Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 6, 1281–1292

Copyright 2005 by the American Psychological Association
0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.6.1281

Word Learning and Individual Differences in Word Learning Reflected in Event-Related Potentials

Charles A. Perfetti, Edward W. Wlotko, and Lesley A. Hart
University of Pittsburgh

Adults learned the meanings of rare words (e.g., *gloaming*) and then made meaning judgments on pairs of words. The 1st word was a trained rare word, an untrained rare word, or an untrained familiar word. Event-related potentials distinguished trained rare words from both untrained rare and familiar words, first at 140 ms and again at 400–600 ms after onset of the 1st word. These results may point to an episodic memory effect. The 2nd word produced an N400 that distinguished trained and familiar word pairs that were related in meaning from unrelated word pairs. Skilled comprehenders learned more words than less skilled comprehenders and showed a stronger episodic memory effect at 400–600 ms on the 1st word and a stronger N400 effect on the 2nd word. These results suggest that superior word learning among skilled comprehenders may arise from a stronger episodic trace that includes orthographic and meaning information and illustrate, how an episodic theory of word identification can explain reading skill.

Keywords: word learning, ERP study, N400, and vocabulary learning

Adult English speakers know the meanings of thousands of words and are vaguely familiar with many more, on one estimate learning about 3,000 new words a year from the beginning of literacy (Nagy & Herman, 1987). Nearly all college students know the meanings of even many low-frequency nouns such as *rubble, flint,* and *abstention*, all of which have printed word frequencies of less than 5 per million words of text in some word counts (e.g., Kučera & Francis, 1967). However, most students do not know the meaning of *gloaming, ibex*, and *agog*, rare words that fail to occur in some of these word counts (e.g., Kučera & Francis, 1967). However, just as many people learn the meaning of *abstention* from some reading or spoken language experience, some will also encounter *gloaming*, *ibex*, or some other word they do not know, and perhaps add its form and something about its meaning to their mental lexicon.

Our interest here is in examining the consequences of learning a new word for subsequent encounters with the word. Of course, if the meaning of the word has been learned, we should observe that the learner can recognize the word and understand its meaning. However, beyond this behavioral outcome, the process of reading the word, the time course of its identification, and meaning retrieval processes also should be affected. Recordings of event-related potentials (ERPs) may expose the consequences of learning in a word-processing task.

Beyond using ERPs to expose the consequences of new word learning, we examine a corollary question about individual differences in reading comprehension skill. Comprehension skill among children and adults is supported by their knowledge of words, including, according to the Lexical Quality Hypothesis (Perfetti & Hart, 2001), the precision of the reader's representation of orthography, phonology, and meaning, as well as the sheer number of known words. Skill in reading comprehension, to the extent that it has a word knowledge component, may also support the ability to learn the meanings of new words. Skilled readers may be better able to take advantage of word training events by remembering a new association between an orthographic form and a meaning. If so, we may observe the consequences of differential learning in an ERP component that reflects memory for recently learned words.

In examining these issues, we exploit two well-established ERP facts. One is that ERPs reveal the differences between "old" and "new" words in recognition memory experiments (Curran, 1999; Rugg & Doyle, 1992; Rugg & Nagy, 1989). A previously encountered word produces a late positive-going wave (peaking at around 600 ms) following the onset of the word compared with a word not previously encountered in the experiment ("new word"). This late positive wave (or P600 component) is thus a marker of an episodic memory trace (see Rugg, 1995 for a review). Thus, if we have recently taught the meaning of a word to a participant, presenting this word should evoke a P600 compared with a word that was not taught.

The second fact is that an ERP component, the N400, reflects the meaning congruence between a word and its previous context. In sentence contexts (Kutas & Hillyard, 1980) and in single-word semantic priming contexts (Nobre & McCarthy, 1994), a word that is incongruent with its context produces a negative-going wave peaking at about 400 ms after the onset of the word, whereas a congruent word produces a reduced N400. Thus, if we test whether a participant has learned the meaning of a word that was taught by presenting the taught word followed by a word that could be related in meaning, we expect to see a reduced N400 in the related case, compared with the unrelated case.

Charles A. Perfetti, Edward W. Wlotko, and Lesley A. Hart, Learning Research and Development Center, University of Pittsburgh.

This research was supported by U.S. Department of Education Grant R305G020006.

We gratefully acknowledge the important contributions of Dayne Grove and Chin-Lung Yang to data analysis and the helpful comments of Erik Reichle and Gwen Frishkoff on a draft of the article.

Correspondence concerning this article should be addressed to Charles Perfetti, Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260. E-mail: perfetti@pitt.edu

In the study reported here, college students first learned the meanings of very rare English words (e.g., *gloaming*). We controlled the selection of rare words for each participant individually such that the rare words were unfamiliar to a given participant prior to the experiment. The learning phase of the study was very simple—the presentation of the rare word on one side of a card and a brief definition on the opposite side. Following learning, participants made meaning judgments about pairs of words, presented one at a time while we recorded the ERPs. In the meaning judgment task, the first word was from one of three categories: the rare words that we had just taught to this participant, rare words that we did not teach to the participant, and familiar, medium-frequency words that the participant had not seen during the experimental session. Each word was followed by a word that was either related or unrelated in meaning, and the participant pressed a "yes" or "no" button to indicate their decision about whether the two words were related in meaning. For example, *gloaming* followed by *twilight* would require a "yes" response.

We hypothesized that following word learning, we would see ERP evidence, during a meaning judgment task, that the participant had become familiar with the words we taught. The evidence for this would come from an indicator of episodic memory, specifically a late positivity (P600) that would show differences between trained words and both untrained rare words and familiar (but not recently viewed) words. We hypothesized also that during the reading of a second word, we would see evidence for a meaning match for trained rare and familiar words, compared with untrained rare words, in a reduced N400. Finally, we sought to test the hypotheses that skilled comprehenders would show more effective learning of rare words and that ERP differences would reflect stronger recognition and meaning match components. In addition to these targeted tests, we assumed an exploratory approach to other ERP results, for example, whether we could observe a meaning response as distinct from an episodic memory response. If so, some ERP component should separate both familiar words and trained rare words from untrained rare words.

Method

Participants

Twenty-four undergraduates from the University of Pittsburgh Psychology Department participant pool provided data for the experiment. Data from 6 additional participants were not analyzed because of hardware malfunction, and data from 2 participants were rejected due to excessive recording artifacts. Participants were invited from a larger pool whose individuals had previously completed a variety of reading-related tasks. Twelve were skilled readers and 12 were less skilled, as determined by performance on the Nelson–Denny comprehension test. The two skill groups were drawn from pools that included the top 20% of those tested (skilled comprehenders) and the bottom 20% of those tested (less skilled comprehenders). All skilled comprehenders exceeded an accuracy of 64% on the Nelson–Denny; no less skilled comprehender exceeded an accuracy of 50%.

Procedure

Word Selection

Participants first completed a paper-and-pencil word detection task. They viewed a list of 250 letter strings, consisting of 135 rare words that did not appear in the Kučera & Francis (1967) corpus of over a million words, 51 relatively common words (a Kučera–Francis frequency rating of 40 words per million or greater), and 64 pseudowords, that is, legally spelled and pronounceable nonwords. They were instructed to mark only the letter strings that they were sure were real words in English. From the results of this task, a stimulus list was constructed for each participant by randomly selecting for training 60 of the rare words that the participant failed to mark as words. The remaining rare words that were not marked as words by the participant became the set of untrained rare words. Those relatively common words that the participant did mark as words became the set of familiar words. Thus, this procedure resulted in three classes of words, two classes of rare words—trained and untrained—and a class of familiar words. Notice that the procedure produced a different set of randomly selected trained words for each participant.

Training

Participants studied the 60 rare training words for 45 min. The experimenter presented them with flashcards containing words on the front and their definitions on the back. For example, *gloaming* was defined as "the twilight period before dark" and *clowder* was "a collection or group of cats." Participants were instructed that they would be given 45 min to learn the 60 words and that they should become as familiar with the words and their definitions as they could, with the understanding that they might not be able to learn all of them.

