Maturational Constraints on the Recruitment of Early Processes for Syntactic Processing

Eric Pakulak and Helen J. Neville

Abstract

■ An enduring question in the study of second-language acquisition concerns the relative contributions of age of acquisition (AOA) and ultimate linguistic proficiency to neural organization for second-language processing. Several ERP and neuroimaging studies of second-language learners have found that neural organization for syntactic processing is sensitive to delays in second-language acquisition. However, such delays in second-language acquisition are typically associated with lower language proficiency, rendering it difficult to assess whether differences in AOA or proficiency lead to these effects. Here we examined the effects of delayed second-language acquisition while controlling for proficiency differences by examining participants who differ in AOA but who were matched for proficiency in the same language. We compared the ERP response to auditory English phrase structure violations in a group of late learners of English matched for grammatical proficiency with a group of English native speakers. In the native speaker group, violations elicited a bilateral and prolonged anterior negativity, with onset at 100 msec, followed by a posterior positivity (P600). In contrast, in the nonnative speaker group, violations did not elicit the early anterior negativity, but did elicit a P600 which was more widespread spatially and temporally than that of the native speaker group. These results suggest that neural organization for syntactic processing is sensitive to delays in language acquisition independently of proficiency level. More specifically, they suggest that both early and later syntactic processes are sensitive to maturational constraints. These results also suggest that late learners who reach a high level of second-language proficiency rely on different neural mechanisms than native speakers of that language. ■

INTRODUCTION

An enduring question in the study of second-language acquisition concerns the relative contributions of age of acquisition (AOA) and ultimate linguistic proficiency to neural organization for second-language processing. Several ERP and neuroimaging studies of second-language learners have found that, while subsystems implicated in on-line semantic processing are relatively invulnerable to delays in second-language acquisition, neural organization for syntactic processing is altered by delays in acquisition as short as 4 years (Kotz, 2009; Abutalebi, 2008; Wartenburger et al., 2003; Hahne, 2001; Hahne & Friederici, 2001; Kim, Relkin, Lee, & Hirsch, 1997; Weber-Fox & Neville, 1996). However, such delays in second-language acquisition are typically associated with lower language proficiency (Weber-Fox & Neville, 1996; Johnson & Newport, 1989), rendering it difficult to assess whether differences in AOA or proficiency lead to these effects. One approach to this problem is to study participants of different proficiency levels matched for AOA. Previously, we used ERPs to examine the relationship between AOA and proficiency by studying on-line syntactic processing in English-speaking adults who, as monolingual native speakers, had the same AOA but varied in their native language proficiency as assessed by standardized measures of English proficiency (Pakulak & Neville, 2010). Results from that study revealed large effects of proficiency on neural organization for syntactic processing. Another approach to this problem is to study participants who differ in AOA but who are matched on proficiency level in the same language. Here we take this approach and test the hypothesis that AOA will also have effects on neural organization for syntactic processing independently of proficiency. To this end, we compare on-line syntactic processing in a group of late learners of English matched for grammatical proficiency with the lower proficiency monolingual participants from our previous study. We used the same standardized measures to assess proficiency and the same ERP paradigm, which allowed for a more direct assessment and comparison of the differential effects of AOA and proficiency. Specifically, we compared the ERP response to auditory phrase structure violations in both groups to test the hypothesis that nonnative speakers of English who learned English later in life recruit different neural systems in order to achieve a level of proficiency comparable to that of some native speakers.

ERP Studies of Language Processing

ERPs provide an on-line, multidimensional index of cognitive processes with a temporal resolution of milliseconds,

University of Oregon

and thus have emerged as one of the more widely used methodologies in examining on-line language processing. Consistent with other methodologies, ERP studies have demonstrated that different linguistic subsystems are mediated by nonidentical neural mechanisms. Numerous studies in both the visual and auditory modalities have found that semantically unexpected words elicit a negative-going potential peaking around 400 msec (N400) compared to contextually appropriate words (Friederici, Pfeifer, & Hahne, 1993; Holcomb & Neville, 1991; Kutas & Hillyard, 1980), leading to the hypothesis that the N400 component indexes semantic processes of lexical integration.

Although the N400 has consistently been related to aspects of semantic processing, at least two components hypothesized to index syntactic processing have been identified. The first of these is a negative-going wave between 100 and 500 msec, often larger over left anterior electrode sites, referred to as the LAN. The LAN has been elicited by a variety of syntactic violation types, such as phrase structure violations (e.g., Gunter, Friederici, & Hahne, 1999; Hahne & Friederici, 1999; Friederici et al., 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991) and morphosyntactic violations (e.g., Coulson, King, & Kutas, 1998; Friederici et al., 1993; Münte, Heinze, & Mangun, 1993). The LAN typically occurs in one or both of two time windows (100–300 and 300–500 msec), which has led some researchers to propose the existence of two distinct components, with the first, termed the early left anterior negativity (ELAN), indexing processes different from those indexed by the LAN (Hahne & Jescheniak, 2001; Friederici & Mecklinger, 1996; Friederici, 1995). Two recently proposed theories of on-line sentence processing account for these components in different ways. Friederici, Hahne, and Saddy (2002) propose that the ELAN is functionally distinct from both the LAN and the N400 components, and that it reflects early and automatic processing of word category violations in a first phase of sentence processing which is autonomous and independent of contextual or semantic influences. Hagoort (2003, 2005) and van den Brink and Hagoort (2004) propose a different model in which semantic and syntactic information are processed in parallel as soon as they are available and posit that the timing differences reported between LAN and ELAN effects are a result of differences in the on-line availability of morphosyntactic and word category information. Thus, although the theories differ with regard to the interaction between semantic and syntactic information in parsing, both view the ELAN as an index of the rapid use of on-line information in syntactic processing.

The second component which has been observed in ERP studies of syntactic processing is a large positive-going wave usually maximal between 500 and 1000 msec over bilateral posterior regions referred to as the P600 (Osterhout, Lee, & Holcomb, 1993). The P600 is consistently elicited by syntactic violations (Hagoort & Brown, 2000; Hahne & Friederici, 1999; Osterhout & Mobley, 1995; Hagoort, Brown, & Groothusen, 1993;

Osterhout, Lee, & Holcomb, 1992) as well as by violations of preferred syntactic structure (Osterhout, Holcomb, & Swinney, 1995; Osterhout & Holcomb, 1993) or in well-formed sentences of higher syntactic complexity (Kaan & Swaab, 2003a, 2003b; Kaan, Harris, Gibson, & Holcomb, 2000). Although the distribution of the P600 is usually posterior, several studies have reported a late positivity with a more frontal distribution to grammatically correct but nonpreferred structures (Kaan & Swaab, 2003a, 2003b; Friederici, Hahne, et al., 2002; Osterhout & Holcomb, 1992). This has led to the proposal that the frontally distributed P600 reflects processing difficulties related to revision in the face of nonpreferred structures, whereas the posteriorly distributed P600 reflects processes related to the failure of a parse and related processes of repair and meaning rescue (Friederici, Hahne, et al., 2002; Hagoort & Brown, 2000) or to syntactic integration difficulty (Kaan et al., 2000).

