Measuring Cross-Linguistic Influence in First- and Second-Generation Bilinguals: ERP vs. Acceptability Judgments

Gita Martohardjono, Ian Phillips, Christen N. Madsen II, Ricardo Otheguy, Richard G. Schwartz, and Valerie L. Shafer

1 Introduction

The study of bilingual populations can shed important light on language processing in general and the processing of multiple languages in particular. It can also contribute valuable information regarding the limits of language learning across the lifespan. For example, a long-standing question in the field of bilingualism and second language acquisition regards age effects on learning and attainment.

The bilingual population in North America conventionally termed “heritage speakers” or “second-generation bilinguals” is an important population to study in this regard, as the context of their incipient bilingualism is largely different from that of other types of bilinguals, e.g., those who are exposed to two languages from the inception of language learning, and use the two languages continuously throughout the lifespan; and different again from bilinguals who learn a second language later in life, to advanced levels of proficiency, but continue using the first-learned language in equal measure. Heritage speakers are typically exposed to a “minority” language in childhood, until they enter school, at around age 6, when they begin immersion in the second, “majority” language, which eventually becomes their dominant language in adulthood.

The focus in heritage language studies is on how the first-learned language changes as a second language begins to encroach upon it, thus investigating the effects of cross-linguistic influence. Importantly, the methodology used in these studies reflects an underlying assumption, namely that heritage speaker bilinguals are a linguistically homogeneous group, and that they can be compared as such to non-heritage speakers (bilingual or not), who presumably also form a linguistically homogeneous group. Hence, the analyses typically involve group comparisons (heritage speakers vs. non-heritage speakers, either monolinguals, second language learners or late bilinguals) and ANOVAs. A typical conclusion is that heritage speakers show more influence of the later-learned language than non-heritage speakers (e.g., Montrul and Bowles 2009).

Cross-linguistic influence in bilingual speakers has been found both in representation and in processing. The study of advanced sequential bilinguals has shown cross-linguistic influence in both directions, that is, from the first-learned language to the later-learned language (a phenomenon known in SLA as “transfer from the L1”), but also from the later-learned language to the first (see Cook 2003). Furthermore, the direction of cross-linguistic influence seems to vary according to level of immersion and amount of exposure to the two languages. In a wide-ranging set of studies on bilinguals of various language pairs, Sorace and her colleagues found that bilinguals immersed in the later-learned language (e.g., Italian-English bilinguals in the U.K.) differ in their interpretation of overt pronouns in Italian, the first-learned language, when compared to Italian monolinguals, but also when compared to Italian-English bilinguals in Italy (Sorace et al. 2009).

Cross-linguistic influence of a later-learned language on a first-learned language has been shown to be present in representation (Tsimpli et al. 2004) as well as in processing strategies (Fernández 2003, Dussias 2004, Dussias and Sagarra 2007). In the processing literature, the tendency to override preferred native language strategies and adopt instead those of the second language has led to the hypothesis that parsing mechanisms are “permeable”, and both level of proficiency as well as amount of exposure to the later-learned language are thought to be important factors.

The goal of this paper is to question the underlying assumptions and the methodologies standardly used in heritage speaker studies, by comparing two approaches and their concomitant analyses of the same set of data drawn from heritage speakers (henceforth second-generation bilinguals) and from non-heritage speakers (henceforth first-generation bilinguals) of Spanish whose later-learned language is English. Specifically, we question the conclusion in the literature, that the later-learned language (English) exerts more influence on second-generation speakers, who are immersed in this language in childhood, than on first-generation speakers of Spanish, who are exposed to English in adulthood.

We use and compare two methodologies: the first follows the assumptions that second-generation bilinguals can be assumed to be a linguistically homogeneous group, and can be compared directly to another presumed homogeneous group, i.e., first-generation bilinguals. Following standard analytical procedures, this methodology uses ANOVAs to compare the two groups. This methodology is described in Section 3.1.1, Approach 1.

The second methodology does not make assumptions of group distinctions, nor of homogeneity, relying instead on individual level speaker variables to identify patterns of cross-linguistic influence. This methodology uses Linear Mixed Modeling as a statistical analysis and identifies significant speaker variables from among a large set of potentially predictive ones, collected through a questionnaire described in Section 3. This methodology is more appropriate if we assume that cross-linguistic influence is not (solely) determined by age of first exposure to English, but by other, potentially influential variables, such as amount of use of the two languages, education, and level of proficiency. This methodology is described in Section 3.1.2, Approach 2.