Posttraining Semantic Decisions

After the training period, participants performed a semantic decision task while electroencephalograms (EEGs) were recorded. On each trial, a word was selected randomly from the total set of words (trained, untrained, and familiar) and presented for 1,000 ms. The offset of this word was followed immediately by a second word, a meaning probe, which remained visible while the participant responded with a button press to indicate whether it was related in meaning to the first word. Each trial was preceded by a fixation cross for 400 to 550 ms. (Variability was added to reduce any impact of nonstimulus-related time-locked electrical activity.) On half of the trials, the probe word was semantically related to the first word; on the other half, the two words were unrelated. Semantically related pairs were created by experimenter judgment. Many trained words were paired with a meaning probe that had occurred as part of the definition (e.g., *clowder–cats*), but others were paired with a word that had not occurred as part of the training definition (e.g., *gloaming–evening*). Semantically unrelated pairs were created by shuffling the word pairs in each individual participant's stimulus list. Each stimulus list contained the pairing of a stimulus word with both a related and an unrelated probe word. The order of stimulus words as well as the order of related and unrelated probe words was randomized for each participant.

Participants were instructed to press the "1" key with their right index finger if the two words in a trial were related in meaning and to press the "2" key with their right middle finger if the two words were not related in meaning. The meaning probe word was removed from the screen when a response was made or after 2,000 ms elapsed, whichever came first. Accuracy feedback (correct or incorrect response) was presented after each trial in the form of a stylized smiling face for correct answers and a frowning face for incorrect answers. The feedback image remained in view for 800 ms prior to the onset of the next trial.

Recordings

Scalp potentials were recorded from 128 sites with Electrical Geodesic, Inc.'s (EGI's) Geodesic Sensor Net with Ag/AgCl electrodes. The potentials were recorded with a sampling rate of 500 Hz and a hardware bandpass filter of 0.1 to 200 Hz. Impedances generally were kept below a threshold of 40 kΩ. A digital low-pass elliptical filter of 30 Hz was applied to the recordings. The ERPs were stimulus-locked averages consisting of a 100-ms baseline and a 1,800-ms epoch defined by the presentation of the stimulus word (Word 1 = 1,000 ms) plus the meaning probe (Word 2 = 800 ms). Bad channels were removed from the recordings and replaced by spherical spline interpolation with data from the remaining channels. Trials containing eye movement, eyeblink, and channel artifacts were rejected and not used in analysis. ERPs were transformed using the average reference. Finally, the ERP segments were corrected relative to a 100-ms baseline. Following rejection of trials with artifacts, two thirds of the participants had 30 trials per condition, and all had at least 20 trials.

Results

Behavioral Results

Tables 1 and 2 show the accuracy and decision time results for the meaning judgment task. The results showed higher accuracies for familiar words (87.3%, $SD = 0.02\%$) and trained rare words (83.7%, $SD = 0.01\%$) compared with untrained rare words (56.4%, $SD = 0.10\%$). For accuracy, an analysis of variance (ANOVA) showed a main effect of word type, $F(2, 44) = 331.00$, $p < .01$, and relatedness, $F(1, 22) = 7.58$, $p = .01$, as well as a Word Type × Relatedness interaction, $F(2, 44) = 50.70$, $p < .01$. The interaction showed a "no" response bias for the untrained rare words only: a tendency to judge an untrained word and its meaning probe as unrelated. There was also a Word Type × Skill interaction, $F(2, 44) = 6.72$, $p < .03$, which reflects the fact that skilled comprehenders were more accurate than less skilled comprehenders for trained rare words (about 10% difference) but not for untrained rare words (<1% difference) or for familiar words (about 1% difference).

As shown in Table 2, response times varied between 700 ms and 900 ms, depending on word type, relatedness, and correctness of response. Responses to untrained rare words were slower than to trained and familiar words, responses to related words were faster than responses to unrelated words, and correct responses were faster than incorrect responses. An analysis of meaning decision times showed a main effect of word type, $F(2, 42) = 8.17$, $p < .01$; relatedness, $F(1, 21) = 10.14$, $p = <.01$, and response accuracy, $F(1, 21) = 57.41$, $p < .01$. However, an interaction showed that the difference between correct and incorrect responses was present for trained and familiar words only, with untrained rare words showing no effect, $F(2, 42) = 20.34$, $p < .01$.

Table 1
Behavioral Results: Percent Accuracy

Condition	Skilled	Less skilled	*M*
Familiar			
Related	87.6	87.1	87.4
Unrelated	88.3	86.3	87.3
M	88.0	86.7	87.4
Trained rare			
Related	89.4	81.6	85.5
Unrelated	87.8	76.0	83.7
M	88.6	78.8	81.9
Untrained rare			
Related	40.4	39.6	40.0
Unrelated	69.6	69.0	69.3
M	55.0	54.3	54.7

Event-Related Potential Results

To take advantage of the high-density recordings across the full epoch, we carried out a temporal principal components analysis (PCA) on the ERPs. The logic of PCA is to use the full set of electrodes and time points to determine the intercorrelations of ERP shifts over time, allowing an overall data-driven view, from which factors emerge for further statistical testing. The temporal variables consisted of 900 time points across two words (1,800 ms at 2-ms samples), with the data consisting of the recording from each electrode in each condition for each participant. Thus, this PCA analyzes an epoch that consists of two words, allowing us to observe not only the ERP effects on the first word—the trained rare, untrained rare, and untrained familiar words—but also on the second word, which participants used to make a decision.

The PCA used a correlation matrix and a promax rotation ($\kappa = 4$).[1] Because the promax rotation does not assume that the factors are orthogonal, it has some advantages for temporally correlated ERP data (Dien & Frishkoff, 2005; Dien, 1998). The results of the PCA showed 6 factors with eigenvalues greater than 10, which we retained for further analysis. To these, we added Factor 11, which corresponded to 400 ms after the onset of the probe word, a time point for which we hypothesized a meaning congruence indicator (N400). Together, these factors accounted for 93.3% of the total variance.

The factor scores for these seven factors, shown in Figure 1, were analyzed with repeated-measures ANOVA. Fifteen scalp locations were chosen for analysis based on the 10/10 system, a set of 10 locations commonly reported in ERP research, plus 5 additional locations. To each of these 15 recording locations (3 central, 3 parietal, 4 temporal, and 2 occipital), we added those channels immediately adjacent. This creates a cluster scheme in which the data from a cluster are the average of between 4 and 8 electrodes, with most clusters having 7 electrodes. The original waveforms for the 15 single electrodes are shown in Figure 2, and the clusters are shown in Figure 3. We based our clusters on standard 10/10 locations in order to have our results more easily compared with

---

[1] PCA involves a decision about the form of the association matrix that calculates relationships between pairs of variables by associating their data points. Matrices are usually of one of two types: covariance and correlation. A significant methodological literature has developed around the consequences of choosing one over the other. A recent analysis concluded that misallocation of variance is more likely with correlation than covariance matrices (Dien & Frishkoff, 2005), whereas other analyses conclude that differences are negligible for most ERP data (e.g., Chapman & McCrary, 1995; Van Boxtel, 1998). We used a correlation matrix, which, because it normalizes variances across variables, allows all variables equal weight in determining the factor structure. Although this can increase the chances that noise will influence the factor solution, it also can help to detect small but theoretically important variables in the PCA, something that seems advantageous in the relatively unexamined word learning question we are studying here. To test the robustness of our conclusions over these matrix alternatives, we also carried out a covariance-based PCA. Its factor solutions converged on those we report here, with only minor differences on the second word that do not affect our interpretation.