Second-language Processing

Based on evidence from the development of sensory and motor systems, Lenneberg (1967) proposed that similar maturational processes might constrain language development, such that there may be sensitive periods during which the effects of language experience are maximal on ultimate linguistic proficiency and neural organization for language. This hypothesis is supported by behavioral data from both first- and second-language acquisition, which suggest that proficiency decreases with delays in language immersion (Mayberry, Lock, & Kazmi, 2002; Mayberry, 1993; Mayberry & Eichen, 1991; Newport, 1990; Johnson & Newport, 1989). This evidence also suggests that different subsystems are differentially affected by delays in language experience, as syntactic processing appears to be more profoundly affected, whereas aspects of semantic processing appear to be relatively invulnerable to such delays. Other evidence suggests that a small number of nonnative speakers who acquire a second language after the end of a hypothesized sensitive period, around the onset of puberty, can attain a level of proficiency in syntactic processing which is similar to that of native speakers (White & Genesee, 1996; Birdsong, 1992), although the question of whether such individuals recruit the same neural mechanisms as native speakers to achieve such a level of proficiency is an open one.

Several ERP studies of bilinguals that have replicated behavioral findings of reduced grammatical proficiency with delays in second-language exposure have provided evidence bearing on differences in neural organization for second-language processing that might underlie the effects of proficiency. In a study of Chinese–English bilinguals, Weber-Fox and Neville (1996) found that systems involved in lexical–semantic processing, as reflected by the N400 response to semantic violations, were not affected by delays in exposure as long as 11 years. In contrast, systems involved in syntactic processing were found to be

sensitive to delays of even 4 years: Although syntactic violations elicited a biphasic response in all groups consisting of an anterior negativity between 300 and 500 msec followed by a P600, the anterior negativity was left-lateralized only in groups with earlier ages of first exposure to English, was bilateral in groups whose first exposure to English was later, and was right-lateralized in participants whose first exposure was after age 16. Two subsequent studies of late bilinguals did not find an anterior negativity to syntactic violations. ERP studies of Japanese–German (Hahne & Friederici, 2001) and Russian–German (Hahne, 2001) late bilinguals reported that semantic violations elicited an N400 in both groups of late learners, whereas syntactic violations failed to elicit an anterior negativity response in either group, although such violations did elicit a P600 in the Russian–German group. Recently, two studies have reported more native-like ERP effects to syntactic violations in second-language learners. In a study of Japanese–English bilinguals of different second-language proficiency levels, Ojima, Nakata, and Kakigi (2005) report that whereas semantic violations elicited an N400 in both late learner proficiency groups, syntactic violations elicited a left-lateralized negativity between 350 and 550 msec only in the native speaker and high-proficiency late bilingual groups. In a study of low- and high-proficiency late learners of German and Italian processing their respective second languages, Rossi, Gugler, Friederici, and Hahne (2006) report that, in response to phrase structure violations, both groups showed an extended bilateral anterior negativity beginning around 100 msec followed by a P600, and in response to verb agreement violations, high-proficiency learners of both languages showed a biphasic LAN–P600 response.

Several positron emission tomography and functional magnetic resonance imaging studies have examined the neural indices of second-language processing (for recent reviews, see Kotz, 2009; Abutalebi, 2008; Indefrey, 2006). Although differences in tasks and paradigms across studies limit the generalizability of the results, overall, the findings with regard to syntactic processing in sentential contexts suggest that the recruitment of neural areas for second-language processing are more dependent on differences in AOA than in proficiency, with late second-language learners recruiting more neural resources either around areas found to underlie first-language syntactic processing (e.g., Rüschemeyer, Zysset, & Friederici, 2006; Rüschemeyer, Fiebach, Kempe, & Friederici, 2005) or in additional areas such as the basal ganglia (Rüschemeyer et al., 2005; Wartenburger et al., 2003). Neuroimaging studies that have specifically examined the role of experience and proficiency have found evidence for a role of both age of exposure and ultimate second-language proficiency in the determination of neural organization for a second language. One study reported that although no differences in neural organization for first and second languages were found for early-acquisition high-proficiency bilinguals, late acquisition (after age 6) bilinguals recruited additional resources in inferior frontal and parietal regions for grammatical pro-

cessing in their second language (Wartenburger et al., 2003). Another study found no differences in neural activation between two groups of highly proficient bilinguals who differed in AOA while participants listened to stories in their second language (Perani et al., 1998), although the use of a story listening paradigm limited the degree of focus on syntactic processes. Although the excellent spatial resolution afforded these methodologies provides valuable information about neural areas subserving first- and second-language processing, these methodologies do not provide the temporal resolution necessary to disentangle early and late processes in language processing or to assess the hypothesis that there may be differential degrees and/or types of maturational constraints on the recruitment of early and later syntactic processes in second-language processing.

ERPs and Proficiency

Data from two ERP experiments suggest significant differences in proficiency in monolingual adults, and are linked to altered neural organization as indexed by ERPs. In a visual sentence processing paradigm, Weber-Fox, Davis, and Cuadrado (2003) compared brain responses to visually presented semantic violations in participants who scored either exceptionally high or in the normal range on four subtests of the Test of Adult and Adolescent Language-3 (TOAL-3), a standardized assessment of English language proficiency. Although no differences were found in early ERP components indexing perceptual processing, high-proficiency participants had an earlier N280 to closed-class words only over left anterior regions, suggesting more rapid lexical access of grammatical words specifically in these participants. We reported results from a study in which we examined differences in the neural response to auditory phrase structure violations in English sentences in two groups of monolingual native speakers of English who were classified as higher or lower proficiency based on their scores on the TOAL-3 (Pakulak & Neville, 2010). Violations elicited a typical biphasic response in both groups, but there were differences in this response between groups. The anterior negativity effect was spatially and temporally more focal in the left hemisphere in the higher proficiency group but was more widely distributed and prolonged in the lower proficiency group. The P600 effect was larger in amplitude and more broadly distributed in higher proficiency participants compared to lower proficiency participants. These effects of proficiency on neural organization for syntactic processing were confirmed by a correlational analysis across a wide range of proficiency scores.

The Present Study

Because numerous lines of evidence suggest that the syntactic subsystem is more vulnerable to differences in language experience, here we focus on this subsystem.

Previously, we studied the effects of proficiency on neural organization for syntactic processing by studying a group of monolingual native speakers, who had the same AOA, but who differed on standardized measures of English proficiency. Here we continue this systematic exploration of the relative contributions of AOA and proficiency to neural organization for syntactic processing by comparing two groups of participants who were matched on English proficiency but had different AOAs. We recruited native speakers of German who had acquired English later in life but who had achieved a level of proficiency that was equal to that of the lower proficiency monolingual group from our previous study based on a standardized measure of English grammatical proficiency. Both groups were run in the same auditory ERP paradigm featuring phrase structure violations in simple, single-clause sentences in English. We hypothesized that the neural response to syntactic violations would be affected by differences in AOA, and that early and late components of this response would be differentially affected. Specifically, we predicted that differences related to AOA would be most strongly reflected in differences in the early anterior negativity, a component hypothesized to reflect early and automatic processing. In contrast, we hypothesized that the P600, a late component thought to reflect more controlled processes, would be more similar in late learners and native speakers.

METHODS

Participants

Thirty-six right-handed adults with normal hearing participated in the study. Nineteen participants (the nonnative speaker group, NNS) were native speakers of German who began learning English between the ages of 10 and 12 years, and had reached a high enough level of proficiency in English to function as undergraduate students, graduate students, or a professor at the University of Oregon. Any participant with a score more than two standard deviations above the mean on any behavioral or ERP measure was removed from the analysis as an outlier; this resulted in the removal of one NNS participant. Seventeen participants (the native speaker group, NS) were monolingual native speakers of English recruited from both the university and general population. These were the same participants who formed the lower proficiency group in our previous study (Pakulak & Neville, 2010) and they had proficiency scores that matched those of the late learners. Groups were matched on sex (NNS: 8 women; NS: 7 women) and age (NNS: $M = 26.30$ years, $SD = 4.58$; NS: $M = 24.65$ years, $SD = 5.15$).