We based our investigation of these issues on the differences in *Comp*-trace effects between Spanish and English. *Comp*-trace effects differ in the two languages in the following ways: In English sentential complement clauses, omission of the complementizer *that* is optional, as in (1). However, in these structures for most forms of Spanish, the complementizer *que* is obligatory, as in (2). Furthermore, subject extraction out of the subordinate clause results in unacceptability in English when the complementizer is present, but is acceptable when the complementizer is absent, as in (3). In standard Spanish, the opposite obtains; the complementizer *que* must be present for the question to be acceptable, as in (4). The opposite effects in Spanish and English regarding the presence of the complementizer *que/that* in *Wh*-questions make these constructions ideal for testing cross-linguistic influence.

(1) John said (that) Mary will leave.
(2) Juan dijo *(que) María va a salir
(3) Who did John say (*that) will leave?
(4) Quién dijo Juan *(que) iba a salir?

To address our questions, we administered an event-related potential (ERP) task (Section 2) to measure processing of Spanish *Comp*-trace effects. During subsequent visits to the lab with the same subjects, we administered a detailed questionnaire (Section 3) to gather information regarding individual level characteristics such as language background, education, and current language usage, and a speeded acceptability judgment task (Section 4) in order to examine the relationship between processing and acceptability judgments for these constructions and relate our findings to the heritage speaker literature.

2 Experiment 1: Event-Related Potentials (ERP)

2.1 Rationale

ERPs are voltage changes at the scalp associated with the synchronous firing of brain cell populations in response to an event (Light et al. 2010). In this study, we recorded ERPs in response to the absence of the complementizer *que* in Spanish *Wh*-questions. The literature on ERP components as linguistic processing signatures shows that this technique has reliably been used to establish the brain's sensitivity to lexical, semantic, and syntactic anomaly in monolingual populations (Friederici 2002, Osterhout 1994) and in bilingual and learner populations (Kotz 2009, Moreno, Rodríguez-Fornells, and Laine 2008). Semantic and lexical anomaly is typically indexed by a negative-going wave largest over centro-parietal regions of the scalp, called the N400 because it peaks around 400ms following the anomaly (Kutas and Hillyard 1980, Swaab et al. 2012). Syntactic anomaly is typically indexed by an increase in positivity over posterior sites and has been called the P600 because it peaks around 600ms following the anomaly. Certain types of morphosyntactic anomaly have also been found to elicit a negativity over anterior sites at similar latencies as the N400, called the left-anterior negativity (LAN) due to a largely left lateralization (Friederici 1995, Friederici, Pfeifer, and Hahne 1993). These well-established ERP components can provide information about the moment-by-moment processes in sentence comprehension that underlie behavioral perception and comprehension, and the comparative examination of ERP responses can potentially reveal between-group differences in the presence or absence of certain processing patterns (via ERP components); in speed of processing (via component latency); and in robustness of processing (via component amplitude).

The ERP technique is ideal for studying processing in the first-learned language of second-generation bilinguals, as it allows auditory presentation of stimuli while accurately registering sensitivity to linguistic contrasts without requiring metalinguistic judgments. Most studies on the Spanish of second-generation bilinguals in the U.S. use explicit and often written metalinguistic tasks, which introduce potential confounds since for the vast majority of this population, the language of instruction and literacy is English from elementary school through college (Beaudrie and Fairclough 2012, City of New York Department of Education 2011, Menken and Kleyn 2010, Valdés et al. 2006). Spanish-language instruments that require metalinguistic judgments of grammaticality, felicity, and logic thus confront this population with a highly unfamiliar, if not downright unnatural, task (Bialystok 1986, Bialystok and Ryan 1985, Birdsong 1989), and second-generation bilinguals tend to show hesitancy and indetermination when asked to make metalinguistic judgments about Spanish even when they show high proficiency and oral fluency in that language (Klein and Martohardjono 2009).

The ERP experiment presented in this section was designed to answer the following research questions: (1) Do first- and second-generation bilinguals exhibit the same or different processing patterns when hearing sentences that are anomalous in Spanish but not in English? (2) In particular, do second-generation bilinguals exhibit weaker or no reactions to a missing complementizer (*que*) that is anomalous in Spanish but acceptable in equivalent English sentences? (3) For all bilingual subjects, are ERP components elicited by anomalous sentences affected by factors such as generation, language use, socioeconomic status (SES), etc.? We address these questions by measuring the ERPs elicited by the *que*-less stimuli in both bilingual groups.

2.2 Subjects

A total of N=38 subjects (N=23 female) completed Experiment 1. All subjects were Spanish-English bilingual adults residing in the New York City area. 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. All subjects gave informed consent and were paid a total of $80 for their participation in this study.