Table 2
Behavioral Results: Semantic Decision Times (in Milliseconds)

Condition	Skilled			Less skilled			
	Correct	Incorrect	*M*	Correct	Incorrect	*M*	*M*
Familiar							
Related	730	897	814	702	778	740	777
Unrelated	763	917	840	743	886	815	828
M	747	907	827	723	832	778	801
Trained rare							
Related	700	909	805	709	823	766	786
Unrelated	771	878	825	758	837	798	812
M	736	894	815	734	830	782	799
Untrained rare							
Related	891	873	882	840	804	822	852
Unrelated	866	904	885	811	856	834	860
M	879	889	884	889	826	828	857
Overall *M*	787	896	842	768	830	796	819

any effects in the previous literature while still taking advantage of the benefits of clustering.[2]

For the ANOVAs, the 15 clusters organized factors corresponding to hemisphere (left, right, midline) and lobe (frontal, central, parietal, two temporal locations, and occipital). Because temporal and occipital locations do not have a "midline," we performed separate ANOVAs for midline and lateral locations. We tested the lateral locations using a Hemisphere (2) × Lobe (6) × Word Type (3) × Relatedness (3) × Skill (2, between subjects) ANOVA. The midline locations were tested with a Lobe (3) × Word Type (3) × Relatedness (2) × Skill (2, between subjects) ANOVA. For factors that did not show sphericity across factor levels, the Greenhouse–Geisser correction was used.

Our analysis spans a time period across the presentation of two words with the factors identified by the PCA. Four factors were associated with the time period of Word 1, the stimulus word: at 500 ms (Factor 2), 234 ms (Factor 4), at 140 ms (Factor 5), and 64 ms (Factor 6) after the onset of Word 1. Three factors peak within the time period of Word 2, the probe word: one peaks at 372 ms (Factor 11) after the onset of Word 2; another (Factor 1) is a broad "slow wave" factor that begins to rise sharply at approximately 300 ms after the onset of Word 2 and peaks at 626 ms after the onset. Factor 3 is a two-word factor, rising from about 600 ms after the onset of the first word and peaking at 32 ms after the onset of Word 2. In reporting the analyses of these factors below, we group them according to whether their peaks occurred during the first word or the second word and refer to them by the latencies of their peaks.

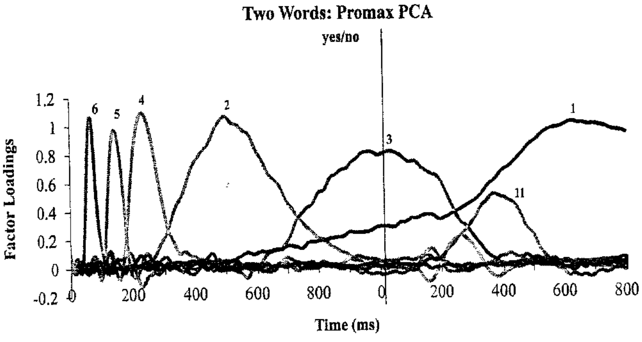


Figure 1. Principal components analysis (PCA) factors retained for analysis. Factors 1, 2, 3, 5, and 11 showed significant effects of experimental variables. Factors 4 and 6 did not show effects of experimental variables, instead reflecting general task effects.

Because two factors (peaks at 64 and 234 ms during Word 1) showed no effects of experimental conditions, we do not report their analysis below. The 64-ms peak appears to reflect exogenous factors associated with the onset of the first word. The 234 ms factor rises again in a less pronounced form at the same latency following the onset of Word 2, suggesting a word-onset related process. Its timing and topography (positive going in frontal sites, negative in posterior) are similar to ERP components that have been interpreted as graphic processing (Liu & Perfetti, 2003; Liu, Perfetti, & Hart, 2003) and graphic–phonological coding (Barnea & Breznitz, 1998) as well as more general attention (Hackley, Woldorff, & Hillyard, 1990), feature detection (Luck & Hillyard, 1994), and short-term storage (Chapman, McCrary, & Chapman, 1978) processes.

Word 1

140 ms. The 140-ms factor showed a main effect of word type in the midline ANOVA, $F(2, 44) = 3.97, p = .03$. Trained words were more negative than untrained words and familiar words. Pairwise tests showed that trained rare words were distinguished from both familiar ($p = .03$) and untrained rare words ($p = .01$), which did not differ. The greater negativity at 140 ms for trained words, although not pronounced, can be seen in all frontal and central electrodes as well as left temporal and parietal sites. The

---

[2] ERP clustering does not yet have a standardized procedure and may not be appropriate for applications to traditional low-density electrode recordings. However, for high-density recordings, clustering, or regional averaging, of electrodes has advantages. It reduces the noise associated with individual electrodes while allowing a large number of electrodes to be used (Dien & Santuzzi, 2005).

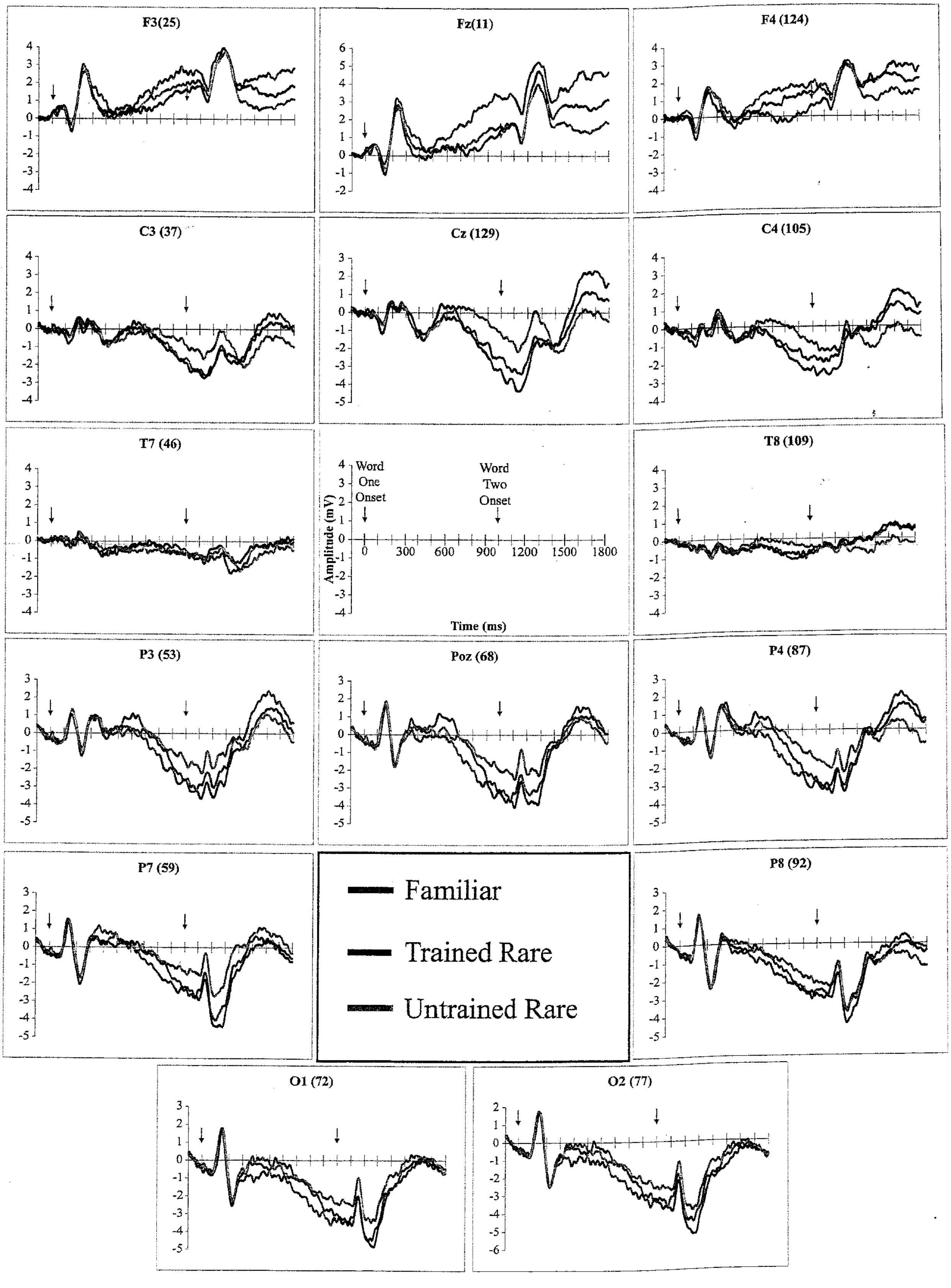
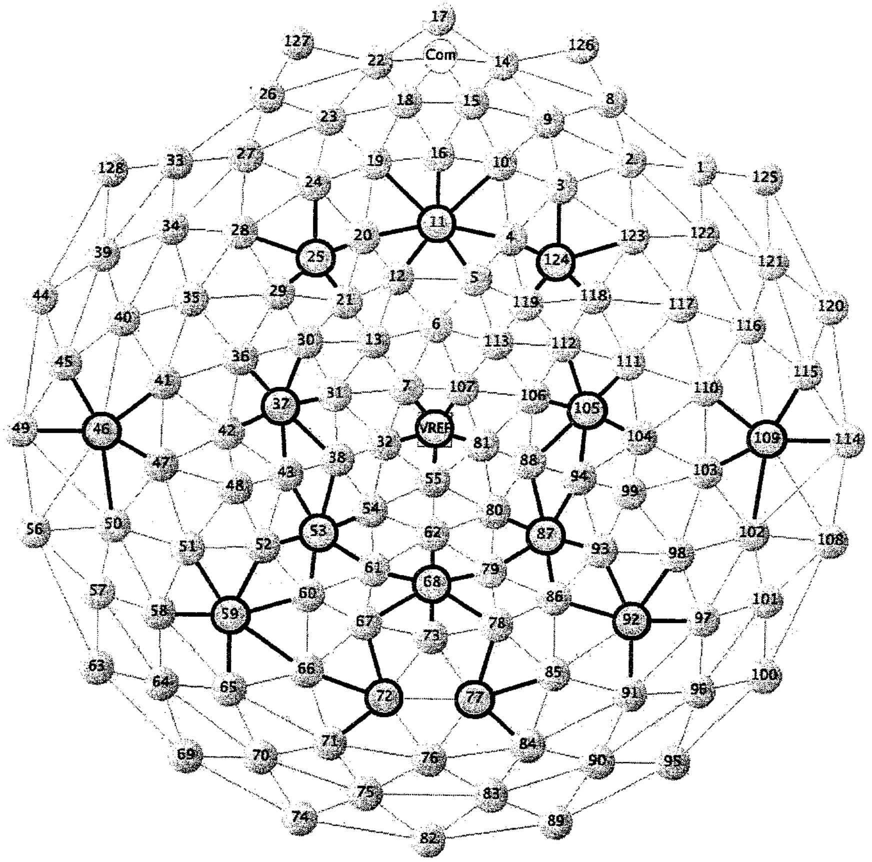


