Journal of Psycholinguistic Research, Vol. 33, No. 2, March 2004 (© 2004)

Who Do You Love, Your Mother or Your Horse? An Event-Related Brain Potential Analysis of Tone Processing in Mandarin Chinese

Sarah Brown-Schmidt[1,2] and Enriqueta Canseco-Gonzalez[1]

In Mandarin Chinese, word meaning is partially determined by lexical tone (Wang, 1973). Previous studies suggest that lexical tone is processed as linguistic information and not as pure tonal information (Gandour, 1998; Van Lanker & Fromkin, 1973). The current study explored the online processing of lexical tones. Event-related potentials were obtained from 25 Mandarin speakers while they listened to normal and anomalous sentences containing one of three types of semantic anomalies created by manipulating the tone, the syllable, or both tone and syllable (double-anomaly) of sentence-final words. We hypothesized N400 effects elicited by all three types of anomalies and the largest by the double-anomaly. As expected, all three elicited N400 effects starting approximately 150 ms poststimulus and continuing until 1000 ms in some areas. Surprisingly, onset of the double-anomaly effect was approximately 50 ms later than the rest. Delayed detection of errors in this condition may be responsible for the apparent delay. Slight differences between syllable and tone conditions may be due to the relative timing of these acoustic cues.

KEY WORDS: Chinese; event-related potentials; lexical tone; N400.

INTRODUCTION

The pronunciation of virtually every syllable of every word in Mandarin Chinese is defined not only by the vowels and consonants but also by its tone (Wang, 1973). Ho and Bryant (1997) describe tone as a suprasegmental feature

We would like to thank Jenn-Yeu Chen and Cyma Van Petten for helpful discussions about these data. Additionally, we would like to thank Cheng Wen for help in stimulus construction and Dr. Michael Sullivan for writing our PsyScope program.

[1] Department of Psychology, Reed College, Portland, Oregon 97202.

[2] To whom all correspondence should be addressed: Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14627. email: sschmidt@bcs.rochester.edu

0090-6905/04/0300-0103/0

that is attached to the rhyme in a syllable (see also Chen, 1999; Leben, 1978; Wan & Jaeger's [1998] description of an autosegmental framework for a different perspective). "Lexical tone" is different from the type of intonation used in English to give emphasis and mark interrogatives (Repp & Lin, 1990; Wang, 1973). In Chinese, lexical tones have phonemic status, meaning that they provide semantic and syntactic information of a word (Ho & Bryant, 1997; Wang, 1973). Errors in the pronunciation of these lexical tones can produce semantic and syntactic errors. For example, in the sentence "Wo ai wo mā" (I love my mother), changing the tone of "mā" could produce the words *horse, hemp*, or the verb *to scold* (see also Chen [1999] for a different perspective and thorough review on the controversial topic of the nature and representation of lexical tones).

THE PROCESSING OF LEXICAL TONE IN THE BRAIN

In addition to understanding how tonal information is produced and perceived, an additional concern is how tonal information is processed in the brain. A common claim in psycholinguistics is that the left hemisphere is dominant for language processing in right-handed individuals (Kimura, 1973, Rasumssen & Milner, 1977). An early contributor to this theory, Paul Broca, collected extensive data through his work with neuropsychological patients. He demonstrated that damage to the left hemisphere was typically associated with severe language impairments whereas damage to the right hemisphere was not (Broca, 1861; see Purves *et al*. [1997] for a review). Converging evidence for the lateralization of language comes from evidence that people respond faster to linguistic stimuli presented to the right visual field than the left (see Elias *et al*. [2000] for one example). Right visual field information travels to the left hemisphere first because of the optic nerve crossover in the optic chiasm. In a similar manner, most (but not all) auditory fibers from the right ear cross over to the left hemisphere (Carlson, 1994).

While specialization in language processing is left lateralized, the right hemisphere is argued to be dominant in the processing of musical, tonal, and intonational stimuli in linguistically devoid contexts (Breier *et al*., 1999; Gootjes *et al*., 1999; Packard, 1986; Repp & Lin, 1990). The fact that lexical tones are both linguistic and tonal yields a conundrum: are lexical tones processed as linguistic information or not? Van Lanker and Fromkin (1973) compared the processing of tonal words in Thai with consonant words and hums. Native right-handed Thai and English speakers listened to three different sets of stimuli in a dichotic-listening situation. One set was composed of Thai words that differed only in lexical tone. A second set was composed of Thai words differing only in the initial consonant. The third set was a series of five hums, produced at the pitches of the five different Thai tones. Results of a word identification task revealed a left hemisphere (right ear) advantage for both the tone and consonant tasks. There was no difference in accuracy between the two ears for the hum task, even though the hums contained tones. These results suggest that when tone is linguistically processed, it is primarily processed in the left hemisphere, as with other linguistic stimuli. In contrast to the Thai speakers, English speakers showed no significant ear advantage for either the tone or hum conditions, but did show a right ear (left hemisphere) advantage for the consonant condition. However, it is important to point out that not all types of linguistic information are left lateralized; right hemisphere dominance has been reported in the processing of speech intonation (Shipley-Brown *et al*., 1998), and damage to the right hemisphere disturbs the production of prosody (see Edmondson *et al*., 1987).

Lexical tones in Mandarin Chinese have been argued to be processed in a similar fashion to lexical stress in English and unlike segmentals (Chen, 1999). While we are unaware of any studies using the event-related brain potential (ERP) methodology to study lexical stress, there have been a few papers on the electrophysiological correlates of prosodic processing. These studies have primarily focused on comparisons between music and language and have found prosodic errors (mismatches between syntax and prosody) to elicit ERP components (e.g., P600) similar to those elicited by syntactic errors signaled by nonprosodic information (Besson, 1998; Steinhauer *et al*, 1999).

A cornerstone of research in the lateralization of language debate comes from research with patients suffering from aphasia. Aphasia is a type of language impairment generally linked to left hemisphere brain damage (Parkin, 1996). If tone production and/or comprehension by native speakers of tone languages were impaired more by right than left hemisphere damage, this would support the hypothesis that the right hemisphere is dominant for lexical tones. However, research with left-hemisphere-damaged aphasics indicates that there is no difference in language impairment between speakers of tonal and nontonal languages (Gandour, 1998; Packard, 1986). After reviewing a large number of neuropsychological findings, Gandour (1998) found no evidence of right hemisphere specialization for lexical tone function in speakers of tone languages. For the purposes of the current study, these findings imply that online processing of lexical tones will likely resemble processing of other kinds of linguistic information, rather than alinguistic or pure tonal processing.

Additional support for the idea that the physical nature of the linguistic information (tonal or otherwise) has little impact on the processing of that information comes from research with deaf aphasics who communicate in sign languages. The domain-specific viewpoint would hold that because sign languages are expressed in the visual and spatial realms, damage to the right hemisphere would cause serious language impairments. As it turns out, damage to the right hemisphere of signers impairs nonlinguistic visuo-spatial abilities whereas only damage to the left hemisphere of signers causes serious language impairment (Corina, 1998; see also Langdon & Warrington, 2000). This is the same pattern found in patients of spoken languages (Parkin, 1996; Zurif, 1995). In the cases of both sign and tone languages, it appears that it is not the physical characteristics of the stimuli, but rather their linguistic nature that determines the specialization of the left hemisphere (but see Bavelier *et al.* [1997, 1998] for evidence of right hemisphere activation during ASL comprehension in neurologically intact deaf individuals).

A remaining question concerns how lexical tone information is integrated with other types of linguistic information. The neuropsychological evidence mentioned above suggests that tonal information is processed no differently than other types of linguistic information. In addition, developmental research with Mandarin-speaking children indicates that tone is acquired early in development, before mastery of segmentals (Hua & Dodd, 2000; Li & Thompson, 1977). Furthermore, in comparison to other types of speech errors, tone errors in adult speech are relatively infrequent (Chen, 1999). Tone information is a salient aspect of a word as it regularly determines which, among a variety of possible meanings, is intended by the speaker. The significance of tone in determining word meaning, along with the fact that it is acquired early, suggests that tone is a central aspect of Chinese. Unlike prosody, which is a less critical cue, lexical tone is essential to ascertain word meaning.

On this basis, we expect the online processing of lexical tone information to be more similar to the processing of phonemic information than to intonational information. On the other hand, it has been argued that lexical tones may be processed more like lexical stress in English than the segmental information in Mandarin (Chen, 1999). Chen (1999) examined a corpus of phone calls made to radio programs and analyzed the tapes for production errors, identifying tone errors, and errors in the production of segmental information (consonants and vowels). The result of the analysis was that tone errors were rare in comparison to the other kinds of errors. In addition, the kinds of errors observed most frequently for tones were different from those associated with segmentals. Tone errors were primarily perseverations, with few anticipatory errors, whereas errors in the segmental elements were frequently anticipatory or movement errors. Chen's conclusion was that the relative lack of, and qualitative difference in tone errors, suggests that tonal information is processed differently from segmental information. Chen concludes that lexical tone in Mandarin may be processed more like lexical stress in English, which (reportedly) is also fairly immune to errors.