Each subject was categorized as either a first-generation bilingual (a Latin American-born, first immigrant generation adult) or a second-generation bilingual (a U.S.-born, second immigrant generation adult) based on criteria commonly used in heritage speaker studies (Benmamoun, Montrul, and Polinsky 2013a, 2013b). First-generation 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). Second-generation bilinguals (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 first-generation caregivers.[1]

To ensure adequate levels of Spanish proficiency for our second-generation bilinguals, who are often English dominant, we tested their auditory comprehension of complex sentences using the *RISLUS Multilingual Syntax Test* (RMST): a computerized picture-pointing task that assesses comprehension of auditorily-presented complex sentences, which has been used in previous research on bilingual populations with limited Spanish literacy (Klein and Martohardjono 2009). We administered this assessment rather than standardized language proficiency tests because it does not rely on advanced reading or writing skills. All second-generation bilinguals performed at ceil-

[1]We define the Spanish-speaking Caribbean in this study to include Cuba, Puerto Rico, and the Dominican Republic. N=4 second-generation subjects were raised by first-generation caregivers born and raised in Puerto Rico and N=14 were raised by first-generation caregivers born and raised in the Dominican Republic.

ing on this test, indicating fluency in Spanish. Our subjects were selected so as to make them similar in their basic socio-demographic and linguistic profiles, while leaving remaining individual differences subject to analysis. We categorized subjects into first- and second-generation groups based on their responses to a questionnaire administered during the recruitment process (see Section 3), which included standard items pertaining to language history (Li, Sepanski, and Zhao 2006) and additional items pertaining to immigration, educational history, and SES.

2.3 Materials and Procedure

The data examined in this paper were collected as part of a larger experiment. The stimuli that we used to test the processing of *Comp*-trace effects in Spanish consisted of complex *Wh*-questions like those shown in (5–6); (7) shows the base sentence from which the questions were derived.

(5) *que*-less	*Qué	hermana	confesó	Inés	∅	||había	comdio	la	tarta?
	what	sister	confessed	Inés		had	eaten	the	cake
	What sister did Inés confess had eaten the cake?								
(6) *que*-full	Qué	hermana	confesó	Inés	**que**	||había	comdio	la	tarta?
	what	sister	confessed	Inés	that	had	eaten	the	cake
	What sister did Inés confess had eaten the cake?								
(7) base sentence	Inés	confesó	que su	hermana	había	comido	la	tarta.	
	Inés	confessed	that her	sister	had	eaten	the	cake	
	Inés confessed that her sister had eaten the cake.								

The target stimulus sentences (N=20) contained a syntactic anomaly, in that the subordinate clause was not introduced by the complementizer *que*, as in (5). The control stimuli (N=20) were grammatically well-formed sentences, identical to these target stimuli except that they contained the complementizer *que* at the embedded clause boundary, as in (6). 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 context sentence, as in (7), followed by a *Wh*-question related to the context sentence, either a target sentence (5) or a control sentence (6). These trials were interspersed with an additional N=270 filler items presented in the same trial format, and were distributed evenly over five blocks (N=62 trials each) and pseudorandomized so that the same condition never appeared in consecutive trials. Each block lasted approximately 15 minutes and subjects were given short breaks between each block.

Subjects completed Experiment 1 seated in a comfortable chair, 70cm from a computer monitor and external speakers in a shielded IAC booth. Prior to the task, each subject listened to recorded instructions in English and Spanish and completed a practice session. Each trial was presented as follows: at 0ms a fixation cross appeared centered on the monitor; at 250ms a 1000Hz, 100ms tone sounded; at 500ms the context sentence was auditorily presented, followed by 1000ms of silence; then the target stimulus was auditorily presented, followed by 3000ms of silence. Following 40% of the trials, a visual cue appeared that prompted subjects to answer a comprehension question about the trial. The experiment resumed 1000ms after subjects answered the question. Subjects were not asked to make any metalinguistic judgments during the ERP experiment.

2.4 Data Recording and Reduction

EEG data were recorded for 32 Ag/AgCl sintered electrodes mounted in a QuikCap, positioned according to the 10-20 International Electrode System (Jasper 1958).[2] All electrodes were referenced online to the nose, and mastoids were recorded for offline re-referencing. Eye movements were monitored by electrodes attached to the right and left outer canthi and above and below the left orbit. Continuous EEG data were amplified with a Neuroscan SynAmps^2 amplifier, using a DC–100Hz low-pass filter, a 1000Hz sampling rate, with electrode impedances below 15kOhm. Continuous EEG data were saved with condition codes for off-line data reduction and analysis.

[2] We collected data for the following electrodes: 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.