Behavioral Language Inventories

The groups were matched for proficiency based on their scores on the Speaking/Grammar subtest of the TOAL-3

(Hammil, Brown, Larsen, & Wiederholt, 1994). The TOAL-3 Speaking/Grammar subtest requires participants to repeat exactly the sentences said by the examiner as the sentences increase in syntactic difficulty. In order to receive a correct score, the participant must repeat the item without any changes in syntax or morphology. Two additional tests were administered to assess linguistic proficiency: the Listening/Grammar subtest of the TOAL-3 and the Saffran and Schwartz Grammaticality Judgment Test (Linebarger, Schwartz, & Saffran, 1983). The TOAL-3 Listening/Grammar subtest requires participants to determine, out of three sentences presented auditorily, which two sentences have similar meaning. The Saffran and Schwartz Grammaticality Judgment Test is a 78-item assessment in which participants must recognize a variety of syntactic violations, adapted for purposes of this study into the auditory modality. In order to assess WM capacity, participants were also given the Carpenter Span Reading Test (Daneman & Carpenter, 1980), a widely used assessment in which participants must recall the final word of two or more sentences after reading them consecutively. Participants also filled out a questionnaire which gathered information on education level and socioeconomic status of origin (SES) as measured by the Hollingshead Four Factor Index of Social Status (Hollingshead, 1975).

Bilingual Questionnaire

In order to explore the role of different aspects of language experience in second-language acquisition, NNS participants were given an additional questionnaire. This questionnaire included questions about participants' amount of English exposure throughout their lives; sources of this exposure; first exposure to English instruction and amount of time spent studying English; amount of time spent living in an English-speaking country; relative helpfulness of different activities in learning English; relative frequency of English use throughout their lives in school, home, and other environments; and self-ratings of German and English proficiency in spoken, written, and overall language.

Stimuli

In the ERP paradigm, participants heard both English sentences and Jabberwocky sentences, in which open-class words were replaced with pronounceable nonwords to greatly reduce the semantic context; only the results for the English stimuli are presented here. The English stimuli were sentences which were canonical (50%) or which contained an insertion phrase structure violation in which an additional closed-class word was inserted in a sentence-final prepositional phrase. In all cases, the phrase structure violation clearly occurred at the onset of either a demonstrative (50%) or possessive (50%) pronoun directly following the inserted pronoun. The ERPs to the onset of

the target word (underlined below) in the canonical and violation (*) sentences were compared:

English: Timmy can ride the horse at <u>his</u> farm.
        *Timmy can ride the horse at my <u>his</u> farm.

A number of measures were undertaken in order to provide prosodic variability as well as to ensure that subjects listened fully to the sentences and did not focus only on the location of the critical violation. In 5% of the experimental sentences, an additional prepositional phrase was added to the beginning of the sentences, and in 20% of the experimental sentences, an adjective was placed directly after the target word so that the target word was not invariably in the penultimate position in the sentence. In addition, filler sentences and probe questions were constructed. Filler sentences contained a permutation phrase structure violation in which a main verb and the determiner of the object noun phrase were reversed. Probe questions took the form, "Did you hear the word (<u>blank</u>)?" Participants heard 62 sentences of each condition. Twenty-eight filler sentences (10% of total) were pseudorandomly interspersed between the experimental sentences, as were 16 probe questions, such that filler sentences and probe questions occurred equally across quarter stimulus blocks and were always separated by at least two experimental sentences.

All sentences were recorded using SoundEdit 16 Version 2 with 16-bit resolution and a 16-kHz sampling rate then transferred to a PC for presentation. The sentences were spoken by a female with natural tempo and prosody, and critical word onsets were identified and coded by three trained coders using both auditory cues and visual inspection of sound spectrographs for increased accuracy. Any sentences in which codes differed by more than 20 msec between coders were recoded by all three coders together until a consensus was reached by all three to ensure reliability.

Procedure

Most participants were tested in one 3-hr session, with the standardized tests of language administered immediately before ERP testing. A subset of participants in both the NS group ($n = 5$) and the NNS group ($n = 7$) was given the behavioral measures and ERP testing in separate sessions. In each ERP session, a 32-channel electrode cap (Electro-Cap International, Eaton, OH) was applied while the participant completed an information sheet which included questions about education, SES, handedness, neurological history, and language habits. NNS participants also completed the questionnaire assessing their acquisition and current usage of English. In the third part of each session, subjects were seated in a comfortable chair in an electrically shielded sound-attenuating booth. Sentences were presented via a speaker placed centrally on a monitor 70 in. from the participant. Participants were given auditory instructions including exam-

ples of both sentence types and emphasizing the need to judge the sentences based on grammatical, and not semantic, correctness. On each trial, participants pushed one of two response buttons to play a sentence. While the sentences were playing, participants were asked to refrain from blinking or moving their eyes as a box with a central fixation cue ("*") was displayed. After each sentence, participants were cued to make a judgment with a display of "Yes or No?" on the screen. The judgment was made with a button press with either the left or right hand, counterbalanced across participants. Participants proceeded at their own pace and were given two regularly scheduled breaks and additional breaks as requested.

EEG Equipment and Analysis

The EEG was recorded using tin electrodes mounted in an appropriately sized elastic cap (Electro-Cap International) over 29 scalp sites based on Standard International 10–20 System electrode locations: F7/F8, F3/F4, FT7/FT8, FC5/FC6, T3/T4, C5/C6, CT5/CT6, C3/C4, T5/T6, P3/P4, TO1/TO2, O1/O2, Fp1/Fp2, Fz, Cz, and Pz. Scalp electrode impedances were kept below 3 kΩ. Data from all scalp electrodes were referenced on-line to the EEG from an electrode placed over the right mastoid and later referenced off-line to the mathematical average of the left and right mastoids. Horizontal eye movements were monitored using electrodes placed at the outer canthus of each eye and referenced to each other, whereas vertical eye movements were monitored using an electrode placed beneath the right eye and referenced to the right mastoid. The raw EEG signal was collected at a sampling rate of 250 Hz and was amplified using Grass amplifiers with high- and low-pass filter settings of 0.01 and 100 Hz, respectively.

Only trials on which subjects responded correctly were included in the ERP analyses. The individual-trial EEG data for each participant were examined for eye movements, muscle artifact, and amplifier saturation and drift, and any trials contaminated by these artifacts were excluded from final data analyses. There were no differences between groups in the number of trials remaining after artifact rejection [NS: $M = 77.34\%$; NNS: $M = 68.23\%$; $t(33) = 1.19$, ns]. ERPs were computed for 1200 msec after the onset of the target word relative to a 100-msec prestimulus baseline. ERP waveforms were measured within time windows determined by visual inspection of individual and group averages; specific time windows are described in the Results section. Based on a priori hypotheses from previous results and on visual inspection of the effects, the anterior negativity effect was characterized by analyzing the 12 anterior electrode sites and the P600 by analyzing the 12 posterior electrode sites. Mean voltage amplitude was measured within each time window and analyzed using ANOVAs with repeated measures, including two levels of condition (C: canonical, violation), two levels of hemisphere (H: left, right), three levels of

anterior–posterior [A: frontal, fronto-temporal, temporal (anterior sites); central, parietal, and occipital (posterior sites)], and two levels of lateral–medial (L: lateral, medial), as well as a between-subjects factor AOA, with two levels (N: native speakers; nonnative speakers). Following omnibus ANOVAs, additional analyses were performed in step-down fashion such that follow-up analyses were performed to isolate any significant interactions, collapsing across factors with which an interaction was not found. When significant between-group interactions were found, separate ANOVAs were performed for each group to better characterize group differences. Greenhouse–Geisser corrections were applied for all ANOVAs with greater than one degree of freedom.

