Published in final edited form as:
Dev Neuropsychol. 2010 ; 35(4): 376–403. doi:10.1080/87565641.2010.480915.

Lexical Quality in the Brain: ERP evidence for robust word learning from context

Gwen A. Frishkoff[1],
University of Pittsburgh

Charles A. Perfetti, and
University of Pittsburgh

K Collins-Thompson[2]
Carnegie Mellon University

Abstract

We examined event-related potentials (ERPs) before and after word learning, using training contexts that differed in their level of contextual support for meaning acquisition. Novel words appeared either in contexts that were semantically constraining, providing strong cues to meaning, or in contexts that were weakly constraining, that is, uninformative. After each sentence, participants were shown the word in isolation and were asked to generate a close synonym. Immediately after training, words trained in high-constraint contexts elicited a smaller left temporal negativity ($N300_{FT7}$) compared with words trained in low-constraint contexts, and both types of trained words elicited a stronger medial frontal negativity ($N350_{Fz}$) relative to familiar words. Two days after training the $N300_{FT7}$ disappeared and was replaced by a later, left parietal ($P600_{Pz}$) effect. To examine robust learning, we administered a semantic priming test two days after training. Familiar words and words trained in high-constraint contexts elicited strong N400 effects. By contrast, words trained in low-constraint contexts elicited a weak N400 effect, and novel (untrained rare) words elicited no semantic priming. These findings suggest that supportive contexts and the use of an active meaning-generation task may lead to robust word learning. The effects of this training can be observed as changes in an early left frontal component, as well as the classical N400 effect. We discuss implications for theories of "partial" semantic knowledge and for robust word learning and instruction.

INTRODUCTION

Word knowledge is often acquired gradually, over many exposures to a word in different contexts (Beck, McKeown, & Kucan, 2002; Biemiller, 2003). Thus, at any time, a learner may have partial or incomplete knowledge of a word's meaning. For example, a person may recognize that the word *supercilious* has a negative connotation and refers to a human trait, but fail to select the correct meaning from a set of possible definitions (Brown, Frishkoff, & Eskenazi, 2005). It is also well-known that people passively comprehend many more words than they actively use (Waring, 1999; Durso & Shore, 1991; Wesche & Paribakht, 1996). Together, this evidence suggests that knowledge of word meanings can be partial and fairly

Gwen Frishkoff, Ph.D., Department of Psychology, 140 Decatur Street, Urban Life, Rm 712, Georgia State University, Atlanta, Georgia 30303-3083, gfrishkoff@gsu.edu.
[1]Now at Georgia State University
[2]Now at Microsoft Research

shallow, allowing for passive recognition of a word in a particular context, but failing to support active and flexible use of the word across diverse contexts.

Recognizing that word knowledge can be partial, researchers have attempted to characterize the depth as well as the breadth of a person's vocabulary— that is, not just the number of words that a person recognizes, but also the quality (Perfetti & Hart, 2001, 2002; Dale, 1965; Stahl, 1986), completeness (Waring, 1999; Durso & Shore, 1991; Wesche & Paribakht, 1996; Brown, et al., 2005), and robustness (Beck, et al., 2002) of word semantic representations. The goal is to discover what it means to have robust word knowledge and the steps that a learner takes from complete lack of knowledge — that is, the inability to discriminate a real word from a made-up word — to robust knowledge of a word’s meaning.

A related goal is to determine what learning and instructional techniques lead to robust word knowledge. This issue has stimulated considerable debate in reading education, particularly with regard to the efficacy of “incidental” (nondirected) learning of vocabulary during reading (see Stahl & Fairbanks, 1986 and Swanborn & de Glopper, 1999 for reviews). Some researchers have suggested that word learning is best achieved through extensive reading. And, indeed, research has shown that words can be acquired through exposure to words in a variety of contexts without direct instruction, although gains are typically small (Swanborn & de Glopper, 1999). Exposure to words in written contexts may be important, because low-frequency words occur rarely, it at all, in spoken language (Nagy, Herman, & Anderson, 1985; Nagy, Anderson, & Herman, 1987). At the same time, it is precisely these low-frequency (literary or “Tier II”) words that are key to unlocking meaning in harder texts (Beck, McKeown, & Kucan, 2002). It has also been suggested that incidental learning is the main source of word learning gains for good reason: there are too many words to learn and too little time for direct instruction. Researchers have estimated that children acquire ~3,000 words per year on average; however, voculabulary instruction in schools targets several hundred words per year at best (Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985). Finally, it has been suggested that encounters with a word across a variety of contexts may promote a rich set of semantic associations, including connotative as well as denotative understanding (Beck, et al., 2002). The cumulative impact of these exposures may result in a deeper understanding of a word’s meaning, particularly subtle features that may be hard to convey through direct instruction, but may be integral to the correct use of words in context. In this way, seeing words in context may be part and parcel of building robust word knowledge.

Given this body of evidence, most researchers would agree that exposure to words in context — and written contexts in particular — is important for word learning. On the flip side, incidental learning may result in knowledge that is partial or inaccurate. The focus during reading is typically on comprehending the overall gist, rather than guessing the meaning of an unfamiliar word. Further, even if word learning is explicitly targeted, not all contexts are equally informative — and in fact, some contexts are misleading (Beck, McKeown, & McCaslin, 1983; Frishkoff, Collins-Thompson, et al., 2008). To make matters worse, a learner could go many weeks without seeing a word more than once, allowing the word to fade from memory; this is particularly the case for the kinds of low frequency, literary (or "Tier II") words that are most in need of active learning (Beck, et al., 2002). Thus, incidental learning may be unreliable and inefficient as the sole context for word learning. By contrast, active and extended engagement with words over time has been linked to robust gains in vocabulary (Beck, et al., 2002). These findings suggest that learners may profit from active, targeted word learning activities, in addition to gains that come about through incidental learning.

The present study contributes to these issues by examining event-related potentials (ERPs) to novel words before and after training, and by manipulating the training contexts to bring about either partial or robust word knowledge. The use of ERPs affords an opportunity to study word learning processes that take place within tens and hundreds of milliseconds — the timescale at which cognition actually unfolds during learning. It also enables us to separate different neurolinguistic processes that are engaged in learning. In this way, we hope to observe the component processes that take place in the learner's mind as they process newly learned words.

In the present paper, we report changes in ERPs from pre- to post-training in two conditions. In one condition, very rare words (such as "nutant" and "kippage") were presented in contexts that were highly constraining, providing strong cues to the word's meaning. In the second condition, contexts were weakly constraining, that is, relatively uninformative. Given that trained words occurred with the same frequency in both conditions, we reasoned that ERP differences after training would be more likely to reflect word semantic learning, as opposed to changes in familiarity with word forms (i.e., phonology or orthography). More specifically, our goal was to address two questions:

1. How do cortical responses to novel words differ before and after learning?
2. Are there ERP correlates of partial versus robust meaning acquisition?

We anticipated word learning effects for the following ERP components: the centroparietal $N400_{Pz}$ and $P600_{Pz}$, the mid-frontal $N350_{Fz}$, and the left fronto-temporal $N300_{FT7}$. The following section provides backfound on each of these patterns and outlines our specific hypotheses. In naming these components, we follow the recommendation of Dien (2008) and reference the topographic distribution (e.g., "Fz" for mid-frontal patterns), as well as approximate peak latency (e.g., "350" for ~350 ms) and scalp polarity ("P" for positive, "N" for negative) for each ERP component.

Temporoparietal effects in semantic comprehension and meaning acquisition

The posterior $N400_{Pz}$ is by far most widely studied ERP marker of semantic processing. The so-called "N400 effect" is seen as an increased negativity over centroparietal electrodes between ~350–550 ms (average peak difference, ~400 ms) when a word's meaning is harder to access from memory, harder to integrate with the preceding context, or both (see Federmeier & Kutas, 2000 for a review). The precise timing and topography of the N400 effect vary across studies, and some studies have suggested that the time window that has been traditionally been associated with the $N400_{Pz}$ may actually comprise several distinct patterns (Nobre & McCarthy, 1994; Hill, et al., 2002; Frishkoff, Tucker, Davey, & Scherg, 2004; Frishkoff, 2007). Source modeling of the $N400_{Pz}$ has identified possible sources in the left temporal lobe (Silva-Pereyra, Rivera-Gaxiola, Aubert, Bosch, Galan, & Salazar, 2003; Frishkoff, et al., 2004; Van Petten & Luka, 2006; Frishkoff, 2007). These results are consistent with evidence from fMRI and other imaging methods that shows modulation of temporal lobe activity during semantic comprehension (Binder, et al., 1997; Bookheimer, et al., 2002).

Although there is widespread agreement that the $N400_{Pz}$ reflects semantic — as opposed to nonlinguistic, phonological or syntactic — processing, its precise intepretation remains controversial. For example, it is unresolved whether the $N400_{Pz}$ is modulated by automatic or "implicit" semantic priming, as claimed by some authors (e.g., Deacon, Hewitt, Yang, & Nagata, 2000; Kiefer, 2002; Kiefer & Brandel, 2006), or whether it is modulated only under conditions of "explicit" or controlled semantic processing (Silva-Pereyra, Harmony, Villaneuva, Fernandez, Rodriguez, Galan, et al., 1999; Ruz, Madrid, Lupianez, & Tudela, 2003). Despite these controversies, modulation of the $N400_{Pz}$ (in one form or another) is seen in most semantic ERP paradigms. The robust nature of this pattern makes it a good candidate ERP for tracking changes in word semantic knowledge.

Two previous publications have described $N400_{Pz}$ effects of word learning (Perfetti, Wlotko, & Hart, 2005; Mestres-Misse, Rodriguez-Fornells, & Munte, 2007). In Perfetti, et al. (2005), participants studied novel (very rare) words and their definitions using a flash-card method. Subsequently, they completed a semantic priming task while their EEG was measured. In the post-training task, the trained rare words were presented, followed by a familiar word, and participants were asked to decide if the two words were semantically related. ERPs distinguished trained rare words from both untrained rare and familiar words at two times: there was an early increased centroparietal negativity at ~140 ms (at the time of the occipital P100) and an increased positivity (P600) effect later in the epoch, which was interpreted as a marker of episodic memory (specifically, recollection-based memory). In addition, there was an $N400_{Pz}$ effect to the second (primed) word for trained, but not for untrained, words, which was seen as a reduced negativity to related versus unrelated words. The relatedness effect is consistent with successful learning of the trained words.

The results of Perfetti, et al., (2005) show several semantic and memory-related effects for words after training, providing an important foundation for subsequent ERP studies of word learning. At the same time, the study design does not allow for separation of effects due to meaning acquisition, as opposed to familiarity with word forms (i.e., phonological and/or orthographic representations). By contrast, a recent study conducted by Mestres-Misse, et al. (2007) shows specific effects of meaning acquisition. Mestres-Misse, et al. (2007) presented novel words in three different sentences that either supported inferences about novel word meanings, or not. Results showed a reduction in the $N400_{Pz}$ component for words that were trained in the supportive versus unsupportive contexts. After the learning block, participants completed a semantic priming task, where they made meaning judgments to words that were semantically related or unrelated to the trained words. The newly learned words elicited an $N400_{Pz}$ semantic relatedness effect, as expected. In addition to this $N400_{Pz}$ post-training effect, Mestres-Misse et al. (2007) also observed a reduced $N400_{Pz}$ to words during training: that is, the $N400_{Pz}$ to trained words decreased incrementally as subjects acquired new semantic knowledge about the word with each additional context.

To summarize, in two previous experiments (Perfetti, et al., 2005; Mestres-Misse, et al., 2007), word learning has been shown to result in $N400_{Pz}$ effects, both during and immediately following training. These results suggest that the $N400_{Pz}$ can be used to detect semantic learning. In the present experiment, we sought to replicate and extend these prior results. To this end, we had learners perform a modified version of the Perfetti, et al., (2005) paradigm two days after training. The goal was to test transfer of learning to a different task context, in particular, one that could be used to link new findings to previous results for the well-studied $N400_{Pz}$ effect.

The mid-frontal $N350_{Fz}$ and activation of frontal cortex during word comprehension

While ERP studies of semantic comprehension have consistently described posterior ERP effects (i.e., the semantic N400), descriptions of frontal ERPs during these same tasks have been less consistent (see Dien et al, 2008, for a recent review). In the present study we used a meaning-generation task, which requires participants to actively retrieve the meanings of novel words that have recently been trained. We expected that this task would engage frontal networks that have previously been implicated in effortful processing of word meanings (e.g., Gabrieli, Poldrack, & Desmond, 1998; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). We expected in turn to see modulation of ERPs that have previously been modeled with frontal cortical sources.