After recording, each continuous EEG file was pre-processed using Neuroscan Edit software. Each subject's continuous EEG file was visually inspected to check for blocked or disconnected channels. For N=3 subjects, a single blocked channel due to a damaged electrode was interpolated in Neuroscan by creating a linear derivation based on neighboring electrodes. Following visual inspection, each EEG file was bandpass filtered at 0.1–30Hz using an FIR digital filter. Next, the Neuroscan Spatial SVD (ICA) and Spatial Filter functions were applied to the EEG to correct for large ocular artifacts (i.e., blinks).[3] Following this procedure, the EEG file was segmented into epochs of 1000ms starting at the onset of the embedded verb in each target (5) and control (6) stimulus (indicated with "||" in (5–6)). The resulting epochs were corrected using a 200ms pre-stimulus baseline. Epochs containing artifacts that exceeded a +/- 70µV threshold at any single electrode during the baseline region or between 0–900ms were removed from further analysis. At this point, the data for four subjects (N=3 second-generation, N=1 first-generation) who had fewer than 60% accepted trials over the experimental condition and all fillers were excluded from further analyses. For the remaining N=34 subjects, the artifact rejection procedure resulted in 88.06% retention of the total number of experimental trials.[4] Data for these individual trials were exported to R for statistical analysis in order to determine the effects of absence/presence of the complementizer *que* on the processing of the target sentences. Data were downsampled in R to 250Hz and grand mean centered by time point, by item, and by participant.

3 Subject Questionnaire

Participants completed a 38-question survey to collect data on demographics and language background, including education, SES, self-reported abilities in English and Spanish, current language usage patterns related to different interlocutors and social situations, languages used during childhood and adolescence, and the amount of exposure to English over their lifetime.

The number of questions exceeded the number of predictors that we were able to enter into a Linear Mixed-Effects Model (LMM) due to the number of participants in the study and collinearity of the variables. To explore inter-relatedness, we examined the distribution of questionnaire responses and correlations between them. Related and correlated items were combined to create theoretically- and sociolinguistically-motivated composite variables. For example, a number of items probe language preferences with nine different interlocutors, which cannot all be included in the model, so we grouped these items into two composite variables: *language used with family* and *language used with others*. For the composite variable *language used with family,* we took the average of father, mother, siblings, children, and partner variables. For the composite *language used with others,* we took the average of friends, boss, coworkers, and classmates variables. Following this procedure, we examined the distribution and correlation of the reduced set of composite variables to confirm that correlation was avoided and to ensure that the variables, especially language use and exposure to English, varied between and within generation. In the model, we included both variables that were entered as gathered and composite variables.

3.1 Results and Discussion

3.1.1 Approach 1: Group Comparison with Visual Inspection of Individual Electrodes

This approach followed standard methods in heritage speaker research, grouping participants into first- and second-generation bilinguals according to their questionnaire responses, and compared potential effects across the two groups following common practices in ERP research, selecting electrode sites typically associated with N400 and P600 components (centro-parietal sites for N400; posterior sites for P600).

For the first-generation group, no effect consistent with a P600 was found. However a significant effect consistent with an N400 was found over sites CZ, CPZ, and CP3. A dependent measures *t*-test was run in R over the 200-400ms time window. Mean amplitude over these sites

---

[3]These functions utilize an independent component analysis (ICA) of each subject's average blink profile and point-by-point subtraction to correct individual blinks.

[4]Retention rates were similar for target *que*-less trials (88.67%) and control *que*-full trials (87.35%).

for *que*-less sentences (M=-1.11, SD=0.96) was significantly more negative than mean amplitude over the same sites for *que*-full sentences (M=-0.47, SD=0.88), $t(18)=6.62$, $p<.001$, $r=.84$. This is illustrated in the left-hand plot in Fig. 1, which shows mean amplitude for the three electrode sites with 95% confidence interval bars. The blue (upper) wave indicates mean response to sentences with *que* (the control condition) and the red (lower) wave indicates mean responses to sentences without *que* (the anomalous condition). Figure 1 shows a negative going deflection for the *que*-less (anomalous) condition which peaks between 200–400ms in the centro-parietal region, consistent with an N400 (Kutas and Hillyard 1980, Swaab et al. 2012). The early peak of this component is likely due to the auditory presentation of the stimuli (rather than written, as in most ERP studies on syntactic processing), in which co-articulatory cues in the offset of the preceding word may indicate presence or absence of *que*.[5]

