Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers

Christine M. Weber-Fox and Helen J. Neville
University of Oregon, Eugene

Abstract

Changes in several postnatal maturational processes during neural development have been implicated as potential mechanisms underlying critical period phenomena. Lenneberg hypothesized that maturational processes similar to those that govern sensory and motor development may also constrain capabilities for normal language acquisition. Our goal, using a bilingual model, was to investigate the hypothesis that maturational constraints may have different effects upon the development of the functional specializations of distinct subsystems within language. Subjects were 61 adult Chinese/English bilinguals who were exposed to English at different points in development: 1-3, 4-6, 7-10, 11-13, and after 16 years of age. Event-related brain potentials (ERPs) and behavioral responses were obtained as subjects read sentences that included semantic anomalies, three types of syntactic violations (phrase structure, specificity constraint, and subjacency constraint), and their controls. The accuracy in judging the grammaticality for the different types of syntactic rules and their associated ERPs was affected by delays in second language exposure as short as 1-3 years. By comparison the N400 response and the judgment accuracies in detecting semantic anomalies were altered only in subjects who were exposed to English after 11-13 and 16 years of age, respectively. Further, the type of changes occurring in ERPs with delays in exposure were qualitatively different for semantic and syntactic processing. All groups displayed a significant N400 effect in response to semantic anomalies, however, the peak latencies of the N400 elicited in bilinguals who were exposed to English between 11-13 and >16 years occurred later, suggesting a slight slowing in processing. For syntactic processing, the ERP differences associated with delays in exposure to English were observed in the morphology and distribution of components. Our findings are consistent with the view that maturational changes significantly constrain the development of the neural systems that are relevant for language and, further, that subsystems specialized for processing different aspects of language display different sensitive periods.

INTRODUCTION

Our goals in this study were to assess the hypothesis that alterations in the timing of language experience have more pronounced effects on some aspects of language than others. Several studies have shown that the nature of sensory input significantly affects the development of the visual, auditory, and somatosensory systems (Freeman & Thibos, 1973; Knudsen, 1988; Wiesel & Hubel, 1963, 1965; Knudsen & Knudsen, 1990; Kaas, 1991; Kaas, Merzenich, & Killackey, 1983). Altered input can also significantly impact sensory and motor maps in adult mammalian brains (Kaas, 1991; Kaas et al., 1983). However, many such experience dependent changes only occur during specific developmental windows, so called critical and sensitive periods. Moreover, even within the primate visual system different brain areas and functions display different sensitive periods (Harwerth, Smith, Duncan, Crawford, & von Noorden, 1986). Less is known about the neural bases of critical and sensitive period phenomena in humans, however, clinical evidence suggests that analogous constraints on development are likely to occur in the human brain (Maurer & Lewis, 1993). In humans, some maturational changes have been found to occur over developmental periods as long as 16-20 years (Lenneberg, 1967; Huttonlocher & de Courten, 1987; Conel, 1939). Several investigators have suggested that the ongoing elaboration of neural circuitry during postnatal maturation may be relevant to understanding critical periods and the ability of young animals to recover from brain injuries that would cause severe and permanent deficits in older animals (Pomeroy, LaMantia, & Purves, 1990; Rakic, Bourgeois, Eckenhoff, Zecovic, & Goldman-Rakic, 1986). However, little evidence exists on this point.

It is likely that the general principles that govern neurophysiological development in sensory systems may also be relevant to some degree for cognitive develop-
ment including language. Lenneberg (1967) interpreted several lines of evidence as showing that there is a critical period for normal language acquisition that depends upon diminished brain plasticity and, in turn, determines related capabilities for reorganization in the postpubescent adult. Several behavioral studies examining the effects of delays in primary and secondary language acquisition lend support to the general hypothesis that the ultimate proficiency attained for various linguistic skills, including phonological, semantic, grammatical, and syntactic skills, depends in part on the age of exposure to the language (Curtiss, 1992; Johnson & Newport, 1989; Newport, 1988; Oyama, 1976; Seliger, Krashen, & Ladefoged, 1975; Mayberry & Fischer, 1989; Flege & Fletcher, 1992; Oyama, 1982; Patkowski, 1980; Johnson & Newport, 1991; Liu, Bates, & Li, 1992). Results from behavioral studies also indicate that the acquisition of different linguistic processes is not uniformly affected by delays in language exposure (Johnson & Newport, 1989; Newport, 1988; Mayberry & Eichen, 1991; Mayberry & Fischer, 1989; Flege & Fletcher, 1992). The acquisition of vocabulary appears least vulnerable to delays in exposure, while other linguistic structures related to phonological and syntactical processing appear to be most affected.

Important evidence on the nature and neural bases of language processing has come from recordings of event-related brain potentials (ERPs). ERPs provide a means for assessing brain electrical activity, measured at the scalp, that is time-locked to a particular stimulus including different types of linguistic constructs (Hillyard & Picton, 1987). For example, the ERP response to violations of semantic expectations (N400) is a broadly distributed increase in negativity that is maximal at posterior scalp locations and peaks at approximately 400 msec post-stimulus onset (Kutas & Hillyard, 1980; Van Petten & Kutas, 1990; Holcomb & Neville, 1991). Recent ERP evidence suggests that different functional specializations are utilized for syntactic processing, and that the ERP responses for at least some types of syntactic processing may be generated within anterior regions of the left hemisphere (Neville, Mills, & Lawson, 1992; Munte, Heinze, & Mangun, 1993; Rosler, Putz, Friederici, & Hahne, 1993; Friederici, Pfeifer, & Hahne, 1993; Kluender & Kutas, 1993; Osterhout & Holcomb, 1993; Osterhout & Holcomb, 1992; Neville, Nicol, Barss, Forster, & Garrett, 1991). In addition, ERPs have been used successfully to show the effects of altered early experience on functional specializations by comparing responses of normal hearing subjects to those of individuals who are congenitally deaf (Neville, 1990; Neville et al., 1992; Neville, Coffey, Lawson, Fischer, Emmorey, & Bellugi, 1996). Results of one such study suggest that semantic processing may be relatively impervious to delays in exposure (Neville et al., 1992). In that study, ERPs were recorded while deaf and hearing subjects read English sentences. The ERPs of deaf subjects to semantically anomalous sentences and to open class words (words which primarily convey referential meaning) were similar to those of hearing subjects despite the fact that the deaf subjects acquired English late in development and imperfectly (Neville et al., 1992). In contrast, ERPs elicited in response to closed class words (words that primarily provide structural information in a sentence) were markedly different in the deaf and hearing populations. Specifically, the N280 response (localized over anterior, left hemisphere regions), elicited by closed class words in hearing subjects, was absent in the ERPs of deaf individuals. These results suggest that functional specialization for grammatical processing may be more sensitive to the timing of exposure to a language than is semantic processing. Additional evidence along these lines comes from ERP evidence in a study of deaf and hearing individuals who were exposed to American Sign Language (ASL) early or late in development (Neville et al., 1996). In that study, the distribution of ERP responses was found to distinguish native ASL learners (both deaf and hearing subjects) from hearing interpreters who acquired ASL as adults.

Very few investigators have utilized ERP recording techniques for determining effects of delayed language exposure in bilingual populations. One study reports age of exposure effects for bilinguals who were required to identify whether each of a series of auditorially presented words was French or English (Genesee et al., 1978). The latencies of the N1 response were shorter over the left hemisphere for infant and early childhood learners, and shorter over the right hemisphere for the third group who had learned their second language as adolescents. Two studies of N400 responses in bilinguals have linked ERP changes to fluency with a language. Ardal, Donald, Meuter, Muldrew, and Luce (1990) found the latency of the N400 response to semantic anomalies of bilinguals to be longer than that of monolinguals. Further, the N400 latency in response to anomalies in the second language was longer than that for the first language of bilinguals. No effects of age of exposure on N400 latencies or amplitudes were found. Kutas and Kluender (1993) also reported a longer N400 peak latency for processing in the bilinguals' less fluent language. In addition, they reported an overall N400 reduction for processing the bilinguals' less fluent language. The relationship of the N400 latency and amplitude to the age of second language exposure was not specified. Differences in the presence of components of ERP responses to morphological violations were reported for monolingual and bilingual Spanish speakers. No age of exposure effects were reported.

Our goal in the current study was to investigate, using a bilingual model, the hypothesis raised by previous studies of deaf and bilingual subjects that the functional specializations of different aspects of language processing may be differentially impacted by delays in exposure. The language stimuli utilized in the present study were...
the same as those employed in a previous study of semantic and syntactic processing in monolinguals (Neville et al., 1991). That investigation reported the typical posterior, bilateral N400 in response to mid-sentence semantic anomalies. In contrast, violations of phrase structure elicited an enhanced N125 over anterior sites within the left hemisphere and increased negativity between 300 and 500 msec over temporal and parietal sites within the left hemisphere, followed by a broadly distributed late positive shift. ERPs were also recorded to two violations of movement constraint, i.e., rules that specify the types of noun phrases from which WH phrases can be extracted. Violations of specificity constraints elicited a sustained negativity beginning around 125 msec that was localized over anterior regions of the left hemisphere. ERPs elicited by subjacency constraint violations were broadly distributed increases in positivity beginning around 200 msec. The current study utilized these stimuli with bilinguals who were grouped according to their age of exposure to English, and was designed to provide information regarding the nature of sensitive periods associated with semantic and different types of syntactic processing (see Tables 1 and 2).