Figure 2. Grand average waveforms for 15 electrodes.



Cluster Locations	
10/10 System	**EGI 128**
F3	25
Fz	11
F4	124
C3	37
Cz	VREF
C4	105
T7	46
T8	109
P3	53
Poz	68
P4	87
P7	59
P8	92
O1	72
O2	77

Figure 3. Electrode clusters. Fifteen electrodes from the 10–10 were defined as cluster centers. EGI = Electrical Geodesic, Inc. sensor; VREF = voltage-reference electrode.

general topography for this early time window is a negativity at frontal and central sites, with a positivity at parietal sites, midline lobe effect, $F(2, 44) = 6.96$, $p < .01$, Greenhouse–Geisser $\varepsilon = .717$. The lateral analysis also showed this pattern, bilaterally negative at frontal and central locations and positive at parietal, posterior temporal, and occipital locations, lobe effect $F(5, 110) = 11.25$, Greenhouse–Geisser $\varepsilon = .264$.

500 ms. In this time period, trained words were more positive than other word types in posterior locations. This trained-word factor appears to correspond to the old–new memory effect (P600 or late positive complex [LPC]) that is found for recently viewed and recognized items. Furthermore, this trained-word factor was more pronounced for skilled comprehenders than less skilled comprehenders. These conclusions are supported by Word Type × Lobe interaction, $F(10, 15) = 4.73$, $p < .01$, Greenhouse–Geisser $\varepsilon = .43$, and a Word Type × Lobe × Skill interaction, $F(10, 15) = 3.76$, $p < .01$, Greenhouse–Geisser $\varepsilon = .786$ in the lateral ANOVA; a Word Type × Lobe interaction also occurred in the midline ANOVA, $F(4, 88) = 5.24$, $p < .01$, Greenhouse–Geisser $\varepsilon = .703$. Figure 4 shows the Word Type × Lobe × Skill interaction and the waveforms for the Poz cluster that reflects the interaction. The more pronounced positivity is visible in this central parietal cluster as well as in individual parietal electrodes (and Cz) of Figure 2.[3]

32 ms After Onset of Word 2. The factor that spans the first and second word appears to be sensitive to a separation of all three word types. This factor rises from 600 ms of the first word, spilling over into the presentation of the probe word. Word Type × Lobe interactions were present for both the lateral ANOVA, $F(10, 220) = 7.71$, $p < .01$, Greenhouse–Geisser $\varepsilon = .36$ and midline ANOVA, $F(4, 88) = 22.75$, $p < .01$, Greenhouse–Geisser $\varepsilon = .333$. The midline analysis showed the frontal sites to be more positive for familiar words than for rare words (both trained or untrained), whereas the central and parietal sites distinguished both familiar and trained rare words (more negative) from untrained rare words. The lateral analysis also showed that familiar and trained words (more negative) were distinguished from untrained words (less negative) in the central sites, whereas trained rare words were less positive than familiar and untrained rare words in frontal sites. However, this pattern was modified by a Word Type × Lobe × Hemisphere interaction in the lateral ANOVA, $F(10, 220) = 3.59$, $p < .01$, Greenhouse–Geisser $\varepsilon = .443$. Trained words were positive only in left frontal sites, whereas

---

[3] In addition to the analysis based on the two-word epoch reported here, we carried out an analysis separately for each word, 1,000 ms for Word 1 and 1,000 ms for Word 2. The trained-word effect actually is seen even more clearly in the separate word analysis. In general, however, the two-word and separate word analyses showed very similar patterns, except that only in the two-word epoch analysis can one see a factor that overlaps the offset of the first word and the onset of the second word.

