Getting the Gist is not Enough: An ERP Investigation of Word Learning from Context

Arielle Borovsky (aborovsk@cogsci.ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive #0515
La Jolla, CA 92093-0515
Jeff Elman (elman@cogsci.ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive #0515
La Jolla, CA 92093-0515
Marta Kutas (kutas@cogsci.ucsd.edu)
Department of Cognitive Science, 9500 Gilman Drive #0515
La Jolla, CA 92093-0515

Abstract

We examined contextual effects on word learning using event-related brain potential (ERP) methodology. Adult participants first read known and unknown novel words in strongly and weakly constraining sentence contexts. Then, to assess acquisition of word meaning, they rated the plausibility of probe sentences in which these same words served as objects of a transitive verb. Plausibility effects were observed in the N400 component to the verb only when the upcoming novel word object had initially appeared in a strongly constraining context. These results demonstrate differences in electrical brain activity related to rapid, one-shot contextual word learning.

Keywords: word learning; N400, event-related brain potentials; language learning

Introduction

Humans have an amazing ability to learn new words throughout their lifespan. While there is an enormous body of research on word learning, much of it has focused either on the factors involved in early language learning in children, or on second language learning in adults. As a result, we know surprisingly little about word learning by adults in their native language. In addition, much of what is known about the neural and cognitive correlates of word learning (in children) comes to us from studies involving early object learning in explicit language training paradigms. However, the majority of words that we know are not objects and their meanings are not learned explicitly. Instead, most new words are acquired incidentally, via surrounding words in sentence contexts (Nagy & Herman, 1987; Sternberg, 1987). The goal of the present work is thus to study the cognitive and neural underpinnings of incidental word learning by adults in their native language. More specifically, in this study we explore the role of contextual constraint in word learning by measuring how an event-related brain potential (ERP) component – the N400, can index knowledge of word meaning and usage.

Background

Research on the neural mechanisms of word learning has recently become a topic of much interest, especially for investigators studying word learning in toddlers (i.e., Friedrich & Friederici, 2004; Friedrich & Friederici, 2005a; Friedrich & Friederici, 2005b; Mills, Coffey-Corina, & Neville,1997; Mills, Plunkett, Part & Schafer, 2005) and in adult second language learners (McLaughlin, Osterhout & Kim, 2004; Osterhout, McLaughlin, Kim, Greenwald & Inoue, 2004; Stein, Dierks, Brandeis, Wirth, Strik, & Koenig, 2006; Ojima, Nakata, Kakigi, 2005). Developmental research has revealed the emergence of adult-like patterns of brain activity during lexical processing from 13 to 20 months. Additionally, research in adults finds evidence for rapid neural change due to word learning in both L2 and L1. For example, McLaughlin and colleagues (2004) find that a few months of college language instruction can yield brain responses to L2 words in a lexical decision task that are indistinguishable from native speakers of the same language. This work demonstrates that changes in neural activation reflect lexical acquisition and that the brain appears to process word meanings that are acquired in adulthood and childhood similarly -- even though the adults in this study did not show native-like behavioral performance. L1 word learning also can result in rapid neural change (Perfetti, Wlotko & Hart, 2005; Mestres-Missé, Rodriguez-Fornells & Münte, in press). For example, Mestres-Missé and colleagues (in press) found that three presentations of a novel word in meaningful sentence contexts can result in significant change in neural responses to words.

These ERP studies of word learning have focused upon one particular brainwave component, the N400. The N400 is a negative going wave with a centroparietal maximum that peaks approximately 400ms after the onset of any potentially meaningful stimulus (Kutas & Hillyard, 1980). Its amplitude has been found to decrease when a word is more expected or when features associated with its meaning are more easily integrated within its surrounding context (Kutas & Hillyard, 1980; Van Petten & Kutas, 1990; Federmeier & Kutas, 1999). For example, Kutas and Hillyard (1980) recorded brainwave responses to sentence completions that were either congruent or incongruent with the context of the preceding sentence. In a sentence like: “I drink my coffee with cream and *sugar*” where the sentence ending was congruent given the sentence context, the evoked N400 response was much lower than to sentences like "I drank my coffee with cream and *dog"* where the sentence completion was not congruent to the sentence context. It has also been found that one of the best predictors of N400 amplitude to words in sentential context is cloze probability (Kutas & Hillyard, 1984). Cloze probability is measured by determining the probability that a particular word is given in a context on a sentence completion task. For words with low cloze probabilities, the N400 is large, and the N400 decreases accordingly as cloze probability increases. This suggests that the N400 is not only enhanced to semantically anomalous words, but also is related to a word's degree of expectancy or ease of which its meaning may be integrated with the surrounding context. Additionally, the N400 for pronounceable nonwords is larger than that of real words (or about the same as for low frequency words), but is not present for true nonwords that do not have orthographically legal spellings, or are unpronounceable (Bentin, 1987; Kutas & Hillyard, 1980). Therefore, N400 amplitude is associated with a word's meaningfulness in a given context, ranging from very small in amplitude when a word is very easily integrated or understood, to very large when the meaning of a word is unknown.