One frontal ERP component that has been observed in studies of word-level semantic processing is a negativity peaking between ~300 and 500 ms, variously called an "N300" (Franklin, et al., 2007), an "N300-Fz" (Dien, 2009), and a "medial frontal negativity" or "MFN" (Tucker, et al., 2003; Frishkoff, 2007; Frishkoff, Perfetti, & Westbury, 2008). This pattern has a mid-frontal distribution (sometimes overlapping with the left-lateralized $N300_{FT7}$; see below). Because it typically peaks between ~300 and 400 ms, we refer to it here as the $N350_{Fz}$. Like the $N400_{Pz}$, the $N350_{Fz}$ is enhanced (i.e., more negative over mid-frontal electrodes) when a word's meaning is more difficult to process. For example, in a recent ERP study examining effects of "partial" word knowledge (Frishkoff, et al., in press) the $N350_{Fz}$ amplitude was inversely related to word familiarity: the more unfamiliar the words, the stronger the $N350_{Fz}$. This finding suggests that the $N350_{Fz}$ may be related to the "fN400" pattern that has been observed in studies of familiarity-based memory (Curran, 2000; Rugg & Curran, 2007). Given the mid-frontal distribution of this component and its tentative localization to the mid-frontal gyrus and anterior cingulate (Frishkoff, et al., 2004; Frishkoff, 2007; Frishkoff, et al., in press), it is probably not language specific (see also Curran & Cleary, 2003, for evidence that nonlinguistic stimuli also evoke an fN400 familiarity effect).

Interestingly, Mestres-Misse et al. (2007) report that the novel (newly learned) words elicited a semantic relatedness effect with a fronto-central distribution, as compared with the more typical, posterior $N400_{Pz}$ effect for familiar words. The ERP priming effect was also delayed and more sustained for the novel (trained) words. It is not clear how this frontal effect may be related to the $N350_{Fz}$ pattern that has been observed in response to single words. Nonetheless, this result provides evidence that early stages of word semantic learning may evoke ERP effects that are more frontal than the semantic priming effects that are typically observed for well-known words.

The $N300_{FT7}$ and engagement of left frontal cortex during meaning-generation

Neuroimaging studies of semantic processing have also shown robust effects over left inferior frontal cortex, with evidence for specific responses to semantic processing in anterior regions of LIFG, such as BA47 (Bookheimer, 2002). Typically, anterior LIFG is more active when there are increased demands on memory or attention during semantic processing (Gabrieli, et al., 1998; Thompson-Schill, et al., 1997). There is also evidence that words acquired later in life (after ~age 8–10) may engage anterior LIFG more strongly than words learned in early childhood, even when frequency effects are controlled (see Hernandez & Li, 2007 for a recent review). This result may have strong relevance for word learning, as it suggests that late-acquired words may place greater demands on frontal cortical networks.

Written words typically evoke a sequence of surface-negative patterns over ventral temporal regions, beginning with the left occipito-temporal N1 (or N170), which peaks at ~170–200 ms, and terminating with a left frontotemporal pattern, the $N300_{FT7}$, which begins as early as 220 ms and is sometimes extended over 200 ms or more, depending on the nature of the task (Dien, et al., 2003; Frishkoff, et al., 2004; Frishkoff, 2007; Frishkoff, et al., 2009). The left anterior $N300_{FT7}$ may be related to early word-level semantic processing (Abdullaev and Posner, 1997, 1998), but so far there is too little evidence to link this pattern to a specific stage of word recognition.

The $N300_{FT7}$ appears to be the same pattern that was described in an early dense-array study of ERP responses in semantic processing (Nobre & McCarthy, 1994). In Experiment 3, the authors describe a left lateral negativity peaking at ~316 ms. This pattern was clearly distinguished from the later, parietal "N400" both in terms of its spatiotemporal features and in the direction of its response to semantic priming. Semantic priming was measured by subtracting ERPs to two consecutive words (designated as “prime” and “target”) when the two words denoted concepts from the same semantic category. Interestingly, the ERP to the target showed an enhanced left frontal negativity compared to the prime word. This effect is opposite to the direction of the N400 effect, which is reduced (less negative) to primed words. While they did not say so explicitly, it is important to note that the left lateral negativity in their experiment was distinct from a mid-frontal negativity, which occurred at the same time (~300–330 ms), but was not modulated by priming (ibid., p. 242, Figure 14). That is, the mid-frontal negativity (similar to what we have termed the $N350_{Fz}$) occurred with equal magnitude in response to the prime and target words, whereas the $N300_{FT7}$ was greater (more negative) in response to semantically related targets (see also Dien, et al., 2003 for similar findings).

Another probable manifestation of the $N300_{FT7}$ was seen in Abdullaev and Posner (1997, 1998) in a series of ERP studies of cortical responses during meaning-generation. In this task, participants were presented with nouns denoting concrete objects, such as “hammer,” and were asked to say the object’s use (e.g., “pound”). In the control condition, subjects simply repeated the noun. This use-generation task had been successfully employed with Positron Emission Tomography (PET) imaging, revealing activation in left inferior frontal, as well as left temporal, cortex during semantic retrieval (Snyder, et al., 1995). In the ERP studies, the meaning-generation task evoked an increased left anterior positivity compared with the control condition, beginning at ~200 ms and extending to ~400 ms. Dipole modeling tentatively localized this ERP effect to the LIFG, consistent with earlier PET studies using the same paradigm. These results are consistent with source localization of the $N300_{FT7}$, which has identified the left inferior frontal gyrus and left anterior temporal pole as likely generators (Dien, et al., 2003;Frishkoff, Tucker, Davey, & Scherg, 2004;Frishkoff, 2007;Frishkoff, Perfetti, & Westbury, 2008). To our knowledge, the studies by Abdullaev and Posner (1997, 1998) are the only ones to report ERP responses during meaning-generation. Thus, the specific modulation of the left anterior $N300_{FT7}$ during their task has strong relevance for the present study, where a meaning-generation task was used to probe semantic learning.

Goals and hypotheses for the present experiment

In the present experiment, we sought to isolate effects of partial vs. robust meaning acquisition by comparing ERPs for words that were trained in high or low constraint contexts. In both conditions, each word occurred in three single-sentence contexts, which were interleaved (i.e., participants saw a different word and a different context on each trial). Spaced vs. massed practice has been shown to maximize word learning and retention (Frishkoff, et al., 2008). Interleaving also minimized the likelihood that participants would simply repeat the same response to a word across trials.

High-constraint contexts were expected to support robust learning of the target word meanings. Low-constraint contexts were expected provided few if any clues to the meaning of the target. These contexts were expected to provide some information (e.g., part of speech and some semantics), but much less than the high-constraint sentences, leading to only partial word learning. An important feature of this paradigm, like that of Mestres-Misse, et al. (2007), is that it enables us to isolate ERP effects of meaning acquisition, separate from task-related effects that may reflect increased familiarity with word forms, but do not correlate with semantic learning.

To promote robust learning, we asked learners to generate the meaning of the target word after each context during training, as well as during the pre- and post-training blocks. Previous work has demonstrated that the use of an active, meaning-generation task leads to effective learning and retention of words (Frishkoff, et al., 2008a). In our experience, these gains are substantial when compared with the more modest word learning gains that are typically seen during "incidental" or passive learning (Nagy, Anderson, & Herman, 1987). It is well known that generative tasks promote "deeper" or more elaborated processing of stimuli: retrieving information from memory requires more active processing, because one needs to select the response from a very large set of possible answers. In turn, active processing has been linked to enhanced learning and retention (for a review, see Roediger & Karpicke, 2006). To test retention of learned words, participants returned two days later and repeated the meaning-generation task, followed by a semantic priming task that was designed to measure transfer of learning to a different task context.

We hypothesized that partial versus robust word learning would elicit differences in the four ERP patterns discussed previously: the centroparietal $N400_{Pz}$ and $P600_{Pz}$, left frontal $N300_{FT7}$, and mid frontal $N350_{Fz}$. We predicted that the use of an active meaning-generation task would evoke particularly strong frontal components ($N300_{FT7}$ and $N350_{Fz}$), and that the amplitude of these components would be modulated by word learning and semantic retrieval. By contrast, we predicted that exposure to new words in less supportive (low-constraint) contexts would lead to partial word semantic knowledge that would be evident in short-term measures of word comprehension, but would prove less robust when studied two days after training using an N400 semantic priming task.

METHODS

Subjects

Participants were 15 adult native English speakers (7 males). All subjects were right-handed, with normal or corrected-to-normal vision, no neurological deficits and no history of a reading or language disorder. Mean age was 19 (± 1) years.

Twelve of the fifteen participants completed a battery of reading and language prescreening tests. The battery included a test of reading comprehension ability (Brown, et al., 1993), a test of vocabulary skill (Brown, et al., 1993), and a test of nonverbal reasoning (Raven, 1960), in addition to a variety of other tests of reading and language ability and reading experience (see Frishkoff, Perfetti, & Westbury, 2008 for a complete description of the subtests in this battery). The mean composite score on the comprehension test was 20.70 (s.d., 5.95), which is similar to the mean test score for a group of "Average" readers (20.88 ± 6.63), that were included in a previous study (Frishkoff, Perfetti, & Westsbury, 2008). This score is derived from accuracy (85% ± 8%) and percentage of items attempted within the 15-minute time limit (75% ± 12%). The mean composite score on the vocabulary test was 40.90 (s.d., 10.58). By comparison, the mean vocabulary score for "Average" readers was reported as 43.84 (±9.71) in the earlier study (Frishkoff, Perfetti, & Westsbury, 2008). Finally, percent accuracy on the test of nonverbal reasoning was 65% (±18%), which is also roughly equivalent to the mean scores for "Average" subjects in Frishkoff, et al. (2008). Thus, scores on standardized tests of vocabulary, reading comprehension, and nonverbal intelligence for the present group of subjects was similar to that of "Average" adult readers, as defined in previous work.

Subjects were recruited through introductory courses in Psychology at the University of Pittsburgh, and by posting ads in the Pittsburgh campus newspaper. Either payment ($15 per hour) or academic course credit, or a combination, was given in exchange for subject participation.

ERP Experiment Design

Participants completed two study sessions, which were scheduled two days apart. Figure 1 shows the timeline for the ERP tasks in Sessions 1 and 2.

All stimuli were black and were presented foveally on a light grey screen. When the target words appeared in isolation (as events for time-locking of ERPs), they were 26 dpi and subtended a visual angle of between 1.5 and 2.1 degrees. Further details regarding stimuli and stimulus presentation for each task are described in the following subsections.

Stimuli

Words used in Task 1: Task 1 stimuli included two types of words: **Rare Trained** words (n=60) and **Familiar Untrained** words (n=30). Rare trained words included 17 nouns, 9 verbs, and 34 adjectives, with a mean length of 7.1 letters (s.d., 1.1). See Frishkoff, Perfetti, & Westbury (2008) for additional details on how these words were selected. All rare words had a written frequency of less than 1 in a million, according to the Kucera–Francis norms (Francis & Kucera, 1982). Pretesting confirmed that most or all of these words were unfamiliar to subjects prior to training, and participants were at chance levels in selecting the correct meaning of the words (see *Pretest and Posttest measures* for details on vocabulary assessment).

The 30 Familiar Untrained words were taken from a larger set of 60 relatively low-frequency words that had used in a previous ERP experiment on word recognition among adult native-English speakers (Frishkoff, Perfetti, & Westbury, 2008). The mean written word frequency for the subset of 30 Untrained Familiar words was 3.38 per million (s.d., 8.15). The average word length for this set was 6.90 letters (s.d., 1.32), which was not significantly different from the average word length for the Trained Rare words ($p > .7$).

Words used in Task 2: In Task 2, 120 words served as “primes” in the semantic priming task. Prime words included the 60 Trained Rare and 30 Untrained Familiar words. In addition, 30 Untrained Rare words (such as “accolent” and “loxotic”) were included, to establish a baseline for detection of semantic priming effects for trained words. Like the trained rare words, the untrained rare words all had a frequency of less than 1 in a million. The average length of untrained rare words was 7.37 letters (s.d., 0.92).

The target words in Task 2 consisted of 120 familiar words, which were synonyms or near-synonyms of the prime words. Target words had a mean length of 5.00 letters (s.d., 1.25) and a mean written word frequency of 90.65 per million (s.d., 134.51). In the Unrelated condition, the prime–target word pairs were shuffled, so that each participant saw the same target twice, paired once with a closely related prime and once with an unrelated prime.