In the correlational and regression analyses, for each of the 35 participants, the average difference amplitude (violation − canonical) was calculated for each electrode site. Based on the results from the between-group analyses, the average difference amplitude across anterior sites in the 100–300 msec time window was analyzed to capture the anterior negativity and the average difference amplitude across posterior sites in the 300–1000 msec time window was analyzed to capture the P600 effect. Zero-order correlations were then calculated between individual average difference amplitudes and individual working memory span scores. In the regression analyses, average difference amplitudes in these respective windows were regressed on proficiency, working memory span, and AOA using a backward stepwise regression procedure.

RESULTS

Behavioral Results

Behavioral results for all measures of proficiency and working memory are summarized in Table 1. NNS and NS groups were matched on the Speaking/Grammar subtest of the TOAL-3. The resulting mean average scores for the NS ($M = 15.47$, $SD = 4.26$) and NNS ($M = 17.11$, $SD = 3.46$) groups were not statistically independent [$t(33) = 1.566$, ns]. NNS participants scored higher than NS participants on the TOAL-3 Listening/Grammar subtest [$t(33) = 3.373$, $p < .001$]. Although this result seems surprising, a likely explanation involves group differences in working memory span, as this particular subtest likely induces a high working memory load. The NNS group did have a significantly higher working memory span than the NS group [$t(33) = 2.669$, $p < .05$]. The NS group scored higher on the Saffran and Schwartz Grammaticality Judgment Test [$t(33) = 2.525$, $p < .05$]. In the ERP grammaticality judgment task, there was a trend for a higher percentage of correct responses by the NNS group [$M = 97.41$, $SD = 1.93$] compared to the NS group ($M = 94.96$, $SD = 9.94$), which did not reach significance [$t(33) = 1.723$, $p = .094$]. The NNS group also had a higher level of education [$t(33) = 5.948$, $p < .005$] and SES [$t(33) = 3.12$, $p < .005$] than the NS group (although caution is necessary when comparing SES between groups from different countries).

Bilingual Questionnaire

Results from the Bilingual Questionnaire revealed that all NNS participants began learning English in a school setting at around the same age ($M = 11.05$ years, $SD = 1.10$, range = 10–14). Only one NNS participant had parents who spoke English in the home, and only two to three times per month. Participants had spent, on average, 27.7 months total living in an English-speaking country, although after the removal of four outliers, the mean time spent living in an English-speaking country went down to 8.6 months. In order to assess the effect of these outliers on the behavioral measures used, group analyses of all measures were run with and without the outliers; because no significant differences were found for any

Table 1. Behavioral Results

Group	TOAL-3 Speaking Grammar	TOAL-3 Listening Grammar*	Saffran and Schwartz**	Carpenter Span**
Native Speakers (n = 17, 7 F)				
M (SD)	17.06 (3.36)	19.00 (7.98)	74.29 (3.08)	2.79 (.53)
Range	9–22	7–32	67–78	2–4
Percentile	16	25	N/A	N/A
Nonnative Speakers (n = 18, 8 F)				
M (SD)	15.11 (4.07)	28.17 (4.20)	70.61 (5.22)	3.22 (.52)
Range	5–20	21–34	57–77	2–4
Percentile	N/A	N/A	N/A	N/A

*$p < .01$.

**$p < .05$.

of the measures, all of the analyses reported here include all 18 NNS participants. When asked to rate their language skills on a 4-point scale for both English and German, participants rated themselves significantly better in German for listening, reading, writing, and speaking. Participants reported that, on average, they rarely heard English before age 11, and the most common source for those who did have such exposure was the radio. When asked to rate activities in terms of helpfulness in learning English, formal instruction was rated most helpful and socializing second most helpful, with reading rated next most helpful and watching TV much lower. Participants reported almost exclusive use of German throughout primary and secondary school, with use of English increasing only in adulthood, and then most often in a university or work setting.

ERP Results

The ERP data to the critical word in English sentences over all electrode sites are shown for the NS group in Figure 1 and for the NNS group in Figure 2. Visual inspection of the waveforms revealed clear patterns and clear differences between groups. The NS group showed a biphasic response to phrase structure violations in English: an extended, bilateral anterior negativity with onset around 100 msec and a posterior positivity peaking around 600 msec. A different pattern was observed in the NNS group, which showed no anterior negativity but a robust P600 over posterior sites extending to anterior sites.

Early (100–300 msec) Anterior Negativity

A group interaction supported the observation that the negativity was larger in the NS group in this time window [C × N: $F(1, 33) = 4.67, p < .05$; Figure 3].

In the NS group, analyses across anterior electrode sites in the 100–300 msec time window revealed a significant main effect of Condition [C: $F(1, 16) = 14.94, p < .001$], which was largest over anterior-most sites [C × A: $F(2, 32) = 10.41, p < .005$]. Although this effect showed a greater degree of left lateralization over lateral sites [C ×
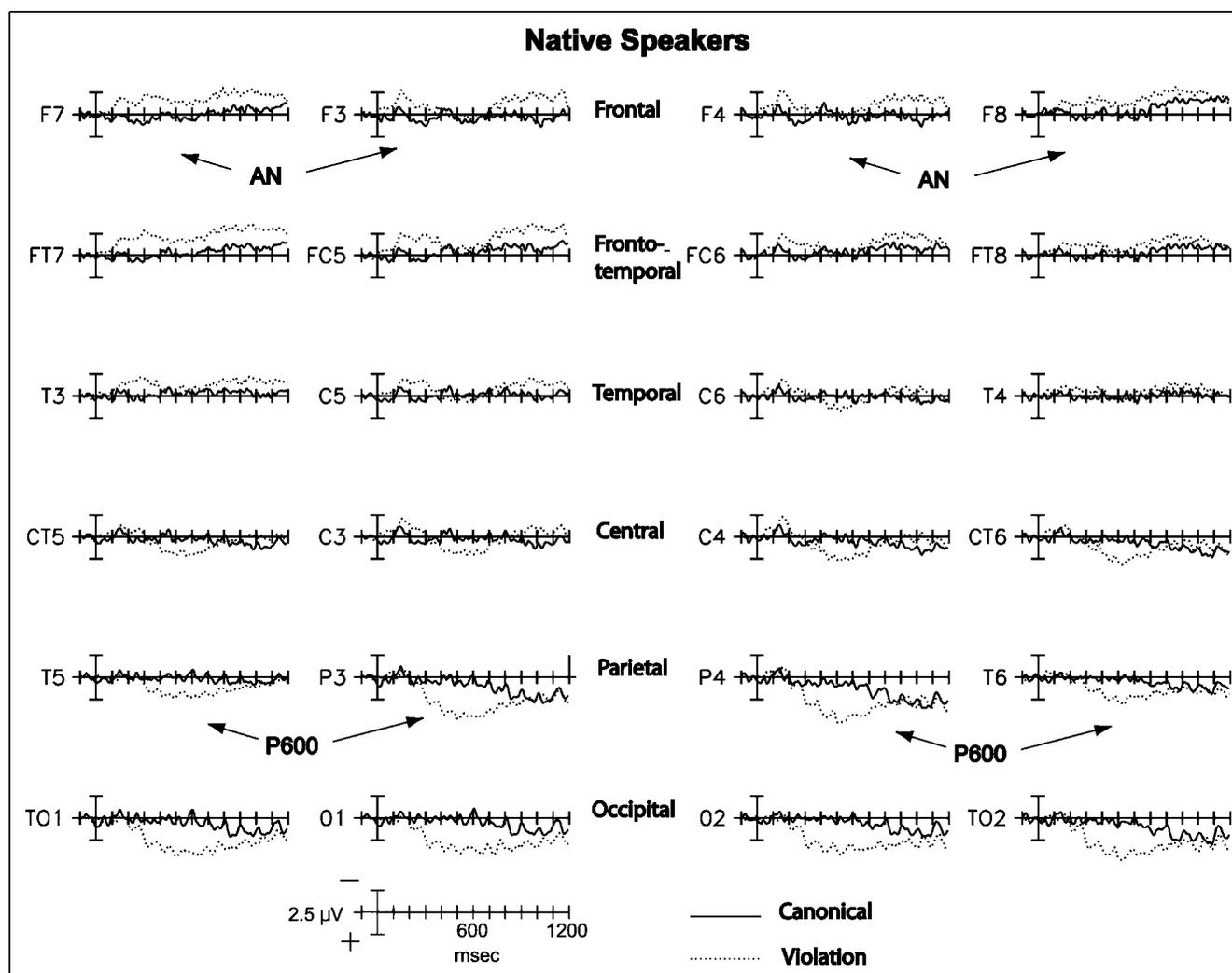