Altogether, these findings suggest that the N400 is likely to vary with the degree to which the meaning of a newly encountered word is known. As described above, N400 amplitude in second language learners is reduced commensurate with experience with the second language (McLaughlin, Osterhout & Kim, 2004). Mestres-Missé and colleagues (in press) also reported that the N400 component amplitude changes as new words are incrementally learned over several more progressively constrained sentences, but this does not happen when the contexts do not make sense of the word. It is therefore likely that changes in contextual constraint may also result in N400 changes that are associated with word learning. In our task, we vary the degree to which a single, precise meaning of a novel word can be predicted from context by varying the cloze probability of the context in which novel words appear. By comparing how words are learned in strongly and weakly constrained contexts, we can better understand how rapid changes in neural activity associated with word learning are related to knowledge of word meaning.

Additionally, unlike previous work, we gauge successful word meaning acquisition in a task that does not require identification of a word's precise definition. We believe that this more closely mimics natural language learning as it is uncommon to generate an explicit definition for a novel word. Instead, it is more likely that in everyday context, subsequent usage and comprehension of a novel word rely upon an appreciation of its appropriate usage. Accordingly, in our task, participants indicate whether novel words are and are not being used appropriately. As reviewed above, words that appear in plausible contexts typically show a reduced N400 when compared to those that appear in implausible contexts. Thus it is possible to gauge how contextual constraint influences acquisition of word meaning by contrasting responses to subsequent plausible and implausible usages of the word. Unlike prior work, in which learning was inferred from changes in the response to the novel word upon its repetition, we examine the consequences of novel (noun) word "learning" on verb processing in a sentence plausibility task. Thus, if the N400 amplitude to the verb is modulated by the acceptability of the upcoming novel word as its object, we can infer that some of its meaning was learned on its initial appearance. We can further infer whether what is learned is more or less precise from an interaction between the plausibility effect and contextual constraint.

Methods

Participants:

26 college students (9 M, 17 F) were given credit or paid $7/hr for their participation. Ages ranged between 18-25 (mean: 19.8). All participants were right handed, native English speakers, and had no significant exposure to another language at least before the age of 12. Participants reported no history of mental illness, learning disability, language impairment, drug abuse, or neurological trauma. All participants had normal hearing and normal (or corrected to normal) vision. An additional 11 participated but were not analyzed: 4 had excessive blinking or motion artifact, 3 because of equipment failure or experimenter error, and 4 reported a characteristic which disqualified them from analysis (3 had significant second language exposure as a child, 1 reported significant illicit drug use.)

Stimuli:

Two types of sentences were selected for the study: 1) Context sentences, which provide strong and weak constraint contexts for real and novel word targets; and 2) Test sentences, which were very short sentences, following the form "Pronoun – Transitive Verb – Article – Target word". 80 strongly constrained and 80 weakly constrained Context sentences were selected from Federmeier and Kutas (2005). Strongly constraining sentence contexts were paired with the endings obtaining the highest cloze probability, and these same 80 words were then paired with the weakly constraining contexts to yield plausible, low cloze probability endings. Each sentence pair was assigned one of 80 pronounceable nonwords to serve as an alternate ending for each sentence. This arrangement yielded four main Context sentence conditions with 40 sentences each: 1) High constraint sentences with real word endings (High/Real), 2) High constraint sentences with novel word endings (High/Novel) 3) Low constraint sentences with Real word endings (Low/Real) and 4) Low constraint sentences with Novel word endings (Low/Novel) Table 1 presents illustrative examples of sentence stimuli in each condition. Across all versions of the experiment, sentence-final target words in the two constraint conditions were counterbalanced, such that the same sentence and target ending did not appear twice in any version, but all possible combinations of sentence / target word pairs appeared in all versions. The purpose of this was to ensure that any differences in performance due to the properties of the words were balanced out across conditions.