Sentence contexts: A total of 720 sentences were constructed, and each sentence was constructed to be either high or low constraint. This classification was based initially on experimenter intuition and was subsequently validated in a cloze completion task (Taylor, 1953), which involved a separate group of 60 adult, monolingual English-speaking participants. In the cloze task, the rare words in each sentence were replaced by an underscore (blank space), and participants were asked to provide a one-word completion (“cloze” response) for each sentence stem. Consider examples (1) and (2) below.

1. The _______ firefighter ran into the burning house.
2. His friends did not consider him a very _______ man.

The target word for both examples was ‘impavid’. Intuitively, context (1) provides strong clues to the meaning of the target word “impavid” (*brave*). The cloze task results reflect this intuition: 64% of respondents gave “brave” as the response to context (1). By contrast, sentence (2) is fairly low-constraint: practically any adjective can be used in this context, as long as it can refer to a personal trait. In our cloze experiment, the most common response for sentence (2) was “nice,” with a relatively low cloze probability of 9%.

To capture the average level of constraint for each sentence context, we examined several measures related to the cloze probability for the target word meaning. First, we analyzed the cloze probability for the modal response, that is, the response that was given most frequently for each sentence stem. Second, we examined the cloze probabilities for two near-synonyms of the rare (target) words. For example, near-synonyms for the word “impavid” were “brave” and “fearless.” Third, we calculated a composite cloze score, which summed over semantically related responses. For example, in response to one sentence stem, 34% of the respondents provided the word “silent” and another 21% gave the word “quiet”; thus, the composite cloze score for this context was 0.55 (55%). Finally, we computed “sentential constraint,” defined as the number of unique cloze completions for each context. The results of these analyses are summarized in Table 1.

According to each of these measures, the high-constraint sentences were more semantically constraining than the low-constraint sentences. On this basis, we expected that the high-constraint sentences would provide stronger support for learning new words as compared with the low-constraint contexts.

There were an average of 10.22 words per sentence (s.d.=1.50), and the high and low constraint sentences did not differ in length (mean number of words/sentences 10.52 for high-constraint, and 9.93 for low-constraint sentences).

Session 1 ERP Tasks—Session 1 consisted of a pre-training section (Task 1A), a training section (Task 1B), and a post-training section (i.e., a repetition of Task 1A).

Task 1A (Meaning-generation task): In the pre- and post-training blocks of Session 1, target words were presented in isolation, and subjects were asked to generate the words’ meaning after a short delay. Each trial began with the presentation of a rectangular outline, which cued the subject to initiate the experimental sequence when ready by pressing the space key. During the intertrial interval, subjects were permitted to blink or adjust their position. Once the subject initiated a trial with the space key, a central fixation (‘+’ symbol) appeared for 500 ms, followed by the target word, which was presented foveally for 250 ms. The target then disappeared and was replaced by a central fixation (‘+), which remained on the screen for 1250 ms. Participants were instructed to wait until the target word appeared a second time, together with a symbol (‘?’), which was presented below the target word, indicating the need for a response. The stimulus onset asyncrhony (SOA) between the first and second presentation of the target was 1,500 ms. At the response probe (‘?’), the subject was asked to type in a synonym or near-synonym of the target word. Instructions emphasized that participants were to provide a response on every trial, even for words that were unfamiliar or partially known.

Task 1B (Training task): In the training block (Task 1B), subjects viewed rare words, embedded in meaningful sentences. On each trial, the subject was asked to read the sentence and then press a button when they were ready to continue. After 10 seconds, if there was no button press, the program automatically advanced. The target word was then presented in isolation. As in the pre- and post-training blocks, the subject was instructed to wait until the target appeared together with a response probe (‘?’). At the probe, the subject was asked to type in a synonym or near-synonym of the target word. A response was elicited after each training context.

Each of the 60 trained rare words was presented in 3 different contexts during training (Task 1B), for a total of 180 training contexts per subject. Assignment of words to training condition (high vs. low constraint) was randomized across study participants, and the ordering of the 3 contexts for each target was randomized within training condition. All 180 sentences were presented in pseudorandom order, with the constraint that the spacing between the 3 sentences for a particular target word be as wide as possible (minimum spacing was 6, and maximum spacing was 80).

Session 2 ERP tasks—In Session 2, subjects completed Task 1A a third time, so we could examine ERP effects of training after a 2-day delay period. Subjects then completed a new task (Fig. 1, lower right panel), as described in the following section.

Task 2 (Semantic priming task): In the final ERP task, subjects completed a modified version of the semantic priming paradigm used in Perfetti, Wloko, and Hart (2005). In this task, subjects saw a sequence of two words, and were asked to decide if the two words were semantically related. The first ("prime") word belonged to one of three categories: (1) Rare Trained words (n=60), (2) Familiar Untrained words (n=30), or (3) Rare Untrained words (n=30). The second ("target") word was a familiar word and it was either a near-synonym for the prime word (i.e., was semantically related), or was unrelated. The unrelated word pairs were created by reshuffling the 60 prime–target pairs and then verifying that the resulting pairs were not semantically related. The related and unrelated word pairs were the same for every participant. Presentation order was randomized across sessions.

The stimulus timing for Task 2 was as follows. Prime words were presented for 1,000 ms, followed immediately by a target word that was either semantically related or unrelated. The targets were presented for 1,000 ms, followed by a 500 ms interstimulus interval (blank screen), and then a probe stimulus ("?"). Participants were asked to judge the semantic relatedness of the prime and target words and to indicate their decision with a key press as soon as the probe appeared. They indicated their response by pressing either the "1" key (right index finger) or "2" key (right middle finger) on the right side of the keyboard. "Yes"/"No" assignment of keys was counterbalanced across subjects. Accuracy was emphasized over speed for this task. In contrast with Perfetti, et al. (2005), no accuracy feedback was provided during our version of this task.

Pretest and Post-test Assessments—Written assessments of target word knowledge were administered at three times: (1) at the beginning of the experiment, before the Pre-training block; (2) immediately after the Post-training block in Session 1; and (3) at the end of the experiment, immediately after the priming task in Session 2. The assessment consisted of 90 multiple-choice items (60 trained rare words and 30 untrained familiar words as controls). For each item, subjects were asked to select the word that was closest in meaning to the target word. There were five choices, including 4 distracters that were from the same word frequency range and were the same average length as the correct response. The choices for each item belonged to the same part-of-speech (e.g., noun, verb, adjective). After the synonym judgment, participants were asked to indicate their confidence on a scale from 1 (just guessing) to 3 (very confident). A post-test was administered in the same session, after the ERP experiment task. Items were the same as in the pre-test, but the item order was reshuffled.

It is important to note that the high-frequency words that were used as targets in the semantic priming task were different from the choices on the written, multiple-choice tests. Thus, there was no risk that participants would encounter the same correct responses on different tests.

ERP Data Acquisition and Preprocessing

ERPs were recorded using a 128-channel electrode array, with vertex recording reference (Electrical Geodesics, Inc.). Data were sampled at a rate of 500 per second and were amplified with a .01-hz highpass filter (time constant ~10 seconds). Channels were automatically marked bad for a given trial if blinks or eye movements were detected, if amplitudes >150 μV, if differential average amplitude (i.e., changes in slope) >75 μV, if the channel was flat (had zero variance), or if manual inspection suggested noise specific to that channel (i.e., not affecting surrounding channels). Any channel marked bad for more than 20% of the total trials was interpolated in the raw EEG based on data measured at nearby electrodes. After exclusion of artifacts, remaining trials were segmented into 1,100 ms epochs, starting 200 ms before onset of the target word. Segmented data were averaged across trials (within subjects and within condition) and digitally filtered with a 30-Hz lowpass filter. After further channel and subject exclusion, bad (excluded) channels were interpolated. The data were re-referenced to the average of the recording sites, and were baseline corrected, 200 ms prior to the segmentation event.

ERP Analysis

ERP analyses were conducted separately for three sets of data: Task 1A (Session 1: Pre- and post-training); Task 1A (Session 2: delayed post-training); and Task 2 (Session 2: transfer task). Analysis of ERPs acquired during the training blocks (Task 1B) will be presented in a separate manuscript.

For analysis of Task 1A effects, ERP data were decomposed using a temporal principal components analysis (tPCA). Dien's PCA Matlab Toolbox (v 1.093) was used to implement the analysis (Dien, in submission; see Dien & Frishkoff, 2005 for methods). After factor extraction and retention (10 factors with eigenvalues >1 in each case), we applied a Promax rotation (kappa=4). The factor loadings that result from this data decomposition represent time courses of the latent ERP patterns. Factor scores represent the weighting of each temporal pattern across electrodes (i.e., they are used to plot the topographic distribution of the latent patterns), and also represent weightings across subjects and experiment conditions. It is thus the factor scores for each latent pattern that are entered in statistical analyses, to test for condition differences. Factor waveforms and topographies were plotted for the grand average data, were visually inspected, and were then interpreted and labeled with reference to previously described ERP patterns in word recognition (Frishkoff, et al., 2007).

For examination of topographic effects, we divided channels into six scalp regions-of-interest, or **ROI**: orbitofrontal, frontocentral, anterotemporal, posterotemporal, parietal, and occipital (see Appendix A for channel groupings). Factor scores were averaged over these ROI to examine the topographic distribution of effects for each pattern factor. **Laterality** (right versus left ROI) was considered as a separate factor in each analysis.

Repeated-measures analyses of variance were conducted separately for each ERP factor (pattern). For the Task 1A (Session 1) analysis, there were two experimental factors of interest: **Word** (3 levels: Rare-Trained-HiConstr, Rare-Trained-LoConstr, and Familiar-Untrained), and **Time** (2 levels: Session 1, Session 2). For the Task 1A (Session 2), Word was the only experimental factor. Because Task 1A in Sessions 1 and 2 were acquired on separate days, we anticipated that some ERP components might differ in peak latency. Therefore, we subjected the Session 2 data to a separate temporal PCA and separate analyses of variance. While this does not permit direct comparison with the ERPs from Session 1, we believe it is the more conservative approach: combining ERP data from Sessions 1 and 2 into a single analysis may risk distorting patterns that vary in time course across the two sessions.

Finally, t-tests were conducted at each time sample (n=350 samples, 700 ms) and each electrode (= 129), and the resulting "t-maps" were plotted as scalp-topographic maps, thresholding at a corrected p-value of 0.0000011, $t(14) = 8.14$. The t-maps are used for illustration of condition differences; readers are referred to the analyses of variances for Task 1A results based on inferential statistics.

For analysis of Task 2 effects, we adopted a hypothesis-driven approach. The semantic priming task has been used many times, in contrast with the meaning-generation task (Task 1A). Therefore, we restricted our focus to the $N400_{PZ}$ window, 350–450 ms after onset of the second, target word. A priori regions of interest for the N400 analysis were left and right frontal, and left and right parietal ROI (see Appendix A for channel groupings). Repeated-measures ANOVA were conducted to examine differences in N400 peak latency and mean amplitude within the 350–450 ms window. Only correct responses were entered into the analyses.

RESULTS

Pre- vs. Post-test Assessments of Target Word Knowledge

To characterize participant knowledge of the target words before and after training, we compared scores on the written multiple-choice pretest to scores on the immediate and delayed post-tests. Mean accuracy for the three types of words —rare words trained in high-constraint contexts, rare words trained in low-constraint contexts, and untrained familiar words— are given in Figure 2.

Mean accuracy scores were entered into a 3 (Word) × 3 (Time) repeated-measures analysis of variance. As expected, there were main effects of Word, $F(2, 28) = 267.63$, $p < .001$, and of Time, $F(2, 28) = 66.34$, $p < .001$. The interaction was also significant, $F(2, 28) = 38.21$, $p < .001$, consistent with different learning gains across the three word types. Post-hoc comparisons showed a strong increase from pre- to post-test for rare words trained in the high-constraint condition — $t(14)= 9.64$, $p < .001$ — and more modest increase in the low-constraint condition — $t(14)= 2.43$, $p < .05$. Thus, there was learning in both training conditions, as expected. Importantly, post-test scores were significantly higher for words trained in the high-constraint condition — $t(14)= 6.67$, $p < .001$ (mean difference in accuracy, 27). Thus, accuracy results suggest there was more successful learning from the high- versus low-constraint contexts.

Contrary to prediction, the delayed post-test did not show evidence of forgetting for either of the trained word categories. This result may be partly due to the influence of the semantic priming task just before the final post-test, which could have reminded participants of the target word meanings. The ordering of the two tasks reflects a calculated trade-off: if the delayed post-test had been administered earlier, it may have contaminated results on the semantic priming task, rather than the other way around. In any case, the delayed post-test results show substantial retention of target word knowledge over the two days, even assuming an extra boost from the semantic priming task in Session 2.