RESULTS

Language History—Use Patterns

On a language history questionnaire (Appendix A), subjects reported their relative use of Chinese and English for different points in development: 0-4, 5-12 years (primary school age), 13-18 years (secondary school age), and at the time of testing. For these different age ranges, subjects indicated their relative use of Chinese and English in home and school/work settings. The rating scale used was as follows: 1 = Chinese only, 2 = Chinese frequently, English rarely, 3 = Chinese majority of time with English used at least 1/4 of the time, 4 = Equal use of Chinese and English, 5 = English majority with Chinese used at least 1/4 of time, 6 = English frequently, Chinese rarely, 7 = English Only. Subjects’ reports are summarized in Figure 1. From Figure 1, it can be seen that there is a systematic pattern of relative use for periods before age 5, primary school age, and secondary school age, that could be predicted in part by the age that subjects were immersed in English, in most cases, the age that subjects moved to the United States. Use patterns at the time of testing showed a different pattern, since in the university/work setting, all groups were using English predominantly. However, the bilinguals exposed to English after 16 years of age were using English less than the other groups, about 75% of the time as compared to using English almost exclusively (overall group ANOVA: F(4,56) = 3.14, p = 0.0213, 11-13; >16 group contrast: F(1,21) = 7.93, p = 0.0103). In the home setting, the 1-3 and 7-10 bilinguals reported the highest relative use of English with English used about 75% of the time (1-3; 7-10 group contrast, n.s.). The 4-6 bilingual group reported less use of English and about equal use of Chinese and English [1-3:4-6 group contrast F(1,26) = 4.62, p = 0.0411, 4-6:7-10 group contrast, n.s.]. The 11-13 groups’ relative use of Chinese and English fell between that of the 4-6 and >16 groups and did not differ significantly from either group. The bilinguals exposed to English after 16 years of age on average used English slightly less than 25% of the time at home, the group contrast with the 4-6 group was marginal [F(1,21) = 4.10, p = 0.0556]. Regression analyses were done to determine whether use patterns of Chinese/English at the time of testing were better predicted by age of exposure to, or years of experience with English.1 In the university/work setting, no significant correlations were found for age of exposure or years of experience. In the home setting, both age of exposure and years of experience were significantly correlated with relative

| Table 1. Subject Characteristics for Each Bilingual Group |
|-------------|-------|-------|-------|-------|-------|
|             | 1-3  | 4-6  | 7-10 | 11-13 | >16  |
| N           | 15   | 13   | 10   | 13    | 10   |
| Sex (Female) | 57%  | 46%  | 54%  | 54%   | 54%  |
| Age mean    | 19.3 | 19.1 | 19.8 | 19.9  | 30   |
| (SE)        | (0.35) | (0.24) | (0.42) | (0.40) | (2.14) |
| Number of years English | Mean | 17.9 | 14.3 | 10.8  | 7.8  | 7.6  |
| (SE)        | (0.43) | (0.25) | (0.59) | (0.50) | (0.84) |

1* subjects are grouped according to the age they began using English.

| Table 2. Examples of Control (C) and Violation (V) Sentences a |
|-------------|-------|-------|-------|-------|
| Semantic/pragmatic control and violation: | C#1. | The scientist criticized Max’s proof of the theorem. | V#1. | The scientist criticized Max’s event of the theorem. |
| Phrase structure control and violation: | C#2. | The scientist criticized a proof of the theorem. | V#2. | The scientist criticized Max’s proof of the theorem. |
| Specificity constraint controls and violation: | C#3. | What did the scientist criticize a proof of? | V#3. | What did the scientist criticize Max’s proof of? |
| Subjacency constraint control and violation: | C#4. | Was the proof of the theorem criticized by the scientist? | V#4. | What was a proof of criticized by the scientist? |

a Italicized words indicate comparison points for control and violation sentences.

Weber-Fox and Neville 233
Figure 1. Subjects' reports of relative usage of Chinese and English for various periods of development for both home and school or university/work settings: before school age (<5 years), primary school age (5-12 years), secondary school age (13-18 years), and at the time of testing. Reports are grouped according to age of exposure to English: 1-3, 4-6, 7-10, 11-13, and >16 years. Rating scale used: 1 = Chinese only, 2 = Chinese frequently, English rarely, 3 = Chinese majority with English used at least 1/4 of the time, 4 = Equal use of Chinese and English, 5 = English majority with Chinese used at least 1/4 of the time, 6 = English frequently, Chinese rarely, 7 = English only (refer to Appendix A for more detail).

use and were not significantly different from one another ($r = -0.5044$ and $r = 0.3771$, respectively).

Self-Rated Comfort and Proficiency

Subjects indicated whether they were more comfortable using English, both English and Chinese equally, or Chinese. The results are illustrated in Figure 2. The majority of subjects who were exposed to English <11 years reported they were more comfortable using English, but with increasing age of exposure, there was a slight increase in the percent of reports of feeling equally comfortable using English and Chinese. With delays of 11-13 years, approximately half the subjects reported feeling more comfortable using English and half felt equally comfortable using Chinese and English. No subjects exposed to English before the age of 14 reported feeling more comfortable using Chinese. In contrast, the >16 bilinguals who had a similar number of years of experience with English as the 11-13 group (Table 1) most often reported they were more comfortable using Chi-

Figure 2. Percent of subject reports for which language they felt more comfortable using, English, Chinese, or both equally. Reports are grouped according to age of exposure to English: 1-3, 4-6, 7-10, 11-13, and >16 years.
nese (60%) and the remaining subjects reported feeling equally comfortable using English and Chinese. Regression analyses\(^2\) revealed that age of exposure to English was a better predictor of subjects' comfort ratings compared to years of experience with English \((r = -0.6705\) and \(r = 0.5102\), respectively, \(t = 2.501, p < 0.02\)).

Subjects also rated their current proficiency in comprehension, speaking, reading, and writing in Chinese and English: 4 = perfectly, 3 = well, 2 = sufficiently, 1 = scarcely (refer to Language History Questionnaire in Appendix A for details). These proficiency ratings are illustrated in Figure 3. Group effects were significant for each of the ratings in Chinese and English (Table 3). The ratings for the first 3 groups were similar for 7 of the 8 measures, only one group contrast between the 1-3, 4-6, and 7-10 bilinguals was significant (Table 3). The 7-10 group rated their proficiency slightly higher for reading Chinese, which probably reflects their exposure to formalized education in Chinese that the other two groups did not have. Compared to these earlier learning groups, the 11-13 group rated themselves higher in Chinese proficiency and lower in English proficiency in comprehension, speaking, reading, and writing (Table 3). Finally, compared to the 11-13 group, the latest learning group, >16 years, rated themselves even higher in Chinese and lower in English proficiency on all measures except reading English (Table 3).

The relative ratings in Chinese and English within each group also reflected a relationship to age of exposure to English. The 1-3, 4-6, and 7-10 groups rated their proficiency higher for English than Chinese for every measure (Table 4). In contrast, the 11-13 groups rated themselves equally proficient in Chinese and English for comprehension, speaking, and reading (Table 4). The >16 group's reports of relative proficiency in Chinese and English were markedly different from the 11-13 group's despite the fact they had similar years of experience with English. The >16 groups rated themselves more proficient in Chinese on every measure. The results of regression analyses indicated significant relationships be-
Table 3. Group ANOVA Results for Self-Rated Proficiency: Summary of F and p Values

<table>
<thead>
<tr>
<th>Group effects</th>
<th>Understand</th>
<th>Speak</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>9.74 0.0000</td>
<td>14.67 0.0000</td>
<td>44.37 0.0000</td>
<td>25.66 0.0000</td>
</tr>
<tr>
<td>English</td>
<td>7.49 0.0001</td>
<td>13.67 0.0000</td>
<td>7.76 0.0000</td>
<td>16.04 0.0000</td>
</tr>
</tbody>
</table>

Individual group contrasts

<table>
<thead>
<tr>
<th>1-3:4-6:7-10</th>
<th>Chinese</th>
<th>NS</th>
<th>NS</th>
<th>4-6:7-10 F(1,21) = 5.33 p = 0.0312 NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>

7-10:11-13

| Chinese     | 4.03 0.0577 | 6.68 0.0173 | 10.96 0.0033 | 4.12 0.0551 |
| English     | 4.54 0.0450 | NS a | 6.56 0.0182 | 4.57 0.0446 |

11-13:16

| Chinese     | 8.40 0.0086 | 10.02 0.0047 | 10.80 0.0035 | 12.46 0.0020 |
| English     | 4.22 0.0526 | 8.19 0.0093 | NS            | 14.19 0.0110 |

a Contrast between 4-6:11-13 was significant for this measure however, F(1,24) = 9.85, p = 0.0045.

between each of the measures of self-rated proficiency and variables of timing and duration of exposure to English (Table 5). While age of exposure resulted in higher correlations than years of experience for each of the eight measures taken, only the correlations for writing in Chinese were significantly different (t = 2.7596, p < 0.01).

Standardized Test Performance

The behavioral performances of bilinguals on four standardized tests is summarized in Figure 4. The double dashed lines on each graph indicate the performance of monolinguals who were tested previously (mean +/- the standard error). Bilingual group performances that dif-

Table 4. Statistics for the Relative Self-Rated Proficiencies in Chinese and English within Each of the Bilingual Groupsa

<table>
<thead>
<tr>
<th></th>
<th>1-3 (df) 14</th>
<th>4-6 12</th>
<th>7-10 9</th>
<th>11-13 12</th>
<th>&gt;16 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched-t</td>
<td>-3.87 0.0017</td>
<td>-5.20 0.0002</td>
<td>-3.87 0.0038</td>
<td>NS 0.0000</td>
<td>8.51 0.0000</td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>-8.92 0.0000</td>
<td>-5.33 0.0002</td>
<td>-4.13 0.0026</td>
<td>NS 0.0000</td>
<td>8.57 0.0000</td>
</tr>
<tr>
<td>Matched-t</td>
<td>-16.75 0.0000</td>
<td>-10.12 0.0000</td>
<td>-5.46 0.0004</td>
<td>NS 0.0002</td>
<td>6.13 0.0000</td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>-19.86 0.0000</td>
<td>-14.50 0.0000</td>
<td>-7.57 0.0000</td>
<td>-3.95 0.0019</td>
<td>3.00 0.0150</td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched-t</td>
<td>-19.86 0.0000</td>
<td>-14.50 0.0000</td>
<td>-7.57 0.0000</td>
<td>-3.95 0.0019</td>
<td>3.00 0.0150</td>
</tr>
</tbody>
</table>

a These ratings reveal for each group, whether they considered themselves more proficient in Chinese or English.
Table 5. Significant Correlations (Two-Tailed, \( p < 0.05 \)) Between Self-Rated Proficiency in Chinese and English and Age of Exposure and Years of Experience with English

<table>
<thead>
<tr>
<th></th>
<th>Understand</th>
<th>Speak</th>
<th>Read</th>
<th>Write</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age of exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>0.5750</td>
<td>0.6450</td>
<td>0.8150</td>
<td>0.7468</td>
</tr>
<tr>
<td>English</td>
<td>-0.4900</td>
<td>-0.6694</td>
<td>-0.5024</td>
<td>-0.7013</td>
</tr>
<tr>
<td><strong>Number of years experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>-0.4524</td>
<td>-0.5507</td>
<td>-0.7067</td>
<td>-0.5872</td>
</tr>
<tr>
<td>English</td>
<td>0.4061</td>
<td>0.5861</td>
<td>0.5209</td>
<td>0.6164</td>
</tr>
</tbody>
</table>

Scores that differed from monolinguals' accuracy are indicated on each graph with a plus or asterisk symbol (**\( p < 0.01 \), *\( p < 0.05 \), +\( p < 0.10 \)). The group effect for the CELF: Processing Word and Sentence Structure subtest was significant [\( F(5,92) = 4.83, p = 0.0006 \)]. Group comparisons revealed that the 1-3 and 4-6 bilingual groups performed similarly to monolinguals on the CELF subtest. The bilinguals who were exposed to English between 7-10, 11-13, and >16 years scored lower than monolinguals [7-10: pooled \( T(45) = 2.07, p = 0.0440 \), 11-13: pooled \( T(48) = 3.15, p = 0.0028 \), >16: pooled \( T(45) = 3.76, p = 0.0005 \)]. A group effect was also significant for the Saffran and Schwartz Grammaticality Judgment subtest [\( F(5,94) = 2.90, p = 0.0178 \)]. As can be seen in top right panel of Figure 4, the standard error associated with bilinguals exposed between 4 and 6

![Graphs showing performance on standardized tests](image-url)
years was large resulting in a significant Levene F for variability test [Levene F(50) = 4.55, p = 0.0373]. Examination of individual subject scores indicated this large variability was primarily due to one subject who's score was 50. A Welch separate T test (which does not assume equal population variances) indicated the difference in performance between monolinguals and this group was not significant [Welch separate T(13.4) = 1.55, p = 0.1444]. The scores of bilinguals exposed to English performance between monolinguals and this group was not significant [Welch separate T(13.4) = 1.55, p = 0.1444]. The overall group effect for the TROG was not significant, however, a group comparison revealed that the >16 group performed less accurately compared to monolinguals [pooled T(48) = 2.47, p = 0.0171]. Finally, there was a significant group effect for performance on the Carpenter Reading Span test [F(5,94) = 3.06, p = 0.0133]. Only the 11-13 group scored lower than monolinguals on the Carpenter Reading Span Test [pooled T(50) = 2.81, p = 0.0071]. There was trend for the >16 group to score lower [pooled T(47) = 1.71, p = 0.0929].