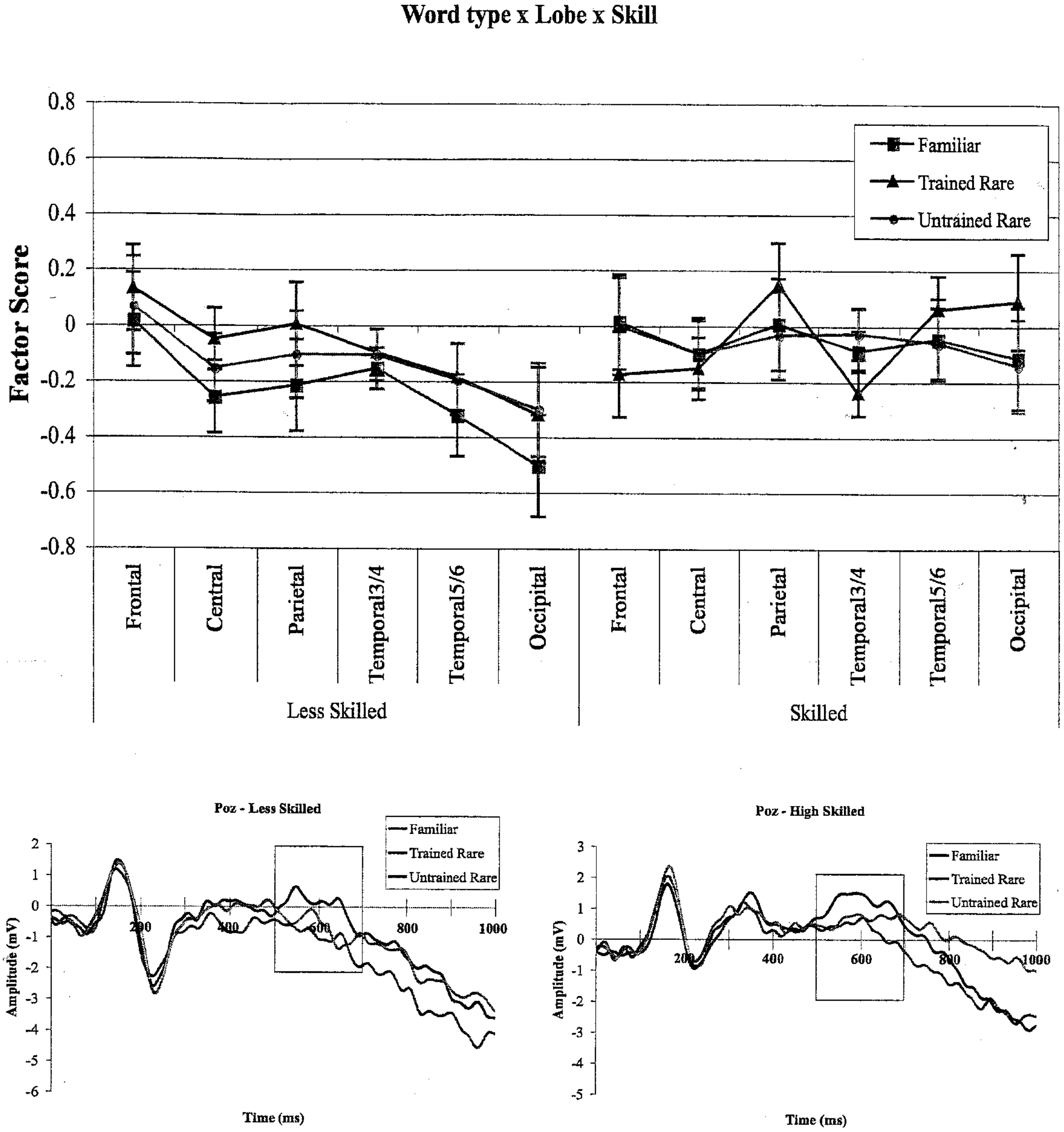


Figure 4. Word Type × Lobe × Skill interaction for Word 1, 500-ms factor (top) and Poz clusters for skilled and less skilled readers showing the interaction 500–600 ms after the onset of Word 1 (bottom).

untrained words were bilaterally positive, with familiar words showing an intermediate pattern, bilateral but more positive in left than right sites. This pattern can be seen in the waveforms in Figure 2. Finally, both the midline and lateral analyses showed a Word Type × Skill interaction, respectively, $F(2, 44) = 5.57, p < .01$, Greenhouse–Geisser $\varepsilon = .935$, and $F(2, 44) = 4.74, p = .02$, Greenhouse–Geisser $\varepsilon = .920$. The distinction among the three word types was more pronounced for skilled comprehenders. As can be seen in the midline interaction shown in Figure 5, skilled comprehenders were sensitive to differences among all three word types in this time window, whereas less skilled comprehenders showed less sensitivity, especially to the difference between familiar and trained words.

Word 2

372 ms. Because the task required a semantic decision, we hypothesized an N400 relatedness effect as the participant read the second word; accordingly, we tested the 372 ms factor, although its eigenvalue was less than 10. The hypothesis predicts relatedness effects according to whether Word 2 was related or unrelated in meaning to Word 1. The lateral analysis showed a significant relatedness effect, $F(1, 22) = 4.67, p = .04$, and significant interactions of Word Type × Relatedness × Hemisphere, $F(2, 44) = 3.97, p = .03$, and Word Type × Lobe × Hemisphere × Skill, $F(10, 220) = 3.12, p = .02$, Greenhouse–Geisser $\varepsilon = .41$. The midline ANOVA showed interactions of Word Type × Re-

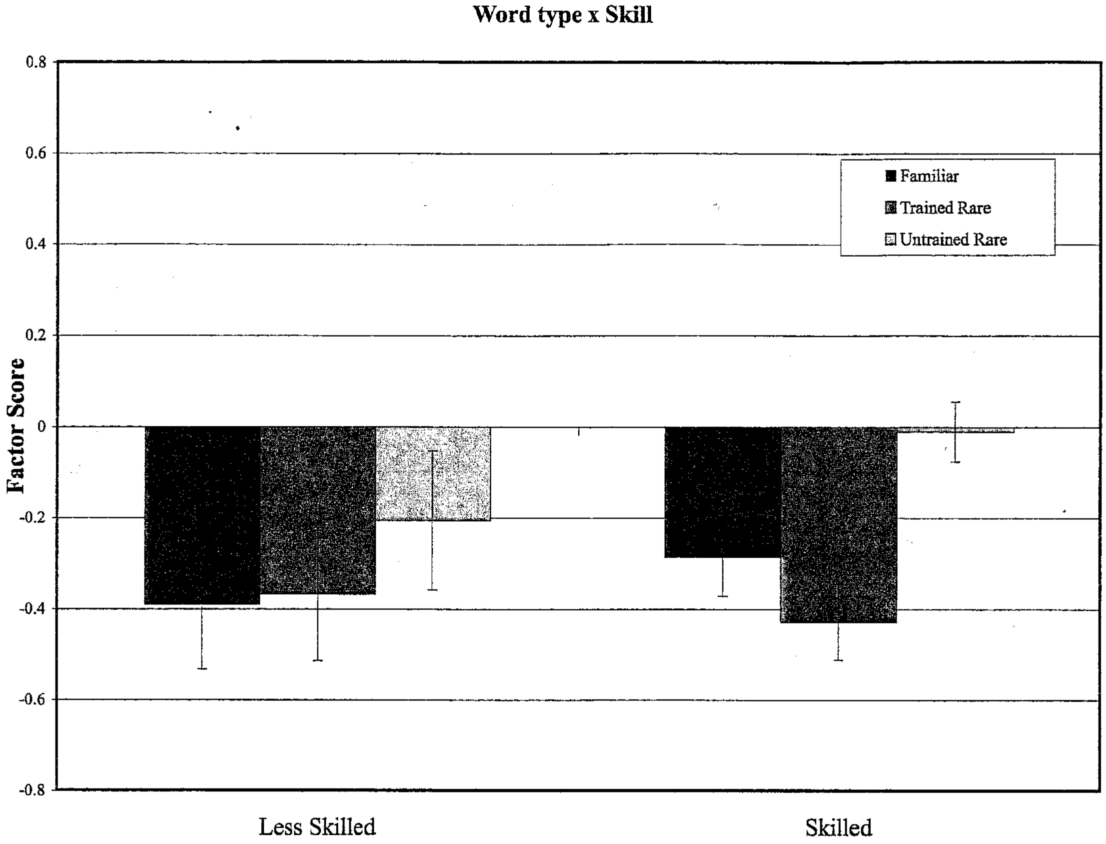


Figure 5. Word Type × Skill, two-word factor, with peak at 32 ms after onset of Word 2.

latedness, $F(2, 44) = 6.74, p < .01$, and Word × Lobe, $F(4, 88) = 8.76, p < .01$. (See Figures 6 and 7.) The interactions confirmed that an N400 effect, which was larger in the right than the left recording locations, was present for familiar words and trained words, but not untrained words. The interactions also reflected a larger N400 effect for skilled comprehenders than for less skilled comprehenders. As can be seen in Figure 6, skilled comprehenders showed a stronger relatedness effect, especially in midcentral and parietal sites. Figure 7 shows that the word type patterns for the two skill groups were similar in right hemisphere sites (although frontal sites were more positive for skilled comprehenders); however, in the left hemisphere temporal (T3) cluster, less skilled comprehenders showed no separation of word types, whereas for skilled comprehenders, familiar and trained words were separated clearly from untrained rare words. Figure 8 shows the basic N400 effect in the average waveforms from the right parietal (P4) cluster, where the reduced negativity for related words can be seen for familiar and trained words. Notice the lack of an effect for untrained rare words, for which one expects their relatedness to be undetected.