Behavioral effects in semantic priming task

Accuracy and reaction times on the meaning judgment task were examined using 4 (Word) × 2 (Relatedness) repeated-measures analyses of variance. Subjects performed close to ceiling on accuracy (mean 97%) for familiar words relative to rare words. In addition, there was a strong "no" (unrelated) response bias, as evidenced by the relatively high accuracy (Table 1) and faster response times (Table 2) for unrelated versus related targets. Accuracy analyses revealed main effects of Relatedness — $F(1, 12) = 22.59$, $p < .001$ — and Word — $F(3, 36) = 92.29$, $p < .001$. There was also a significant interaction, Word × Relatedness, $F(3, 36) = 12.25$, $p < .001$. Post-hoc paired comparisons confirmed that the differences in accuracy for all four categories were significant (p < .01 for all comparisons). Reaction time analyses indicated a main effect of Relatedness, $F(1, 12) = 7.29$, $p < .05$. Overall, subjects were faster to indicate that targets were unrelated. The effect of Word approached significance, $F(3, 36) = 2.79$, $p = .056$.

ERP Results

ERP results are described separately for the following three analyses:

1. Task 1A (Session 1) — Word (3) × Time (2) × ROI (6) × Laterality (2)
2. Task 1A (Session 2) — Word (3) × ROI (6) × Laterality (2)
3. Task 2 (Session 2) — Word (4) × Relatedness (2) × ROI (2) × Laterality (2)

ERP Results for Session1, Task 1A—Figure 3 shows a 129-channel plot of the ERP waveforms for Trained Rare words before and immediately after training, collapsing over the high- and low-constraint conditions to show the main effects of training.

A prominent feature of this waveplot is a negative-going dip that peaks around 300–350 ms over left anterotemporal electrodes and is enhanced (i.e., more negative) after training. Figure 4 shows the corresponding F300 factor (Figure 4, fourth row), which has at a peak loading at ~300 ms. The timing and topography of this factor are prototypical of the $N300_{FT7}$ pattern.

Another salient feature is a mid-frontal negativity, which peaks slightly later than the $N300_{FT7}$ pattern and is also enhanced (more negative) immediately following training. This feature is captured by tPCA factor F440 and correspond to the medial frontal negativity (“MFN”) described in prior work, or $N350_{Fz}$ pattern as it is termed here. Note that the mid-frontal negativity in the present study peaks somewhat later than the typical $N350_{Fz}$ pattern.

The mean amplitudes for each of the 5 factors shown in Figure 3 were entered into a 3 (Word) × 2 (Time) × 6 (ROI) × 2 (Laterality) ANOVA. Results are summarized in Table 3.

There are three effects that are particularly worthy of note. First, we observed an early training effect during the time of the P100 component (peak latency, ~130 ms), similar to the training effect that was described in Perfetti, et al. (2005).

Second, after training, the Rare words elicited an increased mid-frontal negativity, which started at ~280 ms and persisted until the end of the epoch. Interestingly, the Untrained Familiar words also showed an increased mid-frontal negativity at the time of the N350Fz (Figure 5; cf. Figure 3, fifth row). Moreover, this mid-frontal effect did not differentiate words trained in the high and low constraint conditions.

Finally, the major difference for ERP responses to rare words trained in high- vs. low-constraint contexts was seen over left fronto-temporal and left orbitofrontal sites (Fig 4; effect labeled as “3” in right-hand panel). Rare words in both conditions elicited an increased left inferior frontal negativity, starting at around 300 ms and peaking at around 400 ms (an $N300_{FT7}$). This increased negativity was stronger in the low-constraint, as compared with the high-constraint, condition after training.

ERP Results for Session2, Task 1A—Figure 6 shows the ERPs in Session 2 for the meaning-generation task. Comparison with Figure 5 shows that the left inferior $N300_{FT7}$ condition difference has disappeared. Interestingly, however, the difference for words in the high- versus low-constraint conditions is now seen over left parietal regions, with the low-constraint words eliciting an increased positivity (P600) at about 550–600 ms relative to the high-constraint words.

The tPCA for these data revealed two main factors of interest: an F340 (an analog of the $N300_{FT7}$ pattern in Session 1), and an F540, a left parietal positivity that was largest for Familiar Untrained words.

The mean amplitudes for each factor were entered into a 3 (Word) × 6 (ROI) × 2 (Laterality) ANOVA. For Factor F340 (the $N300_{FT7}$), the interaction between Word and ROI was significant, $F (10, 140) = 4.12$, $p < .001$. Consistent with the upper left-hand plot in Figure 6, there is no longer any difference between the Trained Rare words in the two conditions. Rather, both types of Trained Rare words showed an enhanced negativity over frontal regions, as compared with the Untrained familiar words. For Factor F540, there was a 3-way Word × ROI × Laterality interaction, $F (10, 140) = 3.28$ $p < .01$. Consistent with the upper right-hand plot in Figure 6, rare words trained in the low-constraint condition were more positive during this P600 time window.

ERP Results for Session2, Task 2 (Transfer Task)—The ERP response to the second, "target" word in the semantic judgment task (Session 2) showed a classical N400 effect, that is, an enhanced right parietal negativity to unrelated versus related targets (Figure 7).

Analysis of N400 mean amplitudes revealed a 3-way interaction, Word × Relatedness × ROI, $F(15, 225) = 2.36$, $p< .01$. Post-hoc paired comparisons showed that the semantic relatedness effect was significant for familiar words over frontal ($p < .01$) and parietal ($p < .001$) sites. Similarly, there rare words trained in high-constraint contexts elicited priming effects over frontal ($p < .01$) and parietal ($p < .001$) sites. By contrast, there was a much smaller relatedness effect for rare words trained in low-constraint contexts over parietal sites ($p < .05$), and no relatedness effect whatsoever over frontal sites ($p > .8$). There was no effect of semantic relatedness for untrained words over frontal or parietal regions ($p > .3$ for both comparisons).

Analysis of N400 latencies was restricted to the three conditions that displayed significant relatedness effects over parietal sites: Familiar Untrained words and the two categories of Rare Trained words (RareHiConstr, RareLoConstr). The main effect of Word just missed significance ($p = .06$), as did the Word × Laterality interaction ($p = .06$). Mean latencies are plotted in Figure 7 (right panel).

SUMMARY

In the present study, we examined ERPs to rare words (such as "nutant" and "impavid") before and after training. Words were presented in contexts that were either high- or low-constraint (Figure 1). We predicted that words in high-constraint contexts would show more robust learning and that this effect would be indexed by frontal ERPs during meaning-generation, immediately post-training. Learning was evaluated independently with a written pre-/post-test design. The post-test was administered shortly after training (Session 1) and after a 2-day delay period (Session 2). In addition, an N400 priming task ("transfer task") was administered in in Session 2 to evaluate robust learning.

As expected, participants demonstrated larger gains in word knowledge for the high- versus low-constraint training condition. These differential gains were seen on the written pre- and post-test assessments (Figure 2). They were also evident in the meaning-generation task. Words trained in both types of contexts showed increased frontal negativities ($N300_{FT7}$ and $N350_{Fz}$) immediately after training (Session 1). In addition, words trained in low-constraint contexts showed a stronger increase in the left inferior frontal $N300_{FT7}$ (Figures 3–5).

In Session 2, the ERPs for words trained in high- and low-constraint contexts no longer showed a left inferior frontal $N300_{FT7}$ effect. Instead, words in the low-constraint condition now elicited larger positivities over left parietal regions from ~500–700 ms (Figure 6).

Finally, rare words trained in high-constraint contexts elicited strong N400 priming effects during Session 2 over frontal, as well as parietal, sites. By contrast, low-constraint words elicited smaller effects over parietal sites, but little or no priming over frontal sites (Figure 7).

DISCUSSION

The present study aimed to address two main questions:

1. How do cortical responses differ to words before and after learning ?
2. Are there ERP correlates of partial versus robust learning?

Two early effects are worthy of note before we review the major effects of partial versus robust meaning acquisition. First, we replicated the effects of word training that were observed by Perfetti and associates (Perfetti, et al., 2005) at approximately 130–140 ms, during the time of the visual-evoke “P100” component. After training, there was an enhanced negativity over centroparietal electrodes and an enhanced positivity over occipital regions to trained words. In our study, this effect was insensitive to the semantic manipulation (i.e., training in high versus low constraint contexts), providing support for the idea that this effect reflects modulation of visual or early orthographic processing. The condition contrast appeared to extend into the N170 epoch (Table 3).

Second, beginning at around 220 ms, there was an enhanced negativity to familiar versus rare words over left posterior temporal sites. Because familiar words were not presented during training, it is impossible to separate effects of lexicality and training in interpreting this effect. Furthermore, the effect was relatively weak (note that it does not show up in Figure 3, where we applied fairly stringent correction for multiple comparisons).

Anterior LIFG in word learning and semantic retrieval

The key contrast in the present study was between rare words that were presented in high- versus low-constraint contexts before and after training. This contrast revealed an $N300_{FT7}$ effect over left inferior frontal (LIF) regions immediately after training. The $N300_{FT7}$ differentiated Rare Trained words from Familiar Untrained words, and also differentiated rare words trained in high- versus low-Robust learning constraint contexts. This $N300_{FT7}$ effect is unlikely to reflect changes in familiarity with the target word forms (i.e., orthographic representations), since trained words occurred equally often in the high- and low-constraint conditions. It seems more likely that the $N300_{FT7}$ effect reflects word-level semantic processing. It is also possible, however, that the $N300_{FT7}$ effect is related to the demands for effortful retrieval that are evoked by our meaning-generation task. Abdullaev & Posner (1997, 1998) described a similar, LIF pattern during meaning-generation. The effect was accounted for by an equivalent dipole located in left inferior prefrontal cortex, consistent with activations seen in PET and fMRI studies of semantic processing (Gabrieli, et al. 1998;Thompson-Schill, et al., 1997). Prior work suggests that LIF is specifically enhanced during more effortful encoding and retrieval of word meaning (e.g., Addis & McAndrews, 2006).

A few recent studies have studied the neurocognitive processing of words acquired early in life (< 6 or 7 y.o.), versus late-acquired words (see Hernandez & Li, 2007 for a review). Fiebach and associates (Fiebach, et al., 2003) examined correlation of the BOLD response with “age of acquisition” (AOA) during a lexical (word vs. nonword) decision task, and found increase activation for late- versus early-acquired words in left inferior frontal and anterior insular regions (BA45/47 and BA12). These are the same subregions of the LIFG have previously been linked to effortful retrieval of semantics (Gabrieli, et al. 1998; Thompson-Schill, et al., 1997). The proposed mechanisms underlying AOA effects have been a subject of recent discussion (Hernandez & Li, 2007). One interesting proposal is that word meanings that are acquired later in life may be less stable, and therefore more prone to interference during semantic retrieval.

ERP markers of partial versus robust learning

In the present experiment, the $N300_{FT7}$ effect increased from pre- to post-training in both the high- and low-constraint conditions, consistent with increased task engagement for both training conditions. Importantly, however, words trained in the low-constraint condition showed a larger $N300_{FT7}$ compared with the $N300_{FT7}$ to words trained in high-constraint contexts. This result is consistent with partial learning of words trained in low-constraint contexts: although knowledge of word meanings increased (on average) from pre- to post-test, the larger $N300_{FT7}$ suggests that newly learned meanings for the low-constraint words were more difficult to retrieve from memory.

Interestingly, when the meaning-generation task was repeated two days after training, the $N300_{FT7}$ effect was no longer present. Instead, there was a posterior P600 effect, which was not seen in Session 1. This frontal-to-posterior shift is intriguing, particularly given similar findings that early learning engages more frontal cortical areas, whereas posterior areas are more typically engaged after memories have had time to become more stable and better consolidated (Friedman, 1990; Luu, Tucker, & Stripling, 2007). It is therefore possible that this frontal-to-posterior shift in meaning representation could emerge as one marker of robust semantic learning.