Regressions between total score correct on each of the standardized tests with age of exposure and years of experience revealed that performance on the Carpenter Reading Span Test was significantly correlated with years of experience with English and not with age of exposure (years: r = 0.4076, age: n.s.). Of the three tests that examined knowledge of English grammar, performance on the CELF test was the only one that resulted in significant correlations with exposure variables. For this test, the regressions with age of exposure and years of experience did not differ significantly (r = -0.4523 and r = 0.3220 respectively, t = 1.418, n.s.). As is evident from Figure 4, the three tests designed to test knowledge of English grammar (CELF, SSG, and TROG) showed different degrees of decline in proficiency for later learners. This is most likely due to the fact that each of the tests targeted different aspects of linguistic knowledge and the number of errors associated with different linguistic structures was not uniform. There were not enough items for the different linguistic structures for separate more detailed analyses.

Accuracy in Judgments of Experimental Sentences

Semantic Judgments

A significant group effect was found for the accuracy of judging semantic anomaly and control sentences [Group: F(5,95) = 6.86, p = 0.0000]. Semantic judgment accuracy of each bilingual group relative to monolinguals' performance is shown at the top of Figure 5. (Note that for all graphs in Figure 5, percentages correct were based on a total of 60 test sentences, 30 control and 30 violation.) Refer to Table 2 for example sentences (C#1 and V#1). As can be seen in Figure 5, with age of exposure greater or equal to 13 years, performance was similar to monolinguals. Only the >16 group performed less accurately than monolinguals on judgments of semantic sentences [Welch separate T(9.2) = 2.55, p = 0.0307]. Additional analyses revealed that the >16 group's decreased accuracy was due to failure to reject semantic violations [Welch separate T(9.2) = 2.57, p = 0.0298]. No differences in accuracy were found for judgments of the control sentences. Regression analyses revealed that age of exposure was a better predictor of performance than years of experience with English (r = -0.5142 and r = 0.2882, respectively, t = 2.9776, p < 0.01).

Phrase Structure Judgments

Group performances differed for judgments of sentences testing phrase structure rules [F(5,95) = 6.18, p = 0.0001]. Judgment accuracy of phrase structure constructs is shown for each bilingual group relative to monolinguals' performance at the top right of Figure 5. Refer to Table 2 for example sentences that tested knowledge of phrase structure (C#2 and V#2). With an age of exposure <4 years, subjects performed similarly to monolinguals. There was a trend for subjects exposed between 4 and 6 years to perform less accurately than monolinguals [Welch separate T(12.7) = 1.85, p = 0.0883]. With exposure ages >6 years, performance accuracy was reliably below that of monolinguals [7-10: pooled T(48) = 2.55, p = 0.0142, 11-13: Welch separate T(14.6) = 2.71, p = 0.0164, >16: Welch separate T(9.3) = 2.72, p = 0.0232]. Reduced judgment accuracy for the 7-10 and 11-13 groups was due to failure in rejecting sentences that violated phrase structure rules [7-10: pooled T(48) = 2.31, p = 0.0250, 11-13: Welch separate T(14.1) = 2.56, p = 0.0225]. Judgment accuracy of control sentences for these groups was similar to monolinguals' performance. The >16 group showed reduced accuracy in both accepting control sentences and rejecting violation sentences [control: Welch separate T(9.5) = 2.31, p = 0.0449, violation: Welch separate T(7.2) = 2.49, p = 0.0341]. Phrase structure judgment accuracy was significantly correlated with age of exposure, but not with years of experience (age: r = -0.4390, years: n.s.).

Specificity Constraint Judgments

Group performances were also found to differ for judgment accuracy of specificity constraint rules [F(5,95) = 12.70, p = 0.0000] (Fig. 5, bottom left). Refer to Table 2 for example sentences that tested knowledge of specificity constraint rules (C#3 and V#3). Decreased performance accuracy relative to monolinguals' accuracy was found for delays of exposure equal to or greater than 4 years [4-6: pooled T(51) = 2.30, p = 0.0253, 7-10: pooled T(48) = 2.92, p = 0.0053, 11-13: pooled T(51) = 4.75, p = 0.0000, >16: pooled T(48) = 7.36, p = 0.0000]. Regression analyses indicated that age of exposure to
English and years of experience with English were similar in predictive value for judgments of specificity constraint \((r = -0.6346 \text{ and } r = 0.5346, \text{ respectively, } t = 1.486, \text{ n.s.})\).

Informal subject reports revealed that rejection of sentences testing specificity constraint rules as ‘good English sentences’ were to some extent due to the fact that these sentences ended with a preposition (refer to Table 2, C*3, V*3). This basis for judgment was evident for all groups except the bilinguals with delays in exposure to English > 16 years (Table 6). The rejection of sentences based on their preposition ending appeared even in the monolingual group, as evidenced by the lower scores in judgments of this conditions’ control sentences compared to other control sentences. The effect appeared to be strongest for the 7-10 and 11-13 groups since accuracies within these groups were less for the control sentences than for the violation sentences. However, bilinguals with delays in exposure >16 years tended to accept the violation sentences as ‘good English sentences,’ which contributed most to their decreased accuracy. This confound in the behavioral results for this condition thus limits a strict interpretation of these data as reflecting knowledge of specificity constraint rules. Additional information regarding the extent to which different groups of subjects distinguish specificity constraint violations from controls may be gleaned from the ERP data (reported below).

Table 6. Mean Judgment Accuracy for Specificity Constraint Control and Violation Sentences for Monolinguals and Each of the Bilingual Groups

<table>
<thead>
<tr>
<th></th>
<th>Monolinguals (%)</th>
<th>Bilinguals (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (C*3)</td>
<td>Violation (V*3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-3</td>
<td>4-6</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>68</td>
</tr>
</tbody>
</table>

WeberFox and Neville 219
Subjacency Constraint Judgments

A group effect for judgment accuracy of sentences testing knowledge of subjacency constraint rules was significant \(F(5,95) = 4.74, p = 0.0007\). The judgment accuracy of each bilingual group relative to monolinguals’ performance is shown at the bottom right of Figure 5. Refer to Table 2 for example sentences testing knowledge of subjacency constraints (C#4 and V#4). For the subjacency constraint condition even the earliest exposed bilingual group performed less accurately than monolinguals and was due to decreased accuracy in accepting control sentences [Welch separate \(T(18.9) = 2.31, p = 0.0321\) and Welch separate \(T(16.8) = 2.72, p = 0.0146\), respectively]. Table 7 summarizes the judgment accuracies separately for the control and violation subjacency constraint condition. The lower overall score for the 4-6 group did not reach significance \((p = 0.14)\). However, similar to the 1-3 group, the accuracy level for acceptance of control sentences was below that of monolinguals [Welch separate \(T(12.6) = 2.37, p = 0.0542\)]. The 7-10 and 11-13 groups performed less accurately compared to monolinguals and similarly to the 1-3 group [7-10: pooled \(T(48) = 2.37, p = 0.0220\), 11-13: Welch separate \(T(14.8) = 2.12, p = 0.0518\)]. In both cases, reduced accuracy was due again to failure to accept control sentences [7-10: pooled \(T(48) = 2.46, p = 0.0175\), 11-13: pooled \(T(51) = 3.79, p = 0.0004\)]. Finally the >16 group’s accuracy was well below that of monolinguals [pooled \(T(48) = 5.33, p = 0.0000\)]. Judgment accuracy for both control and violation sentences was lower compared to monolinguals (control: pooled \(T(48) = 3.05, p = 0.0037\), violation: pooled \(T(48) = 4.28, p = 0.0001\)]. Regression analyses revealed a weak but significant relationship between age of exposure and judgment accuracy in the subjacency constraint condition \((r = -0.2909)\). The regression between performance and years of experience was not significant.

Event-Related Brain Potentials for Processing Experimental Sentences

Semantic Processing

Consistent with ERP patterns observed previously in monolinguals, the amplitude and latency of the N125 component in bilinguals were not affected by sentence type (control vs. semantic anomaly). Similarly, as previously reported for monolinguals, bilingual groups displayed increased negativity to semantic anomalies with a maximum at approximately 400 msec, characteristic of the previously described N400 (Kutas & Hillyard, 1980). Figure 6 illustrates the responses for monolinguals and each of the bilingual groups for electrode locations over left and right anterior temporal and parietal sites. Significant sentence type effects on the mean amplitude of the negativity occurring in the latency window of 300-500 msec were found for all bilingual groups except the

| Table 7. Mean Judgment Accuracy for Subjacency Constraint Control and Violation Sentences for Monolinguals and Each of the Bilingual Groups |
|---------------------------------|---------|---------|---------|---------|---------|---------|
|                                 | Monolinguals (%) | Bilinguals (%) |
|                                 | 1-3 | 4-6 | 7-10 | 11-13 | >16 |
| Control (C#3)                  | 95  | 86  | 81   | 90    | 88    | 87    |
| Violation (V#3)                | 79  | 72  | 84   | 69    | 68    | 52    |

>16 group for which a significant effect was found within a broader latency window of 300-600 msec. The 1-3 group showed a pattern of responses previously described for monolinguals, that is, an increased negativity to the violation that has maximal amplitude toward posterior regions with a slightly larger effect over the right hemisphere (the asymmetry was significant over frontal and anterior temporal electrode locations) [sentence type \(F(1,14) = 5.12, p = 0.0401\); sentence type \(\times\) hemisphere \(F(1,14) = 5.01, p = 0.0420\); sentence type \(\times\) hemisphere \(\times\) electrode \(F(4,56) = 7.15, p = 0.0008\)]. The responses of bilinguals exposed between 4-6 and 7-10 years were also characterized by increased negativity with maximal amplitude toward posterior electrode sites, however, no differences between hemispheres were detected [sentence type \(\times\) electrode \(F(4,48) = 4.00, p = 0.0466\) and \(F(4,36) = 16.05, p = 0.0000,\) respectively]. The responses of the 11-13 group displayed the same pattern with a main effect of sentence type [sentence type \(F(1,12) = 6.83, p = 0.0397\); sentence type \(\times\) electrode \(F(4,48) = 5.0, p = 0.0247\)]. The responses for the late learning group, significant for the broader latency window of 300-600 msec, were also characterized by increased negativity to semantic violations with larger mean amplitudes occurring at posterior electrode locations [sentence type \(\times\) electrode \(F(4,36) = 4.32, p = 0.0486\)].