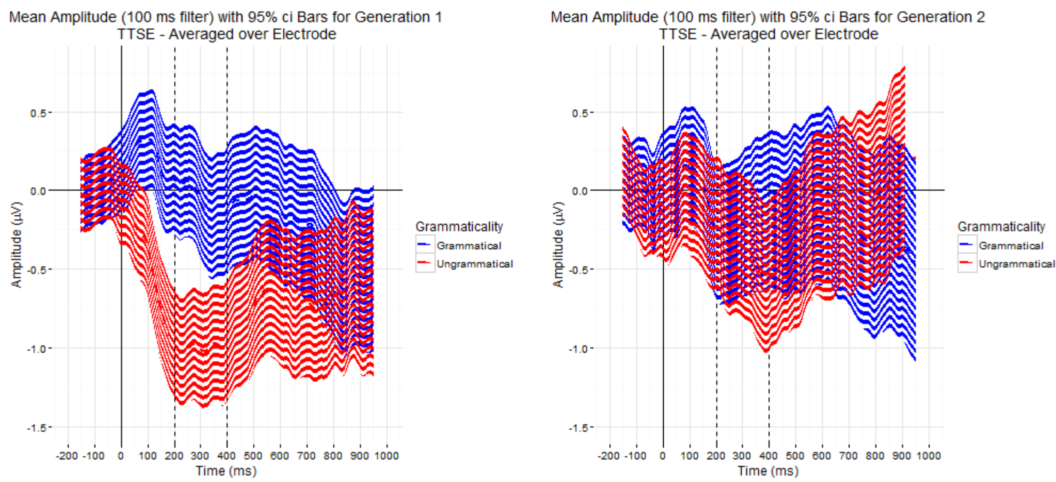

Figure 1: Mean amplitude over electrodes CZ, CPZ, CP3 for first- (left-hand plot) and second-generation (right-hand plot) groups, with 95% CI bars. Positive is plotted upwards; data were downsampled to 250Hz before analysis and smoothed with a 100ms moving average.

In contrast to the results for the first-generation group, no negative-going deflection consistent with an N400 was found for the second-generation group. A dependent measures *t*-test was run in R in the 200–400ms time window. Mean amplitude for the region of interest for *que*-less sentences (M=-0.40, SD=0.99) was not significantly different from the mean amplitude for the same region of interest for *que*-full sentences (M=-0.28, SD=1.24), $t(14)=0.30$, $p=.77$, $r=.08$. This can be seen in the right-hand plot in Fig. 1.

These results appear to show a difference between the generations in the way *que*-less (anomalous) sentences are processed. Since an N400 is a reliable indicator of surprise at the lexical level, the N400 effect seen for the first-generation group can be interpreted as evidence of processing difficulties in the absence of the complementizer *que* in these sentences. The absence of an ERP component for the second-generation group can be interpreted as evidence that the missing complementizer *que* did not affect processing for this group. Since the equivalent (grammatical) English sentences do not contain a complementizer, the contrast between first- and second-generation bilinguals could be ascribed to larger influence of English on the second-generation bilinguals.

3.1.2 Approach 2: Individual-Level Comparison with Principal Component Analysis

Approach 2 modeled N400 responses for all participants and then examined the effects of individual speaker variables (including generation) on component amplitude. The statistical analysis involved three step: (1) a principal component analysis (PCA) of ERP amplitude, (2) individual speaker variable selection (discussed in Section 3), and (3) Linear Mixed-Effects Modeling.

[5]Other studies have also found the N400 effect to occur earlier in the auditory modality, e.g., Holcomb and Neville 1990, Swaab et al. 2012.

To isolate a principal/rotated component consistent with an N400, a PCA was performed for ERP subtraction waves for all participants during the 200–400ms time window, using the **`psych`** package for R. The subtraction wave was calculated by-participant and by-item by subtracting amplitude for the *que*-full sentences from amplitude for the *que*-less sentences. For missing items, the participant-condition average was used. After calculating the subtraction wave, a spatial PCA with a promax rotation extracting 6 rotated components using the Cattell's scree test as the stopping criterion was performed for subtraction wave amplitude by time point between 200–400ms. The subtraction wave amplitude by-electrode by-participant by-item was scaled by the weightings for the fourth rotated component, which had the highest weightings over the centro-parietal region. Scaled subtraction wave amplitude, aggregating over the 200ms–400ms window for all electrodes for both groups was modeled using the **`lme4`** package in R. Approximate degrees of freedom and *p*-values for predictor specific t-values were obtained using the Satterthwaite approximation implemented in the **`lmerTest`** package in R.

To test the hypotheses about the effect of generation on processing, we included generation as a predictor variable in the LMM. To explore the effect of SES on processing, we included a composite SES variable as a predictor. To explore the effect of current English use we included composite variables related to the domains where English is used, who participants use English with, and how much English is used when consuming media (watching television, listening to music, and reading) as fixed effects in the model. To explore the effect of language experience, we included language use during childhood and adolescence and the proportion of the participant's life spent in an English dominant environment as fixed effects. As random effects, we had intercepts for participant and items. Models with by-participant and by-item random slopes did not converge.