Mestres-Misse, et al. (2007) have made a similar suggestion: they noted that ERP priming effects to newly learned words evoked a more frontal “N400” effect, as compared with familiar words. Although they do not report the frequency of the familiar words in their experiment, examples such as *Autobus* (Sp. “bus”) and *Cuchilo* (Sp. “knife”) suggest that they were probably medium-to-high frequency. In our study, by contrast, both familiar words and rare words trained in high-constraint contexts elicited broadly distributed effects over frontal and posterior sites in our Session 2 transfer task. The broad distribution of “N400” priming effects in our experiment (encompassing frontal as well as posterior sites) is unsurprising given that familiar words in our task fell on the low-frequency end of the “familiar” spectrum. Likewise, the rare words trained in high-constraint contexts were still fairly novel. It would be interesting to compare these results with N400 priming effects to high-frequency words and to rare words that have been over-trained. It is possible that we might then see a pattern similar to that of Mestres-Misse, et al. (2007). More difficult to square with this theory is the finding that rare words trained in low-constraint contexts elicited priming over parietal, but not over frontal, sites. It is unlikely that this effect can be explained by thesholding: Figure 7 (lower right panel) shows that the frontal ERP effect is disproportionately reduced for low- versus high-constraint words, compared with the parietal N400 effect. This is an effect that will need to be replicated and accounted for in future work.

Implications for Vocabulary Learning and Instruction

Evidence for partial semantic learning in the present study is relevant for questions concerning the efficacy of word learning from written contexts. A large body of work suggests that words can be learned "incidentally" during reading (Nagy, Anderson, &

Herman, 1987; Nagy, Herman, & Anderson, 1985). At the same time, there is considerable controversy about the optimal use of contexts to support word learning (Beck, et al., 2002; Biemiller, 2003). While some workers have proposed that frequency and breadth of reading is the key to vocabulary development (Nagy, et al., 1985, 1987), others have suggested that "incidental" learning from may be unreliable, because many words occur infrequently and because contexts may be only weakly supportive, or even misleading (Beck, et al., 1983; Frishkoff, et al., 2008). According to Beck and associates (Beck, et al., 2002), acquisition of words that are more abstract or literary ("Tier II" words) requires rich and extended instruction (Beck, et al., 2002). While this approach takes time, it has been shown to promote more robust learning when compared with instructional methods that rely soley on incidental learning.

One solution may be to develop a combined strategy for vocabulary training that involves multiple exposures to words in context, together with activities that promote deep and active processing of meaning. For example, in the present study, subjects were exposed to words in various contexts. At the same time, they were asked to generate the meaning of the target word after each context, to engage deep and active processing of semantics. We note that written assessments showed impressive learning in the high-constraint condition (with ~10/30 words successfully learned). This is particularly encouraging given that training consisted of only three exposures to words in context. These findings contrast with previous estimates that 12 or more exposures may be needed to promote robust word learning (e.g., McKeown, et al., 1985) and suggest that meaning-generation may be useful for promoting efficient and effective learning of new word meanings. Additional work will needed to determine the efficacy of word learning from contexts under different instructional and task conditions, and to understand the role of left inferior frontal cortical networks in the learning and retention of word meanings. In the long run, our goal is to develop more fine-grained models of word learning and to link these models to meaningful outcomes in reading and vocabulary acquisition (see also Frishkoff, White, & Perfetti, 2009).

Conclusion

The importance of research on word semantic learning is two-fold. First, evidence for multiple stages of meaning acquisition could have implications for theories of word representation: if there are qualitative differences in the representation of words that are only partially known versus those that are linked to robust semantic knowledge, it may be important to understand the processes that account for these shifts in knowledge representation. In the present experiment, ERPs to newly trained words evoked a left inferior frontal negativity that was larger for words trained in low-constraint contexts. Two days after training, this experiment effect shifted from frontal to posterior regions of the scalp. It is possible that this pattern reflects decreased engagement with frontal control networks over the course of learning (see also Friedman, 1990; Chein & Schneider, 2005). It is also possible that this frontal-to-posterior shift may reflect changes in the quality or robustness of lexical representations. In either case, the implication is that partial and robust word representations may engage different neurocognitive networks.

Second, evidence for distinct cognitive stages of word learning may have implications for instruction: the optimal approach to word learning may be different for words that are more or less familiar prior to training (Frishkoff, et al., 2008b). Thus, an improved understanding of partial versus robust word learning could contribute to the design of effective methods for word learning and instruction, as well as to theories of semantic representation in the mind and brain.

Acknowledgments

This research was supported by the Institute of Educational Sciences (Department of Education), grant number R305U030004. We thank Erika Taylor, Laura Halderman, and Rhonda MacLain for their assistance with development of sentence stimuli. We thank Bob Frank for assistance with matlab scripts for generating 129-channel ERP waveform plots.

REFERENCES

Abdullaev YG, Posner MI. Event-related brain potential imaging of semantic encoding during processing single words. NeuroImage 1998;7(1):1–13. [PubMed: 9500833]

Addis DR, McAndrews MP. Prefrontal and hippocampal contributions to the generation and binding of semantic associations during successful encoding. NeuroImage 2006;33:1194–1206. [PubMed: 17023179]

Beck, I.; McKeown, M.; Kucan, L. Bringing words to life: Robust vocabulary instruction. New York: Guilford Press; 2002.

Beck I, McKeown M, McCaslin E. Not all contexts are created equal. Elementary School Journal 1983;83:177–181.

Biemiller, A. Teaching Vocabulary in the Primary Grades: Vocabulary Instruction Needed. In: Baumann, J.; Kame'enui, E., editors. Reading Vocabulary: Research to Practice. New York: Guilford Press; 2003. p. 28-40.

Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J. ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. Journal of Cognitive Neuroscience 1999;11(3):235–260. [PubMed: 10402254]

Binder JR. Neuroanatomy of language processing studied with functional MRI. Clinical Neuroscience 1997;4(2):87–94. [PubMed: 9059758]

Bookheimer S. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annual Review of Neuroscience 2002;25:151–188.

Brown, J.; Frishkoff, G.; Eskenazi, M. Automatic question generation for vocabulary assessment. Proceedings of the Annual Human Language Technology Meeting; October 6–8; Vancouver, B.C. 2005.

Brown, JI.; Fishko, VV.; Hanna, G. Nelson-Denny Reading Test. New York: Riverside Publishing Co; 1993.

Chein JM, Schneider W. Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. Cognitive Brain Research 2005;25(3): 607–623. [PubMed: 16242923]

Curran T. Brain potentials of recollection and familiarity. Memory & Cognition 2000;28(6):923–938.

Curran T, Cleary AM. Using ERPs to dissociate recollection from familiarity in picture recognition. Cognitive Brain Research 2003;15(2):191–205. [PubMed: 12429370]

Dale E. Vocabulary measurement: Techniques and major findings. Elementary English 1965;42:82–88.

Deacon D, Hewitt S, Yang C, Nagata M. Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. Brain Res Cogn Brain Res 2000;9(2):137–146. [PubMed: 10729697]

Dien. The Neurocognitive bases of reading single words as seen through early latency ERPs: A model of converging pathways. Biological Psychology. 2008 in press.

Dien J. The ERP PCA Toolkit: An Open Source Package For Facilitating Multivariate Decomposition of Event Related Potential Data. (submitted).

Dien J, Beal DJ, Berg P. Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. Clin Neurophysiol 2005;116(8):1808–1825. [PubMed: 15996897]

Dien, J.; Frishkoff, GA. Introduction to principal components analysis of event-related potentials. In: Handy, T., editor. Event-related potentials: A methods handbook. Cambridge, MA: MIT Press; 2005. p. 189-208.

Dien J, Frishkoff GA, Cerbone A, Tucker DM. Parametric analysis of event-related potentials in semantic comprehension: Evidence for parallel brain mechanisms. Brain Res Cogn Brain Res 2003;15(2):137–153. [PubMed: 12429366]

Durso FT, Shore WJ. Partial knowledge of word meanings. Journal of Experimental Psychology: General 1991;120:190–202.

Francis, WN.; Kucera, H. Frequency analysis of English usage. Boston: Houghtonn Mifflin Company; 1982.

Franklin MS, Dien J, Neely JH, Huber E, Waterson LD. Semantic priming modulates the N400, N300, and N400RP. Clinical Neurophysiology 2007;118(5):1053–1068. [PubMed: 17336145]

Friedman D. ERPs during continuous recognition memory for words. Biological Psychology 1990;30(1):61–87. [PubMed: 2223937]

Frishkoff GA. Hemispheric differences in strong versus weak semantic priming: Evidence from event-related brain potentials. Brain & Language 2007;100(1):23–43. [PubMed: 16908058]

Frishkoff GA. Fractionating the N400 Effect: Separation of frontal and posterolateral responses in semantic priming. Clinical Neurophysiology. 2008 in revision.

Frishkoff, Collins-Thompson, Perfetti, Callan. Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. Behavioral Research Methods. 2008a

Frishkoff GA, Perfetti CA, Westbury C. ERP measures of partial semantic knowledge: Left temporal indices of skill differences and lexical quality. Biological Psychology. 2008 in press.

Frishkoff GA, Tucker DM, Davey C, Scherg M. Frontal and posterior sources of event-related potentials in semantic comprehension. Cognitive Brain Research 2004;20(3):329–354. [PubMed: 15268912]

Gabrieli, JD.; Poldrack, RA.; Desmond, JE. The role of left prefrontal cortex in language and memory; Proceedings of the National Academy of Sciences, U.S.A. 1998. p. 906-913.

Kiefer M. The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. Cognitive Brain Research 2002;13(1):27–39. [PubMed: 11867248]

Kiefer M, Brendel D. Attentional modulation of unconscious "automatic" processes: evidence from event-related potentials in a masked priming paradigm. Journal of Cognitive Neuroscience 2006;18(2):184–198. [PubMed: 16494680]

Kutas M, Federmeier KD. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Science 2000;4(12):463–470.

Luu P, Tucker DM, Stripling R. Neural mechanisms for learning actions in context. Brain Research 2007;1179:89–105. [PubMed: 17936726]

McKeown M, Beck I, Omanson RC, Pople MT. Some effects of the nature and frequency of vocabulary instruction on the knowledge of use of words. Reading Research Quarterly 1985;20(5): 522–535.

Mestres-Misse A, Rodriguez-Fornells A, Munte TF. Watching the brain during meaning acquisition. Cerebral Cortex 2007;17(8):1858–1866. [PubMed: 17056648]

Nagy WE, Anderson RC, Herman PA. Learning words from context during normal reading. American Educational Research Journal 1987;24(2):237–270.

Nagy WE, Herman P, Anderson R. Learning words from context. Reading Research Quarterly 1985;19:304–330.

Nobre AC, McCarthy G. Language-related ERPs: Scalp distributions and modulation of word type and semantic priming. Journal of Cognitive Neuroscience 1994;6(3):233–255.

Perfetti, CA.; Hart, L. The lexical basis of comprehension skill. In: Gorfien, DS., editor. On the consequences of meaning selection: Perspectives on resolving lexical ambiguity. Washington, D. C: American Psychological Association; 2001. p. 67-86.

Perfetti, CA.; Hart, L. The lexical quality hypothesis. In: Verhoeven, L.; Elbro, C.; Reitsma, P., editors. Precursors of functional literacy. Vol. Vol. 11. Amsterdam: John Benjamins; 2002. p. 67-86.

Perfetti CA, Wlotko EW, Hart LA. Word learning and individual differences in word learning reflected in event-related potentials. Journal of Experimental Psychology: Learning, Memory, & Cognition 2005;31(6):1281–1292.

Raven, JC. Guide to the Standard progressive matrices. London: H.K. Lewis; 1960.

Rugg MD, Curran T. Event-related potentials and recognition memory. Trends in Cognitive Science 2007;11(6):251–257.

Ruz M, Madrid E, Lupianez J, Tudela P. High density ERP indices of conscious and unconscious semantic priming. Cognitive Brain Research 2003;17(3):719–731. [PubMed: 14561458]

Silva-Pereyra J, Harmony T, Villaneuva G, Fernandez T, Rodriguez M, Galan L, et al. N400 and lexical decisions: Automatic or controlled processing? Journal of Clinical Neuropysiology 1999;110(5):813–824.

Silva-Pereyra J, Rivera-Gaxiola M, Aubert E, Bosch J, Galan L, Salazar A. N400 during lexical decision tasks: a current source localization study. Clinical Neurophysiology 2003;114(12):2469–2486. [PubMed: 14652107]

Snyder AZ, Abdullaev YG, Posner MI, Raichle ME. Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. Proceedings of the National Academy of Sciences 1995;92(5):1689–1693.

Stahl SA. Three principles of effective vocabulary instruction. Journal of Reading 1986;29(7):662–668.

Stahl SA, Fairbanks MM. The effects of vocabulary instruction: A model-based meta-analysis. Review of Reading Research 1986;56:72–110.

Swanborn MSL, De Glopper K. Incidental Word Learning While Reading: A Meta-Analysis. Review of Educational Research 1999;69(3):261–285.

Swanborn MSL, De Glopper K. Impact of reading purpose on incidental word learning from context. Language Learning 2002;52(1):95–117.