Table 1. Sentence examples in each condition

Context Sentences	
High/Real:	I just installed a new word processing program on my COMPUTER
High/Novel:	I just installed a new word processing program on my PRESE.
Low/Real:	His seat in the small classroom was next to the COMPUTER.
Low/Novel:	His seat in the small classroom was next to the PRESE.
Test Sentences	
Plausible:	He *moved* the COMPUTER / PRESE.
(Real/Novel)	She *saw* the COMPUTER / PRESE.
Implausible:	He *sipped* the COMPUTER / PRESE.
(Real/Novel)	She *greeted* the COMPUTER / PRESE.

In addition to each context sentence, four additional Test sentences were created for each high/low constraint sentence pair for use in a plausibility judgment task. Two of these sentences presented an implausible usage of the target word, and two were plausible. Plausibility of the Test sentence contexts was confirmed in a separate norming study, where participants rated each sentence and target word combination for plausibility. All four sentences were used in each version, but not repeated, and they were counterbalanced across all versions of the study. Table 1 shows examples of Test sentence contexts used for plausibility ratings.

Procedure:

Participants were tested in a single experimental session conducted in a soundproof, electrically-shielded chamber and were seated in a comfortable chair in front of a monitor. Participants were instructed to read the context sentences for comprehension and to do their best to try to comprehend the sentence even when "nonsense" words appeared on the screen. For the plausibility task, they were instructed to decide if each real and novel word was being used appropriately based upon how they had seen it used previously. Participants were not instructed to generate an explicit meaning for novel words, and the pace of the task required decisions to be made quickly so as to minimize explicit naming strategies.

The presentation of each sentence was preceded by a series of crosses (500 ms duration with a stimulus-onset-asynchrony varying randomly between 300 and 800 ms) to orient the participant toward the center of the screen. Sentences were then presented one word at a time, each for a duration of 200 milliseconds with a stimulus-onset-asynchrony of 500 milliseconds. Participants were asked to minimize blinking and movement as much as possible during sentence presentation. The final, target word appeared on the screen for 1400 ms, and was immediately followed by RSVP presentation of each plausibility sentence. Each plausibility sentence was preceded by a series of question marks (400ms duration with a stimulus-onset asynchrony varying randomly between 100 and 300ms) to prepare the participant to make a plausibility decision at the end of the sentence. Plausibility sentences were presented with identical timing as context sentences, with target words remaining on the screen for 1400ms. Participants were asked to make their decision as soon as possible after the appearance of each final target word in the plausibility sentence. There was short break after every 40 sentences, about every 10 minutes. At the conclusion of the recording session, participants were compensated and debriefed.

Recording

Scalp potentials were continuously recorded from 26 geodesically arranged sites using an ElectroCap with tin electrodes. A left mastoid reference was used. Potentials were digitized at a sampling rate of 250 Hz and hardware bandpass filter of 0.1-100Hz with Grass Amplifiers. Impedances were kept below 5 kΩ. The ERPs were stimulus-locked averages consisting of a 100-ms baseline and a 920 post-stimulus interval.

Data Analysis:

Data were re-referenced offline to a average mastoid reference. Trials contaminated by eye movements, blinks, excessive muscle activity, or amplifier blocking were rejected off-line before averaging. ERPs were computed for epochs extending from 100 milliseconds before stimulus onset to 920 milliseconds after stimulus onset. Averages of artifact-free ERP trials were calculated for target words in the four sentence-final conditions (High/Real, High/Novel, Low/Real, Low/Novel) as well as to verbs in all four plausibility conditions (Plaus/Real, Plaus/Novel, Implaus/Real, Implaus/Novel) after subtraction of the 100 millisecond pre-stimulus baseline.