The amplitude and latencies of the N400 to semantic anomalies, calculated as the difference between the response to the violations minus the controls, were compared to the monolinguals’ responses previously described (Neville et al., 1991). Although the average amplitude of the N400 appeared somewhat reduced for the latest learning group (>16 years) over all except the occipital electrode sites (refer to Fig. 6), group effects for the mean amplitudes, tested for both the 300-500 and 300-600 msec windows, were not significant. Also, no differences were found for the peak latency of responses for the 1-3, 4-6, and 7-10 bilingual groups compared to monolinguals, however, the peak latencies of the responses of the 11-13 and >16 groups were longer compared to monolinguals [11-13: group \(F(1,51) = 4.82, p = 0.0327\), and >16, electrode \(\times\) group \(F(4,192) = 6.70, p = 0.0007\)]. The longer latencies occurred for processing semantic violations, no group effects were found in
Figure 6. Averaged ERP waveforms over left and right anterior temporal and parietal sites for monolinguals and each of the bilingual groups. Responses indicated by dashed lines were elicited by violations of semantic expectation (underlined in example sentences). In this and subsequent ERP figures, traces in solid lines indicate responses to control words and negativity in plotted upward. The shading highlights the 300-500 msec poststimulus window in traces where increased negativity was elicited by semantic violations.

responses to the control words [violations: $F(5,95) = 2.26, p = 0.0383$]. Additional analyses of correct trials only were carried out for the 11-13 and >16 groups. Two subjects in the 11-13 group and three subjects in the >16 group were excluded due to an insufficient number of correct trials. The results were similar to the analyses for all trials. The mean latencies of the N400 for trials in which semantic violations were correctly judged were 407.46 msec for monolinguals, 431.42 msec for the 11-13 bilingual group, and 431.60 msec for the >16 bilingual group.

Regression analyses of the peak latencies of the N400 responses (averaged across electrode sites for each subject) revealed significant and linear relationships to age of exposure and years of experience ($r = 0.3068, p = 0.0153$ and $r = -0.2549, p = 0.0456$, respectively). The strength of the correlations was not significantly different for the two predictors ($t = 0.73$ n.s.).
In summary, similar to monolinguals all bilingual groups showed increased negativity to violations of semantic expectations compared to control sentences. The N400 peak latency was slightly longer, however, in the groups of subjects who were exposed to English between 11–13 and after 16 years of age.

**Processing Phrase Structure**

The ERPs elicited by phrase structure control and violation conditions are shown in Figure 7 for monolinguals and each of the bilingual groups. Two electrode locations, anterior temporal and the more posterior parietal site from over both hemispheres, are shown. The ERP response to phrase structure violations previously described in monolinguals was an enhancement of the N125 component at anterior regions of the left hemisphere, followed by increased negativity between 300 and 500 msec maximal over temporal and parietal regions, and a large sustained positivity over all electrode locations beginning around 500 msec.

**N125.** There was a significant interaction for the 1–3 bilingual group’s N125 peak amplitude [sentence type × hemisphere × electrode: F(4,56) = 3.11, p = 0.0229]. Unlike monolinguals, this interaction was due to an asymmetry in the N125 component that occurred over frontal electrode sites (F7 and F8) for the control sentence (LH > RH) but not the violation condition. There were no sentence type effects on the peak amplitude of the N125 responses in the 4–6 and 7–10 groups. For the 11–13 and >16 groups, the increase in N125 to the violations was enhanced over both hemispheres and this effect was largest over the right hemisphere [11–13: sentence type × hemisphere F(4,48) = 8.81, p = 0.0117 and >16: sentence type F(1,9) = 6.56, p = 0.0306, sentence type × hemisphere F(1,9) = 6.56, p = 0.0306].

The N125 effect measured in the ERP difference waves (violation minus control) was compared to those previously described in monolinguals. Significant group effects were found for measures of both latency and peak amplitude [latency: group F(5,95) = 4.56, p = 0.0009, amplitude: group F(5,95) = 2.67, p = 0.0267]. Group contrasts between monolinguals and each of bilingual groups revealed that these group effects were due to longer latencies and greater enhancement of the N125 component in the difference ERPs of the 11–13 and >16 groups [11–13, latency: group F(1,51) = 12.06, p = 0.0011, amplitude: group F(1,51) = 4.83, p = 0.0325, and electrode × group F(4,204) = 4.32, p = 0.0262; >16, latency: group F(1,48) = 4.74, p = 0.0345; amplitude: group F(1,48) = 5.66, p = 0.0214, electrode × group F(4,192) = 3.75, p = 0.0398]. Comparison of monolinguals and the >16 group’s N125 amplitudes measured in the difference ERPs also resulted in marginal group by hemisphere interactions [amplitude: hemisphere × group F(1,48) = 3.88, p = 0.0548, hemisphere × electrode × group F(4,192) = 2.88, p = 0.0545]. Regression analyses (which included all of the bilingual subjects) revealed that years of experience with English may be a better predictor of the latency of the N125 response over the left hemisphere (years: r = −0.3928, age: n.s.). Similar predictive values were found for locations over the right hemisphere (years: r = −0.4497, age: r = 0.3829, t = 0.88, n.s.). The changes in peak amplitude of the N125 were not linearly related to age of exposure or years of experience.

**P250.** The effect of sentence type (control vs. phrase structure violation) on the 180–300 msec positive component was analyzed. In monolinguals previously studied, no effect of sentence type was found for this component. For both the 1–3 and 4–6 bilingual groups, the positivity was enhanced over anterior electrode sites for the violation compared to the control [1–3: sentence type × electrode: F(4,56) = 4.73, p = 0.0235; 4–6: sentence type: F(1,12) = 12.02, p = 0.0047]. No significant effects were found for the 7–10 and 11–13 groups. Finally, for the >16 group, a decrease in positivity was associated with the violation condition [F(1,9) = 5.24, p = 0.0478]. An analysis of group effects on amplitude and latency measured in the difference waves (violation minus control) did not yield significant differences between groups for the P180–300 response.

**N300–500.** In the 300–500 msec window, the responses of the bilinguals exposed to English between 1 and 3 years displayed a similar pattern to that previously reported for monolinguals: i.e., an increased negativity to phrase structure violations beginning around 300 msec, which was maximal over temporal and parietal sites over the left hemisphere. This sentence type effect approached significance for the 300–500 msec window (p < 0.07), and was significant for a slightly sharper window of 300–450 msec [sentence type F(1,14) = 9.44, p = 0.0083; sentence type × hemisphere F(1,14) = 6.04, p = 0.0276; sentence type × hemisphere × electrode F(4,56) = 3.22, p = 0.0310]. The responses of the 4–6 group also showed a negative enhancement over the left hemisphere that was significant in the 300–500 latency window [sentence type × hemisphere F(1,12) = 6.84, p = 0.0226]. The sentence type effect for phrase structure violations was significant for the 7–10 group, which showed a similar pattern within the sharper latency window [300–500 msec window (p < 0.09), 300–450 msec window: sentence type × hemisphere F(1,9) = 5.35, p = 0.0460]. The responses of the 11–13 group displayed an increased negativity to violations [sentence type F(1,12) = 5.46, p = 0.0377]. For this group, a hemispheric asymmetry approached significance in the sharper latency window [300–450 msec sentence type × hemisphere F(1,12) = 4.68, p = 0.0514]. The >16 group’s responses also displayed in-
Age of Exposure: 1-3 yrs

Age of Exposure: 4-6 yrs

Age of Exposure: 7-10 yrs

Age of Exposure: >16 yrs

Figure 7. Averaged ERP waveforms over left and right anterior temporal and parietal sites for monolinguals and each of the bilingual groups. Responses indicated by dashed lines were elicited by phrase structure violations (underlined in example sentences). The shading highlights the 300-500 msec poststimulus window in traces where increased negativity was elicited by phrase structure violations.

creased negativity to violations within the 300-500 msec window, however, the effect occurred over both hemispheres [sentence type \( F(1,9) = 13.55, p = 0.0051 \)]. In fact in contrast to each of the other groups, sentence type effects on the peak amplitude tended to larger over the right hemisphere [sentence type \( \times \) hemisphere \( F(1,9) = 3.66, p = 0.0882 \)]. Analyses of the difference waves revealed a significant group effect for the peak amplitude of the responses in the 300-500 msec window [hemisphere \( \times \) group \( F(5,95) = 2.55, p = 0.0484 \]). Group contrasts revealed that the >16 groups’ responses displayed increased negativity over the right as well as the left hemisphere unlike the monolinguals’ responses, which were lateralized over the left hemisphere [hemisphere \( \times \) group \( F(1.48) = 7.18, p = 0.0101 \)]. Figure 8 illustrates the reversed asymmetry in the response of the
Figure 8. The asymmetry in the response amplitude over the right hemisphere for the >16 group measured in the 300-500 msec window of the difference waves.

P500-700. A late (500-700 msec window) sustained positivity to phrase structure violations previously described for monolinguals was found in the 1-3, 4-6, and 11-13 bilingual groups [effect of sentence type on mean amplitude for each group: F(1,14) = 11.21, p = 0.0048, F(1,12) = 10.25, p = 0.0076, F(1,9) = 0.1804, p = 0.0022, respectively]. The main effect of sentence type was not significant for either the 11-13 or >16 group. However, for the >16 group, there was an interaction between sentence type and hemisphere [F(1,9) = 6.57, p = 0.0305] indicating an asymmetry for the control condition (mean amplitude left: -3.1 µV, right: -1.6 µV), which did not occur for phrase structure violations (mean amplitude left: -2.57 µV, right: -2.45 µV).

