Semantic processes and individual differences detected through error-related negativities

Article *in* Journal of Neurolinguistics · February 2016

DOI: 10.1016/j.jneuroling.2015.08.002

Erik Benau
University of Kansas

Journal of Neurolinguistics 37 (2016) 82–97

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Semantic processes and individual differences detected through error-related negativities

Michal Balass [a, *], Laura K. Halderman [b], Erik M. Benau [c], Charles A. Perfetti [d, **]

[a] Towson University, United States
[b] Educational Testing Service, United States
[c] University of Kansas, United States
[d] Learning Research and Development Center, University of Pittsburgh, United States

A R T I C L E I N F O

Article history:
Received 6 October 2014
Received in revised form 5 August 2015
Accepted 11 August 2015
Available online xxx

Keywords:
Semantic processing
Semantic categories
ERN
Error detection
Individual differences
Comprehension and lexical skill

A B S T R A C T

We report a study that demonstrates the application of the Error-Related Negativity (ERN), an event-related brain potential that accompanies an overt incorrect response, to the study of semantic processes. Twenty-two adult participants completed a time-pressured semantic categorization task to elicit errors to simple semantic category decisions, e.g., is *Africa* a *country*? Our results indicated an increased negativity between 30 and 80 ms following incorrect category judgments, illustrating that the ERN is associated with errors in semantic processing. Furthermore, the magnitude of the ERN response was modulated by the participants' performance in discriminating between category members and non-members, and by off-line measures of word knowledge and vocabulary. These individual differences, both in task performance and in on-line assessments, emerged especially following correct trials. We suggest this is because feedback from knowledge states affects the ERN on correct trials, thus disproportionately affecting the ERNs of better performers and people with more word knowledge.

Semantic processing is an important part of skilled reading. Although off-line assessments of word knowledge (e.g., vocabulary tests) can indicate general differences in word knowledge, observations of on-line meaning processing are important for understanding dynamic semantic processes that are part of reading skill. According to the Lexical Quality Hypothesis, accurate and fluent reading is supported by reliable access to word representations that link well-specified word forms with meanings (Perfetti, 2007; Perfetti & Hart, 2001; 2002). Support for this hypothesis has been reported in a variety of ERP studies that tested differences in word identification skills and its effects on comprehension. One ERP component frequently observed in these studies is the N400, a negative waveform that is associated with semantic processing, and is sensitive to relatedness. Specifically, the N400 is reduced (less negative) when the target stimulus is semantically related to the context in which it appears (see, Kutas & Federmeier, 2011 for a review). For example, ERP studies of adults have shown that skilled comprehenders are more effective at integrating a meaning when an inference was required for text comprehension (indicated by an earlier onset of the N400) than less skilled comprehenders (Yang, Perfetti, & Schmalhofer, 2005; 2007). In an ERP study that controlled for lexical (decoding) knowledge, Landi and Perfetti (2007) found greater N400 reductions in meaning judgments for both associatively and categorically related prime-target word

* Corresponding author. Towson University, 8000 York Road, Towson, MD 21252, USA.
** Corresponding author. LRDC, University Pittsburgh, 3939 O'Hara St., Pittsburgh, 15260, PA, USA.
E-mail addresses: mbalass@towson.edu (M. Balass), perfetti@pitt.edu (C.A. Perfetti).

http://dx.doi.org/10.1016/j.jneuroling.2015.08.002
0911-6044/

pairs (e.g., snow-ICE) when compared to unrelated pairs (e.g., dust-GLOW), and these reductions were larger for skilled than less-skilled comprehenders. These studies suggest that even amongst college students, subtle differences in semantic processing can manifest in ERP measures.

These ERP studies measure effects that are initiated by viewing a word, a procedure that yields multiple voltage shifts associated with word processing. The latencies of these stimulus-locked shifts vary with individual differences in reaction to the stimulus, including word processing speed. By contrast, the event-related negativity (ERN) is time-locked to the response, and thus is not dependent on the speed of the participant's processing of the stimulus. The ERN is an electrophysiological measure that is associated with error processing (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). Specifically, it is a negative-going ERP component with a fronto-central scalp distribution; it is elicited 0–160 ms following an overt erroneous response, with a typical peak at approximately 80 ms. Such a measure can be exploited to examine error monitoring for a variety of cognitive processes, and in the case of word identification, the ERN may expose individual differences in error monitoring during semantic processing and the relation of these differences to measures of reading comprehension. The ERN is not directly associated with meaning processing, unlike the N400 (e.g., Kutas & Federmeier, 2011). Its use in this study is to expose semantic processes indirectly through their effect on the ERN as a signal of error monitoring. Such a connection may be implied from behavioral studies using semantic tasks. For example, errors in meaning selection during word processing have been related to inadequate suppression of irrelevant semantic information (Gernsbacher, 1993). ERNs, if they are sensitive to semantic errors, may expose semantic processes required by meaning tasks and, perhaps, individual semantic knowledge states that drive these processes.