Results

Behavior

Accuracy: Tables 2 shows the accuracy results for the plausibility judgment task. As seen in this table, accuracy ranged between 69% and 90% depending on condition. A (2 x 2 x 2) ANOVA was carried out, with factors of

Constraint (Low and High), Word Type (Real and Novel) and Plausibility (Plausible and Implausible). This analysis revealed main effects of Constraint [F(1,200)=25.4, p<0.0001] and Word Type [F(1,200)=45.6, p<0.0001], but not Plausibility [F(1,200)=1.79, p=.18]. There was also a Constraint x Type interaction [F(1,200)=5.76, p=0.0173] and Type x Plausibility interaction [F(1,200)=15.82, p<0.0001]. Post-hoc Tukey pairwise analyses revealed that responses to High and Low constraint Real words did not differ from each other, but that Novel/Low words were less accurate than all other conditions, and Novel/High words differed from all other conditions except Real/Low. Additional Tukey analyses revealed no differences in accuracy between Novel words in both plausibility conditions, and Real/Implausible words, but all other combinations were different from each other. These analyses were identical when conducted on arcsin transformed percentages, suggesting these effects were not driven by Real word accuracies being at ceiling.

Table 2. Accuracy of responses to plausibility task

Condition	% Accuracy (SD)
High/Real	
Plausible	90 (2.84)
Implausible	88 (3.20)
Overall	89 (3.02)
High/Novel	
Plausible	80 (4.09)
Implausible	84 (3.84)
Overall	82 (4.02)
Low/Real	
Plausible	89 (2.61)
Implausible	81 (3.52)
Overall	85 (3.55)
Low/Novel	
Plausible	69 (3.12)
Implausible	75 (4.54)
Overall	72 (4.07)

Reaction Time: Response times (shown in Table 3) varied between 603 and 794 ms, depending on condition. A (2 x 2 x 2 x 2) ANOVA was carried out with factors of Constraint (High and Low), WordType (Real and Novel), Plausibility (Plausible and Implausible) and Accuracy (Correct and Incorrect). However, this ANOVA analysis revealed only a main effect of Accuracy [F(1,385)=31.85, p< 0.0001], with Incorrect responses being slower than Correct ones. No other main effects or interactions were observed.

Event-related potentials

Context Sentence Endings: Brain potentials to the terminal words of the context sentences are shown in Figure 1. As seen from this figure, ERP responses to Real words show a difference peaking around 400 ms (N400) and largest at centro-parietal sites, while responses to Novel words do not show a constraint difference in this time region. Additionally, in both Real and Novel target conditions, an extended positivity (LPC) is observed for High constraint items ranging from Central to Posterior sites.

Table 3. Plausibility decision times (in milliseconds)

Condition	Correct (SD)	Incorrect (SD)
High/Real		
Plausible	630 (141.15)	753 (225.46)
Implausible	692 (130.58)	784 (210.6)
High/Novel		
Plausible	637 (155.77)	794 (196.34)
Implausible	667 (108.48)	780 (268.89)
Low/Real		
Plausible	603 (171.32)	769 (249.6)
Implausible	673 (151.31)	725 (160.22)
Low/Novel		
Plausible	657 (139.66)	714 (177.35)
Implausible	701 (103.28)	739 (185.66)

N400 amplitude was measured in a time period ranging from 300 to 500 ms at four central electrode sites where N400 effects are typically largest (RMCe, LMCe, MiCe, MiPa). A (2 x 2) repeated measures ANOVA was carried out on these measurements with factors of Word Type (Real/Novel) and Constraint (High and Low). This analysis revealed main effects of Constraint [F(1,25)=24.92 p<0.0001], Type [F(1,25)=22.44, p<0.0001], as well as a Constraint x Type interaction [F(1,25)=23.45, p<0.0001]. Follow-up posthoc Tukey tests revealed that these effects were driven by Real/High target endings being significantly more positive than every other condition, with no other condition differing from any other.

Since an extend LPC positivity was observed for words of high constraint, a repeated measures (2x2) ANOVA was conducted with factors of Word Type (Real and Novel) and Constraint (High and Low) in a time period ranging from 500 to 700 ms at identical electrode sites previously used for N400 analysis (RMCe, LMCe, MiCe and MiPa). This revealed main effects of Constraint [F(1,25)=8.71 p=0.0068], with High constraint target words showing greater positivity in this time-window, and a main effect of Type [F(1,25)=14.34 p=0.0009], with Real words showing greater overall positivity in this period. An Constraint x Type interaction was also found [F(1,25)=14.18, p=0.0009]. Posthoc Tukey tests revealed that this Real/Low targets were significantly more positive than any other condition in this time window, but no other differences were observed.

Plausibility effects: Difference ERPs to verbs in Plausible and Implausible Test sentences are shown in Figure 2. ERPs to verbs were analyzed since it is at this point in the sentence where all the information needed for a plausibility judgment first becomes available. Additionally, analysis at the verb avoids confounds due to reduction in N400 amplitude associated with repetition of the target word. As can be seen from this figure, difference waves to real words peak between 400 and 500 ms (N400) at central and parietal sites. This effect is also seen in the Novel/High condition, but is not as large for the Novel/Low condition.