Slow wave factor. The "slow wave" factor, a typical component in PCAs for ERP data, can include noncognitive time-locked factors as well as cognitive factors. In these data, this factor reflected a clear cognitive component in both ANOVAs. A word type effect appeared in both the lateral, $F(2, 44) = 9.32, p < .01$, Greenhouse–Geisser $\varepsilon = .780$, and midline analyses, $F(2, 44) = 17.330, p < .01$. Both analyses also showed an interaction of Word Type × Lobe, lateral $F(10, 220) = 5.02, p < .01$, Greenhouse–Geisser $\varepsilon = .415$, and midline $F(4, 88) = 5.70, p < .01$, Greenhouse–Geisser $\varepsilon = .618$. The interaction corresponds to the late separation of the word types, especially in frontal regions, reduced in posterior locations. The general pattern is that familiar words are most positive and untrained rare words are most negative, as can be seen in the waveforms of Figure 2. This pattern may reflect meaning retrieval or verification processes that are stronger for the familiar words than for rare words, especially untrained rare words.

Discussion

Our results demonstrate that ERP measures can be used as indicators of word learning. When people learned the meaning equivalence of a rare word such as *gloaming,* the consequences of this learning were observable when the learners made meaning judgments on the word. Accuracy of meaning judgments was about 84% for 60 words following 45 min of training, comparable to the accuracy on medium-frequency words already familiar to the learners. More interesting is that the effects of learning were observed in ERP records as well as in behavioral measures.

The effect of training was seen in a late positive shift that we identify as an episodic training effect, similar to an old–new P600 observed in recognition memory. Trained words showed the effect, whereas untrained rare words and familiar words did not. We interpret this effect as an episodic memory indicator, that is, that

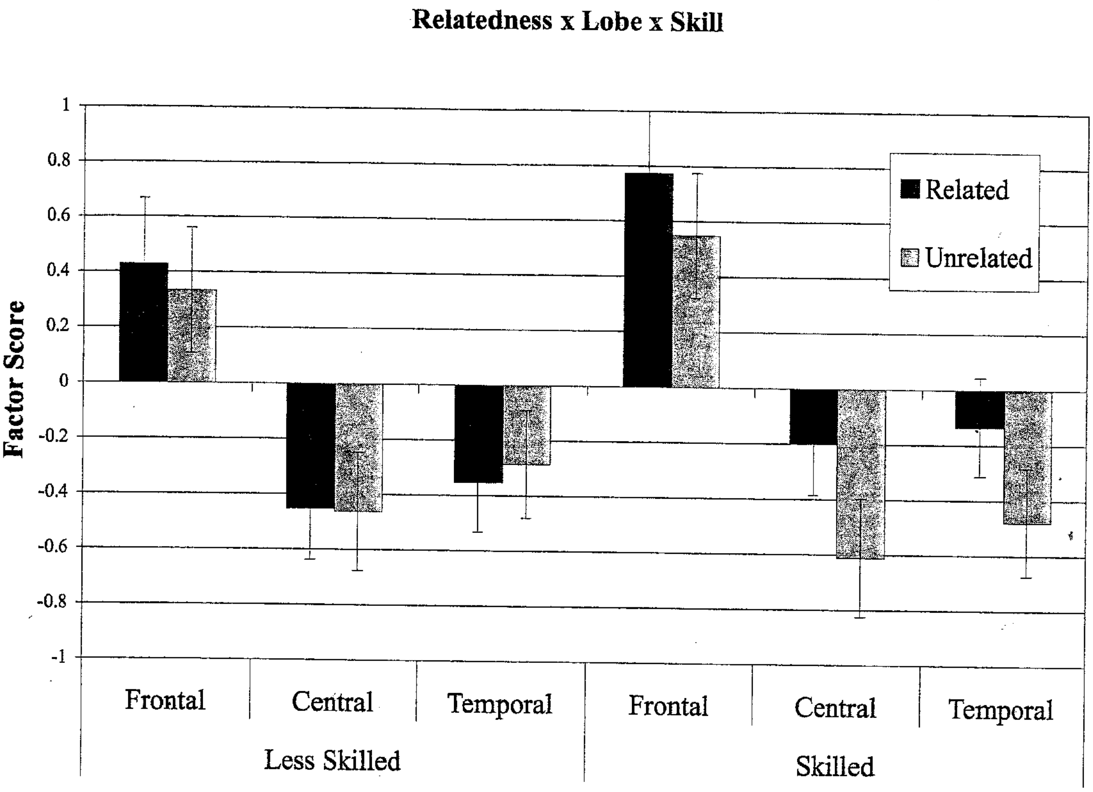


Figure 6. Relatedness × Lobe × Skill interaction for Word 2 at 372 ms.

participants were recognizing the trained words as recently experienced during training. However, the behavioral data indicate that more than mere recognition occurred, because participants were nearly as accurate on the rare trained words as on familiar words.

Moreover, trained rare words separated from both untrained and familiar words even prior to this late positive shift, at around 140 ms after the onset of Word 1. This time window is rather early to be interpreted in the same manner as the late positivity old–new effect. However, it might reflect a process in which visual attention is drawn to features of a word that has been recently viewed, a slightly different form of an old–new effect reflecting something less than full orthographic analysis. Although word-related components, including a sensitivity to word frequency, have been observed around 140 ms in a study by Sereno, Rayner, and Posner (1998), the topography in that study differed from what we found here. Other studies have found early word-processing components at around 170 ms (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999), but again the topography is different. Furthermore, what is distinctive in the present study is the sensitivity of an early component to the episodic status of the word, not its frequency or orthography—a component not dependent on the word itself but on its recent exposure history.

Neither the early nor the later separation of trained words from other words can be taken as an indicator of learning word meanings. For a learning indicator, we have, in addition to behavioral results, ERP data from the N400, which was observed during the presentation of the second word, the meaning probe. In a semantic judgment task, an unrelated probe word is expected to produce a larger N400 than a related probe word, because the unrelated probe is semantically incongruent with the first word. In the present study, unrelated probes for both familiar and trained words produced a large negative deflection in the N400 compared with related probes. In contrast, the unrelated probes to untrained rare words produced no N400 effect, because participants did not know the meanings of the untrained words. This gives us further evidence that participants learned something about the meaning of the trained words that allowed a congruence effect to be observed on a following meaning probe.

We also found several interesting differences between skilled and less skilled comprehenders, evidenced in both behavioral and ERP data. First, skilled comprehenders were reliably more accurate than less skilled comprehenders in meaning judgments on recently trained words. This result may reflect slightly better learning by the skilled readers. On this interpretation, skilled comprehenders, who also have larger vocabularies, are better able to use their word knowledge to add new words to their vocabularies or are simply better at learning new associations or retaining specific episodic information. Although one might consider other explanations, for example, some familiarity for the rare words for skilled comprehenders, we point out again that the words used for training were individually selected as words that a given participant did not recognize as a real word. Furthermore, no skill effects were observed for the equally unfamiliar rare words that were not trained. These facts suggest that the differences between skill groups emerged during encounters with the word during learning that are reflected at testing.