Taylor WL. "Cloze" procedure: a new tool for measuring readability. Journalism Quarterly 1953;30:415.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proceedings of the National Academy of Sciences 1997;94(26):14792–14797.

Tucker, DM.; Frishkoff, GA.; Luu, P. Microgenesis of language. In: Stemmer, B.; Whitaker, HA., editors. Handbook of neuroscience of language. Cambridge, MA: Elsevier; 2007. in press

Tucker DM, Luu P, Frishkoff G, Quiring J, Poulsen C. Frontolimbic response to negative feedback in clinical depression. Journal of Abnormal Psychology 2003;112(4):667–678. [PubMed: 14674878]

Van Daalen-Kapteijns MM, Elshout-Mohr M. The acquisition of word meanings as a cognitive learning process. Journal of Verbal Learning and Verbal Behavior 1981;20(4):386–399.

Van Daalen-Kapteijns MM, Elshout-Mohr M, De Glopper K. Deriving the meaning of unknown words from multiple contexts. Language Learning 2001;51(1):145–181.

Van Petten C, Luka BJ. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language 2006;97(3):279–293. [PubMed: 16343606]

Wagner AD, Koutstaal W, Maril A, Schacter DL, Buckner RL. Task-specific repetition priming in left inferior prefrontal cortex. Cerebral Cortex 2000;10(12):1176–1184. [PubMed: 11073867]

Waring, R. Tasks for assessing receptive and productive second language vocabulary. Swansea, UK: University of Wales; 1999.

Wesche M, Paribakht T. Assessing vocabulary knowledge: depth vs. breadth. Canadian Modern Language Review 1996;53(1):13–40.

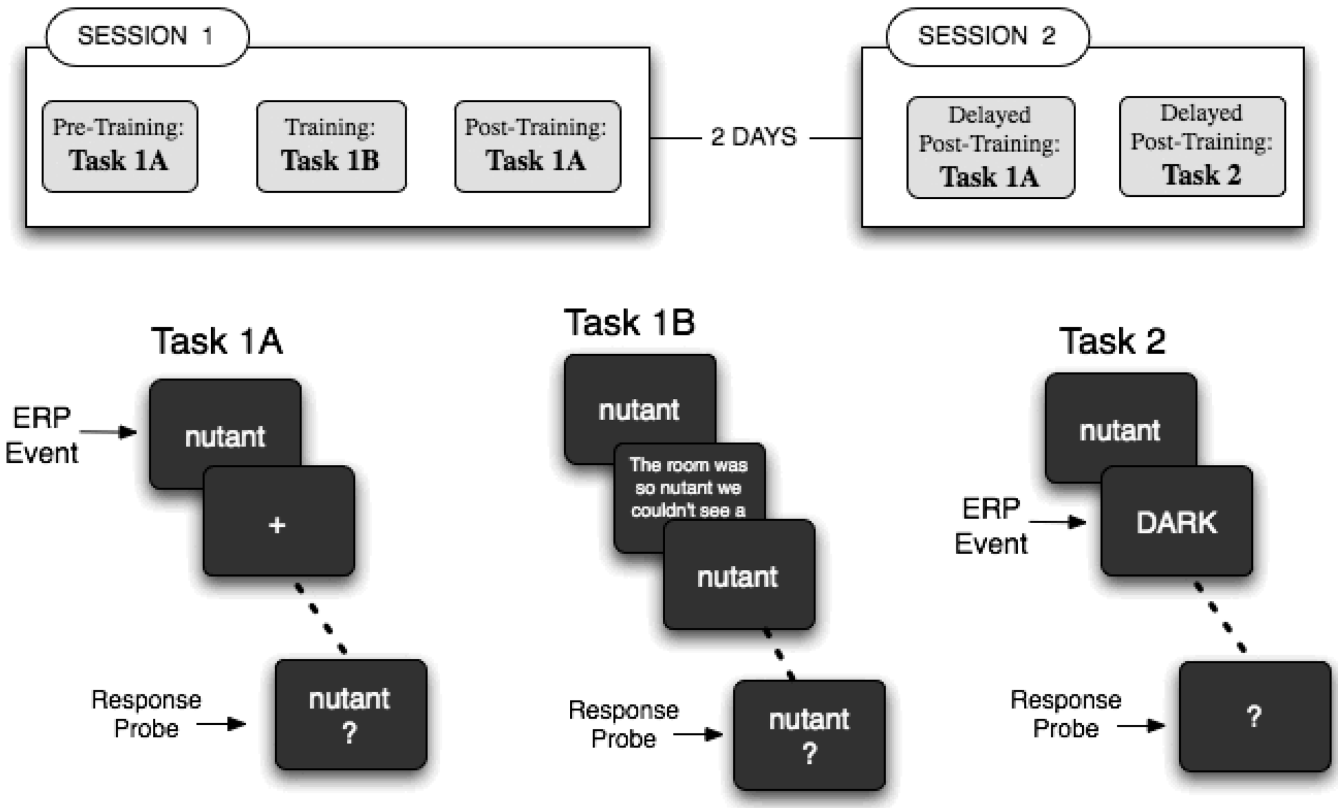


Figure 1.

Timeline for the three ERP tasks in Sessions 1 and 2. **Task 1A** is a *meaning-generation task*: participants are asked to generate the meaning of each word. This task was administered before training, immediately after training, and after a 2-day delay. Comparison of ERP effects at these three times allowed us to examine immediate and delayed effects of training. **Task 1B** is the *training task*: like Task 1A, participants are asked to generate a response on each trial. Just before the meaning-generation probe, the word is presented in a sentence is that is either highly constraining (HiConstr condition) or weakly constraining (LoConstr condition). **Task 2** is a *semantic priming task*: trained words appear, followed by high-frequency words that are either close synonyms or semantically unrelated. This task is designed to elicit classic N400 semantic priming effects, which we expected to vary as a function of training condition.

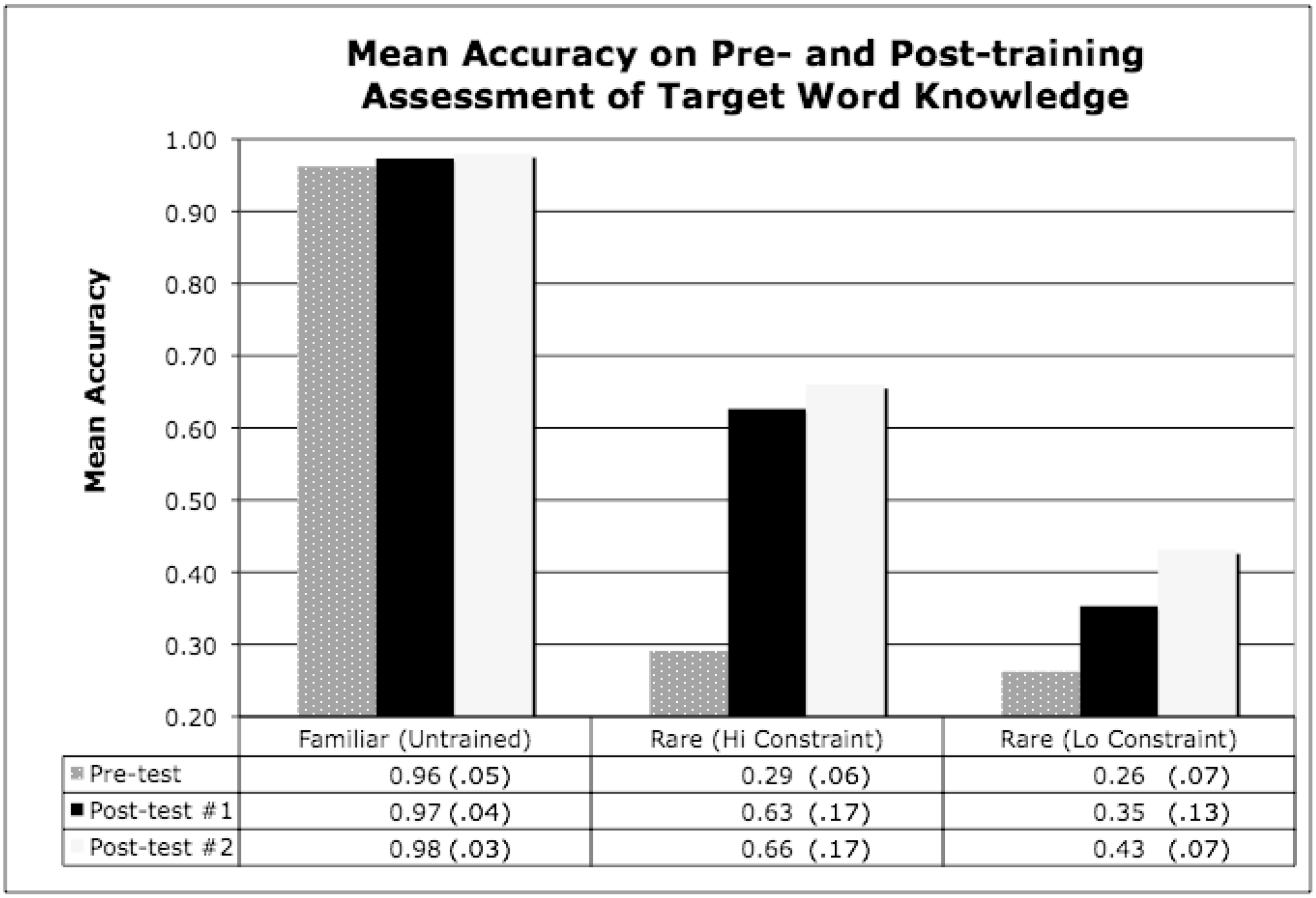


Figure 2.
Means and standard errors for accuracy scores on written tests of word knowledge. Chance (baseline) was .20. Posttest #1 was administered immediately after training. Posttest #2 was administered two days later. Standard deviations are given in parentheses.

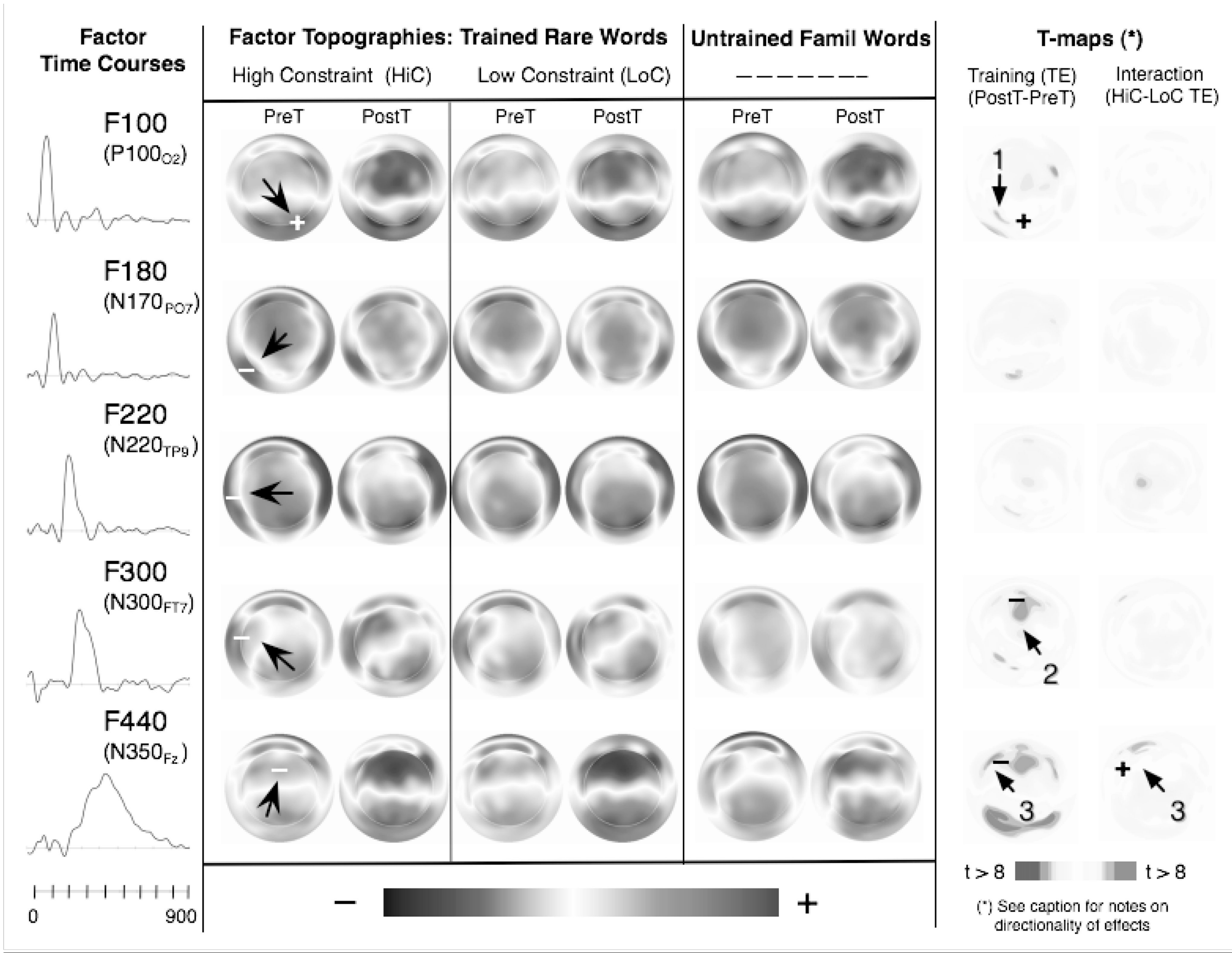


Figure 3.