Yet, it is not known how tonal information is used during online sentence processing. One way of approaching this question is by thinking about when, during the pronunciation of a word, tonal information can determine word meaning. In Chinese, lexical tone is heard on the vowel. The basic unit of speech in Chinese is the morpheme. Generally, each morpheme or syllable has two parts: the initial and final segments. The initial segment is composed of the beginning consonant(s). The final segment is made up of the ending vowel(s) plus the ending consonant(s). This final segment can also be called the rhyme, and it is during the rhyme that lexical tone appears (Ho & Bryant, 1997). For example, the word gǒu (dog) has an initial consonant "g" and rhyme "ǒu," on which the tone is realized. Because words are detected incrementally, with each bit of sound information listeners are making guesses as to what the word might be (Van Petten *et al.* 1999). As a result, during spoken word recognition, tone information is heard slightly later than word-initial-segmental information and likely contributes to word recognition slightly later. In a series of experiments and explicit models of word recognition processes, Allopenna *et al.* (1998) demonstrated that as a word unfolds in time, other words with overlapping sound information become activated in time. Hearing the word "beaker" activates not only "beaker" and cohorts like "beetle," but also activates rhymes like "speaker." Note, however, that the activation of "speaker" is slightly delayed relative to that of "beetle," likely due to the fact that the overlapping sound information between "beaker" and "speaker" occurs milliseconds later than that between "beetle" and "beaker." Given a continuous model of spoken word recognition, we expect that the lexical tones in Mandarin words will influence potential lexical candidates in real-time during the process of word recognition, and that this kind of information may be available slightly later than initial consonant information.

Another piece of information suggesting that syllabic and tonal information are treated differently comes from a series of implicit priming production studies by *Chen et al.* (2002). In their task, Mandarin speakers responded to previously memorized compound word pairs by uttering the second word upon seeing the first word as a cue on a screen. In the learning phase of the three experiments described below, participants memorized four word pairs. Following this, they were tested by flashing a cue word, to which they were required to respond with the corresponding target (second) word. In Experiment 1b, the initial syllable + tone (bolded) of the four target words were the same (see examples below).[3] In Experiment 2a, the first syllable of the four target words (bolded) shared the same syllable but not the tone.

[3] Examples adapted from Appendix 1 of Chen *et al.* (2002).

Experiment 1b	Cue words	Target words
pair1	chi1-bing1	**qing1**-liang2
pair2	zi4-sha1	**qing1**-sheng1
pair3	kun1-chong2	**qing1**-ting2
pair4	shu1-cai4	**qing1**-jiao1
Experiment 2a	**Cue words**	**Target words**
pair1	hang2-kong1	**fei1**-ji1
pair2	su4-shen1	**fei2**-pang4
pair3	zhu1-bao3	**fei3**-cui4
pair4	chou1-yan1	**fei4**-yan2
Experiment 2b	**Cue words**	**Target words**
pair1	hang2-kong1	**fei1**-ji1
pair2	dian4-nao3	**ke1**-ji4
pair3	shui3-guo3	**xi1**-gua1
pair4	shu1-cai4	**qing1**-jiao1

Finally, in Experiment 2b, the first syllables of the four target words (bolded) shared tone but were different syllables.

Speech onset times (SOTs) for target words containing overlapping sound information were compared to SOTs for word pairs that did not contain any overlapping sound information. They found priming (faster SOTs as compared to the baseline unrelated word pairs) for both syllable + tone and syllable alone conditions and no (or sometimes slightly negative) priming for the tone condition. These timing differences suggest that at least in production, the syllable is easier to prepare ahead of time than the tone, and that tone and syllable may be differently processed. Implications for the current study will be discussed in a later section.

ERPs AND LANGUAGE PROCESSING

In recent years, the recording of event-related brain potentials has been a profitable method for research in language processing. The recording of scalp electrical potentials provides a detailed time course of the electrical activity in the brain associated with specific events. One of the best-described brain electrical potentials in the ERP language community is the N400. The term "N400" is used because this component is negative relative to a baseline and peaks at approximately 400 ms following stimulus presentation, although N400 amplitude differences can be seen as early as 50 ms after stimulus presentation (Holcomb & Neville, 1991; cf. Kutas, 1993). The N400 reflects a word's expectancy and degree of association with its context. Presentation of any word will generally elicit an N400, but less-associated and less-expected words tend to elicit N400s with larger amplitudes (Brown & Hagoort, 1993; Kutas, 1993; Kutas & Hillyard, 1984; Kutas *et al*., 1987; Kutas *et al*., 1988).

The distribution of the N400 is typically centro-parietal and often seen with larger amplitudes in the right hemisphere (Kutas & Hillyard, 1980; Kutas & Van Petten, 1994). The N400 and its underlying cognitive processes are thought to be modality independent (Kutas & Kluender, 1994; Kutas *et al*., 1987; Nigam *et al*., 1992). However, modality-modulated effects have been reported; for example, auditorially elicited ERPs are reported to show left (rather than right) hemisphere prominence (Holcomb & Neville, 1990; Kutas & Van Petten, 1994) and tend to onset earlier and last longer than those in the visual modality (Holcomb & Neville, 1991, 1990).

Within the sentence processing community using the ERP technique, the presentation of anomalous sentences is a standard technique used to examine the details of human sentence processing. An "anomaly" or anomalous word is generally defined as a word that deviates from an expected word or set of words in one or more dimensions. For example, the word "run" in the sentence context "Jackie Joiner run fast" deviates from the expected "runs" in its marking for third person. Comparisons of "run" versus "runs" in this sentence context arguably reveals something about the processing of the third person particle. Numerous authors have used anomalous or highly unexpected words in online studies of sentence comprehension (see Kutas & Van Petten [1994] for a discussion).

Kutas *et al*. (1994) demonstrated that N400 amplitude varies as a function of the semantic association between a word and its context. The words *cry* and *drink* are anomalous for the sentence context "The pizza was too hot to . . . ." As a result, ending the above sentence fragment with the words *cry* or *drink* elicits a large N400 in comparison to the N400 elicited by *eat*, which is considered the best completion for that sentence. In addition, the amplitude of the N400 elicited by the word *cry* is larger than that elicited by *drink*. This difference indicates that the N400 is related not only to the presence of an anomaly, but also to the degree of semantic association between a word and its context; less-associated or less-predictable words elicit N400s of larger amplitudes (Holcomb & Neville, 1991; Kutas & Hillyard, 1980; Kutas & Hillyard, 1984; Kutas *et al*., 1984; Van Petten, 1993).

A related finding is that the N400 amplitude grows, in part, as an inverse function of word expectancy (Besson *et al*., 1992; Kutas & Hillyard, 1984; Kutas *et al*., 1987; Kutas *et al*., 1998). Word expectancy is typically quantified by estimating a word's Cloze probability for a particular sentence context. The Cloze probability for a word in a particular sentence context is the probability that participants complete the sentence with that word. A sentence with a high degree of contextual constraint will be highly predictive of a final word; that final word is said to have a high Cloze probability (Bloom & Fischler, 1980; Kutas & Hillyard, 1984; Taylor, 1953). The final words in Sentences

(1a) and (1b)[4] have high and low contextual constraints and Cloze probabilities, respectively:

(1) a. He mailed the letter without a *stamp*.
    b. He was soothed by the gentle *wind*.

Kutas and Hillyard (1984) presented subjects with sentences like (1a) and (1b) and found that N400 amplitude to sentence-ending words decreased as Cloze probability increased. This finding indicates that semantic incongruity is not a necessary condition for the elicitation of an N400; word expectancy is an additional important factor in determining N400 characteristics.

In later research, Kutas (1993) describes the N400 component as sensitive to both contextual constraint and semantic association. Additionally, meaningful sentences that are terminated unexpectedly will also generate N400s (Kutas *et al*., 1987). Finally, larger N400s are elicited by content words when compared to function words (Kutas *et al*., 1987; Kutas *et al*., 1988).

Some authors have proposed that N400s can be elicited by factors other than semantic expectation. For example, presentation of the second word of a non-rhyming pair (in contrast with a rhyming word) elicits a late negative component termed N450 that is largest over temporal sites (Rugg, 1984a). Rugg argues that the N450 belongs to the same family as the N400 but suggests that the N450 reflects processing at the phonological level. Radeau *et al*. (1998) also found the N400 component to be sensitive to phonological processing. In an auditory priming task, N400 amplitudes were attenuated for words preceded by semantic primes; a smaller but significant attenuation was found for phonologically related primes. These priming effects first became significant in the 300–800 ms window following stimulus presentation, with larger priming effects at posterior electrode sites. The findings by Rugg and by Radeau *et al*. are particularly relevant to the current study. Specifically, we wondered whether lexical tone deviations would elicit the typical auditory N400 effect or if we would observe data patterns more similar to those in these phonological processing studies. One difficulty in comparing the current study with the Rugg and the Radeau *et al*. findings is that these latter two studies investigated the processing of isolated words whereas the current study uses words embedded in sentence contexts.

Kutas (1993) argues that the N400 is representative of a default response to potentially meaningful items such as words, pseudowords, lexical strings, line drawings, real pictures (Nigam *et al*., 1992), and ASL handsigns. Importantly, deviations in familiar melodies and physical or grammatical anomalies fail to produce the N400 effect (Kutas *et al*., 1987).

Kutas *et al*. (1987) argue that if N400s are linked to phonological processes, then we ought not to find them in deaf signers. They collected ERP data from

[4] Sentences from Kutas & Hillyard (1984).

three groups of people: hearing adults in a visual presentation, hearing adults in an auditory presentation, and deaf signers with visual presentation of signs. All three groups showed the N400 effect, supporting their claim that the N400 reflects language processing in general and is not necessarily linked to phonological processing. Given these findings, it appears that the N400 and N450 are two different components: the first reflective of linguistic processing and the latter reflective of phonological processing.

The current study was aimed at examining the online processing of lexical tones in Mandarin Chinese. Given previous offline and neuropsychological findings (Packard, 1986; Shipley-Brown *et al*., 1988; Van Lanker & Fromkin, 1973), we expected to find that lexical tones are processed like other types of linguistic information. Because the N400 reflects semantic expectancy violations during linguistic processing, we would expect tone violations to elicit N400s only if they are being linguistically processed. An alternate possibility, given the findings of Rugg (1984a; 1984b) and Radeau *et al*. (1998) is that we may find an N450 for the phonological processing of lexical tones that could appear distinct from the processing of non-tonal-based anomalies. A final motivation for the current study was to see if the different types of deviations (changes in lexical tone, segmental information) between anomalous and expected ending words would cause different ERP effects either in latency or in amplitude.