N400 amplitude differences in plausibility were measured in a time period ranging between 300 and 500ms across four centrally located electrodes (RMCE, LMCE, MiCe, MiPa). A (2 x 2) repeated measures ANOVA was carried out on these measurements with factors of Word Type (Real and Novel) and Constraint (High and Low). This analysis revealed no main effect of Constraint [$F(1,25)=2.74$, $p=0.10$], or Type ($F<1$). However, there was a Constraint x Type interaction [$F(1,25)=5.12$, $p=0.033$]. Pairwise post-hoc Tukey analyses revealed that difference waves were larger in Novel/High than Novel/Low conditions, but no other conditions differ from each other at alpha threshold of 0.05.

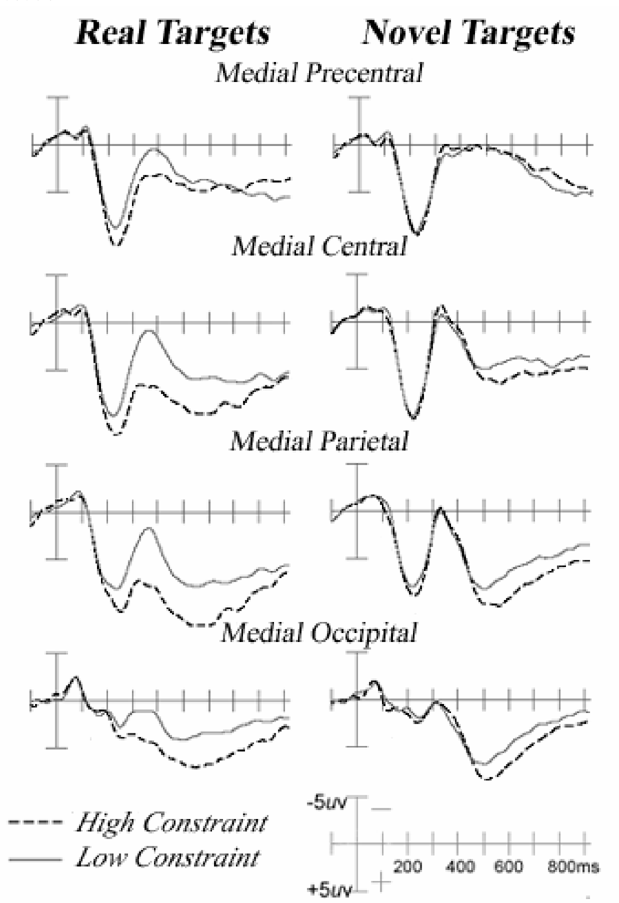

Figure 1. ERPs to Real and Novel target words in context sentences at midline electrode sites

Discussion

In this study, we explored how novel word meanings can be acquired from sentential contexts that are strongly or weakly constraining. Knowledge of word meaning was probed by plausibility judgment, rather than in a task that encourages explicit naming of the novel word or definitional strategies. The aim was to create a task in which novel words are learned implicitly through contextual use, as in natural language.