Illustration of time course (left panel) and topography (middle and right panels) for each of the 5 ERP components in Task 1A (meaning-generation task). *Left panel*, PCA factor loadings for F100 ("$P100_{O2}$"), F180 ("$N170_{PO7}$"), F220 ("$N220_{TP9}$"), F300 ("$N300_{FT7}$"), and F440 ("N350Fz"). *Middle Panel*, Scalp topographic projection of each factor for each experiment condition, scaled at ±4μV. *Right panel*, statistical "t" maps, computed on raw ERP data, illustrating topography of significant condition diffrences. T-maps are shown at peak time for each corresponding factor (component). The three main ERP effects are labeled as follows: (1) early left occipital positivity at ~100 ms, enhanced (more positive after training); (2) mid-frontal negativity at ~300 ms, enhanced (more negative) after training and sustained from ~300–550 ms; (3) left inferior frontal negativity at ~440 ms, enhanced (more negative) after training, with a larger difference for words trained in the low-vs. the high-constraint condition (Interaction Effect).

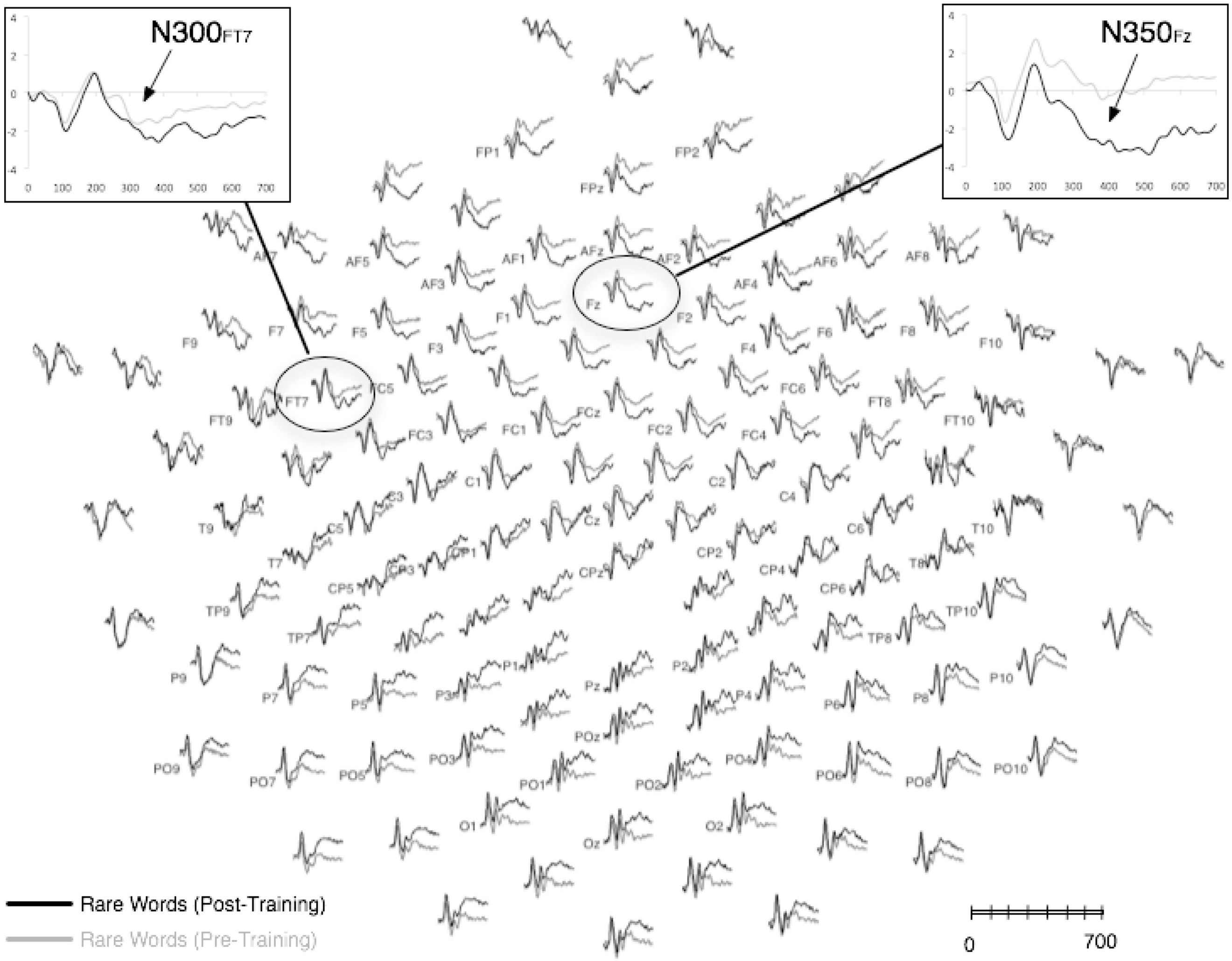


Figure 4.
129-channel ERPs in Session 1 for Trained Rare words (collapsed over training condition to illustrate main effect of training). Positive voltages plotted up in these and all subsequent plots. *Black*, response to word during Task1A (meaning-generation task) before training. *Grey*, response to word during Task1A (meaning-generation task) after training. Upper right-hand panel shows enlarged plot of ERP responses at left frontal electrode AF3: After training, there was an increased negativity during the meaning-generation task, broadly dstributed over frontal sites. Cf. Figure 3, statistical “t” maps showing Main Effect of Training from ~300–500 ms.

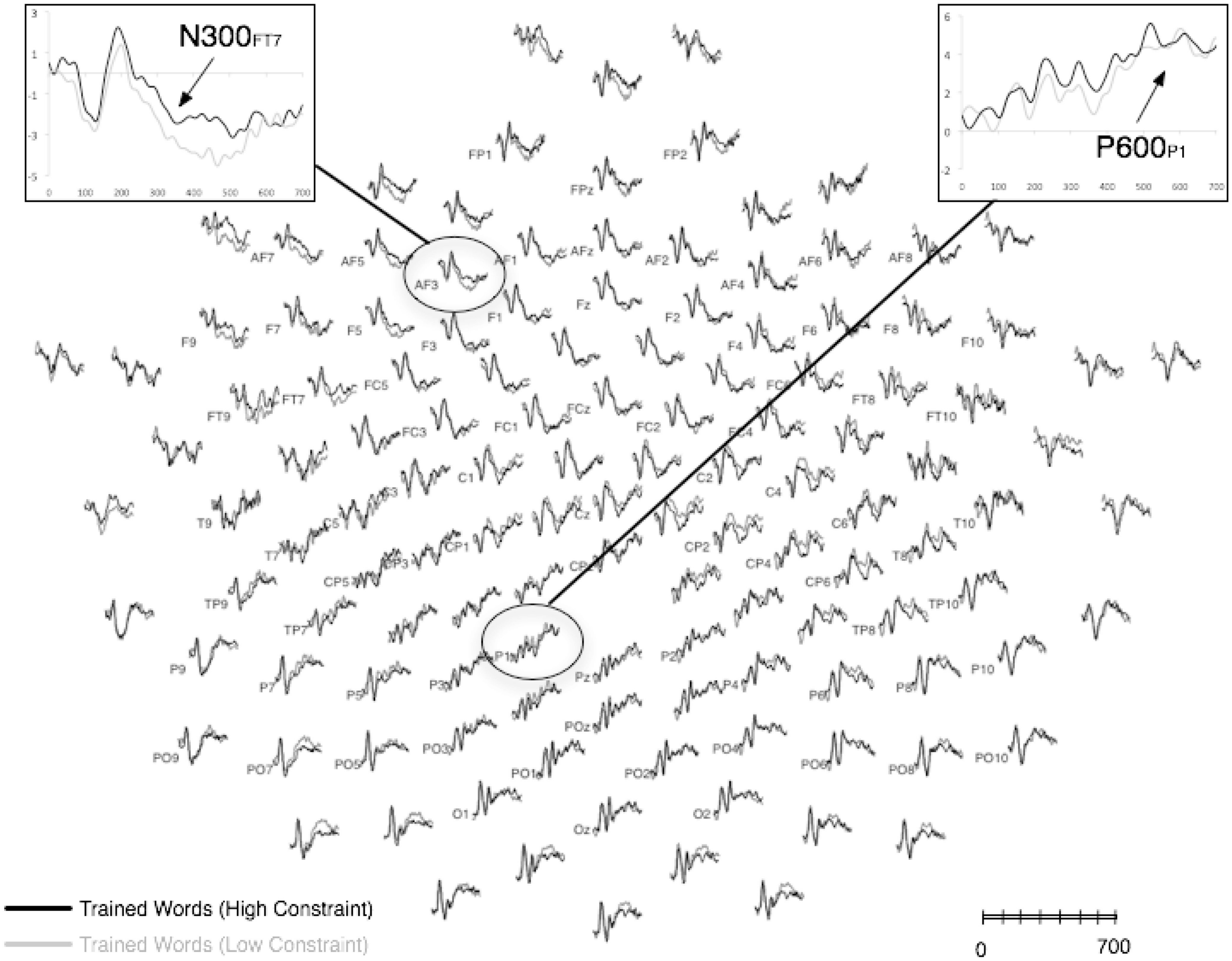


Figure 5.

129-channel ERPs in Session 1 (Task 1A: meaning-generation task) to rare words in two training conditions (HiConstr vs. LoConstr). *Black*, response to words trained in HiConstr condition. *Grey*, response to words trained in LoConstr condition. Upper right-hand panel shows enlarged plot of ERP responses at left frontal electrode AF3: After training, there was an increased left anterior negativity to words in the LoConstr versus the HiConstr condition. Cf. Figure 3, statistical “t” maps showing Interaction Effect at ~450 ms.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