Group analyses of the difference waves revealed a main effect on the mean amplitude of P500-700 [group F(5,95) = 4.10, p = 0.0021]. The amplitude of the 11-13 and >16 groups was significantly reduced compared to that of monolinguals [11-13 group F(1,51) = 8.90, p = 0.0044, >16 group F(1,48) = 10.04, p = 0.0027]. A hemisphere by group interaction was also significant for the contrast between the >16 group and monolinguals due to an asymmetry in the difference ERPs of the >16 group not present for the monolinguals [F(1,48) = 6.42, p = 0.0146]. Regression analyses of P500-700 amplitude averaged across all electrode locations were significant over the right hemisphere only. Age of exposure (r = -0.5065) and years of experience (r = 0.4737) with English were similar in predictive value for the P500-700 amplitude over the right hemisphere.

P700-900. Mean amplitudes of the positive shift over an additional latency window were measured to determine whether the later learning groups’ responses contained significant positive shifts that occurred later in the epoch (well into the next word in the sentence). The responses of the monolinguals showed a large positive shift that was greater for the violations; this effect was larger over the right hemisphere, and over posterior electrode sites [sentence type: F(1,39) = 47.07, p = 0.0000, sentence type × hemisphere: F(1,39) = 4.60, p = 0.0384, sentence type × electrode site: F(4,156) = 31.51, p = 0.0000]. The responses of 1-3, 4-6, 7-10, and 11-13 bilingual groups all showed a greater positive shift for processing the violations [F(1,14) = 15.14, p = 0.016, F(1,12) = 49.12, p = 0.0000, F(1,9) = 27.18, p = 0.0006, F(1,12) = 7.65, p = 0.0171, respectively]. This effect was significantly larger over the right hemisphere in the 1-3 group [F(1,14) = 8.31, p = 0.0120]. By contrast the responses of the >16 group did not show a main effect of sentence type. For this group there was an interaction between sentence type and hemisphere [F(1,9) = 8.92, p = 0.0153] indicating greater asymmetry to the control sentences.

There was considerable variability in the distribution of the P700-900 enhancement to violations in the different groups. Response amplitudes were maximal at frontal and parietal electrodes for the 1-3 group, frontal for the 4-6 group, and parietal for the 11-13 group [sentence type × electrode site: 1-3 group: F(4,56) = 3.51, p = 0.0327, 4-6 group: F(4,48) = 3.80, p = 0.0364, 11-13 group: F(4,48) = 5.05, p = 0.0249]. There were no sentence type by electrode site interactions for the 7-10 or >16 groups.

Analyses of the difference waves revealed group effects on the P700-900 mean amplitude, hemisphere by group effects, and electrode site by group interactions [group F(5,95) = 2.87, p = 0.0187, hemisphere × group F(5,95) = 0.323, p = 0.0097, electrode site × group F(20,380) = 2.35, p = 0.0115]. The results of group contrasts indicated that the >16 positive amplitude effect was significantly reduced compared to monolinguals [group F(1,48) = 9.57, p = 0.0033]. A hemisphere by group interaction was also significant for the contrast between the >16 group and monolinguals due to an asymmetry in the difference ERPs of the >16 group [F(1,48) = 9.09, p = 0.0041]. For the >16 group, the mean amplitude of the effect in the difference ERPs was reduced compared to the other groups over both left and right hemispheres [right: F(1,48) = 14.16, p = 0.0005, left: F(1,48) = 3.18, p = 0.0808]. Comparisons between monolinguals with four of the five bilingual
groups resulted in significant electrode site by group effects \[1-3: F(4,212) = 3.93, p = 0.0218, 4-6: F(4,204) = 3.07, p = 0.0530, 7-10: F(4,192) = 3.25, p = 0.0458, \text{and} >16: F(4,192) = 4.78, p = 0.0118].

In summary, the ERPs of the 1-3 bilingual group did not display the early (125 msec) enhanced negativity previously described for monolinguals. However, their responses did show increased negativity between 300 and 500 msec, which, like monolinguals, was maximal over left temporal and parietal sites, and was followed by a broadly distributed late positive shift beginning around 500 msec. With delays of 4-6 and 7-10 years, the early N125 enhancement elicited by phrase structure anomalies was not present, and the later (300-500 msec) increased negativity was most prominent over the left hemisphere but did not display significant electrode site effects. The late positive shifts observed in the ERPs of the 4-6 and 7-10 groups were similar to monolinguals. The ERPs of the 11-13 group displayed a significant N125 enhancement, however, unlike monolinguals and earlier exposed bilinguals, the effect was larger over the right hemisphere. Further, there was no late positive shift between 500 and 700 msec, but, rather, a very late increase in positivity beginning around 700 msec. Finally, despite the fact that the >16 bilingual group had a similar number of years of experience with English as the 11-13 group, they showed even more marked alterations in their ERPs to phrase structure violations. The responses of the late learning group displayed no left hemisphere specialization for any of the components. The enhancement of the N125 was larger over the right hemisphere and the increased negativity between 300 and 500 msec was bilateral and tended to be larger over the right hemisphere. Further, there was no late positive shift in the responses of the >16 bilingual group.

### Processing Specificity Constraints

Monolinguals' responses to specificity constraint violations were characterized by early and sustained negativity over left anterior regions (Neville et al., 1991). The monolinguals as well as the bilingual groups' ERPs are illustrated in Figure 9.

**N125**. The very early left hemisphere enhancement of the N125 component (found in monolinguals) was not significant for the 1-3 bilingual group. An enhancement of the N125 over the left hemisphere was found to be marginal in the responses of 4-6 group and significant in the responses of the 7-10 group [4-6 group: sentence type × hemisphere \(F(1,12) = 4.59, p = 0.0530\), 7-10 group: sentence type × hemisphere \(F(1,9) = 11.00, p = 0.0090\)]. The effect of specificity constraint violations in the 11-13 bilingual group was enhancement of the N125 over frontal and anterior temporal sites over both hemispheres [sentence type × electrode \(F(4,48) = 8.29, p = 0.0035\)]. For the >16 group, there were no significant effects of sentence type on N125. Group by sentence type effects for the mean amplitude of N1 were significant for the monolingual and >16 group contrast (sentence type × hemisphere × group \(F(1,48) = 6.50, p = 0.0140\), sentence type × electrode × group \(F(3,36), p = 0.0512\)). No group differences were detected in the difference wave analyses and regression analyses indicated no linear relationship with age of exposure or years of experience.

**P180-300.** The responses of the 1-3 and the 4-6 bilingual groups displayed a slight decrease in peak positivity (i.e., relative increase in negativity) over the left hemisphere and slight increase in positivity over the right hemisphere for specificity constraint violations resulting in significant sentence type by hemisphere effects \(F(1,14) = 6.56, p = 0.0244\); and \(F(1,12) = 8.34, p = 0.0136\). Sentence type effects tested separately for each hemisphere were not significant. A sentence type by hemisphere interaction was also significant for the 7-10 group, however, in their responses, the sentence type effect over the right hemisphere (i.e., increase in positivity) was significant [sentence type: \(F(1,9) = 11.24, p = 0.0085\), right hemisphere: sentence type \(F(1,9) = 32.64, p = 0.0003\)]. The 11-13 group's responses to specificity constraint violations in the 180-300 window were characterized by a slight increase in negativity at frontal and anterior temporal electrode sites and an increase in positivity at the other locations [sentence type × electrode \(F(4,48) = 4.25, p = 0.0448\)]. No effects of sentence type on the P180-300 were found for the >16 group. Group differences were not detected in the group ANOVA analyses.

**N300-500.** At the 300-500 msec latency window, the mean amplitude measures of the 1-3, 4-6, and 7-10 groups showed a pattern similar to that of monolinguals: increased negativity to violations over the left hemisphere [1-3: sentence type × hemisphere \(F(1,14) = 12.35, p = 0.0035\); 4-6: \(F(1,12) = 7.71, p = 0.0168\), 7-10: sentence type × hemisphere × electrode \(F(4,36) = 3.27, p = 0.0342\)]. Effects on the mean amplitude in the responses of the 11-13 group indicated trends for increased negativity over the left hemisphere [sentence type × hemisphere \(F(1,12) = 3.65, p = 0.0803\)]. There were no sentence type effects on the mean amplitude of the responses of the >16 group, however, the peak amplitude measures indicated a tendency for a left hemisphere increase in negativity to violations (sentence type × hemisphere \(F(1,9) = 3.46, p = 0.0956\)).

**N500-700.** Finally, consistent with previous findings in monolinguals, an asymmetrical increase in the mean amplitude of the sustained negativity in the 500-700 msec latency window was found for the 1-3, 4-6, and 7-10 groups [sentence type × hemisphere \(F(1,14) = 13.14, p = 0.0028\); \(F(1,12) = 16.43, p = 0.0016\); \(F(1,9) = 7.87\).
The shading highlights the 300-500 msec poststimulus window in traces where increased negativity was elicited by specificity constraint violations.

$p = 0.0205$, respectively]. The average response of the 11–13 group to specificity constraint violations reflected a tendency for an asymmetrical increase in negativity for the 500–700 msec window (see Fig. 9) [sentence type $\times$ hemisphere $F(1,12) = 2.96, p = 0.1112$]. For the >16 group there was no effect of sentence type found for the 500–700 msec window.

In summary, the ERPs elicited by specificity constraint violations in bilinguals exposed to English prior to age 11 were similar to monolinguals responses (with the exception of the N125 for the 1–3 group). With ages of exposure at 11–13 years, the early components (N125 and P180–300) show reduced asymmetry, and sentence type effects approached significance only for the later components (N300–500 and N500–700). Finally, with exposures to a second language >16 years, marked differences in the responses were noted, despite the fact that the number of years of experience with English is
similar to the 11-13 group. No reliable differences between the control and violation sentences were found for the >16 group at any of the latency windows tested.

**DISCUSSION**

Our findings suggest that language proficiency and cerebral organization for language processing are altered by delays in exposure to language, and are consistent with the hypothesis that postnatal maturational processes may underlie critical period phenomena associated with aspects of cognitive development, in this case language learning. Moreover, compatible with previous behavioral studies (Johnson & Newport, 1989; Curtiss, 1992; Newport, 1988; Oyama, 1976; Seliger et al., 1975; Mayberry & Fischer, 1989; Flege & Fletcher, 1992; Oyama, 1982; Patkowski, 1980; Johnson & Newport, 1991; Liu et al., 1992), our results on language proficiency of bilinguals who were exposed to English at different ages indicate that age of acquisition is a crucial variable for proficiency attained in a second language and further that different aspects of language display different critical periods. In agreement with our behavioral findings, our ERP results further suggest that the development of specific subsystems for language processing depends in part on the age of exposure to a second language. Aspects of semantic processing studied here appear least vulnerable to delays in second language exposure. By contrast, development of left hemisphere specialization linked with syntactic processing appears to be dependent on the timing of language experience and differs for different types of syntactic processing (see summary Fig. 10).