More recently, the ERN has been reported for linguistic processing tasks, including verbal self-monitoring (Ganushchak & Schiller, 2006; 2008a; 2008b), lexical processing (Hauk, Patterson, Woollams, Watling, Pulvermuller, et al., 2006; Ito & Kitagawa, 2006), word access (Horowitz-Kraus & Breznitz, 2008), semantic processing (Ganushchak & Schiller, 2008a; 2008b), and decisions about spelling (Harris, Perfetti, & Rickles, 2014). Much like the present study, these experiments aimed at using the ERN to expose potential differences in linguistic performance. Our study aims to determine whether the ERN, can reflect error monitoring using a semantic judgment task and, if so, whether, it is linked to other indicators of semantic knowledge and performance.

In addition to the ERN, other electrophysiological measures are used to examine response monitoring, such as the correct response negativity (CRN), which is a smaller, but similar negative-going component following a correct response (Vidal, Hasbrouq, Grapperson, & Bonnet, 2000). For this study, our interest is specifically the ERN, and what it reveals about error detection during semantic processing. The ERN was first observed in simple perceptual tasks that would have been error-free without a response-time restriction imposed by the experiment (Falkenstein et al., 1991; Gehring, Coles, Meyer, & Dochin, 1993). For example, the commonly used flanker paradigms (e.g., Gehring, et al., 1993; Pailing & Segalowitz, 2004; Scheffers & Coles, 2000; Yeung, Botvinick, & Cohen, 2004) would be virtually error-free if participants had ample time to examine the visual display. Time pressure conditions are not necessary to elicit the ERN; several studies have reported ERNs for conditions where no speeded response was required (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Ganushchak & Schiller, 2006; Gehring et al., 1993). However, ERNs that are elicited under time pressure restrictions are relevant to our study for two reasons. First, a limited response time is more likely to increase errors. This has been observed in behavioral studies of semantic processing by Quinn and Kinoshita (2008). Their research found that participants committed 29% more errors in a simple semantic categorization task (e.g., is the moon a planet?) when the task had a short response-time restriction than when that restriction was removed. This result suggests that semantic categorization, like simple perceptual tasks, will produce a greater percentage of errors detectable by ERNs with imposed response time restrictions. Second, the relationship between error monitoring and individual differences in semantic knowledge is important to our research question. Fine-grain differences in knowledge that would normally be difficult to detect under typical time constraints may be more easily observed under speeded conditions. For example, Gernsbacher (1993) demonstrated that skilled readers are faster at selecting the appropriate meaning of a word in context than less-skilled readers.

We also assume that semantic category decisions (i.e., is a *canary* a type of bird? versus is a *feather* a type of bird?) facilitate the conditions for a cognitive conflict when foils (feather) share semantic features with category members (canary). As demonstrated by Meyers and Schvaneveldt (1971), when the meaning of a target word is activated, other associated meanings related to that word are activated as well. The additional activation of associated meanings may create a conflict at the time of the response. This conflict in semantic processing should be reflected in the ERN, according to the predictions made by the conflict-monitoring hypothesis of the ERN. The main assumption of this hypothesis is that the ERN reflects the presence of a response conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung et al., 2004) that may result when multiple responses compete for selection. Alternatively, the mismatch hypothesis argues that the ERN signals a mismatch between a given response and the internal representation of the intended response (See, Falkenstein et al., 1991; Holroyd & Coles, 2002, for a more thorough review). Whether the ERN arises from error monitoring due to conflicting information or a mismatch process is beyond the scope of our study. However, some recent evidence provides compelling support for the conflict-monitoring hypothesis. For example, Ganushchak and Schiller (2008a) compared the size of the ERN in response to semantically related and unrelated auditory distractors in a verbal monitoring task. They found that the ERN was largest following errors that occurred for semantically related distractors. Their findings support that the ERN resulted from the activation of multiple lexical entries related to the target. Priming with a particular semantic context may potentially activate multiple competing responses, thus creating conflict at the time of the response, and increasing the likelihood of error. This interpretation of the ERN reflects the presence of transient mental states that are determined by moment-by-moment performance factors, and it is this fact that is important for our aim of using ERNs to expose differences in semantic processing.