Of the numerous, mutually unintelligible Chinese dialects, Mandarin was used in the current study because it is the major dialect that varies least from region to region and is considered the national language (Wang, 1973). Additionally, of the various dialects, Mandarin Chinese has the least number of tones (Li & Thompson, 1979). Cantonese, for example, has eight tones whereas Mandarin has four (Ladefoged & Buchholz, 1993). Many Chinese speakers who speak another dialect are also fluent in Mandarin. The fact that Mandarin has fewer tones than the other dialects suggests that processing Mandarin tones may be potentially easier for a speaker of a dialect with more tones than vice versa. Evidence from tone perception research indicates that while speakers will be best at perceiving tones of their own dialect, Cantonese speakers may be more versatile in their ability to perceive tones than Mandarin speakers (Lee *et al*., 1996).

To investigate the online processing of lexical tones in Chinese, we conducted an event-related potential study in Mandarin Chinese. A second experiment investigated the plausibility of the sentences presented in Experiment 1. In Experiment 1, native Chinese speakers were auditorially presented with sentences that had congruent or anomalous endings. Three different types of semantic anomalies were created by manipulating the tone, syllable, or both tone and syllable (double-anomaly) of sentence final words. We hypothesized that an N400 effect would be elicited by all three types of semantic anomalies, independently of the specific source of the semantic information (syllable or tone), and that changing simultaneously both the tone and syllable of the expected ending word would elicit the largest negativity.

EXPERIMENT 1: THE ONLINE PROCESSING OF LEXICAL TONE AND SEGMENTAL INFORMATION

Experiment 1 manipulated the tone (suprasegmental information) and syllable (segmental information in single-syllable words) of ending words in Chinese auditory sentences and looked at the effects of these manipulations on time-locked brain electrical potentials. The plausibility of the tone and syllable of sentence ending words was manipulated to observe their independent and conjoined effects on sentence processing. Plausibility was confirmed in a follow-up study (Experiment 2). The dependent variables were the amplitude, latency, and scalp distribution of event-related brain potentials associated with the presentation of the ending (critical) words in experimental sentences.

Experimental sentences were designed to have high Cloze probabilities. A high Cloze probability means that sentence contexts are highly predictive of the last word of the sentence (Taylor, 1953). In the control condition, sentences were designed such that the ending words had a good fit with the sentence; they were rated as highly plausible. In the three other conditions, critical words were designed to end the sentence in a semantically implausible manner; these sentences were rated as highly implausible (see Experiment 2 for plausibility estimates).

Plausibility was manipulated by changing either the tone or the syllable or both tone and syllable of the expected ending word. In all cases, changing the tone, syllable, or both tone and syllable produced sentences that ended with a semantically implausible word for that sentence context. Tone changes involved changing the tone of one word to one of the other three possible tones in Mandarin. Syllable changes involved changing the initial consonant (when there was one) and all consonant-vowel clusters to different consonant-vowel clusters. For example, a word like “ma” could have been changed to a word like “shu” but not to “mu” because “ma” and “mu” share an initial consonant. Additionally, in all but three (out of 160) of the stimulus items, all vowels were changed as well (e.g., “ma” would not be changed to “ba” because they share the same vowel). All implausible ending words were real Chinese words. In sum, we had four experimental conditions determined by the plausibility of the final word. This design allowed us to isolate the effects of tone and syllable errors and to analyze their interaction. The four conditions as well as example sentences (English gloss) are shown below:

(1) *Control condition*: Sentences ended with an expected, plausible syllable with a plausible tone. “When I go to the movies, I always eat some popcorn and *candy*.” (In Mandarin, “candy” = tāng, first tone).

(2) *Tone condition*: Sentences ended with an expected, plausible syllable but an implausible tone. "When I go to the movies, I always eat some popcorn and *government office*." ("government office" = táng, second tone).
(3) *Syllable condition*: Sentences ended with an implausible, unexpected syllable with a tone from the expected ending word. "When I go to the movies, I always eat some popcorn and *book*." ("book" = shū, first tone)
(4) *Double-anomaly condition*: Sentences ended with both an implausible, unexpected syllable and an unexpected tone. "When I go to the movies, I always eat some popcorn and *political party*." ("political party" = dǒng, third tone).

Participants

Twenty-five participants (11 men and 14 women) were recruited from the greater Portland, Oregon, area for the primary study. Mean participant age was 31.6 years (SD = 13.82, range = 18–64 years). All participants had normal or corrected-to-normal vision, no known neurological impairments, and were right-handed according to self-report. Additionally, all participants were native Chinese speakers who spoke Mandarin Chinese fluently. Most participants also spoke English. For the two participants who did not, a Chinese–English bilingual was present at the time of the study in order to translate if necessary. In addition, one of the researchers spoke Mandarin when necessary.

Although bilingualism was not a requirement for participation in this study, participants in both experiments completed a language questionnaire that yielded information about languages spoken, age of language acquisition, degree of bilingualism, and estimations regarding the amount of use of English and Chinese. On two different scales of 1–10, participants rated degree of bilingualism and use of Chinese and English on a daily basis. On the bilingualism scale, a rating of 1 indicated fluency in Chinese only, a rating of 5 indicated equal or comparable fluency in both Chinese and English, and a rating of 10 indicated fluency in English only. Mean rating on bilingualism scale was 4.23 (SD = 1.78, range = 1–7). On the daily use scale, a rating of 1 indicated primary use of Chinese, and a rating of 10 indicated primary use of English. Mean rating on the daily use scale was 5.2 (SD = 2.69, range = 1–9). Mean age of first exposure to Chinese was 0.5 years (SD = 1.24, range = 0–5 years). Mean age of first exposure to English was 11.22 years (SD = 9.06, range = 0–41 years). All participants in the study were fluent, native speakers of Chinese. All spoke Mandarin fluently, but not all spoke Mandarin as their primary dialect.

Participants were paid $12.00 for participation in the experiment. Additionally, bus fare was made available to all participants.

Apparatus

Auditory and visual stimuli were presented using a Macintosh Quadra 840 AV and a program written in PsyScope, Version 1.02 (Cohen *et al*., 1993). Participants were seated in a moderately lit, electrically shielded, sound-attenuated chamber facing a computer monitor. Sentences were presented using the computer's internal speaker, which was to the left of the participants. An intercom speaker to the right was used to communicate with the experimenter during the study. A button box with three different colored response buttons was placed on a table in front of participants. A mesh-covered window on the right-hand side of the chamber allowed the experimenter to monitor the behavior of participants and the progress of the experiment.

Electroencephalogram and Electro-Oculogram Recordings

An Electro-Cap International brand electrode cap was used to record brain activity from six specific sites on the scalps of participants; a seventh electrode on the cap served as ground. Conductive paste was used to lower impedance between the scalp and the electrodes. Two active electrode sites along the midline, Cz and Pz, were located in the electrode cap according to the International 10–20 System (Jasper, 1958). Four additional nonstandard electrode sites were used: a frontal electrode was located approximately over Broca's area (B3), and a symmetrical electrode was located in the homologous site of the right hemisphere, called Broca's right (B4). These electrodes were sewn into the electrode cap midway between F7/8 and T3/4, respectively. A temporoparietal electrode was located approximately over Wernicke's area (WL), and a symmetrical electrode was placed over the homologous site in the right hemisphere, referred to as Wernicke's right (WR). These electrodes were placed in the electrode cap 13% posterior to Cz and 30% lateral of the interaural distance, on each side of the head. These four nonstandard electrode sites are commonly used in ERP language research (Kluender & Kutas, 1993; Pratarelli, 1994). Each of the six electroencephalogram (EEG) electrodes were referenced to a pair of linked mastoid electrodes (Lehtonen & Koivikko, 1971; Stephenson & Gibbs, 1951).[5] In addition to the six scalp sites, eye movements were monitored with a bipolar electro-oculogram (EOG) recording by placing one electrode above the right eyebrow and the

[5] The linked electrodes used as the mastoid reference were created by Robert Ormond, Electronics Support at Reed College. A 12 kΩ resistor was inserted into each of the electrode wires before they joined to form one single pin.

other on the cheekbone below the right eye. An electrode sewn into the electrode cap just posterior to the center of the forehead was used as ground.

P511 Grass amplifiers with filter band-pass settings of 0.01 to 100 Hz were used to amplify the EEG and EOG signals. A 60-Hz notch filter was used throughout the recording. After being amplified, the EEG and EOG signals were sent to an analog-to-digital (A/D) converter; the sampling rate was 200 Hz. A Macintosh Centris 650 was equipped with an NI NB-MIO-16X board, enabling it to digitize and store the EEG data using LabVIEW software. Following EEG recording, trials with artifacts, excessive noise, or amplifier blocking were rejected by hand. High-frequency noise was removed with a 30-Hz filter. Data analysis was performed both before and after the 30-Hz filter. As the general pattern of results did not change by adding in this filter, we will be presenting only the filtered data.

Stimulus Materials

All stimulus materials were presented auditorially. One-hundred sixty experimental sentence frames in Mandarin Chinese were composed by one of the researchers with the help of a female native speaker of Chinese from Beijing. In addition, 80 filler sentences of similar structure to the experimental sentences were constructed. In sum, each participant listened to 240 sentences: 160 experimental and 80 filler. Experimental sentences were approximately 10 words long. Following each of these sentences, participants were also presented with individual probe words about which they were required to make judgments. Probe words were Chinese words taken either from the stimulus sentence or from a list of about 30 possible probe words not used in the stimulus sentences.