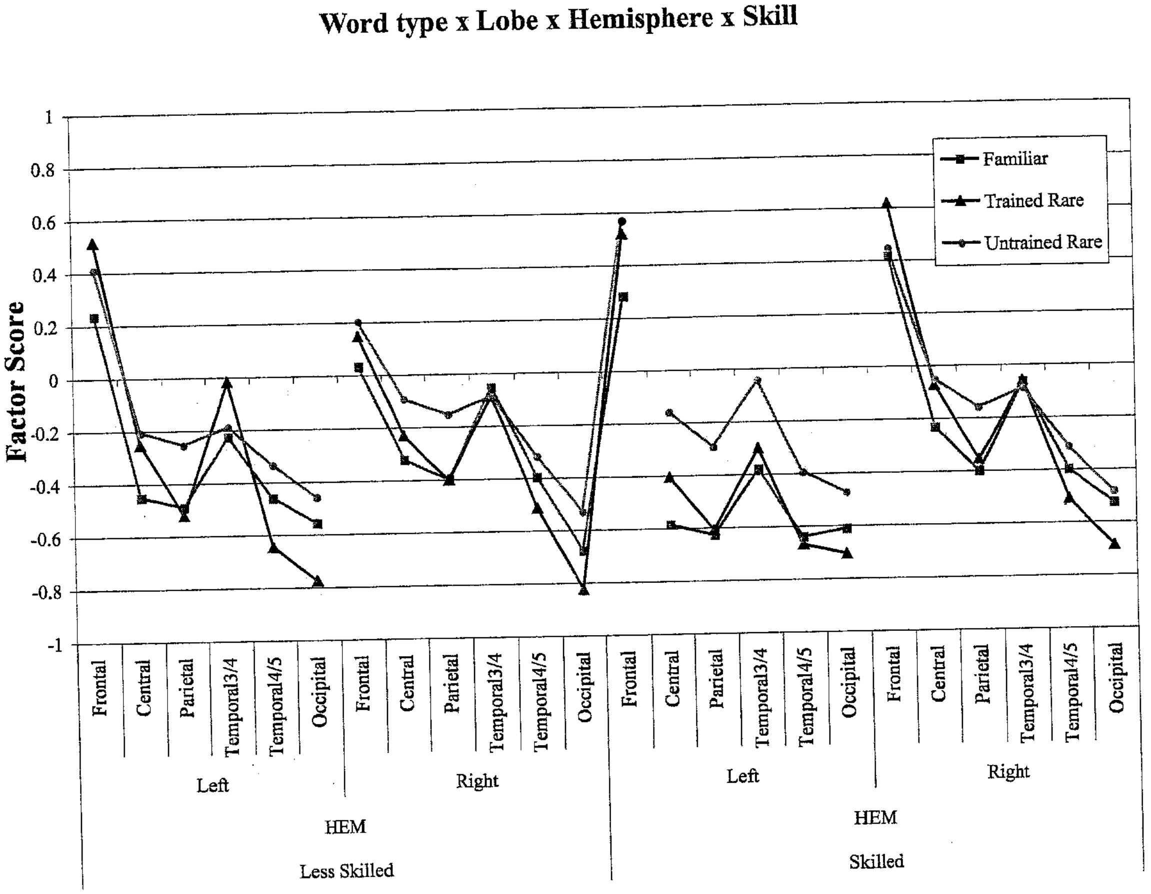


Figure 7. Word Type × Lobe × Hemisphere × Skill interaction for word 2 at 372 ms. HEM = hemisphere.

The 10% accuracy differences between skilled and less skilled comprehenders (and the absence of differences for untrained and familiar words) are mirrored by ERP differences that help explain them. A general result is that ERPs of less skilled readers showed less sensitivity to the differences among the three word types. More specifically, skilled comprehenders showed a larger episodic memory effect, the P600 that distinguished trained rare words from other words. This suggests directly that the training produced a stronger memory trace of the word for skilled comprehenders. A related difference in a P600 effect was reported in a study of recognition memory by Rüsseler, Probst, Johannes, and Münte (2003), who found that a P600 old–new effect was obtained for normal adult readers (more positive for old words in a left parietal electrode) but not adult dyslexics.

It is interesting to note that although the distinction between trained words and other words was visible both very early (140 ms) and later (500 ms), comprehension skill was associated only with the later component. As we suggested above, the earlier component may reflect a general episodic effect that depends not on orthographically based word identification but on some visual attention factor. Comprehension skill may be less relevant to this level of processing, compared with a word-form–based episodic effect that occurs later. It is an open question whether this later episodic trace includes only the orthographic form of the trained word as presented or also the meaning that was associated with it during training. It is possible that both the word form and its associated meaning are part of the episode that is reflected in this later component. Participants knew they would make a meaning judgment. Recalling an associated meaning of the first word would help with that task. However, in research on recognition memory, results suggest that an intention to retrieve information is not necessary for the P600 (Curran, 1999; Paller & Kutas, 1992). Thus, although participants may have been either automatically or intentionally trying to retrieve meaning information associated with the trained word during this 400–600 ms time window, there is no basis to conclude that they were.

It is important to note that our results represent a case in which ERP data help constrain the interpretation of a behavioral result. The skill difference in the late positive shift suggests that skill differences observed in accuracy reflect the strength of the familiarity that resulted from training. However, familiarity is not the end of the story. Skilled comprehenders also showed a larger N400 effect during the presentation of the second word, reflecting a stronger congruence when the second word matched the meaning of the first word. This suggests that skilled comprehenders achieved a better learning of the meaning of the trained word, allowing a related word to show a congruence response. A few other studies have focused on the N400 as capable of distinguishing reading skill. For example, Coch and Holcomb (2003) reported that first-grade children of high reading skill but not low reading

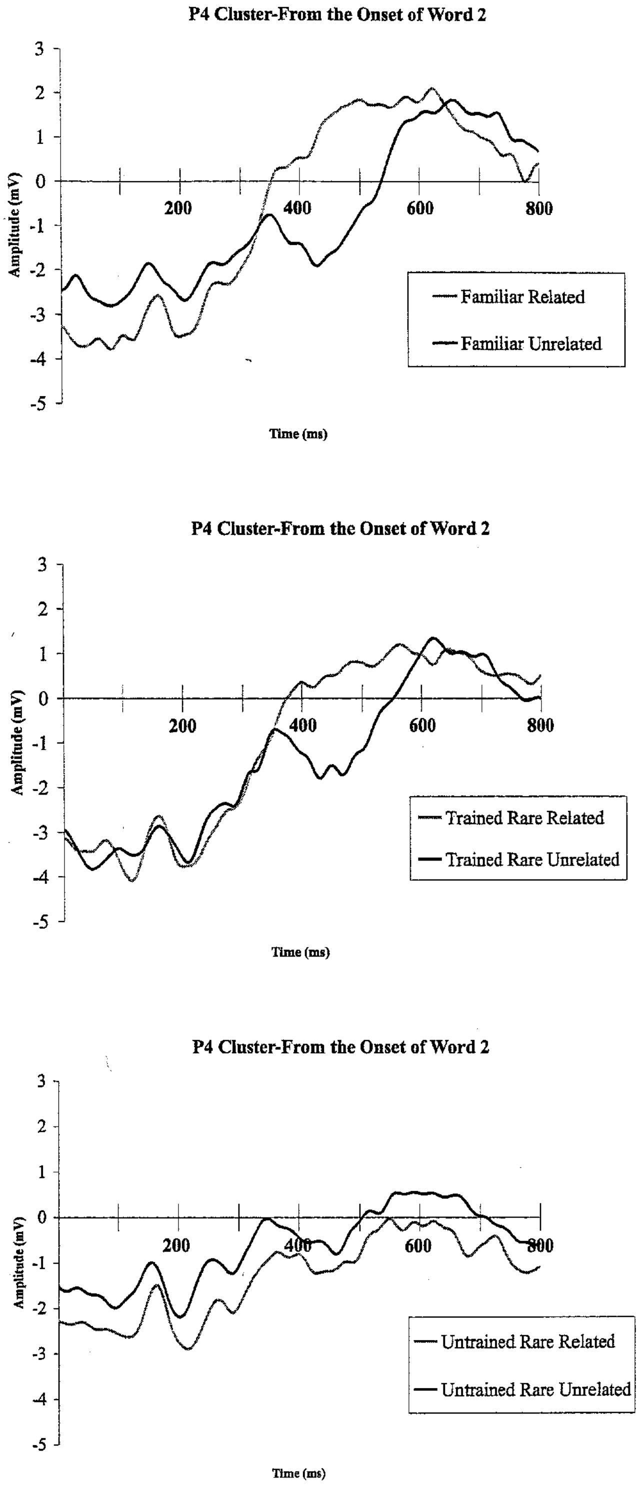


Figure 8. P4 cluster showing word type and relatedness. From 372 ms following the onset of Word 2, there is a separation of related from unrelated words for familiar and trained rare words only.

skill showed N400 responses in passive viewing of words and pseudowords. More related to the present result, Rubin and Johnson (2002) reported that learning-disabled adults showed a longer N400 latency to words in context than did non-disabled readers. However, the specific point for our N400 results is the role of the N400 as an indicator of learning success and the conclusion that skilled comprehenders are more successful at learning new word meanings.