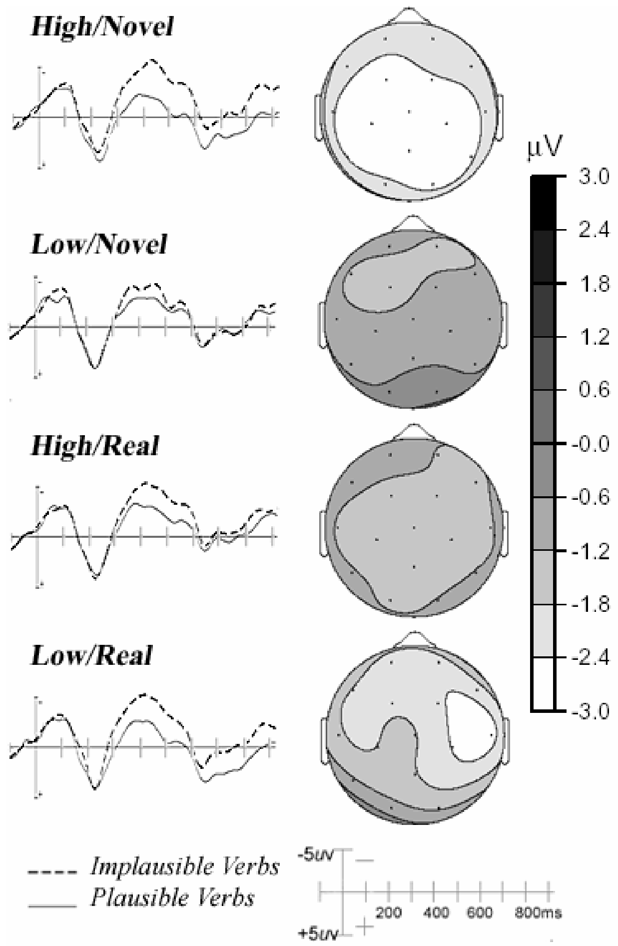

Figure 2. Plausibility effect for Verbs preceding Real and Novel targets. Event-related potentials at right-medial central electrode (RMCe) are plotted in left column for plausible and implausible verbs in each condition. On the right, spline-interpolated isovoltage topographic maps plot the plausibility effect (Implausible-Plausible) from 300-500ms in each condition.

In our task, the initial presentation of novel words in the context of weakly and strongly constraining contexts, unlike real words, did not yield constraint related differences in N400 amplitude to the novel words per se. The consequences of this single exposure, however, were clearly evident in the N400 plausibility effect to the main verb in the probe sentence that immediately followed. N400 amplitudes to probe verbs were significantly smaller if they could plausibly take the novel word as an object than if they could not, but only if the novel word had appeared in a highly constraining context. Although the plausibility judgment was not required or given until after the probe sentence, the N400 modulation to the verb shows that it was "computed" as soon as it could be – at the verb, and before the novel word's second appearance. Clearly, the representation of a novel word was shaped by the initial context in which it occurred so as to influence how it is subsequently understood. This is the first time that changes in brain activity consequent to novel word learning have been observed over such a short duration.

In sum, this study demonstrates differences in neural activity and behavior due to rapid, single-shot, word learning. We find that the context in which a novel word appears matters – and that stronger constraints during learning enhance an individual's ability to judge a word's appropriate usage in subsequent contexts.

Acknowledgments

This work was supported by RO1 AG08313-O1A1 and R01 NICHD22614 to MK and R01 MH60517-04 to JE and a NIH predoctoral training grant to AB.

References

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language, 31*(2), 308-27.

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effect on semantic priming in auditory word processing. *Psychophysiology, 30*, 161-169.

Federmeier, K. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. Journal *of Memory and Language, 41*, 469-495.

Federmeier, K. & Kutas, M. (2005). Aging in context: age-related changes in context use during language comprehension. *Psychophysiology,42*(2), 133-41.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month olds: Processing known and unknown words in picture contexts. *Journal of Cognitive Neuroscience, 16(*8), 1465-1477.

Friedrich, M., & Friederici, A. D. (2005a). Lexical priming and semantic integration reflected in the event-related potential of 14-month-olds. *NeuroReport, 16*(6), 653-656.

Friedrich, M., & Friederici, A. D. (2005b). Semantic sentence processing reflected in the event-related potentials of one and two year old children. *NeuroReport, 16*(16), 1801-1804.

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

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

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, **7**, 703-704.

Mestres-Missé, A., Rodriguez-Fornells, A & Munte, T. F (in press). Watching the brain during meaning acquisition. *Cerebral Cortex.*

Mills, D. L., Coffey-Corina, S. A. & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13-20 months. In D. Thal & J. Reilly (Eds.) *Developmental Neuropsychology, 13,* 397-446

Mills, D. L., Plunkett, K., Prat, C., & Schafer, G. (2005). Watching the infant brain learn words: Effects of vocabulary size and experience. *Cognitive Development, 20*, 19-31.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.) *The Nature of Vocabulary Acquisition*. (pp. 19-35). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Ojima, S., Nakata, H., Kakigi, R. (2005). An ERP study of second language learning after childhood: effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-28.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond.* Psychology Press.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.) *The Nature of Vocabulary Acquisition*. (pp. 89-106). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stein M., Dierks T., Brandeis D., Wirth M., Strik, W., & Koenig T. (2006) Plasticity in the adult language system: a longitudinal electrophysiological study on second language learning. *Neuroimage, 33*(2), 774-83.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18*, 380-393.