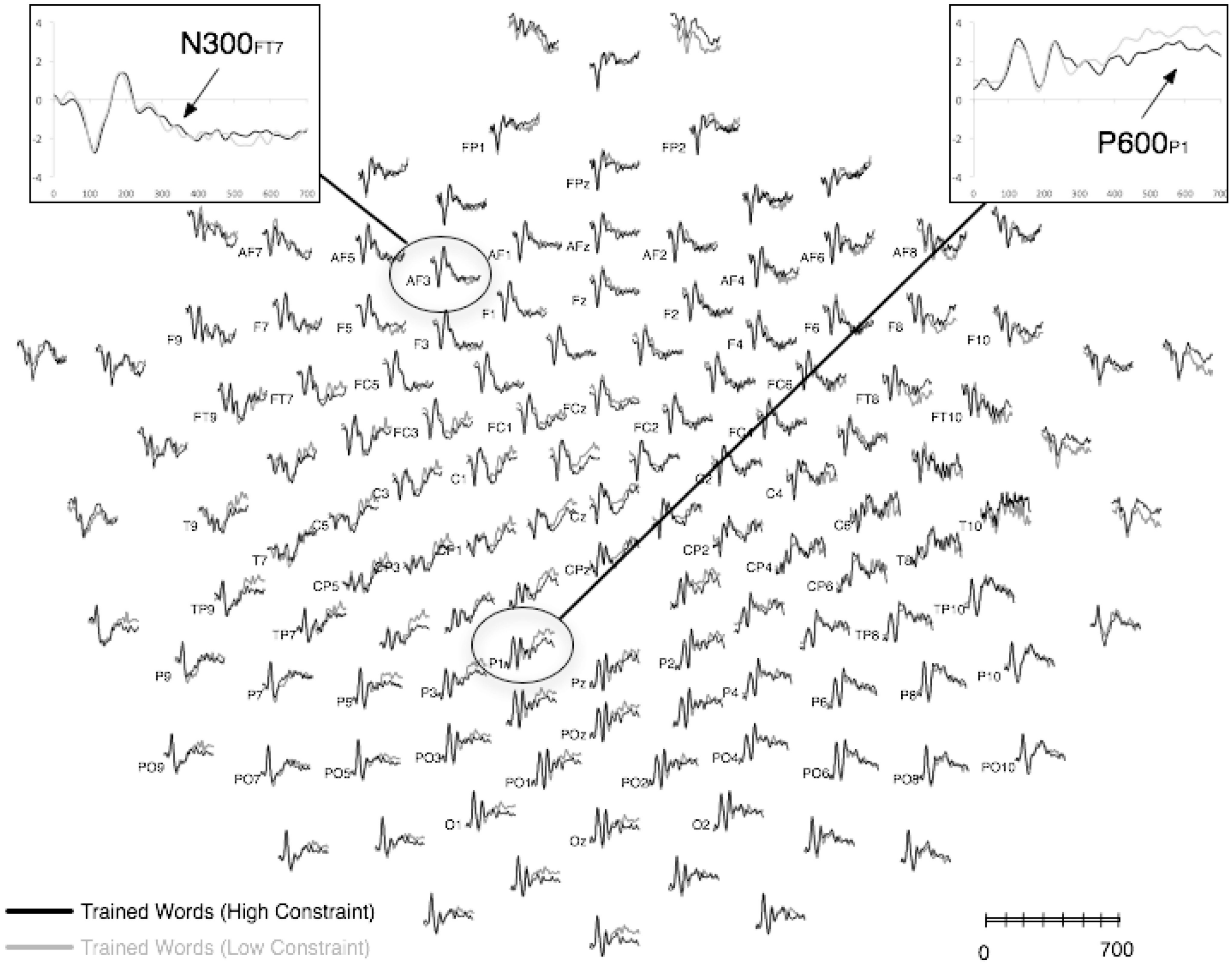


Figure 6.
129-channel ERPs in Session 2 (Task 1A: meaning-generation task) to rare words in two training conditions (HiConstr vs. LoConstr). *Black*, response to words trained in HiConstr condition. *Grey*, response to words trained in LoConstr condition. Upper right-hand panel shows enlarged plot of ERP responses at left frontal electrode AF3: After a 2-day delay, the left anterior effect seen in Session 1 (Fig. 5) is gone. Upper right-hand panel shows enlarge plot of ERPs at channel P1: there is an increased P600 positivity to words trained in the HiConstr condition versus words in the LoConstr condition, an effect that was not present immediately after training (cf. Fig 5).

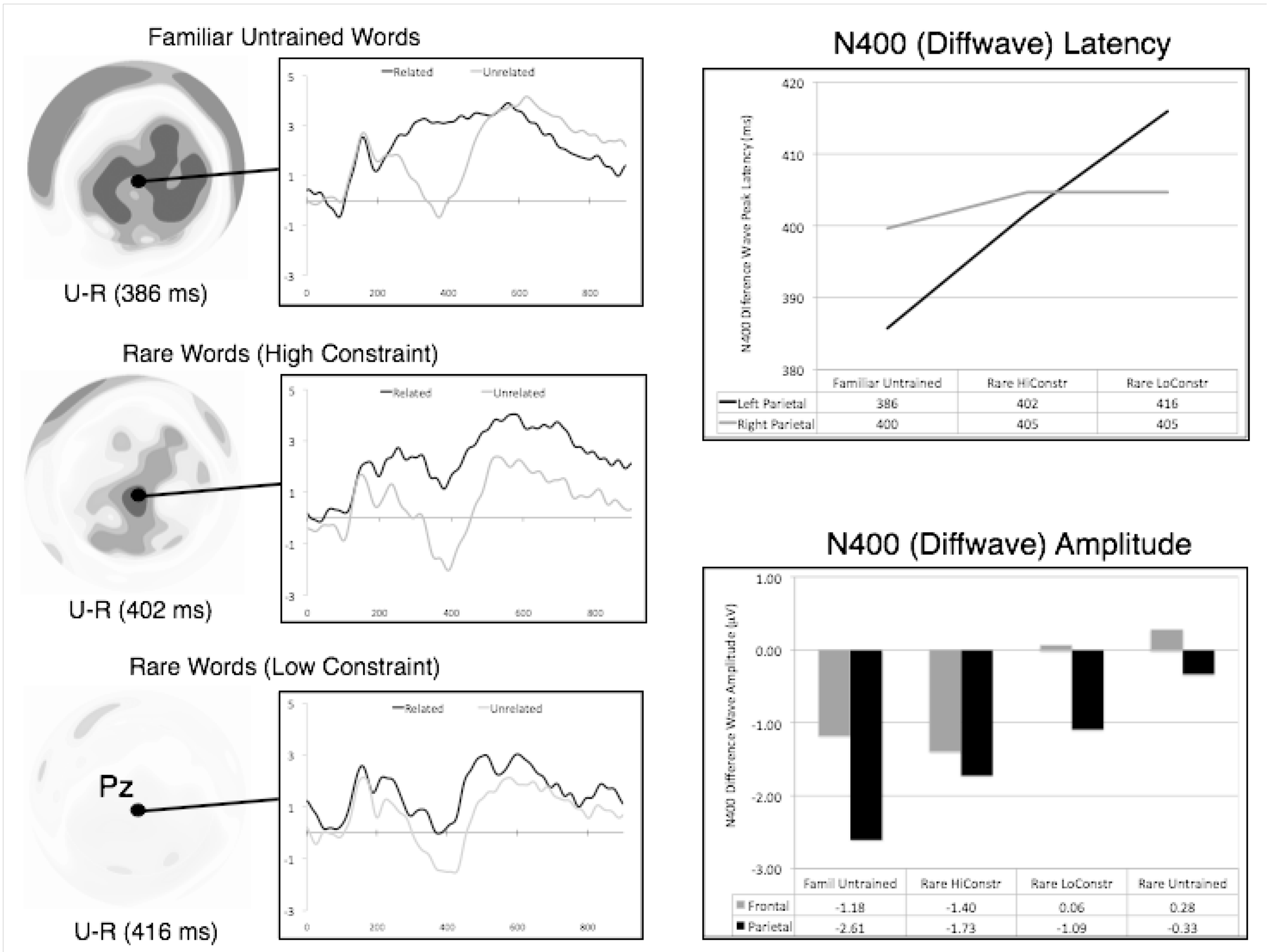
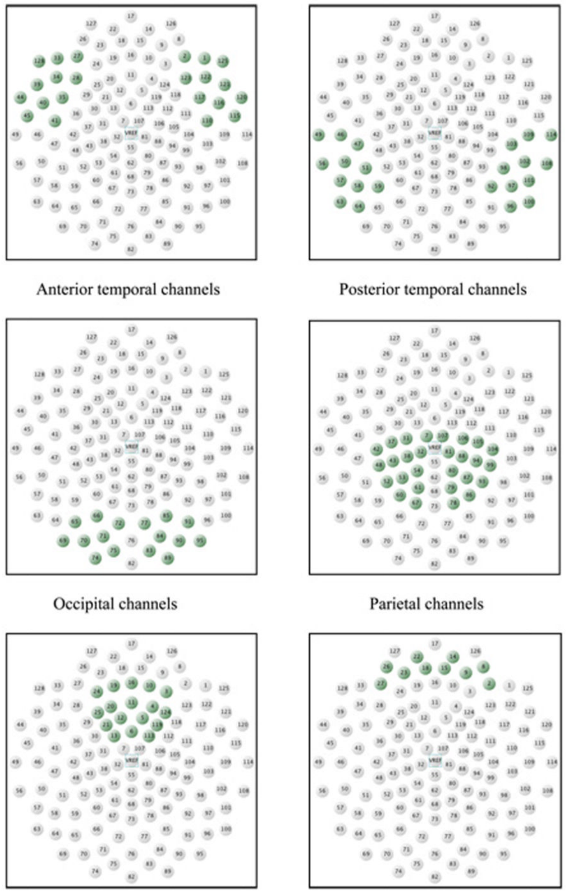


Figure 7.

Summary of ERP results for the N400 semantic priming/ transfer task (Task 2, Session 2). *Left*, Topographic “t-maps” plotted at peak latency for Familiar Untrained words and Rare words trained in High and Low Constraint contexts (Rare Untrained words did not elicit an N400 effect). ERP waveforms plotted at channel Pz: Black, related targets; Grey, unrelated targets. *Top right*, N400 latency effects over left parietal ROI (black line) and right parietal ROI (grey line). The main effect of Word and the Word × Laterality interaction both fell just short of significance (p = .06 for both comparisons). *Bottom right*, Differences in the magnitude of the N400 effect (Unrelated – Related) for all four word types, shown separately for frontal ROI (grey bars) and parietal ROI (black bars). There was a significant Word × Relatedness × ROI interaction, p < .01 (see text for details).



Appendix A.
Electrode Clusters

Table 1

Cloze probability and sentential constraint scores for high-constraint (HiConstr) versus low-constraint (LoConstr) contexts from the stimulus validation study. SE given in parentheses.

	HiConstr	LoConstr	t	df	p	
Cloze (Synonym 1)	0.24 (0.15)	0.07 (0.06)	12.37	714	< .001	***
Cloze (Synonym 2)	0.19 (0.17)	0.08 (0.07)	7.65	714	≈.001	**
Cloze (Modal Resp)	0.36 (0.01)	0.20 (0.11)	13.99	714	< .001	***
Cloze (Composite)	0.39 (0.12)	0.03 (0.07)	27.78	714	< .001	***
Sentential Constraint	13.12 (0.26)	19.22 (0.25)	−16.84	714	< .001	***

** p-values < .01.

*** p-values < .001.

Table 2

Mean accuracy and response times for the semantic priming task (Task 2, Session 2). SE given in parentheses.

	Accuracy		Response Times	
	Related	**Unrelated**	**Related**	**Unrelated**
Familiar words	.96 (.01)	.97 (.01)	597 (55)	517 (37)
Rare words (hi-constr)	.67 (.03)	.72 (.05)	737 (95)	606 (50)
Rare words (lo-constr)	.50 (.03)	.74 (.04)	723 (71)	664 (55)
Rare words (untrained)	.35 (.04)	.73 (.04)	720 (73)	690 (71)

Table 3

Summary of condition differences for all factors (patterns) in Session 1, Task 1A (Meaning -generation task).

Pattern	Effect	Omnibus Effects
F110 (P100)	Word	Word*Hemi, F(1.87, 26.13)=13.92, p< .001
	Time	Time*ROI, F(1.93, 27.04)=20.22, p< .001
	Word*Time	-
F170 (N170)	Word	-
	Time	Time*ROI, F(2.19, 30.61)=8.82, p< .001
	Word*Time	-
F220 (N220)	Word	Word*ROI*Hemi, F(4.86, 68.07)=13.92, p< .01
	Time	-
	Word*Time	-
F300 (N300)	Word	Word*ROI, F(2.69, 37.63)=3.94, p< .05
	Time	Time*ROI, F(1.27, 17.80)=5.37, p< .05
	Word*Time	-
F400 (N400)	Word	Word*ROI, F(2.41, 37.84)=3.13, p< .05
	Time	Time*ROI*Hemi, F(3.64, 50.92)=4.42, p< .01
	Word*Time	Word*Time*ROI, F(3.60, 50.45)=3.74, p< .01

Post-hoc Effects
LH: Fam > Rare (p< .001) RH: ns
Post > Pre: OCC, PAR, PTEMP (all p<.001) Pre > Post: ATEMP (p<.001), ORB (p< .01)
- -
- -
Post > Pre: FRONT, ORB (both p<.05) Pre > Post: PTEMP (p<.01), OCC (p< .05)
- -

Pattern	Effect	Omnibus Effects
		LH: Fam > Rare over OCC,PAR,PTEMP (all p< .05) RH: ns for all ROI
		- -
		- -
		Fam > Rare: FRONT (p< .01) -
		Post > Pre: OCC, PTEMP (both p<.01) Pre > Post: ATEMP, ORB (both p<.01)
		- -
		Fam > Rare: FRONT (p< .01) -
		Post > Pre: left OCC, PTEMP, PAR (all p<01) Pre > Post: left FRONT, ORB, ATEMP all p<.01)
		HiConstr > LoConstr: Post only over ORB (p<.05); Pre n.s. -

Famil= familiar (untrained) words. Rare = rare trained words. Pre = pretest. Post = posttest. HiC = rare words trained in high-constraint contexts. LoC = rare words trained in low -constraint contexts. ‘X > Y’ = X more positive or less negative than Y.