Each of the 160 experimental sentence frames was presented with one of four possible ending words, but each participant heard each sentence frame only once. The four possible endings for each sentence frame provided the four different experimental conditions as described in the experimental design. Three of these endings were semantically implausible, and the fourth was a semantically congruent ending. The three implausible endings were created by manipulating the tone, the syllable, or changing both the tone *and* the syllable of the expected word (double-anomaly). All three types of manipulations changed the expected word into an entirely different, but possible, Chinese word.

Care was taken to preserve the syntactic category of the four possible ending words for each sentence frame. For example, if the expected word was a noun, then each of the three implausible endings were also nouns. As a result, all anomalous sentences in this study were anomalous due to purely semantic (not syntactic) reasons. An example of one of the experimental sentences used in the current study is shown below. The markings above the Chinese words indicate the tone with which they are pronounced. In the examples below, *dàn* is fourth tone; *dǎn* is third tone; *qìan* is fourth tone; and *qiān* is first tone:

a. (pinyin): Mā mā de xiǎo niǎo hěn bīng le tā bù xià *dàn.*
(English gloss): Mother [pos.] small bird very sick [act. completed] it not under *egg.*
(translation): (My) mother's chicken is very ill, she doesn't lay *eggs.*

The other three versions of this sentence (presented to three different participants) were (translations given):

b. (My) mother's chicken is very ill, she doesn't lay *gallbladder* (*dǎn*).
c. (My) mother's chicken is very ill, she doesn't lay *moat* (*qiàn*).
d. (My) mother's chicken is very ill, she doesn't lay *footpath* (*qiān*).

Therefore, our design yielded the following four experimental conditions:

(1) A semantically congruent ending; *dàn* [egg], (control condition).
(2) A semantically incongruent ending due to a change in the tone of the syllable; *dǎn* [gallbladder], (tone condition).
(3) A semantically incongruent ending due to a change in the syllable; *qiàn* [moat], (syllable condition).
(4) A semantically incongruent ending due to both a change in tone and a change in syllable; *qiān* [footpath], (double-anomaly condition).

Even though many words in Chinese are polysyllabic compound words, no ending word was a compound word. However, some ending words were polysyllabic because a particle was attached to the main word. In Chinese, some words have particles such as *-zi*, or *-men* that attach onto a root word (similar to *-s* or *-tion* in English). For example, *Háizi* [children], and *tāmen* [them] both have these extra particles. In this design, only the main part of the word would be manipulated (*hái* and *tā* in the above two examples), and the particle would remain the same.

In order to have an equal number of plausible and anomalous sentences, 80 valid filler sentences that were similar in length and content to the experimental sentences were composed, but ERP data were not collected with these sentences. Each participant was therefore presented with a total of 240 sentences; 160 experimental sentences plus 80 filler sentences, with 50% containing an anomaly overall.

There are many homophones in Chinese, so a syllable with a particular tone can have several different meanings, depending on the context. Because the homophones are distinguishable in writing, but not in speech, we attempted to use the most common meaning of the critical ending words. Critical anomalous words that had homophones that were consistent with the sentence context were avoided.

Sentence frames were organized into four pseudorandomized lists. Each list contained the same 80 filler sentences mixed in with 160 experimental sentences.

The experimental sentence frames were the same across the four lists, but ending words for each sentence frame were rotated using a modified Latin squares design. Each list had an equal distribution of the four critical (ending) word types for the experimental sentences. In addition, a pseudorandomization procedure limited the number of times the same condition could appear in a row.

Sentence Recordings

For each sentence frame, critical word, filler sentence, and probe word, a sound file was created using SoundEdit 16. A female native speaker of Mandarin Chinese from Beijing, who spoke clearly and had good articulation, was recruited to record the sound files. All sentences were recorded at a natural speed into the computer using the congruent ending words.[6] Next, the final word of each target sentence was spliced off using SoundEdit (these were not heard by participants). All ending words (both plausible and implausible) were recorded in isolation and stored in separate files. Sentence frames and ending words were presented consecutively with, on average, a 3-ms break between sentence frame and ending words. Because all of the target sound files were created in the same way, there were not differences in the "choppyness" or "naturalness" of the sound files among experimental conditions. Filler sentences were recorded in the same manner as target sentences, but ending words were spliced off the original sentence recording and presented in a separate file[7]. A native Chinese speaker checked over the sound files and found that the splicing of ending words was not disruptive and that the sentences were easy to understand. Probe words were spliced from target and filler sentences using SoundEdit and stored in a separate file. All 240 sentence frames, ending words, and probe words were auditorially presented.

Experimental Procedure

Total experimental session duration was approximately 2 h. Participants were given both oral and written (in both Chinese and English) instructions regarding the study. Participants were instructed to listen to a series of auditorially presented sentences in Mandarin Chinese for comprehension. They were also told that after each sentence, they would hear a single word. Participants

[6] The tone with which some words in Chinese are pronounced changes depending on the tones of neighboring words. This phenomenon is called "tone sandhi." Stimulus materials were designed with these tone changes in mind. Words immediately preceding ending words were always presented as if the canonical ending word was next, regardless of actual ending word.

[7] We have no reason to believe that participants noted any difference between the filler and experimental sentences in this regard. In addition, we did not analyze the electrical activity associated with filler sentences.

were instructed to determine whether this word occurred in the previous sentence or not by pressing one of two buttons on a button box (the assignment of buttons to responses was counterbalanced across participants). Participants were told that some of the sentences may not fully make sense, but that this should not play a role in determining whether the probe word was in the previous sentence or not. Hereafter, this task will be referred to as the "probe word identification task." This task was used in order to require participants to pay attention to the sentences without requiring from them an overt evaluation of the target sentence.

Because eye blinking and muscle contraction can cause artifacts in the recordings of brain activity, participants were asked to sit as relaxed and still as possible and to avoid eye blinks during trials.

A typical trial ran as follows. At the beginning of each trial, the prompt "Please press red button for the next sentence" appeared in capital letters in the center of the computer screen. When the participant pressed the red button, there was a 500-ms delay followed by the presentation of the sentence frame and critical word (two consecutive sound files). After sentence presentation, there was a 1000-ms delay, followed by a 500-ms duration fixation point (a "+" appeared in the center of the computer screen). The auditory probe word was immediately presented after the fixation point disappeared. After participants responded to this probe word, the screen returned to the beginning prompt. This design gave participants the ability to control the rate of sentence presentation, allowing them to rest and blink between trials as necessary. The entire experiment consisted of 12 practice trials, with a break after the first 6, and 240 experimental trials, with a longer break after the 120th experimental trial.

Experiment 1 Results

Behavioral Data

Accuracy. Participants were highly accurate in the probe word identification task. Average accuracy across the four conditions was as follows: control = 89.9%, syllable = 86%, tone = 88%, and double-anomaly = 86.3%. A repeated measures analysis of variance (ANOVA) indicated a trend for a main effect of condition, $F(3,19) = 2.47$; $p = 0.0710$. However, Scheffé post-hoc analyses revealed no significant differences between conditions.

Reaction times. Participants were not asked to respond quickly to the probe word identification task. As a result, average reaction times were slow, and any conclusions derived from them should be interpreted with caution. Responses longer than 5000 ms (1% of all the data) were eliminated before data analysis. Average reaction times by condition, for correct trials only, were as follows: control = 1510.84 ms (SD = 376.36); syllable = 1637.92 ms (SD = 345.15);

tone = 1502.26 ms (SD = 351.64); and double-anomaly = 1561.01 ms (SD = 351.30). An ANOVA revealed a main effect of condition, with significance levels adjusted using the Greenhouse–Geisser adjustment for sphericity of variance: G–G e = 0.809, $F(3,19) = 6.68$; $p < 0.01$.[8] Post-hoc analyses were used to explore this main effect of condition. Scheffé tests for condition only revealed significant differences between tone and syllable conditions, $p < 0.01$ (mean difference = −135.66 ms), and between syllable and control conditions, $p < 0.01$ (mean difference = 127.07 ms). Analyses were also performed on the data set including incorrect trials. The pattern of results was identical.

Event-Related Potential Analysis

Data of five participants were excluded from analysis: two due to equipment failure and three because of excessive electrical noise or eye movement artifacts. In the end, ERP data from 20 participants (13 female) were used in the analysis.

Plotted in Fig. 1 are the grand average ERPs to final words from the 100-ms prestimulus baseline through 1000 ms poststimulus. Visual inspection reveals that in general, in posterior areas, the ERPs are positive-negative-positive going. In anterior areas, the ERPs are positive-negative going, with the exception of the control condition, which only shows the initial positivity and stays relatively flat for the rest of the epoch. It can be observed that all anomalous conditions elicited larger negativities than the control condition, starting around 300 ms. We will be referring to this as the N400 effect. In accordance with previous studies looking at N400 waveforms elicited by auditory stimuli, the N400 effect onsets early (around 200 ms poststimulus) relative to onset times in studies using visually presented stimuli (Holcomb & Neville, 1990, 1991; Kutas & Van Petten, 1994). In general, anterior regions exhibited larger and longer lasting condition effects than posterior regions. Visible distinctions between the waveforms terminate around 800 ms in posterior regions while at anterior electrode sites, the 1100-ms epoch did not capture the entirety of the effect, as there were still clear differences at 1000 ms poststimulus.

Prompted by analyses of latency windows by Holcomb and Neville (1991), Van Petten *et al.* (1999), and differences seen upon visual inspection of our grand average plots, we evaluated the ERP waveforms at a variety of time windows. First, in order to obtain a general sense of the basic effect, condition effects were analyzed in a large-window analysis, which averaged ERP data across the 200–800-ms poststimulus time window. Following this large-window

[8] Hereafter, all comparisons using the Greenhouse–Geisser adjustment for sphericity of variance will be reported with the Greenhouse–Geisser epsilon (G–G e = —) and adjusted p values; F values are unadjusted.