We add some methodological observations to our discussion. First, we note the value of PCA in a study of ERPs, even in the absence of source analysis. Because we had specific hypotheses about a late-latency positive component during the first word (P600) and a midlatency negative component, we could have tested for these components in *t* tests on specific electrodes. However, the PCA allowed a more data-driven approach that could expose other task-related shifts, and indeed it appears to have done so. The PCA revealed a late positive shift that, beyond any noncognitive components it might have contained, differentiated among word types, suggesting a meaning verification process for words whose meanings were either previously known or recently learned. PCA also exposed a component that overlaps the end of the first word's presentation and the start of the second word's presentation. Notably, the effects that are traditionally reported as components (N400, P600) on the basis of the waveforms of selected electrodes were visible in the PCA as well, allowing follow-up tests of waveforms.

Second, we note the potential value of treating two successive stimuli as a single recording epoch. When the interest is in processes that are distributed over two words, the single epoch can expose both word-specific components and shared components, while accurately reflecting the temporal dynamics that go with asking people about relations between successive words.

Finally, we return to the general question of learning word meanings and the role of ERPs in studying this question. We do not assume that native language word learning usually involves the kind of associative training used in our study. (However, there is a similarity to classroom procedures for second-language learning.) We conclude that there are individual differences in the ability to learn the meanings of new words, as evidenced by the behavioral results. Adults who were higher in comprehension skill showed better learning of 60 words from 45 min of training than did those lower in comprehension skill, as evidenced in their performance on a single-word meaning probe following training. The ERP evidence adds to these behavioral results by showing that skilled comprehenders were more sensitive to whether a word had been in the training set. Because words to be trained were chosen so as to be unknown to individual participants, it is likely that this difference in sensitivity reflected a difference in learning that allowed skilled comprehenders to establish stronger episodic traces for trained words. The source of this stronger episodic trace has at least two possibilities. One is that skilled comprehenders were better at learning the meanings they were taught. Thus, presented with *gloaming,* they retrieved the episodic trace that established the association "gloaming means twilight." A related possibility is that skilled comprehenders were better able to encode (and thus recognize) the orthographic word form, *gloaming*. Retrieving whatever was learned about the meaning of a word depends on recognizing the form of the word. Skilled comprehenders generally know more about word forms (orthography and pronunciations) than do less skilled comprehenders (Bell & Perfetti, 1994; Perfetti & Hart, 2001).

Because our study was about word meanings, one might suppose that the observed skill differences reflected the ability to learn or remember word meanings. Certainly, knowledge about word meanings, as distinct from word forms, could be an independent contributor to reading skill (Nation & Snowling, 1999). However, it is important to keep in mind that the learning of meanings is seldom just about meanings but also about the connections between forms and meanings. Word form knowledge and its connection to meaning is the core of the lexical processing, and weakness in this knowledge will negatively affect word-level comprehension. Our results add the finding that among adults, more highly skilled comprehenders are better at learning new meanings and more sensitive to the episodic status of a word.

This result carries an interesting implication: An episodic word process—a memory for a word experience—can produce differences in semantic knowledge. Such a conclusion is compatible with an episodic theory of word identification of the sort proposed by Reichle and Perfetti (2003), in which the development of a mental lexicon is the result of functional encodings of word episodes that lay down form and meaning relations. The key idea is that effective experiences with words—the multiple encounters with a word that lead to an abstracted representation of form and meaning—is what creates reading skill. The present results can be taken to suggest the plausibility of this proposal in accounting for differences in reading comprehension.

References

Barnea, A., & Breznitz, Z. (1998). Phonological and orthographic processing of Hebrew words: Electrophysiological aspects. *Journal of Genetic Psychology, 159*(4), 492–504.

Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology, 86,* 244–255.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11,* 235–260.

Chapman, R. M., & McCrary, J. W. (1995). EP component identification and measurement by principal components analysis. *Brain and Cognition, 27,* 288–310.

Chapman, R. M., McCrary, J. W., & Chapman, J. A. (1978). Short-term memory: The storage component of human brain responses predicts recall. *Science, 202*(4373), 1211–1214.

Coch, D., & Holcomb, P. J. (2003). The N400 in beginning readers. *Developmental Psychology, 43,* 146–166.

Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old/new effects in lexical decision and recognition memory. *Neuropsychologica, 37,* 771–785.

Dien, J. (1998). Addressing misallocation of variance in principal components analysis of evoked potentials. *Brain Topography, 11*(1), 43–55.

Dien, J., & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 189–207). Cambridge, MA: MIT Press.

Dien, J., & Santuzzi, A. M. (2005). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In T. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 57–82). Cambridge, MA: MIT Press.

Hackley, S. A., Woldorff, M., & Hillyard, S. A. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem and cerebral evoked potentials. *Psychophysiology, 27,* 195–208.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99–116.

Liu, Y., & Perfetti, C. A. (2003). The time course of brain activity in reading English and Chinese: An ERP study of Chinese bilinguals. *Journal of Human Brain Mapping, 18*(3), 167–175.

Liu, Y., Perfetti, C. A., & Hart, L. (2003). ERP evidence for the time course of graphic, phonological, and semantic information in Chinese meaning and pronunciation decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1231–1247.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31,* 291–308.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19–36). Hillsdale NJ: Erlbaum.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70,* B1–B13.

Nobre, A. C., & McCarthy, G. (1994). Language-related ERPs: Scalp distributions and modulation by word-type and semantic priming. *Journal of Cognitive Neuroscience, 6,* 233–255.

Paller, K. A., & Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *Journal of Cognitive Neuroscience, 4,* 375–391.

Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.

Reichle, E. D., & Perfetti, C. A. (2003). Morphology in word identification: A word-experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading, 7*(1), 219–238.

Rubin, S. S., & Johnson, C. M. (2002). Lexical access in college students with learning disabilities: An electrophysiological and performance-based investigation. *Journal of Learning Disabilities, 35*(4), 258–267.

Rugg, M. C. (1995). Event-related potential studies of human memory. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 789–801). Cambridge, MA: MIT Press.

Rugg, M. D., & Doyle, M. C. (1992). Event-related potentials and recognition memory for low-frequency and high-frequency words. *Journal of Cognitive Neuroscience, 4,* 69–79.

Rugg, M. D., & Nagy, M. E. (1989). Event-related potentials and recognition for words. *Electroencephalography and Clinical Neurophysiology, 72,* 395–406.

Rüsseler, J., Probst, S., Johannes, S., & Münte, T. F. (2003). Recognition memory for high- and low-frequency words in adult normal and dyslexic readers: An event-related brain potential study. *Journal of Clinical and Experimental Neuropsychology, 25,* 815–829.

Sereno, S., Rayner, K., & Posner, M. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *NeuroReport, 9,* 2195–2200.

Van Boxtel, G. J. M. (1998). Computational and statistical methods for analyzing event-related potential data. *Behavior Research Methods, Instruments, & Computers, 30*(1), 87–102.

Received October 14, 2004
Revision received March 7, 2005
Accepted May 31, 2005 ■