There are two broad classes of mental states that are relevant for the ERN. One is a complex of transient performance states that include participants' knowledge of their own performance and their intentions to perform accurately. The amplitude of the ERN has been correlated with participants' offline reports of perceived inaccuracy in a flanker task (Scheffers & Coles, 2000) and with the participant's level of certainty in his or her choice in letter and tone discrimination tasks (Pailing & Segalowitz, 2004). In both studies, larger ERN negativity was observed for trial responses in which the participants perceived themselves to be incorrect, regardless of their actual performance accuracy. The ERN also depends on instructions that lead the participant to try to perform accurately (Falkenstein et al., 1991; Ganushchak and Schiller, 2008).

A second class of mental states is the more permanent knowledge in long-term memory. Such knowledge, e.g., knowledge of word forms or meanings, could be the source of the transient mental states (conflict or error) that produce the ERN. These two forms of knowledge—performance knowledge and long-term memory knowledge—are not independent, of course. Long-term memory knowledge is needed both for performance and for knowledge of performance. If both performance knowledge and long-term semantic knowledge can be detected in ERNs, we expect that, in semantic judgments, the ERN will reflect both participants' decision processes (reflecting transient states) and their knowledge of semantic category relations (permanent knowledge states) that guide their decisions. A recent study of spelling decisions found evidence for both kinds of relations (Harris et al., 2014). ERN magnitudes of spelling decision errors were correlated with both performance knowledge (accuracy of spelling decisions) and spelling knowledge measured off-line.

The present study exploits these properties of the ERN to observe on-line semantic processes in relation to semantic knowledge and comprehension skill, which are linked in the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2001; 2002). According to this hypothesis, access to high-quality semantic knowledge allows efficient and accurate comprehension. Studies with both children and adults show that high-skilled comprehenders have superior semantic processing skills (Cain, Oakhill, Oakhill, & Bryant, 2004; Frishkoff, Perfetti, & Westbury, 2009; Nation, Snowling, & Clarke, 2007; Ozuru, Dempsey, & McNamara, 2009). We hypothesize that under the conditions that require speeded responses (i.e., paradigms designed to increase errors), gaps in semantic knowledge or unstable semantic knowledge will also reveal differences in the magnitude of the ERN.

Semantic categorization is a useful task to test this hypothesis, because it represents a simple, highly general, and important level of knowledge that is widely shared. Knowledge of category relations is basic, arising through experience with objects and the classifications language gives to them. For English-speaking college students, one expects universal knowledge that a *canary* is a kind of bird, but that a *feather* is not a kind of bird, although birds have feathers. We would expect, given ample time, that no participant would mistakenly say that a canary is not a bird or that a feather is a bird. However, under the pressure of time, and given that the word *feather* activates the word *bird* in memory, an error becomes possible. The mere existence of foils like *feather* creates conflict for the participant for all bird category trials, thus leading to occasional errors on positive instances (e.g., *canary*) as well. This is the logic of the present experiment.

Participants who represented a range of comprehension and lexical skill made semantic category decisions (e.g., is the following a 'bird'?) while their EEGs were recorded. We tested three hypotheses about the ERN and individual differences: First, we hypothesized that an ERN effect would occur when participants made decision errors, because participants would have the appropriate category knowledge, and that errors would reflect transient performance knowledge. Second, given the relation between ERNs and performance accuracy (e.g., Gehring, et al., 1993), we further hypothesized that the size of the ERN would be larger for participants who performed well (high accuracy in semantic decisions) than for participants who performed less well. Finally, because ERNs may also reflect permanent knowledge states (or the decision certainty that comes with high knowledge), we hypothesized that the ERN would be stronger for participants with higher vocabulary knowledge and higher comprehension skill. To help reduce the impact of non-cognitive factors on accuracy, and therefore the interpretation of the ERN, we provided financial performance incentives to participants. The incentive allowed the assumption that participants would want to be correct and thus would be responsive to their errors.

1. Methods

1.1. Participants

Twenty-eight participants were recruited from the University of Pittsburgh undergraduate Introduction to Psychology participant pool. All participants were compensated $10/hour plus monetary bonuses for accurate performance on the semantic categorization task. All participants were native English speakers, right-handed, and with normal-or-corrected to normal vision. Six participants were eliminated from the final data analyses due to unusable EEG recordings.