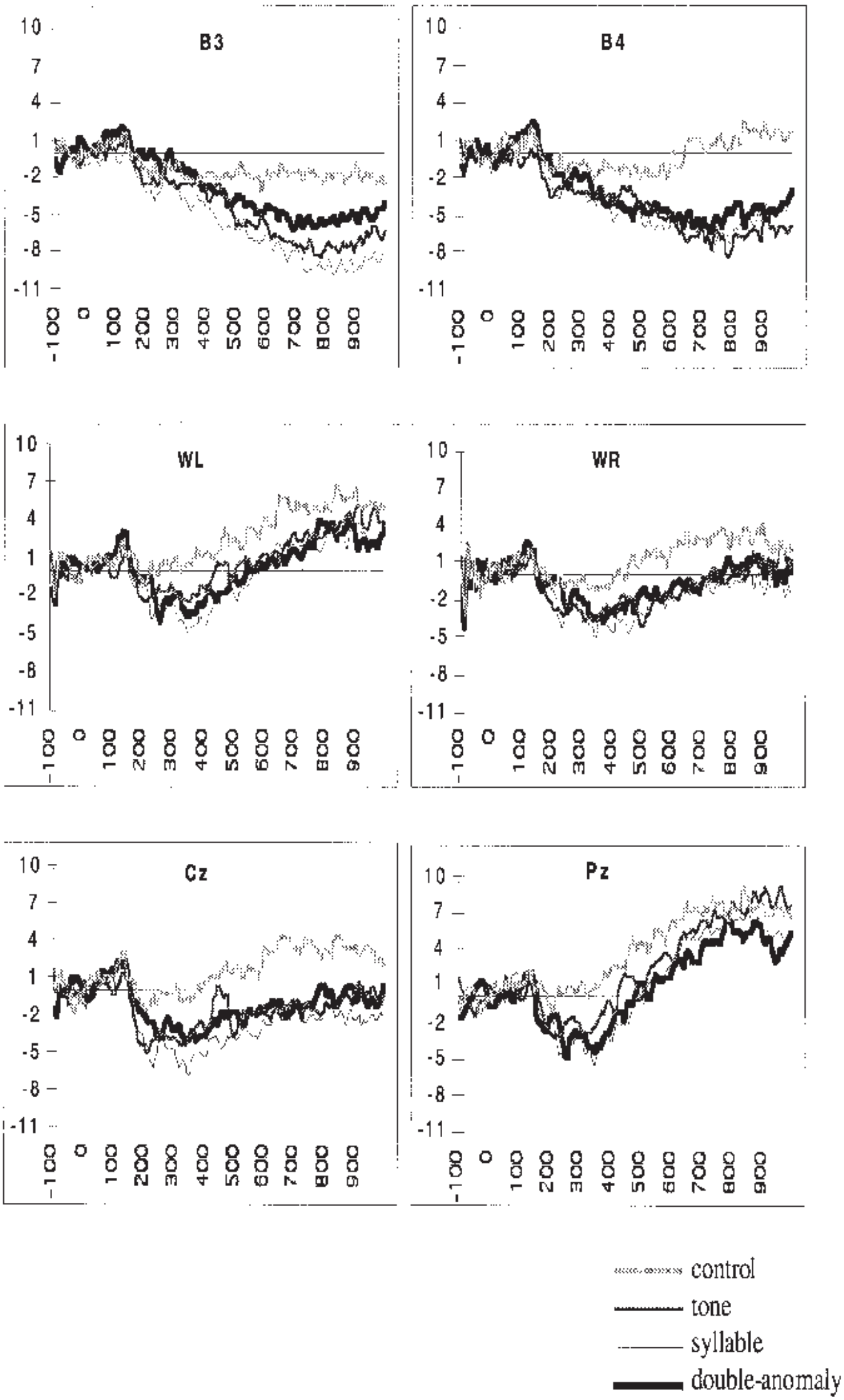

Fig. 1. Grand average ERPs (μV/ms) at six scalp electrode sites (B3, B4, WL, WR, Cz, and Pz) elicited by four different ending word conditions: control, tone, syllable, and double-anomaly. In addition to the 1000-ms poststimulus window, a 100-ms prestimulus baseline is included in the graphs.

analysis, in an attempt to analyze the pattern of effects as they unfolded over time, we separately analyzed condition effects at consecutive 50-ms windows from 150 to 550 ms poststimulus (see Van Petten *et al.* [1999] for a similar analysis). Finally, a late window between 800–1000 ms poststimulus was used to evaluate the offset of the predicted condition effects. In summary, ERP individual averages were initially subjected to three sets of analyses: (1) a large-window ANOVA to determine the basic effects, (2) a set of ANOVAs that were used to ascertain the onset of the condition effects, and (3) an ANOVA to examine the offset of the condition effects.

Large-Window Analyses

Two hundred to 800 ms poststimulus. In general, in this large window, anomalous conditions elicited ERPs with larger negativities when compared to the control condition. Mean voltage differences were calculated by subtracting the mean voltage amplitude for each anomalous condition from the control condition. In Fig. 2, we depict the mean voltage differences for each area and for each condition. The zero line on this graph indicates the amplitude of the control condition. Upon inspection of Fig. 2, the largest differences between the three anomalous conditions appear to be at the B3, WR, and Cz electrodes. To analyze condition and area effects, mean voltages from 200 to 800 ms poststimulus were calculated relative to the 100-ms baseline and submitted to a 4×6 repeated measures ANOVA with condition (levels: control, tone, syllable, double-anomaly) and area (levels: B3, B4, WL, WR, Cz, Pz) as within subject factors. This time window was chosen based on previous

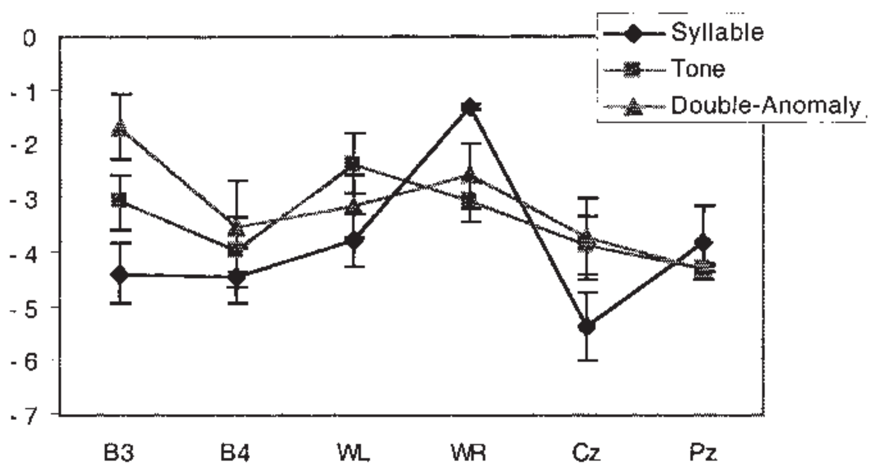

Fig. 2. Amplitude differences between control (μV) and each experimental condition for each area in the 200–800-ms poststimulus window. Error bars denote one standard error.

auditory ERP studies and on visual inspection. The ANOVA confirmed the apparent main effect of condition (G–G e = 0.743, $F[3,19] = 11.92$, $p < 0.001$). A main effect of area was also observed (G–G e = 0.510, $F[5,19] = 12.09$, $p < 0.001$).

Post-hoc analyses were used to explore the main effects of condition and area. A Scheffé test for condition effects revealed significant differences between control and all three anomalous conditions: tone versus control, $p < 0.001$ (mean difference $= -3.44$); syllable versus control, $p < 0.0001$ (mean difference $= -3.84$); and double-anomaly versus control, $p < 0.001$ (mean difference $= -3.16$). There were no differences between the anomalous conditions.

Significant area effects were largely due to larger negativities seen at anterior electrode sites. Scheffé tests (all $p < 0.01$) indicated that B3 and B4 were significantly more negative than all other areas (except Cz). Finally, the ANOVA indicated only a trend for a condition by area interaction (G–G e = 0.331, $F[15,19] = 2.00$, $p = 0.09$).

As predicted, the anomalous conditions elicited larger negativities than the control condition. The lack of a difference between anomalous conditions points to the fact that they all elicited an N400 effect of a similar magnitude. It is only in onset latency (see the early window analyses below) where some differences among the anomalous conditions were observed. Finally, the lack of a significant condition by area interaction indicates that the effect of condition was widely distributed.

Early Effects

In order to verify the differences in condition effects across the time windows, we submitted the first eight 50-ms time windows (beginning at 150 ms poststimulus) to a three-way ANOVA including condition, area, and time window as factors. Replicating the findings of the large-window analysis, we found main effects of both condition (G–G e = 0.795, $F[3,19] = 8.92$, $p < 0.0001$) and area (G–G e = 0.530, $F[5,19] = 4.18$, $p < 0.05$). Additionally, we observed a significant main effect of time window (G–G e = 0.322, $F[7,19] = 3.54$, $p < 0.05$). These main effects were qualified by a significant three-way interaction, ($F[105,19] = 1.34$, $p < 0.05$); additionally, all three two-way interactions were significant.[9] Motivated by this analysis, and in order to better characterize the onset of condition effects, we compared each anomalous condition against con-

[9] While the basic two- and three-way interactions were significant, with the G–G adjustment for sphericity of variance, the three-way interaction (G–G e = 0.105, $p = 0.21$) and two of the two-way interactions were marginal (condition*area G–G e = 0.120, $p = 0.12$: window*condition G–Ge = 0.310, $p = 0.053$; window*area G–Ge = 0.107, $p < 0.0001$). The main effect of window and the borderline window*condition interaction led us to pursue a separate window analyses as indicated in the text.