**Behavioral Findings**

The subjects within each of the age of exposure groups were remarkably consistent in their reports of self-rated proficiency in English and in Chinese. Those subjects who were exposed to English prior to age 11 consistently rated their English proficiency at near perfect and reported that they comprehended and spoke Chinese "well" or "sufficiently." The subjects who were exposed at around the time of puberty (11-13 years) reliably rated themselves equally proficient in Chinese and English, reporting that they comprehended, spoke, and read Chinese and English "well." This group of subjects was closest to being "balanced" in the two languages and it is interesting that they are the only group that did not consider their language proficiency as perfect in either of the two languages. Finally, despite the fact that the latest learning group (>16 years) had a similar number of years experience with English as those subjects exposed between 11 and 13 years of age, this group rated their English proficiency lower (in the sufficient-well range), and their Chinese proficiency as perfect. Thus for our subjects, differences in linguistic proficiency with different ages of exposure to English were reported for both first and second languages and, further, differences

---

Figure 10. Summary of the age of exposure effects on the ERPs and behavioral responses elicited by the experimental sentences. Transverse bars indicate the age of exposure to English at which decreases in linguistic proficiency and changes in ERP patterns were observed. For semantic processing, altered patterns occurred with delays greater than 11 years and ERP changes were characterized by a slight shift in latency. In contrast, altered patterns for processing syntactic constructs were observed with shorter delays in exposure and ERP changes were characterized by amplitude and/or distribution differences. (Note: As explained in the Results section, the subjacency constraint effects on ERPs were uninterpretable due to variable baseline effects across groups and, therefore, are not indicated on this graph.)
in the relative proficiency in Chinese and English varied between the groups. In summary, the self-rated proficiency scores suggest that immersion in a second language before 11 years results in high levels of competence in the language of the new environment and decrement in proficiency of the first language. With immersion in a new language around the time of puberty (11-13 years) there appears to be a substantial but less dramatic decline in competence in the first language, and levels of proficiency in the language of the new environment are reduced compared to those of younger learners. Finally, immersion in a new language after the age of 16 years appeared to have little or no impact on the proficiency of the first language, and the levels of proficiencies attained in the second language were markedly lower than those of bilinguals who were exposed to English at a younger age.

The behavioral responses to the experimental sentences in the current study indicated that much shorter delays in language exposure affected judgment accuracy for syntactic rules (<7 years), and only delays >16 years affected detection of semantic anomalies. Our findings indicate that in comparison to the type of semantic knowledge indexed by the N400, earlier language exposure may be necessary for individuals to achieve native-like proficiency in grammatical aspects of language.

Within syntactic processing, linguistic proficiencies for different types of constructs were affected differently by delays in exposure. Various strengths in the linear relationships between age of exposure and syntactic linguistic proficiency were found (range = -0.29 to -0.63). Previous work has also shown the strength of correlation coefficients depends on the type of linguistic structure tested (Johnson & Newport, 1989). For example, in the Johnson and Newport study, correlations between age of arrival in the United States and linguistic proficiency was as low as 0.29 for competence with third person singular and as high as 0.79 for knowledge of past tense.

The findings from our study indicated that for phrase structure processing, there was a trend for decreased proficiency for delays as early as 4-6 years, with reduced accuracy and similar levels of performance for delays between 7-10 and 11-13, and with maximum decline for delays >16 years.

A different pattern was found for specificity constraint processing. Our results suggested that compared to accuracy in judgments of phrase structure, there is an earlier (4-6 years) and more gradual decline in proficiency (r = -0.6346) for processing specificity constraints. However, as previously noted, sentences in the specificity constraint condition ended in a preposition. Subjects reported that they often judged sentences as ungrammatical based on the preposition endings. Thus interpretation of these behavioral findings was difficult. This confound might account for the differences between the behavioral and ERP findings (discussed below) for this condition. The ERP evidence suggests that the changes in specificity constraint processing related to increases in age of exposure are less gradual than the behavioral results indicate. It will be important for future studies to take into account the potential bias that sentences with preposition endings introduce for grammaticality judgment tasks.

Finally, the accuracies for judgments of subjacency constraints were decreased in even the earliest exposed bilinguals, with a fairly flat level of performance for the subjects exposed to English prior to 14 years. The increased error rates occurring for these bilinguals were found only for judgments of the control sentences. In contrast, late learners (>16 years) showed a sharp decline in accuracy for both control and violation cases. Johnson and Newport (1991) have also found decreased accuracy for subjacency in early second-language learners; however, the decline was more gradual with age of exposure. It is possible that the degree of overall difficulty in the structures of the sentences was different in the two studies, with ours being somewhat more difficult.

One question that arises from the behavioral findings is whether the decline in judgment accuracy for specificity and subjacency constraints in even very early learners of English may be attributable to interference from the first language. Wh-movement, which is constrained by specificity and subjacency rules in English, is not utilized in Chinese. However, a recent study with Spanish-English bilinguals suggests that the decline in performance, at least in learners exposed to their second language as adults, cannot be entirely attributed to the absence of wh-movement in the first language (Johnson, 1996). Nonetheless, late-learning Spanish-English bilinguals in that study did perform better (determined using t test statistics) than the late-learning Chinese-English bilinguals that had been tested by the same experimenter using the same test materials in a previous study (Johnson, 1996; Johnson & Newport, 1991). Those findings suggest that additional influences related to the nature of the first language were exerted on the accuracies in judgments of subjacency constraint rules. Because only late second-language learners were tested in that study, it will be important to determine whether and to what degree decreased accuracy for subjacency judgments occur for early second-language learners whose first language utilizes wh-movement (e.g., Spanish).

Age of exposure effects were also seen in the results of standardized tests assessing knowledge of English grammar; however, the age of exposure that was associated with decreased accuracy varied with the different tests used. This is likely due to the fact that the different tests focused on different linguistic constructs and were perhaps different in their degree of difficulty, which may have resulted in different ceiling effects. Our findings
and the results of others (Johnson & Newport, 1989) strongly indicate that knowledge of different types of linguistic constructs is differently affected by delays in exposure to a language. Thus the use of standardized tests that assess knowledge of different sets of linguistic constructs will, not surprisingly, yield different results in terms of indicating decreased linguistic proficiency.

Finally, the behavioral findings of the present study also suggest that there is a stronger link between age of exposure and specific linguistic knowledge than there is between age of exposure and another type of linguistic-related skill, in this case a skill related to working memory capacity [as tested by the Carpenter Reading Span Test (Danceman & Carpenter, 1980; Carpenter & Just, 1989)]. For the working memory task, years of experience was a better predictor of proficiency than age of exposure to the second language. These preliminary findings suggest that maturational processes may not impose constraints across more general, memory-based types of linguistic-related skills. Consistent with this idea are our findings that indicated that detection of semantic anomalies was least affected by delays in language exposure. These results suggest that compared to syntactic skills, semantic/pragmatic skills assessed in this study may be more closely related to general associative or temporal coincidence learning. This type of semantic knowledge is strongly dependent on learning referential and pragmatic vocabulary, which typically continues undiminished throughout life. It is possible, however, that other types of semantic skills may be more constrained by maturational processes. Such a result has been reported by Mayberry and Eichen (1991). More work is needed to clarify this issue.

**ERP Findings**

**Semantic Processing**

**Latency:** A shift in N400 peak latency (approximately 20 msec) was observed for subjects who were exposed to English >10 years. The degree of latency shift was similar for the 11-13 and >16 groups and age of exposure and years of experience were similar in predictive value for the N400 latency. The similar shift in latency for the 11-13 and >16 groups occurred even though they did not show the same level of accuracy in detecting semantic anomalies, only the >16 group performed less accurately than monolinguals. These results suggest that the latency shift is not closely tied to level of accuracy, but rather may reflect a slight slowing in processing. Previous investigators have linked the N400 latency to fluency (Ardal et al., 1990; Kutas & Kluender, 1993). We did not measure fluency directly, however, our measure of self-rated speaking proficiency was consistent with the idea that “fluency” may be important in determining N400 latency. The bilinguals who were exposed to English <11 years rated themselves more proficient in English than Chinese; further they rated their English proficiency as nearly perfect. These subjects' N400 peak latencies were similar to monolinguals'. In contrast, both the 11-13 and >16 groups indicated reduced proficiency in speaking English and both of these groups showed a later N400 peak latency. On the other hand, in view of the fact that the >16 group's speaking proficiency ratings were reduced more than the 11-13 group, it is interesting that the latency shift was similar for these two later learning groups. Previous results suggest that N400 latency shift is graded depending on fluency (Ardal et al., 1990). However, these two groups' ratings on their English reading proficiency did not differ significantly, and because our task involved reading English sentences, this measure may be a more appropriate predictor for the N400 latency in this study.

**Amplitude.** The amplitudes and distributions of the N400 elicited in bilinguals were not significantly different from the responses of monolinguals. Other studies have also shown that brain systems mediating semantic processing of this type (processes indexed by N400) appear to be quite robust in populations with altered early language experience (Neville et al., 1992). Ardal et al. (1990) reported a “tendency” for reduced N400 amplitude in subjects who used their second language least, but only at anterior electrode sites. In contrast, Kutas and Kluender (1993) reported smaller N400 amplitudes overall for the less fluent language of bilinguals. It is not clear why our results are not consistent with either of these previous findings in bilinguals. More work is needed to determine the exact nature of the N400 amplitude reduction in bilinguals.

The bilingual groups with ages of exposure >3 years did not show the slight asymmetry in the N400 (amplitude over right hemisphere greater than over left), which has often been reported for monolinguals (Kutas, Van Petten, & Besson, 1988). However, as reviewed by Kutas et al. (1988), the finding of an N400 right hemisphere asymmetry has been quite variable in the literature. The variable and evanescent nature of the N400 asymmetry may explain the absence of the hemispheric effect in some of the bilingual groups in the current study. Another possible explanation is the effect of handedness on the N400 asymmetry (Kutas et al., 1988). As noted in the methods (see Note 3), we could not be sure of the effects of familial history of handedness, and to a lesser degree subject handedness due to cultural practices. Ardal et al. (1990) found a trend for the N400 amplitude to be larger over the left hemisphere (parietal site) in bilinguals, despite the fact that the first and second languages varied among the bilinguals (only 3 subjects spoke Chinese). However, they did not include a report of familial histories of handedness, so it is difficult to assess whether the N400 asymmetry in bilinguals is
characteristically different from monolinguals. More work is needed to clarify this issue.

Syntactic Processing

In contrast to the relative stability of the N400 indices of semantic processing, our results suggest that for syntactic processing, the actual presence and distribution of ERPs may be altered by delays in language exposure. These findings are consistent with ERP evidence in deaf individuals that suggests cerebral subsystems linked to grammatical processing are more susceptible to alterations in early language experience (Neville et al., 1992; Neville et al., 1996). The ERP evidence of the present study further suggests that the cerebral subsystems mediating different syntactic constructs are affected differently by delays in exposure.