A backwards elimination procedure was used for all predictor variables, removing in a stepwise fashion the variable with the highest *p*-value in the intermediate models as determined by the *t*-value and the degrees of freedom. 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 model contained only significant predictor variables of N400 amplitude, shown in Table 1 below.

The best fitting model has a significant main effect of English use in the home, social settings, and work; proportion of life in an English dominant environment, and SES ($\chi^2(1)=5.78$, $p<.01$). The mean scaled amplitude difference for 200–400ms decreased as participants' English use increased ($\beta=0.43$, SE(β)=0.15, $t(32.44)=2.86$, $p=.007$), as participants had been in an English-dominant environment longer ($\beta=0.63$, SE(β)=0.13, $t(32.34)=5.05$, $p<.001$), and as participants' SES increased ($\beta=0.01$, SE(β)=0.003, t(32.25)=2,55, $p=.016$).

	Estimate	**Standard Error**	**df**	**t-value**	**p-value**	
(Intercept)	-1.64	0.35	33.07	-4.67	<.001	***
Use - English	0.43	0.15	32.44	2.86	.007	**
SES	0.01	0.00	32.25	2.55	.016	*
Age – English	0.63	0.13	32.34	5.05	<.001	***

Formula in R: Scaled Amplitude ~ Use – English + SES + Age – English + (1 | participant) + (1 | item).

† *p*<.1 *p<.05 ** p<.01 ***p<.001

Table 1: Best fitting model for scaled amplitude across generation for N400.

4 Experiment 2: Speeded Acceptability Judgments

4.1 Rationale

In addition to using ERPs to measure processing of *Comp*-trace effects in Spanish-English bilinguals, we administered an acceptability judgment task during a separate visit to the lab, 10–14 days after Experiment 1. We administered the acceptability judgment task as a secondary task in order to examine the relationship between processing and acceptability judgments for our stimulus sentences and to compare our results to the heritage speaker literature.

4.2 Subjects

The same subjects participated in both Experiment 1 and Experiment 2.[6]

4.3 Materials and Procedure

For Experiment 2, subjects were asked to provide acceptability judgments for a subset of the *Comp*-trace stimuli 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 similar to those used in the ERP task, though subjects were not required to fixate on the screen or answer comprehension questions. For each trial, subjects were instructed to judge the *Wh*-question only as quickly as possible, using a serial response (SR) box to indicate their judgment as soon as they were confident. The experimental trials were combined with N=97 filler items presented in the same trial format and distributed evenly over two blocks. The order of presentation was pseudorandomized so that the same stimulus condition never appeared in consecutive trials. Each block lasted approximately 20 minutes and subjects received a short break between them. Experiment 2 lasted approximately one hour.

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. At the beginning of the experiment, each subject listened to recorded instructions in English and Spanish and completed a practice session. Each trial was presented as follows: at 0ms a blank screen appeared and the context sentence was presented auditorily; the offset of the context sentence was followed by 1000ms of silence and then the presentation of the *Wh*-question; at the onset of the *Wh*-question, a Likert scale appeared on the screen prompting subjects to make a judgment as soon as they were confident. The next trial began 1000ms following the button press.

4.4 Results and Discussion

Both first- and second-generation participants rated the *que*-less and *que*-full sentences as similarly natural. A 2 (generation) x 2 (grammaticality) repeated measures ANOVA run on the participant responses in R using the **`ez`** package shows no significant main effect of grammaticality ($F(1,31) = 0.99$; $p = .33$; $\eta^2 = .008$) or generation ($F(1,31) = 0.006$; $p = .94$; $\eta^2 = .0001$). First-generation participants rated the *que*-full sentences (M=2.02, SD=1.24) similar to the *que*-less sentences (M=2.14, SD=1.27) as did second-generation participants: *que*-less sentences (M=2.17, SD=1.41); *que*-full sentences (M=2.02, SD=1.44).

5 General Discussion

The objective of our study was to investigate processing patterns of anomalous sentences in the home language (Spanish) of first- and second-generation bilinguals, with a view to identifying individual speaker variables that might influence such patterns. We compared two methodologies, one which used a group comparison, looking at first- and second-generation bilinguals separately; and a second analysis, which put all subjects into the same group but used a mixed model analysis to seek out those speaker variables that are significant predictors of specific processing patterns. The results of the first methodology showed a significant difference in the amplitude of an N400 between groups. If taken at face value, this would lead us to conclude that first-generation bilinguals, more than second-generation bilinguals, recognize *que*-less sentences as anomalous. This, in turn, could be taken to mean that second-generation bilinguals are more permeable to English patterns when processing their home language Spanish. It would also suggest that first- and sec-

[6]One second-generation subject did not complete Experiment 2.

ond-generation bilinguals exhibit different linguistic patterns (albeit in our study this is only attested in the receptive, and not the productive, mode) in their Spanish.