Table I. Results of Analyses of Variance with Condition (Experimental vs. Control) and Area as Factors: Main Effect of Condition.

Latency window (ms)	Tone vs. control			Syllable vs. control			Double-anomaly vs. control		
	F(1,19)	MSE	p<	F(1,19)	MSE	p<	F(1,19)	MSE	p<
150–200	3.31	106.7	0.1	2.10	97.2	NS	0.78	23.8	NS
200–250	13.77	267.6	0.01	14.23	392.4	0.01	0.79	25.2	NS
250–300	5.83	244.5	0.05	18.32	495.7	0.001	5.70	255.8	0.05
300–350	12.43	401.8	0.01	19.46	852.5	0.001	3.73	285.9	0.1
350–400	8.96	276.1	0.01	39.32	1037.9	0.0001	8.71	541.0	0.01
400–450	8.25	296.9	0.01	48.92	1204.8	0.0001	7.02	569.1	0.05
450–500	7.73	393.1	0.05	38.10	1283.30	0.0001	12.05	683.1	0.01
500–550	17.72	692.5	0.001	27.31	1234.76	0.0001	7.42	657.7	0.05

Note: The latency windows are calculated with respect to the onset of the sentence final words. MSE = mean squared error.

trol at each of the eight consecutive 50-ms time windows (based on analyses in Van Petten, *et al*., [1999]). The results of these analyses are shown in Table I.[10]

In general, across the time range analyzed, condition effects became stronger the later the time window (as evidenced by larger F values). The tone and syllable conditions onset first, differing significantly from control in the 200–250-ms window. However, the effect for the syllable condition was much larger than that of the tone condition. The double-anomaly condition onset the latest; while it first differed from control at the 250–300-ms window, it was not until the 350–400-ms window that the double-anomaly condition began to consistently differ from control.

In summary, this pattern of data suggests that while condition effects onset early (200 ms poststimulus), these effects were consistent only for the syllable and tone conditions. It was not until a later window (350–400 ms) that all three conditions were consistently and simultaneously different from control. The lack of a condition by area interaction in the large-window analysis suggests that the main effect of condition was widespread; however, visual inspection indicates that the condition effects were largest at Cz and the frontal electrodes. Additionally, the lack of a difference between the anomalous conditions in the large-window analysis suggests a uniform effect across experimental conditions. The differences between them appeared to be primarily one of timing, with the syllable and tone conditions eliciting the earliest significant negativities.

[10] We do not report the area factor because area effects were not directly relevant to our experimental questions.

Eight Hundred to 1000 ms Poststimulus

Analysis at this late window was primarily used to observe when effects seen in earlier windows tapered off. Mean voltages from 800 to 1000 ms post-stimulus were submitted to a similar ANOVA as above. The ANOVA indicated a main effect of condition (G–G e = 0.834, F[3,19] = 5.63, $p < 0.01$) and a main effect of area (G–G e = 0.431, F[5,19] = 21.38, $p < 0.001$]. Although smaller than in earlier windows, the anomalous conditions were still eliciting negativities in comparison to control. Figure 3 shows the mean voltage differences by area and condition.

The main effect of condition was explored in post-hoc analyses. A Scheffé test for condition revealed significant differences between syllable versus control, $p < 0.01$ (mean difference = −4.59) and double-anomaly versus control conditions, $p < 0.05$ (mean difference = −3.46). The difference between tone and control was borderline, $p = 0.06$ (mean difference = −3.30). There were no differences between the anomalous conditions.

In addition, the ANOVA revealed a significant condition by area interaction (G–G e = 0.388, F[15,19] = 4.70, $p < 0.001$). Separate ANOVAs were conducted for each electrode site, revealing significant condition effects at B3 (F = 11.30, $p < 0.001$), B4 (F = 9.10, $p < 0.001$), WR (F = 3.09, $p < 0.05$), and Cz (F = 4.04, $p < 0.05$). Scheffé tests (all effects reported here reached at least < 0.05 of significance) revealed that in this late window, the only condition consistently different from control across these four areas was the syllable condition. The tone condition differed from the control only in anterior areas. Finally, the double anomaly was different from the control only in B4. In summary, while some of the condition effects taper off at this later window, many of the N400 effects are long lasting. The anomaly effect elicited by the syllable condition not only showed the strongest and earliest effects, but it was also the

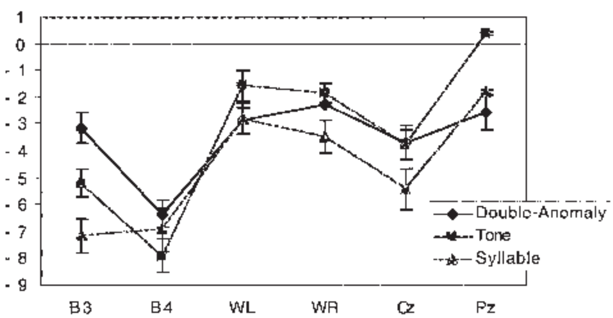

Fig. 3. Amplitude differences between control (μV) and each experimental condition for each area in the 800–1000-ms poststimulus window. Error bars denote one standard error.

longest lasting and the most consistent across areas. The other condition effects remained evident only in anterior areas.

EXPERIMENT 2: SENTENCE PLAUSIBILITY STUDY

Experimental Design

This experiment served as a follow-up to confirm the quality of the stimulus materials used in Experiment 1. Participants in Experiment 2 were asked to rate the plausibility of the 160 experimental sentence frames used in Experiment 1. As in Experiment 1, each of the 160 experimental sentence frames could end with (1) a semantically plausible word, (2) a semantically implausible word due to a tone error, (3) a semantically implausible word due to a syllable error, or (4) a semantically implausible word due to both a tone and syllable error.

Participants

Twenty native and fluent speakers of Mandarin Chinese who were able to read and write in Chinese participated in this experiment. Participants were recruited from the same pool of participants who participated in Experiment 1. Thirteen of the 20 participants in Experiment 2 also participated in Experiment 1.

Stimulus Materials

Participants were given a questionnaire with written versions (simplified Chinese characters) of the 160 experimental sentences from Experiment 1. Participants were asked to rate each sentence on a scale of -4 to 4 for how much the sentence "made sense" by circling one of the numbers. A rating of -4 meant the sentence did not make sense whereas a rating of 4 meant the sentence made perfect sense. Each sentence frame appeared once on each questionnaire. Four different questionnaires were designed so that each sentence frame appeared only once with each of its four possible endings. Questionnaires were delivered and returned through the mail. Participants were paid $5.00 for participation.

Results

Mean plausibility ratings for sentences in each condition were as follows: control = 2.53 (SD = 0.79); tone = -3.74 (SD = 2.01); syllable = -1.10 (SD = 2.08); double-anomaly = -1.170 (SD = 1.89). A one-way, repeated measures ANOVA revealed a main effect of condition, (G–G e = 0.466, $F[3,19] = 41.86$, $p < 0.001$).

We explored the condition effect with post-hoc analyses; a Scheffé test for condition indicated significant differences between all three anomalous conditions and control: tone versus control, $p < 0.0001$ (mean difference = -2.90); syllable versus control, $p < 0.0001$ (mean difference = -3.63); and double-anomaly versus control, $p < 0.0001$ (mean difference = -3.70). None of the anomalous conditions were significantly different from each other.

Discussion

The results of the plausibility study confirm the intended manipulation of sentence endings. That is, all sentences intended to have anomalous endings were rated as such. This was independent of the type of anomaly. It was slightly surprising to find that our control sentences received a mean plausibility rating of 2.53 instead of a 4 (indicating perfect sense). This could be due to a certain degree of "unnaturalness" of our materials as perceived by our participants. Nevertheless, on average our subjects judged control sentences to be plausible and all three types of experimental sentences to be implausible.

GENERAL DISCUSSION

In general, ERPs elicited by sentence ending words were highly reflective of condition. As expected, sentences ending in all three types of anomaly (tone, syllable, and double-anomaly) produced significant N400 effects between 200 and 800 ms poststimulus. This effect was apparent at early windows (starting at 200–250 ms for the syllable and tone conditions) and lasted until the latest window (800–1000 ms). These findings are consistent with the typical N400 findings in previous research. However, the current research is distinct from previous work using the N400 component in that we employed a novel way to generate anomalous sentences. The nature of Mandarin Chinese allowed us to investigate the independent and concurrent effects of tonal and syllabic information on the meaning of Chinese words. A limited number of studies have investigated N400-like brain potentials in Chinese (Lou *et al*., 1989; Valdes-Sosa *et al*., 1993), but to our knowledge, no one has employed the specific manipulation of the current study.

Our results are consistent with previous research on the N400 using the auditory modality (Holcomb & Neville, 1991; Kutas *et al*., 1987). Our N400 effects were significant at all electrode sites (in the large-window analysis), began around 200 ms for some conditions, and lasted through 1000 ms poststimulus. Comparable studies in the visual modality typically find effects beginning around 220–300 ms after word onset (e.g., Kutas *et al*., 1987; Van Petten, 1993). In the current study, additional analyses at 800–1000 ms poststimulus were conducted based on similar work in the auditory modality by Holcomb and Neville (1991), and Kutas *et al*. (1987). Concurrent with the findings in both of these studies, we found large and long-lasting negativities for anomalous conditions. The late plausibility effects (at 800–1000 ms) were strongest at B3, B4, WR, and Cz.