Phrase Structure. The ERPs of the 1–3 bilingual group did not display the early (125 msec) negativity previously described for monolinguals. However, similar to monolinguals, this group's increased negativity between 300 and 500 msec was maximal over left temporal and parietal sites and was followed by a broadly distributed late positive shift beginning around 500 msec. The responses of the 4–6 and 7–10 bilingual groups did not show the early (N125) left anterior negativity previously observed in monolinguals. Their later increased negativity between 300 and 500 msec was lateralized over left hemisphere, but showed no significant electrode site effects. The late positive shifts elicited by phrase structure violations for the 4–6 and 7–10 groups were similar to those previously observed in monolinguals. With 11–13 years delay in second language of exposure, the early enhanced negativity (N125) was larger over the right hemisphere and the asymmetry in the 300–500 msec window was not significant. Further, there was no late positive shift between 500–700 msec, but rather a very late increase in positivity beginning around 700 msec. Finally, despite the fact that the >16 bilingual group had a similar number of years of experience with English as the 11–13 group, it showed even more marked alterations in its ERPs to phrase structure violations. The responses of the late learning group displayed no left hemisphere specialization for any of the components. The enhancement of the N125 was larger over the right hemisphere and the increased negativity between 300 and 500 msec was bilateral and tended to be larger over the right hemisphere. Further there was no late positive shift in the responses of the >16 bilingual group.

Thus, none of the bilingual groups displayed an early (125 msec), left anterior enhanced negativity, which was previously elicited in monolinguals. It is not clear why this effect was not present even in the earliest exposed bilingual group; however, upon inspection of the original monolingual data, it was noted that the effect was quite variable. It is possible that this effect is difficult to obtain in groups with fewer subjects (compared to the monolingual group of 40 subjects). Interestingly, the later learning bilingual groups (11–13 and >16) did show early but bilateral increases in negativity for phrase structure violations. It is unlikely that these groups actually processed the phrase structure violations faster than earlier learning groups. One possibility is that the very early differences in the ERPs of the later learning groups were delayed effects in response to the differences in word classes of the preceding words in the sentences (i.e., nouns in control sentences and possessives in the sentences containing phrase structure violations, refer to Table 2). All of the other groups (monolingual and bilinguals) clearly displayed differences to these word types prior to the onset of the phrase structure violation. In contrast, such differences were not apparent in the 600 msec prestimulus window of the ERPs elicited in the two later learning groups.

In summary, even with short delays in language exposure, decreased accuracy (4–6 years) and altered ERPs (1–3 years) are observed for phrase structure processing. With increased delays in second language exposure, phrase structure processing is associated with even further reduced judgment accuracies and reduced asymmetry in sentence type effects on the N125 and N300–500 components and the absence of the 500–700 msec positive shift. The changes in ERP asymmetries suggest that for specific types of linguistic processing, increased delays in language exposure may be associated with reduced specialization in left hemisphere language processing subsystems and may include increased right hemisphere involvement. The late positivity (500–700) of the phrase structure response in monolinguals, also observed in earlier (<11 years) exposed bilinguals, is similar to the widely distributed syntactic positive shift described by Osterhout and Holcomb (1992, 1993). This late positivity (500–700) has been hypothesized to reflect an attempt to recover the meaning of the sentence (Neville et al., 1991). With exposures >10 years, the amplitude of this response was reduced, suggesting that later learners may not attempt to "patch up" the anomalous sentence. It is also possible that an attempt to recover the meaning of the sentence was slower in these later learning groups and would have been detected at a later latency in the ERP epoch. The results of the analyses of the 700–900 msec latency window suggest that the 11–13 group does show a later positive shift and thus may in fact be slower in attempts to recover the sentence. However, for the >16 bilingual group, even the very late positivity in the 700–900 was significantly reduced relative to monolinguals, suggesting that perhaps these late learners are either much slower in attempts to recover the sentence or do not do so in the same manner.

Specificity Constraint. The behavioral results for this condition suggested that specificity constraint process-
ing may be impacted by delays of only 4-6 years; however, as discussed above, these findings are difficult to interpret due to subjects basing their judgments on the preposition endings of the sentences in this condition. In contrast, the ERPs that are time locked to the onset of violations (controls) in specificity constraint rules suggest that bilinguals exposed to English prior to 11 years processed these sentences similarly to monolinguals. Their ERPs displayed a left hemisphere specialization characterized by increased and sustained negativity over left anterior electrode sites. The responses of the 4-6 and 7-10 groups displayed a significant increase in negativity at the very early (125 msec) latency window typical of monolinguals. This effect was not significant for the 1-3 group, however. It is not clear how to interpret these findings. The \( N_{125} \) was enhanced in the ERPs of the 11-13 group, however, their responses displayed less asymmetry, suggesting that, similar to phrase structure processing, with increased delays in language exposure there is reduced left hemisphere specialization, and possibly greater right hemisphere involvement. The ERPs elicited in the bilinguals who were exposed to English after the age of 16 years (but who had the same number of years of experience with English as the 11-13 group) did not reliably distinguish between controls and violations of specificity constraint rules. There were no significant effects measured for this group in any of the latency windows tested.

**Language Learning after Puberty**

Another aspect of the present results is that subjects who arrived in the United States after the age of 16 years consistently performed less accurately than monolinguals and consistently displayed the most extreme differences in ERP patterns compared with those observed in monolinguals. These subjects who were exposed to English after puberty demonstrated decreased performance relative to monolinguals on each of the standardized tests and for each of the experimental test sentences. This group of subjects showed a unique pattern in their self-rated proficiency scores. They were the only group who rated themselves more proficient in Chinese compared with English. Also, this group of subjects was the only one that reportedly felt more comfortable using Chinese than English. The pattern of behavioral and ERP findings for this group is different from even the group of subjects who arrived in the United States around the time of puberty (11-15 years) despite the fact that both groups had a similar number of years of experience with English (refer to Table 1). Our findings are compatible with other behavioral results that also suggest maturational processes may affect language learning in a markedly different way in the postpuberty time period (Johnson & Newport, 1989).

In summary, our findings suggest that delays in exposure to language even in early childhood impact linguistic proficiency and the development of the functionally specialized processing subsystems that underlie these skills. Further these results are compatible with the idea suggested by Lenneberg's original hypothesis (Lenneberg, 1967) that puberty may mark a significant point in development for language learning capacity and for the reorganizational capability of the specialized cerebral systems important for linguistic processing.

**METHODS**

**Subjects**

Sixty-one Chinese-English bilingual speakers (33 females) between the ages of 18 and 44 years participated in the experiment. Participants were recruited through the University of California, San Diego and The Salk Institute for Biological Studies. Subjects spoke either Mandarin (\( N = 28 \)), Yue (Cantonese, \( N = 17 \)), Mandarin and Yue (\( N = 8 \)), or Mandarin and Min (southern dialect Taiwanese, \( N = 7 \)). All subjects were right-handed according to self-report and as measured on the Edinburgh Inventory for assessment of handedness and 6 of 61 subjects reported a family history of left-handedness (Oldfield, 1971). Subject grouping was based on the age at which subjects began using English. For those subjects who either were born in the United States or arrived before school age, grouping was based on the subjects' reports of their language use histories, that is, the age that subjects reportedly began using English. For subjects who arrived in the United States after school age, grouping corresponded to the age at which they moved to the United States. The grouping ages were 1-3, 4-6, 7-10, 11-13, and >16 years. All subjects included in the study had lived in the United States for a minimum of 5 years and attended school or worked during that period. A summary of subject characteristics for each group is shown in Table 1. Note that for each group, the mean number of years of experience with English is also shown in Table 1. As would be expected, the number of years of experience was correlated with age of exposure to English (\( r = 0.7764 \)); however, it should be noted that the number of years of experience with English did not differ for the two later learning groups, 11-13 and >16 years [\( F(1,21) = 0.09, p = 0.7645 \)].

**Stimuli**

The stimuli were the set of 240 sentences that was employed by Neville et al. (1991) in a previous study of monolingual English speakers. Half of the sentences contained medial words that violated semantic expectations or syntactic rules including phrase structure, specificity constraint, and subjacency constraint (see Table 2). The remaining 120 sentences were semantically and syntactically appropriate. Comparison words for control and violation sentences were at the point of linguistic deviation in the anomalous sentences. The comparison words,
to which ERPs were recorded, are set in italics in the example sentences in Table 2. Note that for the specificity constraint condition, two control sentences were available for comparison, C#1 (same control used for semantic violations) and C#3. For purposes of comparing changes in ERPs due to the specificity constraint violation, C#1 proved more appropriate because the words preceding the comparison points were identical for the control and violation. As can be seen in the sentence examples in Table 2, the words preceding the comparison points in control #3 and violation #3 are of a different class (article and possessive proper name). Using a -600 to -540 msec prestimulus baseline, it was found that some groups (in particular the >16 group) showed a marked and prolonged difference in their responses to the word types preceding the control and violation comparison points. Therefore, the alternate control #1 was used since the words preceding the comparison points in control #1 and violation #3 are identical. For a complete description of each sentence type and complete listing of all sentences refer to Neville et al. (1991).

**ERP Recording**

Scalp electrical activity was recorded with 14 electrodes secured in an elastic cap (Electro-cap). Electrodes were positioned over the same relative locations of the two hemispheres according to the criteria of the International 10-20 system (Jasper, 1958). The specific locations were midline sites Cz and Pz, medial frontal (F3, F4), frontal (F7, F8), anterior temporal [L22, R22: 50% of the distance between F7(8) and T3(4)], temporal [L41, R41: 33% of the interaural distance between F7(8) and T3(4)], parietal (WL, WR: 50% of the interaural distance lateral to a point 13% of the nasion-inion distance posterior to Cz), and left and right occipital (O1, O2) regions. The electrooculogram was recorded from the right outer canthi and left inferior orbital ridge. All recordings were referenced to the linked mastoids. The electrical signals were amplified within a bandpass of 0.01-100 Hz and digitized on-line at 250 samples/sec.

**Procedure**

The experimental task and ERP recording procedures were explained to the subjects. Subjects completed a consent form, the Edinburgh Inventory for assessment of handedness (Oldfield, 1971), and a language history questionnaire (Appendix A), which was modified from Lanza (1988).

Once recording electrodes were in place, subjects were seated comfortably in a copper-shielded room, 57 in. in front of a 23-in. monitor. Subjects were given specific instructions about the task and were asked not to blink or move when the stimuli were being presented. Trials were self-paced; subjects pressed a button to initiate each trial. Following a 1-sec delay, a computer-generated rectangular border (subtending 8 × 4º visual angle) was outlined on the screen. After an additional delay of 750 msec, a stimulus sentence was presented one word at a time in the center of the rectangular border. The randomized sentence stimuli were generated by an IBM-PC computer. Words were 300 msec in duration with a 240-msec interstimulus interval. The visual angles subtended by the word stimuli were 0.5 to 5º horizontally and 0.5º vertically. The last word in each sentence was presented with the appropriate punctuation (period or question mark). The border remained illuminated on the screen for 2 sec following the presentation of the last word in a sentence. Subjects were instructed not to blink or move as long as the rectangle remained. When the border disappeared, subjects pressed one of two buttons on a response box labeled “Yes” and “No” to indicate whether or not the sentence was “a good English sentence.” Subjects were randomly assigned to use of either the right or left hand to respond “Yes.” Subjects were given a break after every 60 trials. The 240 sentence stimuli were contained in two sentence lists, “A” and “B.” Subjects were randomly assigned to receive presentation of list “A” in the first half of the experiment.