The second approach, however, showed that these conclusions are premature. A more fine-grained analysis of the various factors that might contribute to an increase in N400 amplitude shows that generation per se was in fact not a significant predictor. Instead, variables related to English use were significant: amount English use in the home, social settings, and work; proportion of life in an English dominant environment, and socio-economic status, which is likely to be an indicator of English use as well. Importantly, these factors cut across the two generations, suggesting that (a) regardless of generation, usage factors significantly determine processing patterns; and (b) age of first exposure to English may not be as significant a factor as previously thought. The questionnaire revealed first-generation speakers' first age of exposure to English to be 18, and that of second-generation speakers' to be 6. Under a Critical Period Hypothesis view, where age of exposure is predicted to be the determining factor, the variable "generation" should have come out as significant. Contrary to this, our results indicate that usage factors play a more important role. This result is consistent with studies such as Flege and Liu 2001 showing that second language (L2) proficiency in both syntax and phonology are affected more by amount of use of L2 than by age of arrival. Interestingly, the results of Experiment 2 show that judgment tasks are not good indicators of processing, and that having to appeal to metalinguistic knowledge may mask intrinsic knowledge. That is, when asked explicitly, both first- and second-generation subjects found *que*-less sentences to be acceptable, even though the N400 effect clearly indicated recognition of anomaly. One possibility for this result may be the "type" of the anomaly, which in our stimuli consisted of the absence of a very small utterance, *que*. Since the stimuli were presented auditorily, it is possible that some subjects may have "repaired" the *que*-less sentences, rendering them more natural sounding.

References

Beaudrie, Sara M., and Marta Fairclough, ed. 2012. *Spanish as a Heritage Language in the United States: The State of the Field*. Washington, D.C.: Georgetown University Press.

Benmamoun, Elabbas, Silvina Montrul, and Maria Polinsky. 2010. Prolegomena to Heritage Linguistics. Unpublished *White Paper. University of Illinois at Urbana-Champaign and Harvard University*.

———. 2013a. Defining an 'Ideal' Heritage Speaker: Theoretical and Methodological Challenges Reply to Peer Commentaries. *Theoretical Linguistics* 39:259–94.

———. 2013b. Heritage Languages and Their Speakers: Opportunities and Challenges for Linguistics. *Theoretical Linguistics* 39:129–81.

Bialystok, Ellen. 1986. Factors in the Growth of Linguistic Awareness. *Child Development* 57:498–510.

Bialystok, Ellen, and E. B. Ryan. 1985. A Metacognitive Framework for the Development of First and Second Language Skills. In *Metacognition, Cognition, and Human Performance: Theoretical Perspectives*, ed. D. L. Forrest-Pressley, G. E. MacKinnon and T. G. Waller, 1:207–52. Orlando, FL: Academic Press.

Birdsong, David. 1989. *Metalinguistic Performance and Interlinguistic Competence.* New York, NY: Springer-Verlag Publishing.

Cho, Grace, Kyung-Sook Cho, and Lucy Tse. 1997. Why Ethnic Minorities Want to Develop Their Heritage Language: The Case of Korean-Americans. *Language, Culture and Curriculum* 10:106–12.

City of New York Department of Education. 2011. The 2010-11 Demographics of New York City's English Language Learners - Manhattan Comprehensive Night and Day High School.

Cook, Vivian. 2003. *Effects of the Second Language on the First*. Clevedon, UK: Multilingual Matters.

Dussias, Paola E. 2004. Parsing a First Language like a Second: The Erosion of L1 Parsing Strategies in Spanish-English Bilinguals. *International Journal of Bilingualism* 8:355–371.

Dussias, Paola E., and Nuria Sagarra. 2007. The Effect of Exposure on Syntactic Parsing in Spanish–English Bilinguals. *Bilingualism: Language and Cognition* 10:101–116.

Fernández, Eva M. 2003. *Bilingual Sentence Processing: Relative Clause Attachment in English and Spanish*. Philadelphia, PA: John Benjamins Publishing.

Flege, James Emil, and Serena Liu. 2001. The Effect of Experience on Adults' Acquisition of a Second Language. *Studies in Second Language Acquisition* 23:527–52.

Friederici, A. D. 1995. The Time Course of Syntactic Activation during Language Processing: A Model Based on Neuropsychological and Neurophysiological Data. *Brain and Language* 50:259–81.