Figure 1. ERPs to English phrase structure violations for the NS group. Midline and eye electrodes (not included in analysis) are not shown.
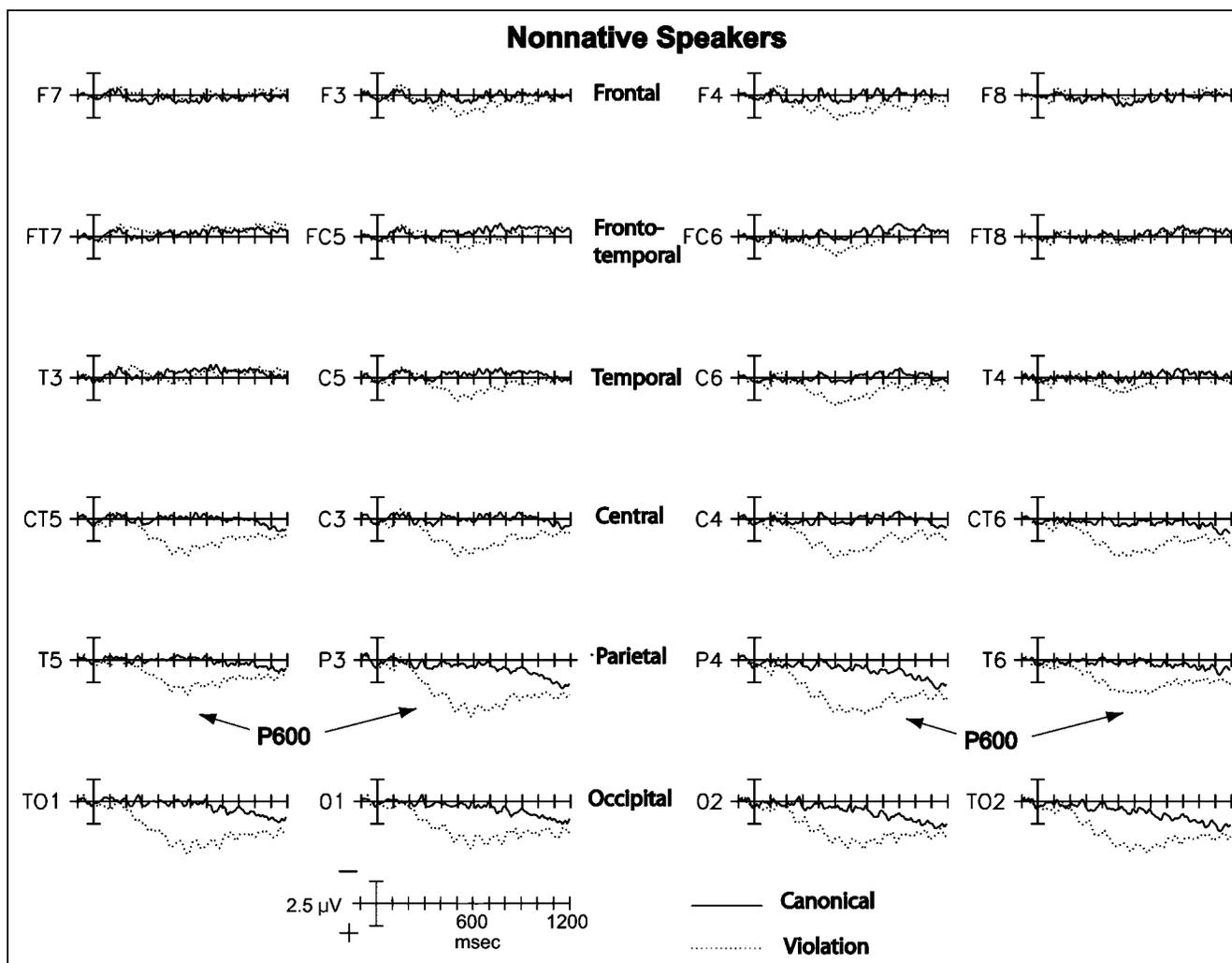
Figure 2. ERPs to English phrase structure violations for the NNS group. Midline and eye electrodes (not included in analysis) are not shown.

H × L: $F(1, 16) = 4.65, p < .05$], overall it was bilateral [C × H: $F(1, 16) = 1.81, ns$] and evenly distributed across lateral and medial sites [C × L: $F(1, 16) = 0.53, ns$].

In the NNS group, analyses across anterior electrode sites in the 100–300 msec time window in the NNS group revealed no main effect [C: $F(1, 17) = 0.69, ns$] and no significant interactions with Condition.

Later Anterior Negativity

300–700 msec. A group interaction supported the observation of a difference in effects between groups [C × N: $F(1, 33) = 6.15, p < .02$; Figure 4].

In the NS group, analyses across anterior sites in the 300–700 msec time window revealed a significant negativity largest over anterior-most [C × A: $F(2, 32) = 12.79, p < .0001$] and lateral [C × L: $F(1, 16) = 12.28, p < .005$] sites.