An additional note is that the effects we observed were in general strongest at Cz and frontal areas. The N400 effect is typically described as having a centro-parietal distribution. To be sure, we did find significant condition effects in all areas, and the lack of an area by condition interaction in the large-window analysis suggests our effects were widespread. Additionally, the timing of the effect is consistent with other N400 studies using auditorially presented stimuli (Holcomb & Neville, 1990, 1991). A possible explanation for this atypical distribution is that we are observing the superposition of the N400 with another effect with a more frontal distribution. One suggestion is that we are observing the effects of a left anterior negativity (LAN), which is typically thought to be associated with language tasks that tax working memory (see Rōsler *et al*, 1998). However, a working memory account of our findings is not clear, and the broad distribution of our effects (with a bilateral anterior maximum) do not clearly support a LAN account. We wish to conclude tentatively that our results are consistent with an N400 interpretation, but we cannot rule out other interpretations. What is important to draw from our data is the fact that all three anomalous conditions elicited negativities that were significantly different from control, and that we observed some timing differences between anomalous conditions.

With respect to the long-lasting nature of the negativities we observed, Kutas *et al*. (1987) suggest that variations in auditory stimulus length may create a "jitter" in auditory ERP effects, resulting in longer lasting effects in auditory, as compared to visual, conditions. This type of jitter may as well have been responsible for the long-lasting negativities we found. The grand average plots reported in their paper show a similar pattern to ours, with condition effects lasting longest primarily in anterior areas.

The fact that our effects onset so early suggests that we are not simply seeing effects of phonological processing in the form of an N450 as reported by Rugg (1984a, 1984b). If we were looking at N450 effects, we would expect the data to show larger amplitude differences between anomalous conditions, as they are quite different from one another on a phonological level. Additionally, our N400s differed from those seen in Rugg (1984a), and the phonological word priming study by Radeau *et al*. (1998) both in their timing and distribution. While our effects were large at all electrode sites, we observed the largest negativities at anterior sites and Cz. In contrast, the N450 described in Rugg (1984a) was largest over temporal sites. A cautionary note is that the Rugg study did not have lateral frontal electrodes; the only frontal electrode was Fz (which we did not use). Given the differences in electrode placement, comparisons between the two studies in terms of distribution of effects is difficult. Future work should use more electrode sites and compare the phonological type effects seen by Rugg with our tone and syllable anomalies.

In contrast to the Radeau *et al.* study, our earliest effects onset around 200 ms poststimulus whereas the onset of their phonological priming effect was around 300 ms (though upon inspection of their waveforms it looks as if the semantic priming effect may have started slightly earlier). In the current study, condition effects were uniform and widespread; the condition by area interaction was not significant in the large-window analysis. However, the interaction was significant at selected windows. In particular, the N400 effect was longest lasting at anterior electrodes (B3 and B4), WR, and Cz. On the contrary, in the Radeau *et al.* study, the priming effects were strongest at posterior electrode sites. However, comparisons between these studies should be taken with caution as the stimulus materials were quite different. Most importantly, the stimuli in the current study were within sentence contexts whereas those in Radeau *et al.* were isolated words.

Kutas *et al.* (1987) also found early effects for N400s elicited by auditory stimulus materials. They attribute early onsets to coarticulatory cues present in natural speech which would give listeners early access to upcoming words (see also Holcomb & Neville [1990; 1991] and McCallum *et al.*, [1984] for an attempt to control for coarticulation). This explanation seems plausible, especially in light of findings by Van Petten *et al.* (1999) who found that the initial sounds of words provided enough information to see N400 effects for contextually inappropriate (in comparison to appropriate) words. In this study, they first used a gating paradigm to identify the isolation points of several words. The isolation point is the average time at which people can uniquely identify a word. Van Petten *et al.* (1999) demonstrated an N400 effect approximately 200 ms before this isolation point. However, an important distinction must be made between the stimulus materials used in the Van Petten *et al.* study and those used in the current paper. In the current experiment, the sentence frames used in the experimental sentences were identical across conditions and the final words spliced off (and not heard by participants). The ending words that participants actually heard were produced in isolation and presented immediately following this sentence frame. As a result, the only condition in which coarticulatory (as well as semantic) information was consistent with the ending word was in the control condition. However, all ending words were similarly awkward sounding as they were all generated in the same way, and thus “didn’t flow” with the sentence in the same way (see “Stimulus Materials”, above). Even so, the Van Petten *et al.* (1999) results are still relevant to the current study, as they suggest that early acoustic information in the ending words can quickly influence semantic processing, even before the uniqueness point (see also Allopenna *et al.*, 1998). Additionally, coarticulatory cues present in words before the ending word in our study could have increased the expectation for the correct ending word. This increase in expectation (similar to a high Cloze probability) may have amplified and extended our effects.

Given that all three types of anomalous sentences were nonsensical, we hypothesized that they would elicit similar N400 effects. If there were to be any differences among the anomalous conditions, we expected the double-anomaly condition to elicit the strongest N400 effect. Our prediction was based on the notion that ending words in the double-anomaly condition contained more cues that they were the wrong word for that sentence context. Previous research has demonstrated that less-associated and less-expected words tend to elicit N400s with larger amplitudes (Brown & Hagoort, 1993; Kutas, 1993; Kutas & Hillyard, 1984; Kutas *et al*., 1987; Kutas *et al*., 1988). However, in spite of slightly lower plausibility ratings for the double-anomaly, this condition elicited the least consistent and most delayed effect.

The finding that the syllable condition elicited the strongest and earliest negativities is consistent with results by Chen *et al*. (2002), who used an auditory priming technique to look at differential priming effects of the syllable, the tone, and the tone + syllable in Chinese compound words (see earlier description of this study in the "Introduction"). They found priming for the syllable and the tone + syllable conditions, but not for the tone condition. This result indicates that speakers may be better able to prepare for the syllable, but not for the tone alone. Because the task was very different, it is difficult to compare their study with ours, but we would simply like to note that the stronger effects for the syllable condition in our study may be related to their finding that speakers were able to prepare for the syllable ahead of time. One possibility is that the syllable unit is more anticipated (in both production and comprehension) than the tone unit. Further investigations of the differential nature of the tone and syllable in Chinese could exploit some of the differences (phonetic, or production based) between tonal and syllabic information to better understand how they are processed. For instance, given that the types of speech errors for tones and syllables tend to be qualitatively different (Chen, 1999), it would be fruitful to replicate our study, using syllable and tone errors that are either consistent or inconsistent with the typical patterns of syllabic and tonal speech errors. If listeners distinguish the different types of tonal and syllabic errors, this might indicate that they are sensitive to fine-grained details of the tones and syllables.

Although these considerations may explain the differences in the strength of the effect between the syllable and tone conditions, we are still left with the problem of accounting for the slight delay in the N400 elicited by the double anomaly. One possibility is that this could be due to differences in plausibility ratings. However, the ending words in the double-anomaly condition were actually rated as least plausible in the offline sentence rating task (the mean plausibility for the double-anomaly was the lowest, although it was not significantly different from the other anomalous conditions). Therefore, the delayed N400 effect for the double-anomaly condition is unlikely to be due to plausibility differences.

We can think of another, unlikely yet possible explanation for the slight delay in the onset of the effect for the double-anomaly condition. We need to present the argument in two parts. First, there are a large number of homophones in Chinese. If you look in a Chinese dictionary at a particular syllable + tone combination, you are likely to see listed at least two different potential meanings for that sound. In our experimental situation, we talk about *the* "expected word," but individual subjects may have considered a different or multiple other possible words as the ending word for the sentence frame. In addition, the large number of homophones in Chinese increases the likelihood that a participant could briefly expect a different ending word with a similar onset to the presented ending word.[11] While experimental sentences were designed to be highly predictive of final words, this design characteristic alone does not rule out the possibility that when listening to the sentence context and the initial sounds of an ending word, a participant would be expecting a different (and perhaps valid) ending word. In this situation, the participant may be slow to recognize the anomaly.

Second, the possibility that participants anticipated a different ending word than *the* expected ending word does not explain by itself the apparent delayed onset in the double-anomaly condition because this possibility exists for all four of the experimental conditions. However, although we do not have a firm idea of why we observed a delayed effect for the double-anomaly condition, the explanation could relate to the process of incremental word recognition and the detection of errors. More specifically, in the tone and syllable conditions, there was at least some overlapping sound information between *the* expected word and the one presented (syllable and tone information, respectively). In these cases, the listener may have been more likely to think of *the* expected word because of phonological priming between the presented and expected word. Recall the findings by Allopenna *et al.* (1998) that sound information at both the beginning and endings of words activates lexical candidates. Increased activation of *the* expected word may have made subjects more aware of this intended ending word, thus directing attention to the error. Noticing the discrepancy between the expected and presented words may have triggered the earlier error detection seen in the tone and syllable conditions. In contrast, in the double-anomaly condition, the presented word sounded nothing like *the* expected word (onset, rhyme, and tone were different). In this case, participants may have been slower to identify the ending word as anomalous because they were entertaining other possible ending words

[11] Although an important experimental design note is that we avoided as much as possible the use of ending words that had homophones that were sensible endings for the sentence fragments.

that shared features (or were homophones with) the presented, erroneous ending word but were consistent with the sentence context. The essence of the argument is that it may take longer to identify something as erroneous when it is completely unrelated as compared to situations when the erroneous word is similar enough to the target word to be identified as a deviation from it.

CONCLUSIONS

In conclusion, the data gathered in the current study are consistent with findings in previous research using the N400 in the auditory modality. We replicated previous findings that auditorially elicited N400s onset earlier and last longer than those elicited in the visual modality. We demonstrated a robust N400 effect for all three types of anomalies in Mandarin Chinese. It is now clear that tonal information is used in online sentence processing and is quickly used to identify word meaning (within 200 ms of word onset) and to identify inconsistencies in sentential semantic structure (revealed by an early N400 response to tonal anomalies). Given the range of ideas about how tonal information might be processed (as linguistic information or as pure tonal information), these results clearly suggest that lexical tones are processed as linguistic information.