———. 2002. Towards a Neural Basis of Auditory Sentence Processing. *Trends in Cognitive Sciences* 6:78–84.

Friederici, Angela D., Erdmut Pfeifer, and Anja Hahne. 1993. Event-Related Brain Potentials during Natural Speech Processing: Effects of Semantic, Morphological and Syntactic Violations. *Cognitive Brain Research* 1:183–192.

Holcomb, Phillip J., and Helen J. Neville. 1990. Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-Related Brain Potentials. *Language and Cognitive Processes* 5:281–312.

Jasper, Herbert Henri. 1958. The Ten Twenty Electrode System of the International Federation. *Electroencephalography and Clinical Neurophysiology* 10:371–375.

Klein, E., and Gita Martohardjono. 2009. Understanding the Student with Interrupted Formal Education (SIFE): A Study of SIFE Skills, Needs and Achievement: Phase II. New York, NY: NYC DOE Office of English Language Learners.

Kotz, Sonja A. 2009. A Critical Review of ERP and fMRI Evidence on L2 Syntactic Processing. *Brain & Language* 109:68–74.

Kutas, Marta, and Steven A. Hillyard. 1980. Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science* 207:203–205.

Li, Ping, Sara Sepanski, and Xiaowei Zhao. 2006. Language History Questionnaire: A Web-Based Interface for Bilingual Research. *Behavior Research Methods* 38:202–10.

Light, Gregory A., Lisa E. Williams, Falk Minow, Joyce Sprock, Anthony Rissling, Richard Sharp, Neal R. Swerdlow, and David L. Braff. 2010. Electroencephalography (EEG) and Event-Related Potentials (ERP's) with Human Participants. *Current Protocols in Neuroscience / Editorial Board, Jacqueline N. Crawley ... [et Al.]* CHAPTER (July): Unit-6.2524.

Menken, Kate, and Tatyana Kleyn. 2010. The Long-Term Impact of Subtractive Schooling in the Educational Experiences of Secondary English Language Learners. *International Journal of Bilingual Education and Bilingualism* 13:399–417.

Montrul, Silvina, and Melissa Bowles. 2009. Back to Basics: Incomplete Knowledge of Differential Object Marking in Spanish Heritage Speakers. *Bilingualism: Language and Cognition* 12:363–383.

Moreno, Eva M., Antoni Rodríguez-Fornells, and Matti Laine. 2008. Event-Related Potentials (ERPs) in the Study of Bilingual Language Processing. *Journal of Neurolinguistics* 21:477–508.

Oldfield, Richard C. 1971. The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia* 9:97–113.

Osterhout, Lee. 1994. Event-Related Brain Potentials as Tools for Comprehending Language Comprehension. In *Perspectives on Sentence Processing*, ed. C. Clifton Jr., L. Frazier, and K. Rayner, 15–44. Hillsdale, NJ: Erlbaum.

Rumbaut, Rubén G. 1991. The Agony of Exile: A Study of the Migration and Adaptation of Indochinese Refugee Adults and Children. In *Refugee Children: Theory, Research, and Services*, ed. F.L. Ahearn Jr. and J.L. Athey. Baltimore, MD: Johns Hopkins University Press.

Sorace, Antonella, Ludovica Serratrice, Francesca Filiaci, and Michela Baldo. 2009. Discourse Conditions on Subject Pronoun Realization: Testing the Linguistic Intuitions of Older Bilingual Children. *Lingua* 119:460–77.

Swaab, Tamara Y., Kerry Ledoux, C. Christine Camblin, and Megan A. Boudewyn. 2012. Language-Related ERP Components. In *Oxford Handbook of Event-Related Potential Components*, ed. S.J. Luck and E.S Kappenman, 397–439. Oxford: Oxford University Press.

Tsimpli, Ianthi, Antonella Sorace, Caroline Heycock, and Francesca Filiaci. 2004. First Language Attrition and Syntactic Subjects: A Study of Greek and Italian near-Native Speakers of English. *International Journal of Bilingualism* 8:257–77.

Valdés, Guadalupe, J. Fishman, R. Chávez, and W. Pérez. 2006. Secondary Spanish Heritage Programs in California. In *Developing Minority Language Resources the Case of Spanish in California*, ed. Guadalupe. Valdés, J. Fishman, R. Chávez, and W. Pérez, 140–86. Clevedon, UK: Multilingual Matters.

Linguistics Program
The Graduate Center, CUNY
365 Fifth Avenue
New York, New York 10016 USA
gmartohardjono@gc.cuny.edu
iphillips@gradcenter.cuny.edu
cmadsen@gradcenter.cuny.edu
rotheguy@gc.cuny.edu
rschwartz@gc.cuny.edu
vshafer@gc.cuny.edu