In the NNS group, analyses in this time window over anterior sites revealed a significant positivity largest over fronto-temporal and temporal [C × A: $F(2, 32) = 12.04,$
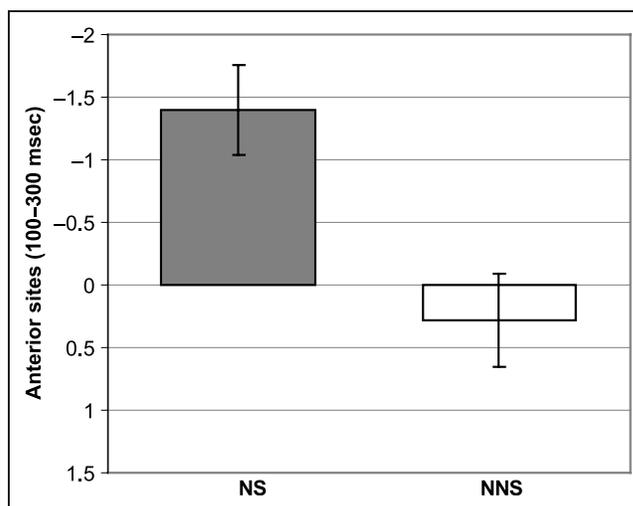

Figure 3. Average difference amplitude (violation − canonical, in μV) across the 100–300 msec time window over anterior sites for NS and NNS groups.
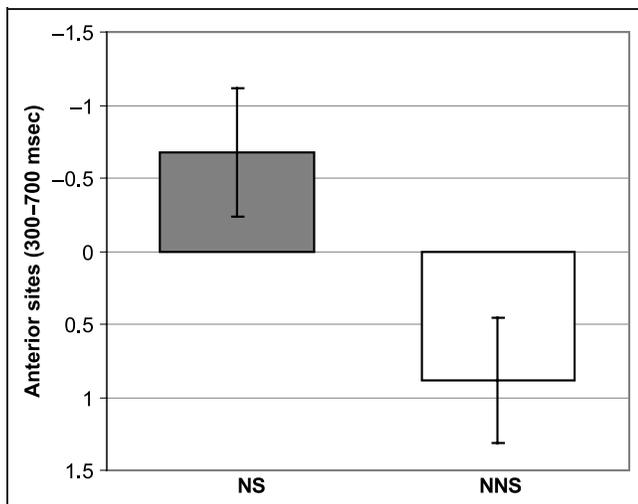
Figure 4. Average difference amplitude (violation − canonical, in μV) across the 300–700 msec time window over anterior sites for NS and NNS groups.

$p < .005$] and medial sites [C × L: $F(1, 16) = 19.56$, $p < .0001$].

700–1200 msec. A group interaction supported the observation of a difference in effects between Groups [C × N: $F(1, 33) = 6.53, p < .02$].

In the NS group, analyses across anterior sites in the 700–1200 msec time window revealed a significant main effect of Condition [C: $F(1, 16) = 5.39, p < .05$], a negativity which was largest over anterior-most sites [C × A: $F(2, 32) = 5.07, p < .05$].

In the NNS group, analyses in this time window over anterior sites revealed a significant positivity largest over temporal [C × A: $F(2, 32) = 3.72, p < .05$] and medial [C × L: $F(1, 16) = 16.47, p < .005$] sites.

Posterior Positivity (P600)

A near-significant group interaction revealed a trend for the P600 to be larger in the NNS group than in the NS group [C × N: $F(1, 33) = 3.14, p = .084$]. Although this interaction did not reach significance, because it reached the trend level, separate group analyses were still performed.

In the NS group, analyses over the three posterior rows of electrodes in the 300–1000 msec time window revealed a main effect of Condition [C: $F(1, 16) = 15.55, p < .005$], a positivity which was largest over posterior-most sites [C × A: $F(2, 32) = 11.80, p < .0001$]. In the NNS group, analyses in this time window over posterior sites revealed a main effect of Condition [C: $F(1, 17) = 26.65, p < .0001$].

As visual inspection suggested that the P600 was longer in duration in the NNS group, an analysis was conducted in the 1000–1200 msec time window. A significant group interaction revealed that the P600 was larger in the NNS

in this time window, with the difference maximal over central and parietal rows [C × A × N: $F(2, 68) = 4.33$, $p < .05$].

Correlational Analyses

In order to explore the possibility that working memory differences may have affected the results, additional correlational and regression analyses were performed.

Anterior negativity. No significant correlation between working memory span and average mean amplitude differences over anterior sites in the 100–300 msec time window was found ($r = .109, ns$).

Average mean amplitude differences over anterior sites in the 100–300 msec time window were regressed on proficiency, working memory span, and AOA using a backward stepwise regression procedure. The resulting best fit model [$R^2 = .124, F(1, 34) = 4.66, p < .05$] retained AOA as the sole predictor (β = .352, $p < .05$).

P600. No significant correlation between working memory span and average mean amplitude differences over posterior sites in the 300–1000 msec time window was found ($r = .128, ns$).

Average mean amplitude differences over posterior sites in the 300–1000 msec time window were regressed on proficiency, working memory span, and AOA using a backward stepwise regression procedure. The resulting best fit model [$R^2 = .093, F(1, 34) = 3.39, p = .074$] retained AOA as the sole predictor (β = .306, $p = .074$).

DISCUSSION

In this study, ERPs elicited by phrase structure violations were examined as two groups of English speakers listened to simple sentences in English. Groups consisting of either native speakers of English (NS) or nonnative speakers who did not begin acquiring English until around age 11 (NNS) were matched on a standardized measure of English grammatical proficiency. Analyses revealed differences in neural organization for syntactic processing between the two groups. In the NS group, consistent with their lower proficiency status, violations elicited a bilateral and prolonged anterior negativity with onset at 100 msec followed by a P600. In contrast, in the NNS group violations elicited only a P600 which was more widespread spatially, extending to more anterior sites, and temporally, extending to 1200 msec, compared to the NS group. The P600 in the NNS group also tended to be larger than in the NS group. Below we discuss possible functional interpretations of these results and their implications for theories of second-language acquisition, and discuss future directions for research into the relative contributions of AOA

and proficiency in determining neural organization for language.

Proficiency Matching

Groups were matched for English proficiency using the Speaking/Grammar subtest of the TOAL-3. This measure was chosen, in part, because it requires elicited imitation under time pressure, and tests which use elicited imitation are considered to be good measures of implicit language knowledge (Erlam, 2006; Munnich, Flynn, & Martohardjono, 1994; Dale, 1976). This measure was also chosen because it is relatively independent of working memory demands, which was desirable because the NNS participants had a higher working memory span than the NS participants. Although efforts were made to match the groups on working memory span, this proved to be difficult for several reasons. As discussed in our previous study (Pakulak & Neville, 2010), in the group of English native speakers, working memory correlated with proficiency, although the correlational analyses in that study showed that proficiency effects on neural organization for language were independent of working memory differences. The NNS participants were recruited from the University of Oregon population; as individuals who were able to work or study at a foreign university using primarily their second language, they had achieved a high enough level of proficiency to match lower proficiency native speakers. However, the use of participants from the university community, including graduate students and one professor from higher SES backgrounds, made it difficult to match this group on working memory span with a group of lower proficiency native speakers. This underscores the difficulty of conducting such research in a small university community in the United States. In future studies seeking to replicate the present results, it will be fruitful to recruit participants from larger communities with a wider range of individuals with good second-language proficiency, although the use of larger communities also presents potential problems, such as increased likelihood of differences in early second-language exposure. Although the groups were not matched on working memory span, we used a correlational approach to confirm that there was no relationship between working memory span and the ERP results.