Following the ERP recording session, four standardized tests were administered. One test was the Processing Word and Sentence Structure subtest of the Clinical Evaluation of Language Functions (CELF) developed by Semel-Mintz and Wiig (1982). This subtest assesses knowledge about English grammar and syntax including noun phrase structure (plurals, possessives, and modification), verb phrase tense, prepositional phrases, pronouns, negation, passives, wh-interrogative, indirect object, and relative clauses. A subtest (78 items) of the Saffran and Schwartz Grammaticality Judgment Test was also administered to assess knowledge of English grammar skills including subcategorization, particles, prepositions and particle movement, extraction of noun phrases, phrase structure, tag questions (verb copying), and reflexives including gender, person, and number (Linebarger, Schwartz, & Saffran, 1983). The Test for Reception of Grammar (TROG), another tool for assessing knowledge of English grammar, tested awareness of structures such as reversible active and passives, personal pronouns, postmodified subject, X but not Y, relative clause, and embedded sentence (Bishop, 1982). Finally, the Carpenter Reading Span Test was also administered. This test was designed to test working memory and has been found to correlate with language comprehension skills (Daneman & Carpenter, 1980; Carpenter & Just, 1989).

**Data Analysis**

Self-rated proficiency measures obtained for the bilingual groups were analyzed for between-group effects using a one-way analysis of variance with five groups (1-3, 4-6, 7-10, 11-13, and >16 bilinguals groups); between-group
contrasts were tested with one-way analyses of variance with two groups (1-3:4-6, 4-6:7-10, 7-10:11-13, and 11-13:16). Self-rated proficiency measures were also analyzed for within-group effects to determine if measures differed for ratings on Chinese and English for a particular age of exposure using a matched T analysis. The behavioral performances of bilinguals on the standardized tests and in judgments of the experimental sentences were compared to data previously obtained in monolinguals (N = 40) by Neville et al. (1991). Analyses of the behavioral performances on each of the four standardized tests and judgment accuracy of experimental sentence types (control plus violation) were completed using a one-way analysis of variance with six groups (monolinguals, 1-3, 4-6, 7-10, 11-13, and >16 bilingual groups). In cases of significant group effects, planned group comparisons on behavioral performance between monolinguals and each of the bilingual groups were carried out using a two sample pooled T test. Welch’s separate T test, which provides a more conservative estimate of group differences, since it does not assume equal population variances for the two groups, was used when the Levene F for variability was significant for the group comparisons (Welch, 1947). Regression analyses were performed to determine whether performance measures (including relative use patterns, self-rated proficiency, standardized tests, and judgment accuracy of experimental sentences) were linearly related to (1) age of exposure to and (2) years of experience with English. Separate regressions (as opposed to one multiple regression) were calculated for age of exposure to and years of experience with English because these variables are highly correlated (r = 0.77). Obtaining separate regression values allowed for testing the difference between the two nonindependent regressions. A specialized t test (two-tailed, 60 df) was utilized that includes a term for the degree to which the two tests themselves are correlated (Howell, 1992).

For analyses of ERP responses, trials with excessive eye movement (approximately 10%) were excluded. Remaining trials associated with each of the eight sentence types were averaged for each subject. The averages were triggered 100 msec prior to the onset of the comparison words in the sentences and included 924 msec after the trigger. The EEG 100 msec prior to the onset of each comparison word was used as baseline for all sentence type comparisons.

Measurements of peak amplitudes and peak latencies were quantified relative to the baseline voltage in each subject’s averages for temporal windows of 50-250 msec (N125), and 180-500 msec (P250). For the broader components occurring later in the responses, the mean amplitude and peak latencies were measured. The latency windows included 300-500 and 500-700 msec. These measurement windows were in part selected so that direct comparisons with previous monolingual data could be made. In some cases, however, the latency window did not capture a response for a particular bilingual group, and additional measurements were made within more appropriate latency windows. These additional latency windows are reported in the results section. To determine effects of semantic and syntactic violations on processing for each of the bilingual groups, three-way analyses of variance with repeated measures were performed for each of the control-violation type pairs for each group: 2 sentence types (control, violation) x 2 hemispheres x 5 electrode sites (frontal, anterior temporal, temporal, parietal, occipital).

Group ANOVAs were performed on the ERP difference waves that isolate the effects of anomalous vs. appropriate linguistic stimuli. The difference waves were computed for each subject by subtracting a control ERP from its violation condition ERP. Measures of difference-wave characteristics included peak latency, peak amplitude, and mean amplitude for particular latency windows. The specific latency windows used are described in detail for each condition in the Results section. Measures obtained from the difference waves were compared to previously acquired data from monolinguals in a three-way analysis of variance with repeated measures: 6 groups (monolinguals, 5 bilingual groups) x 2 hemispheres x 5 Electrode Sites (frontal, anterior temporal, temporal, parietal, and occipital). Central electrode sites were not included in the present analyses. For all repeated measures with greater than 1 degree of freedom, the more conservative Greenhouse-Geisser adjusted p-values were used. Regression analyses (df = 59, two-tailed, p < 0.05) were also performed to determine whether changes in ERP measures were linearly related to (1) age of exposure to English, and/or (2) years of experience with English. As for the regressions used in the analyses of behavioral performance, separate regressions were calculated for age of exposure to and years of experience with English due to the highly related nature of these two variables (r = 0.77). As explained above, obtaining separate regression values allowed for testing the difference between the two nonindependent regressions (Howell, 1992).

Initial analyses were based on all artifact free trials, whether or not they resulted in correct or incorrect behavioral responses. This was done because we did not want to exclude subjects who may not have had an adequate number of correct trials for a given condition. From behavioral results, it is clear that a performance-based criteria would have biased subject inclusion unevenly and most profoundly for the later learning groups. In cases where significant group effects on ERP responses were found, additional analyses were performed to determine whether inclusion of correct trials only yielded similar results. Subjects included in these further analyses were only those with a minimum of 20 of 30 correct for both control and violation sentences. This level of accuracy indicated that the correct trials were not due to guessing or chance and provided an adequate number of trials for ERP averaging of individual subject.
data. Results of analyses of correct trials only of the semantic and phrase structure conditions yielded findings similar to those based on all trials (and all subjects). Therefore only results based on all trials are reported. Due to the relatively high rate of errors for many subjects in judgment of specificity constraint conditions, it was not possible to carry out statistical analyses of the ERPs associated with correct trials only for this condition.

Appendix A

Bilingual Questionnaire

Name: __________________________ Sex: __ Age: ___

Where were you born? __________________________

Where were your parents born? __________________________

If you were not born in the United States, how old were you when you moved to this country? ______ How many years have you lived in the United States? ______

At what age did you start using English? _____________

How did you learn Chinese? __________________________

Which dialect of Chinese did you learn? _____________

How did you learn English? __________________________

In response to the questions below, circle the relative frequency with which you used Chinese and English at different ages and in different situations. Use the following scale:

1 = Chinese only
2 = Chinese frequently, English rarely
3 = Chinese majority with English used at least 1/4 of the time
4 = Equal use of Chinese and English
5 = English majority with Chinese used at least 1/4 of the time
6 = English frequently, Chinese rarely
7 = English only

1. As a young child, before starting school?
   In school
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

2. As a child, primary school age?
   At home
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

3. As a teenager, secondary school age?
   In school
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

   At home
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

4. As an adult?
   In the University or at work
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

   At home
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

   In other places
   Chinese only
   English only
   Equal use—
   Chin. & Engl.
   1 2 3 4 5 6 7

Circle the one that best answers the following questions:

How well do you understand Chinese and English?
   Chinese: perfectly well sufficiently scarcely
   English: perfectly well sufficiently scarcely

How well do you understand what you read in Chinese and English?
   Chinese: perfectly well sufficiently scarcely
   English: perfectly well sufficiently scarcely

How well do you speak?
   Chinese: perfectly well sufficiently scarcely
   English: perfectly well sufficiently scarcely

How well do you write?
   Chinese: perfectly well sufficiently scarcely
   English: perfectly well sufficiently scarcely

Which language do you feel more comfortable using?
   Chinese    English    Chinese and English equally
Acknowledgments

We thank Mrs. Marta Kutas and Debra Mills for their comments on an earlier version of this paper. This research was supported by Grants DC00054 and DC00128.

Reprint requests should be sent to Dr. Helen Neville, Psychology Dept., 1227 University of Oregon, Eugene, OR 97403-1227.

Notes

1. In these and subsequent regression analyses, separate regressions (as opposed to one multiple regression) were calculated for age of exposure to and years of experience with English because these predictors are themselves highly correlated (r = 0.77). This allowed for testing the differences between the resulting two nonindependent regressions. A specialized t test that includes a term for the degree to which the two tests themselves are correlated was utilized (Howell, 1992).

2. Correlations were calculated by assigning a numerical value to each subject’s comfort ratings: Chinese = 1, Both = 2, English = 3.

3. The very late positive shift (beginning around 800 msec poststimulus onset) in the specificity constraint condition (Fig. 9) was probably due to the sentence ending on the following word in the violation case.

4. Subjects reported that up until recently, it was common in their culture to train all children to be “right-handed.” Therefore, it is possible that some subjects and family members would have been left-handed without intervention. Certainly, the report of familial left-handedness (<10%) was much lower for the Chinese-English bilinguals than for the group of monolinguals tested previously (50%) (Neville et al., 1991).

5. The mean age of the latest learning group (>6) was significantly greater than that of the other bilingual groups. To ensure that differences in behavior and ERPs observed in this group were not due to the chronological age of the subjects, a subgroup of 10 monolinguals with a mean age of 29.5 (standard error = 0.91) was formed. No effects of chronological age on the ERPs or behavioral accuracies were observed for this set of older monolinguals.

6. As discussed by Neville et al. (1991), the comparison points in the control and violation sentences for the subjacency constraint condition were preceded by different word classes, open and closed class words, which are known to elicit different ERP waveforms. A baselining procedure was used with the monolingual data to take this into account. The baselining procedure was not adequate for use with the bilingual data, however, because the ERP responses to the preceding words varied for the different groups of bilinguals. Comparisons of subsequent points in the waveforms were therefore uninterpretable and the bilingual data for this condition are not presented. Further investigation of the ERP responses of bilinguals to open versus closed class words as a function of age of second language exposure is in preparation.

REFERENCES


Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. Biometrika, 34, 28-35.