While lexical tones are similar to pure tonal information in form, the former clearly influence semantic processing. Though there may be some similarities between English lexical stress and Mandarin's lexical tones, there are important differences between the two (see Chen [1999] for some of their similarities). It is clear that the stress patterns in English influence segmentation processes (Cutler & Norris, 1988) and that contrastive stress can influence the earliest moments of reference resolution (Sedivy *et al.*, 1999). However, even these dramatic effects pale in comparison to the effect of lexical tone in Mandarin. Unlike stress in English, in Chinese many words have minimal pairs that can be distinguished only by lexical tone. Additionally, changes in lexical tone can cause a change in meaning (e.g., *mother* for *horse*) but also syntactic category information (e.g., *mother* for the verb *to scold*). In future work, a comparison of the processing of lexical stress and lexical tone might yield interesting insights about how these two cues are related and could potentially lend support to Chen's theory about the similarities between lexical stress and lexical tone.

The current study thus supports the wide range of offline and neuropsychological evidence in favor of the idea that lexical tones are processed as linguistic information. Even more exciting than the replication of the N400 effect in Chinese and our findings about lexical tone is our new approach to creating sentential anomalies. We manipulated different features of words (tonal, onset, and rhyme information) and found that all of them affect the processing of semantic information. We can see direct extensions of this work into the study of when and how different features of words are used to guide our predictions about the meaning and form of sentences.

Differences in the identification of anomalous words as such, or so-called jittery detection of errors, is likely directly related to the use of auditorially presented stimuli. Auditory stimuli may be more likely than visual stimuli to reveal effects of the incremental processing of words (see Allopenna *et al*., 1998; Van Petten *et al*., 1999). The delayed effect for the double-anomaly condition we observed is likely due to a jittery or delayed detection of the anomaly in this condition. Nevertheless, the fact is that all three anomalous conditions elicited a long-lasting and widespread N400 effect.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence from continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., Rauschecker, J. P., Braun, A., Turner R., & Neville, H. J. (1998). Hemispheric specialization for English and ASL: Left invariance-right variability. *Neuroreport, 9*(7), 1537–1542.

Bavelier, D., Corina, D., Jezzard, P., Padmanabhan, S., Clark, V. P., Karni, A., Prinster, A., Braun, A., Lalwani, A., Rauschecker, J. P., Turner, R., & Neville, H. J. (1997). Sentence reading: A functional MRI study at 4 Tesla. *Journal of Cognitive Neuroscience, 9*, 664–686.

Besson, M. (1998). Meaning, structure, and time in language and music. *Cahiers de Psychologie Cognitive, 17*(4–5), 921–950.

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4*(2), 132–149.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition, 8*(6), 631–642.

Breier, J. I., Simos, P. G., Zouridakis, G., & Papanicolaou, A. C. (1999). Lateralization of cerebral activation in auditory verbal and non-verbal memory tasks using magnetoencephalography. *Brain Topography, 12*(2), 89–97.

Broca, P. (1861). Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphémie. *Bulletin de la Société d'anatomie, 5*, 330–357.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34–44.

Carlson, N. R. (1994). *Physiology of Behavior*, Paramount: Needham Heights, MA.

Chen, J-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics, 20*(2), 289–301.

Chen, J-Y., Chen, T-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751–781.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers, 25*, 257–271.

Corina, D. (1998). Aphasia in users of signed languages. In P. Coppens, Y. Lebrun, & A. Basso (Eds.), *Aphasia in Atypical Populations* (pp. 261–309). Mahwah, NJ: Lawrence Erlbaum.

Cutler, A., & Norris, D. (1998). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113–121.

Edmondson, J. A., Chan, J-L., Seibert, G. B., & Ross, E. D. (1987). The effect of right-brain damage on acoustical measure of affective prosody in Taiwanese patients. *Journal of Phonetics, 15*, 219–233.

Elias, L. J., Bulman-Fleming, M. B., & McManus, I. C. (2000). Linguistic lateralization and assymetries in interhemispheric transmission time. *Brain & Cognition, 43*(1–3), 181–185.

Gandour, J. (1998). Aphasia in tone languages. In P. Coppens, Y. Lebrun, & A. Basso, (Eds.), *Aphasia in Atypical Populations*. London: Lawrence Erlbaum.

Gootjes, L., Raji, T., Salmelin, R., & Hari, R. (1999). Left-hemisphere dominance for processing of vowels: A whole-scalp neuromagnetic study. *Neuroreport. 10*(14), 2987–2991.

Ho, C. S. & Bryant, P. (1997). Development of phonological awareness of Chinese children in Hong Kong. *Journal of Psycholinguistic Research, 26*(1), 109–126.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes, 5*, 281–312.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology, 19*(4), 286–300.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*(2), 3–42.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.

Kimura, D. (1973). The asymmetry of the human brain. *Scientific American, 228*(3), 70–78.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196–214.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes, 8*(4), 533–572.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(11), 203–204.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Kutas, M., & Kluender, R. (1994). What is who violating: A reconsideration of linguistic violations in light of event-related brain potentials. In H-J. Mūnte, T. F., Mūnte, & G. R. Mangun, (Eds.), *Cognitive Electrophysiology*. Boston: Birkhauser.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*. San Diego, CA: Academic Press.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory States and Processes*. Hillsdale, NJ: Lawrence Erlbaum. (pp. 217–237).

Kutas, M., Neville, H. J. & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing, In R. J. Ellingson, N. M. F., Murray, & A. M. Halliday, (Eds.), *The London Symposia, Electroencephalography and Clinical Neurophysiology* (suppl. 39) (pp. 325–330). Amsterdam: Elsevier.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*, 218–233.

Ladefoged, P., & Buchholz, T. (Eds.), (1993). *A Course in Phonetics*, (pp. 254–556,564). Fort Worth, TX: Harcourt Brace.

Langdon, D., & Warrington, E. K. (2000). The role of the left hemisphere in verbal and spatial reasoning tasks. *Cortex, 36*(5), 691–702.

Leben, W. R. (1978). The representation of tone. In V. A. Fromnkin (Ed.), *Tone: A linguistic survey* (pp. 177–219). New York: Academic.

Lee, Y., Vakoch, D. A. & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25*(5), 527–542.

Lehtonen, J. B., & Koivikko, M. J. (1971). The use of a non-cephalic reference electrode in recording cerebral evoked potentials in man. *Electroencephalography and Clinical Neurophysiology, 31*, 154–156.

Li, C. N., & Thompson, S. A. (1977). A mechanism for the development of copula morphemes. In C. Li (Ed.), *Mechanisms of syntactic change.* Austin: University of Texas Press.

Li, C. N., & Thompson, S. A. (1979). Chinese: Dialect variations and language reform. In T. Shopen (Ed.), *Languages and their status* (pp. 295–335). Cambridge, MA: Winthrop Publishers, Inc.

Lou, L-G., Fan, S-L., & Kuang, P-Z. (1989). Event-related brain potentials (ERPs) reflect mismatch between Chinese character and its mental template. *Acta Psychologica Sinica, 21*(3), 321–327.

McCallum, W. C., Farmer, S. F., & Pocock, P. K. (1984). The effects of physical and semantic incongruities on auditory event related potentials. *Electroencephalography and Clinical Neurophysiology, 59*, 477–488.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15–22.

Packard, J. L. (1986). Tone production deficits in nonfluent Aphasic Chinese speech. *Brain and Language, 29*, 212–223.

Parkin, A. J. (1996). *Explorations in Cognitive Neuropsychology*. Cambridge, MA: Blackwell.

Pratarelli, M. (1994). Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain and Cognition, 24*, 137–157.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A-S., & McNamara, J., Eds. (1997). *Neuroscience*. Sunderland, MA: Sinauer Associates.

Radeau, M., Besson, M., Fonteneau, E., & Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: Behavioral and electrophysiological evidence. *Biological Psychology, 48*, 183–204.

Rasmussen, T., & Milner, B. (1977). Clinical and surgical studies of the cerebral speech areas in man. In K. J. Zulch, O. Creutzfeltd, & G. C. Galbraith, (Eds.), *Cerebral localization* (pp. 238–255) New York: Springer-Verlag. .

Repp, B. H., & Lin, H-B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18*, 481–495.

Rōsler, F., Pechmann, T., Streb, J., Rōder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language, 38*(2), 150–176.

Rugg, M. D. (1984a). Event-related potentials and the phonological processing of words and nonwords. *Neuropsychologia, 22*(4), 435–443.

Rugg, M. D. (1984b). Event-related potentials in phonological matching tasks. *Brain and Language, 23*, 255–240.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*(2), 109–147.

Shipley-Brown, F., Dingwall, W. O., Berlin, C. I., Yeni-Komshian, G., & Gordon-Salant, S. (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain and Language, 33*, 212–223.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191–196.

Stephenson, W. A. & Gibbs, F. A. (1951). A balanced noncephalic reference electrode. *Electroencephalography and Clinical Neurophysiology, 3*, 237–240.

Taylor, W. L. (1953). “Cloze” procedure: A new tool for measuring readability. *Journalism Quarterly, 30*, 415.

Valdes-Sosa, M., Gonzalez, A., Xiang, L., Xiao-Lei, Z., Yi, H., & Bobes, M. A. (1993). Brain potentials in a phonological matching task using Chinese characters. *Neuropsychologia, 31*(8), 853–864.

Van Lanker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and “tone”: Evidence from Thai. *Journal of Phonetics, 1*, 101–109.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes, 8*(4), 485–531.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: LMC, 25*(2), 394–417.

Wan, I-P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology, 15*, 417–461.

Wang, W. S-Y. (1973). The Chinese language. *Scientific American, 228*, 50–60.

Zurif, E. B. (1995). Brain regions of relevance to syntactic processing. In *An Invitation to Cognitive Science: Language, Vol. 1*. L. R. Gleitman, M. Liberman, & D. N. Osherson (Eds.), Cambridge, MA: MIT Press.