Another important point with regard to proficiency matching is that the NS group was significantly higher on the Saffran and Schwartz Grammaticality Judgment task. Although having groups which were matched on this measure as well would have been ideal, it is also unlikely that this had a profound effect on the results. First, while the average score on this measure for the NNS group was lower than that for the NS group, NNS participants still scored an average of 90% correct. This, combined with the high performance of the NNS participants on the grammaticality judgment task in the ERP paradigm (97%), suggests that this group difference did not reflect a profound difference in proficiency which would potentially confound the results. Additionally, the NNS group actually outperformed the NS group on one measure of proficiency, the TOAL-3 Listening/Grammar subtest. Taken together, the behavioral results show that with one exception the NNS participants in this study scored at comparable or slightly higher levels on the proficiency measures used than did NS participants, adding a degree of confidence that the measures used accurately reflected a group of late learners of English with proficiency matching or exceeding that of the English native speakers.

Anterior Negativity

Phrase structure violations in English elicited an anterior negativity in the NS group which began around 100 msec and was robust and widespread, extending to 1200 msec over anterior sites bilaterally. In the NNS group, violations did not elicit a significant negative effect over anterior sites, suggesting marked differences in the degree to which resources indexed by the early anterior negativity were recruited by NNS participants. Regression analyses in the early time window controlling for possible effects of proficiency and working memory span provided additional evidence that these differences were driven by differences in AOA. The early anterior negativity to word category violations has been hypothesized to index early and automatic processes in which a word is integrated into the phrase structure of the preceding sentence fragment (Friederici, 2002). These results suggest that individuals who acquire a language later in life rely primarily on different, more controlled, neural mechanisms to achieve a level of proficiency comparable to that of some native speakers. This also suggests that the development of early and automatic processes hypothesized to be indexed by the early anterior negativity may be governed by maturational constraints consistent with a sensitive period.

Results from several previous ERP studies of syntactic processing in second-language learners support this interpretation. Syntactic violations in the nonnative language of late learners either failed to elicit an anterior negativity (Kotz, Holcomb, & Osterhout, 2008; Hahne, 2001; Hahne & Friederici, 2001) or elicited a negative effect in a later time window (Ojima et al., 2005; Weber-Fox & Neville, 1996), although one study reported a bilateral and extended early anterior negativity effect to word category violations in this time window in nonnative speakers, even in those of lower proficiency (Rossi et al., 2006).

The extended bilateral negativity in the NS group was larger over lateral than medial sites in the 300–700 msec time window. This reduced negativity over medial sites in this time window is likely an interaction with the P600 extending to anterior sites, as has been shown previously (Pakulak & Neville, 2010). This extended bilateral negativity in the NS group was also significant in the 700–1200 msec time window. Results from a recent study using ERP and fMRI data gathered from the same participants in

the auditory syntactic processing paradigm used here provide evidence that a generator or generators in left inferior frontal gyrus contribute to this extended negativity across multiple time windows and hemispheres, suggesting a single unitary component (Pakulak, Dow, & Neville, 2009).

Posterior Positivity

Phrase structure violations elicited a robust posterior positivity in the NS group, part of a biphasic response which is consistent with much previous ERP research examining the neural response to syntactic violations in native speakers. Violations also elicited a robust posterior positivity in the NNS group. This is consistent with previous research examining syntactic processing in late second-language learners, as several studies have reported a P600 to syntactic violations in such groups (Kotz et al., 2008; Rossi et al., 2006; Hahne, 2001; Weber-Fox & Neville, 1996), and suggests that processes reflected in the P600 are less sensitive to maturational constraints than those reflected in the early anterior negativity. However, two ERP studies of syntactic processing in late learners do not report a P600 to syntactic violations (Ojima et al., 2005; Hahne & Friederici, 2001). One study which did not report a P600 (Hahne & Friederici, 2001) attributed the finding to differences in second-language proficiency: Although participants in that study had an error rate of around 20% in an on-line grammaticality judgment task, participants in the study using similar stimuli in which a P600 was found for late learners (Hahne, 2001) had an error rate of 8%. Proficiency differences also likely played a role in the other study which did not report a P600 in late learners (Ojima et al., 2005), as the groups of high- and low-proficiency late learners had error rates of 13% and 33%, respectively, in an off-line acceptability judgment task of stimuli consisting of three-word sentences featuring straightforward subject–verb agreement violations.

Although violations elicited a P600 in both groups, the P600 in the NNS group was more widespread spatially, extending across anterior sites, and also tended to be larger compared to the NS group. The P600 has been hypothesized to reflect more controlled processes involved with a failure to parse and related processes of repair (Friederici, Hahne, et al., 2002; Hagoort & Brown, 2000) or difficulty in syntactic integration (Kaan et al., 2000). Evidence from recent studies suggests that the P600 is not specific to syntax, as it is elicited by nonsyntactic violations such as semantic (van Herten, Kolk, & Chwilla, 2005) and meter (Schmidt-Kassow & Kotz, 2009) violations. Several researchers suggest that the P600 is thus a more general integration mechanism and/or reprocessing mechanism modulated by rule-based building up of expectancies, elicited when an initial analysis has to be rejected (e.g., Schmidt-Kassow & Kotz, 2009). Our results thus suggest that late L2 learners rely more on more general, controlled, rule-based reprocessing mechanisms in experimental

conditions which place demands on second-language processing which more closely approximates those in everyday life (see below). Such general mechanisms may be less likely to be subject to maturational constraints, consistent with the previous results from late learners discussed above.

Thus, the present results suggest that late learners may rely more on these controlled processes to achieve a level of proficiency comparable to some native speakers. Interestingly, this more widespread distribution of the P600 is reminiscent of the effect in the higher proficiency group in our previous study. This raises the tentative hypothesis that this more widespread positivity may reflect a compensatory mechanism which interacts with maturational constraints. In this hypothesis, late learners are not only more reliant on more controlled processes, but may recruit additional controlled processes less sensitive to maturational constraints, as reflected in the widespread P600, in order to compensate for an absent or reduced recruitment of processes which are more sensitive to maturational constraints, reflected in the early anterior negativity. Although this hypothesis is necessarily speculative, it provides an interesting direction for future study.

The P600 was temporally more focal in the NS group, whereas it extended to 1200 msec in the NNS group. This result suggests subtle differences in the use of the resources reflected in the P600. It is possible that this might reflect the more efficient use of resources important for syntactic integration and reanalysis in the NS group as a result of more experience with English, although this hypothesis is necessarily preliminary and requires further research.

Methodological Differences

The basic question which this and the many other studies discussed above seeks to address is whether a person's grammatical knowledge is represented and/or processed differently if it is learned late as a second language as opposed to early as a first (or second) language. However, it is important to note that our measures of how grammatical knowledge is represented/processed vary and that methodological differences could, and likely do, account for differences in results across studies. By definition, a discussion of language knowledge is a discussion of language knowledge as measured by a given experimental paradigm, and given the rather large degree of methodological differences across studies of second-language grammatical processing, it is likely that these differences are a nontrivial source of noise.

For example, several studies which report anterior negativities in late learners use paradigms in which the violations are presented in short, simple, active sentences with no variation in violation position (Rossi et al., 2006; Ojima et al., 2005), either with filler sentences without a second violation type (Rossi et al., 2006) or with no filler sentences and a slow visual presentation rate (Ojima et al., 2005). Rossi

and colleagues attribute the finding of an anterior negativity in late learners of both low and high proficiency to the use of simple, active sentence structures with only two violation types, which likely allowed participants to concentrate on the processing of a limited amount of syntactic rules. Results from artificial language models which elicit an early anterior negativity in participants trained to a high degree of proficiency (Friederici, Steinhauer, et al., 2002) have also been cited as evidence that native-like neural organization for syntactic processing can be achieved in late learners, but such language models are, by definition, highly constrained by the use of a small set of "words" and rules, and thus the paradigms used to assess processing also feature short, simple sentences with a high degree of predictability.

