) and grammatical (M=1.93, SD=1.52) items as natural. In exploring the ratings by group, late bilingual subjects (M=1.92, SD=1.51) and heritage speaker subjects (M=1.94, SD=1.55) rated grammatical items similarly natural. Additionally, heritage speaker subjects (M=2.10, SD=1.63) and late bilingual subjects (M=1.98, SD=1.45) rated the ungrammatical items as similarly natural.

Exploring only ungrammatical items, we looked at unexpected ratings. Out of a total of N=182 ratings for ungrammatical items, there were N=146 unexpected ratings (80.22%), which were distributed across late bilingual (N=86; 58.90%) and heritage speaker subjects (N=60; 41.10%). Unexpected ratings for ungrammatical items were modeled with a logistic binomial mixed-effect model with by-subject and by-item random intercepts. The final null model did not include any variables of language use or language exposure,
\( \chi^2(1)=3.81, p=.05 \). Across groups, all subjects were more likely to accept ungrammatical comp-trace sentences than reject them, \( \beta=2.15, SE(\beta)=0.57, z=3.75, p<.001 \). A subject’s likelihood of accepting an ungrammatical item was crucially also not predicted by the generation variable.

![Comp-Trace Acceptability Judgment Task](image)

**Figure 3:** Mean acceptability ratings for comp-trace items by group (1, 2) and across groups (1 & 2).

## 4. Discussion

Our primary research question asked, “Do heritage speakers differ from late bilinguals in the processing of anomalous sentences in the first-learned language?” Our results clearly showed that the two speaker groups reacted the same, in terms of implicit responses (ERPs), where we found an N400 for the comp-trace condition and a P600 for the complex noun phrase condition in both late bilinguals and heritage speakers. Furthermore, the linear mixed modeling analysis found the speaker variable “generation” to be non-predictive, again pointing to similarity, rather than differences between the groups.

Our secondary research question asked whether individual speaker variables modulate responses in the bilingual. Individual differences were measured in terms of the amplitude of the ERP component, and here we found that the P600 for the complex noun phrase violation was not modulated by any speaker variable. This suggests that when a constraint holds in both of the languages of a bilingual, usage factors do not play a role. For the N400 response in the comp-trace condition, on the other hand, we found that amplitude was modulated by amount of exposure and use of English: The more a bilingual is exposed to and uses the later-learned language, where the constraint does not hold, the smaller the response to this constraint in the first-learned language. Thus we see a clear manifestation of cross-linguistic influence in implicit responses, and importantly, we see this in both late bilinguals and heritage speakers.
This has implications for how we think about heritage speakers and heritage languages. A standard claim in studies on heritage speakers, whom we have been referring to as heritage speaker bilinguals, is that they are speaking a different variety of their first-learned language from that of late bilingual speakers. Furthermore, divergences from the variety spoken by the latter have often been labeled as “deficits” in the heritage language in the experimental, psycholinguistic literature.

The fact that our study did not find differences in processing between late bilinguals and heritage speakers points to less of a clear-cut separation between the two groups. Instead, the effect of English language use and exposure to Spanish constraints suggests that all bilinguals, regardless of generation can implicitly be affected by a later-learned language. This in turn would suggest that heritage languages, rather than being a deficit version of varieties spoken by non-heritage speakers, may simply be the product of language change, occurring in language contact situations, and driven by all bilinguals, not just heritage speakers, who have extensive exposure and use of the later-learned language.

With regard to the methodological question of comparing implicit ERP results to explicit judgments of the same structures in the first-learned language, our results show that these two measures do not always coincide. In particular, we can say that in the complex noun phrase condition, acceptability judgments reflected underlying processing signatures well for late bilinguals, whose clear separation of acceptable and unacceptable sentences aligns well with the P600 effect found for this group.

For heritage speakers explicit and implicit measures did not align quite as well. Note that this group also showed a clear P600, unaffected by any speaker variables, yet they showed less determinacy in their explicit judgments, which tended to be clouded by uncertainty and hovered around 3 (“possibly natural”). We interpret this discrepancy as confirming the general observation that heritage speakers are insecure about their command of the language, and that this insecurity is reflected in their metalinguistic judgments of the language. The implication here is that explicit measures, requiring reflection and conscious manipulation of the language are not suitable for this population.

A surprising result in our study was that in the comp-trace condition, acceptability judgments did not reflect underlying processing mechanisms for either group. ERP results showed an N400 effect for ungrammatical que-less sentences compared to grammatical que-full sentences, while acceptability judgments were not differentiated for the two sentence types, showing high acceptance rates for both. This discrepancy again speaks to the non-compatibility of implicit and explicit measures. The fact that this occurred for both groups, combined with the nature of the ERP response (N400, no P600), would suggest that the absence of the complementizer que constitutes a lexical, rather than a structural anomaly, and that lexical anomalies are more easily reparable than structural anomalies, which may have led to the attested acceptability ratings. In general, the discrepancies we found between acceptability judgment scores and ERP components indicating clear recognition
of anomaly lends more credence to the observation that judgment tasks are inappropriate measures for heritage speakers and may not reflect underlying knowledge, at least for this population.

References