Although it is always unlikely that a processor will be taxed in an experimental paradigm to the same degree as the demands of functioning in a second language on a daily basis, it could be argued that paradigms which attempt to tax the processor to a degree which more closely approximates these demands are closer to a true measure of second-language syntactic processing. In the present study, we used a paradigm specifically designed to tax the processor to a greater degree by using naturally spoken speech in participant-paced auditory presentation, using two intermixed conditions (English and Jabberwocky), varying sentence length and violation position, using filler sentences with a different violation type and position, and using two different tasks to create additional attentional demands and limit the degree to which participants could use strategies which focused on one violation point. Although it could be argued that the phrase structure violations used were quite salient, by varying the predictability of these violations in this manner our experimental design likely taxed the processor to a degree more closely approximating everyday language use than other paradigms in the literature.

Another methodological issue is the use of a grammaticality judgment task. As early anterior negativity elicited by phrase structure violations has been shown to be insensitive to task differences, while the P600 has been shown to be sensitive to such differences (Hahne & Friederici, 2002), it is unlikely that the use of a grammaticality judgment task affected our main finding. However, it is possible that the use of a passive comprehension task would have affected the P600, although this question awaits future study.

Another issue concerns the salience of the violation type used. Although it is difficult to assess differences in saliency of violation, especially across studies, it is possible that such differences also play a role. However, if phrase structure violations, such as those used in the present study are more salient than other violation types, such as agreement violations, and still no early anterior negativity was elicited in late learners, then it is unlikely that our results would have been different with the use of violations which might be considered less salient. However, this is an empirical question which awaits future study.

Although, based on an examination of the literature, the issue of methodological differences appears to be an important one, it is also important to note that this is still an empirical question. Relevant research systematically comparing the effects of these methodological factors on second-language processing still remains to be done.

Implications and Future Directions

These results also provide several other directions for future research. First, it will be important to further explore the degree to which second-language proficiency can impact neural organization for syntactic processing in late learners. Although the results presented here provide evidence that certain processes important for syntactic processing are sensitive to maturational constraints, the group of late learners studied here were, on average, of relatively low proficiency compared to native speakers. The expansion of this study to include late learners of higher proficiency would shed more valuable light on this question. Also, data from a wide range of late second-language learners of varying proficiency levels could allow for a more comprehensive correlational analysis which would, in turn, allow for a more comprehensive investigation of the factors which affect neural organization for syntactic processing in second-language learners. Another important future direction is the use of fMRI in conjunction with ERPs to more fully characterize the effects of both AOA and proficiency on the recruitment of specific neuroanatomical regions in syntactic processing; this is a current line of research in our laboratory (Pakulak et al., 2009).

Although the results discussed above shed light on the role of AOA in the determination of neural organization for syntactic processing, there remains a degree of inconsistency across studies. Methodological differences between laboratories, both specific to the ERP paradigms used as well as with regard to measures of proficiency, make between-studies interpretation and comparison difficult. Of particular importance will be the development and use of better measures of proficiency. Here a higher degree of cooperation between laboratories would greatly help the field overall in this regard, as many of the laboratories actively pursuing this line of research are in different countries with researchers who are speakers of different native languages, using paradigms for which extensive data on native speakers already exist. This is an obvious opportunity for cooperation between laboratories, either at the level of collaborative studies or at a lower level of cooperation featuring the exchange of proficiency and stimulus materials. Such cooperation using paradigms in different languages also raises the tantalizing possibility of directly comparing ERPs from the same participants while processing their native and their second language. The field would also benefit from the establishment of guidelines with respect to the characterization of participants, in particular, a

more comprehensive characterization of second-language proficiency that could be used across laboratories. Taking such factors into consideration as the field moves forward can only lead to stronger results and a better understanding of the role of AOA and proficiency in neural organization for second-language processing.

REFERENCES

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologia, 128,* 466–478.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language, 68,* 706–755.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13,* 21–58.

Dale, P. S. (1976). *Language development: Structure and function* (2nd ed.). New York: Holt Rinehart and Winston.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450–466.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27,* 464–491.

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,* 259–281.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6,* 78–84.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31,* 45–63.

Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research, 25,* 157–176.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1,* 183–192.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 529–534.

Gunter, T. C., Friederici, A. D., & Hahne, A. (1999). Brain responses during sentence reading: Visual input affects central processes. *NeuroReport: For Rapid Communication of Neuroscience Research, 10,* 3175–3178.

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage, 20,* S18–S29.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9,* 416–423.

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

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38,* 1531–1549.

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

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

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

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13,* 339–356.

Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research, 11,* 199–212.

Hammil, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (1994). *Test of adolescent and adult language* (3rd ed.). Austin, TX.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology, 19,* 286–300.

Hollingshead, A. (1975). *Four factor index of social status.* Unpublished manuscript, New Haven, CT: Yale University Department of Sociology.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning, 56(Suppl. 1),* 279–304.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language & Cognitive Processes, 15,* 159–201.

Kaan, E., & Swaab, T. Y. (2003a). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research, 17,* 621–635.

Kaan, E., & Swaab, T. Y. (2003b). Repair, revision, and complexity in syntactic analysis: An electrophysiological

differentiation. *Journal of Cognitive Neuroscience, 15,* 98–110.

Kim, K. H. S., Relkin, N. R., Lee, K.-M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature, 388,* 171–174.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing [Review]. *Brain and Language, 109,* 68–74.

Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologia, 128,* 514–527.

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

Lenneberg, E. H. (1967). *Biological foundations of language.* New York: Wiley.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13,* 361–392.

Mayberry, R. I. (1993). First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language. *Journal of Speech & Hearing Research, 36,* 1258–1270.

Mayberry, R. I., & Eichen, E. B. (1991). The long-lasting advantage of learning sign language in childhood: Another look at the critical period for language acquisition. *Journal of Memory and Language, 30,* 486–512.

Mayberry, R. I., Lock, E., & Kazmi, H. (2002). Linguistic ability and early language exposure. *Nature, 417,* 38.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other* (pp. 227–245). Mahwah, NJ: Erlbaum.

Münte, T. J., Heinze, H., & Mangun, G. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience, 5,* 335–344.

Neville, H. J., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3,* 155–170.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science, 14,* 11–28.

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

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

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language & Cognitive Processes, 8,* 413–437.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1995). Brain potentials elicited by garden-path sentences: Evidence of

the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20,* 786–803.

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

Pakulak, E., Dow, M., & Neville, H. J. (2009). *An investigation of the neural generators of ERP indices of syntactic processing using proficiency-related ERP modulations in an ERP–fMRI paradigm.* Paper presented at the Cognitive Neuroscience Society, San Francisco, CA.

Pakulak, E., & Neville, H. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22,* 2728–2744.

Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121,* 1841–1852.

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18,* 2030–2048.

Rüschemeyer, S. A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping, 25,* 266–286.

Rüschemeyer, S. A., Zysset, S., & Friederici, A. D. (2006). Native and non-native reading of sentences: An fMRI experiment. *Neuroimage, 31,* 354–365.

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience, 21,* 1693–1708.

van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16,* 1068–1084.

van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22,* 241–255.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron, 37,* 159–170.

Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language, 85,* 231–244.

Weber-Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8,* 231–256.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12,* 238–265.