2. Materials

All participants completed a semantic categorization task. Semantic categories and their members were selected from an updated norms list of the Battig and Montague (1969) norms (VanOverschelde, Rawson, & Dunlosky, 2004). From these norms, a total of 48 semantic category types and their associated token category members were selected; yielding a total of

353 members across all 48 categories. The average number of category members per semantic category was 7.35 ($SD = 2.39$) with a range of 5–12 members.

The non-member tokens (foils) were generated for each semantic category type by the experimenters. The generated foils were semantically related to the categories, but were not tokens of the category. For example, for the semantic category "BIRDS", category members included, 'eagle', 'robin', 'blue jay', 'hawk', and 'crow', whereas non-category foils included 'nest', 'feather', 'beak', 'wing', and 'claw'. The number of foils constructed for each semantic category matched the total number of members within that category; a total of 353 foils were generated across all 48 categories. Non-member foils and category members were also matched on word frequency and word length. The average word length for category members was 5.79 letters ($SD = 2.05$), with a range of 3–10 letters, and 5.52 letters ($SD = 1.74$), with a range of 3–12 letters for non-category member foils. Word frequency counts were generated using the SUBTLEXUS corpus (http://subtlexus.lexique.org/; Brysbaert & New, 2009), which has been shown to provide more accurate counts of word frequency (per 1 million words) than Kucera and Francis (1967) norms. The average word frequency for category members was $M = 31.79$, $SD = 55.08$ and $M = 21.75$, $SD = 48.19$ for non-category foils. The complete list of semantic categories types, and their corresponding category member and non-member items are listed in Appendix A.

2.1. Procedure

2.1.1. Comprehension and lexical skill assessment

All participants completed a series of off-line tests measuring their comprehension skill, lexical skill (orthographic, phonological, semantic), and general knowledge. Orthographic knowledge was measured using a spelling test for which participants had to select which letter strings from a list of correctly (e.g., naïve) and incorrectly (e.g., essense) spelled words were real English words. Similarly, for the phonological knowledge test, participants made decisions on which letter strings when pronounced sounded like real English words (e.g., teech versus bape). The hit and false alarm rates were used to calculate participants' d-prime scores for each test.

Semantic knowledge was assessed using the Nelson-Denny Vocabulary Test. Participants were presented with a list of target words and were asked to choose one word from five choices that fit the target words' meaning. This test was a 7 ½ minute timed test. Each participant received two scores for their performance; an accuracy score (the number of items answered correctly) and a "speed" score (the number of items of 100 that was attempted.) To reflect both speed and accuracy in a single measure, a composite vocabulary score was constructed. A fifth of a point penalty was assessed for every target item that was answered incorrectly and subtracted from the total number of correct items, which yielded a final vocabulary composite score. Thus, the composite vocabulary score = number of correct items − (0.20 * (number of correct items − number of incorrect items)).

The Nelson-Denny Comprehension Test, Version E (Nelson & Denny, 1973) was used to assess comprehension. The test included eight passages of varying length (approximately 250–700 words) on different topics, and eight sets of multiple-choice questions (a total of 36 questions) that followed after each passage. This particular version of the Nelson-Denny allowed participants 15 min to complete the comprehension assessment. For this test, accuracy and speed were also combined into a composite comprehension score. The composite comprehension score formula was: comprehension score = number of correct items − (0.20 * (number of correct items − number of incorrect items)).

The Author Recognition Test (ART) was used to assess general literacy exposure and Raven's Advanced Progressive Matrices (Set II) assessed participants' nonverbal ability. For the ART, participants were presented with a list of full names and asked to select the names from the list that they knew to be authors or writers of novels, short stories, magazine articles, etc. Author names were generated from Acheson, Wells, and MacDonald's (2008) updated Author Recognition Test. D-prime scores were computed from the ratio of the number of items selected correctly ('hits') and the number of items selected incorrectly ('false alarms'), providing an approximate measure of text exposure. For the Raven's test, participants were presented with 12 test items, each a three-by-three array of nine patterns with the ninth pattern omitted. The 12 test items were selected from the Raven's original set of 36 items for their discriminative power among college students (Bors & Stokes, 1998). Participants were asked to correctly identify the missing pattern required to complete the array from six possible choices. The items on the test progressively increased in difficulty; thus, each item required greater cognitive processing than the previous item. The test was scored according to how many items were answered correctly.

2.2. Experimental task: semantic-categorization

Following the completion of the comprehension and lexical battery, all participants had their EEGs recorded while they made semantic category decisions on single words. For this task, each experimental block consisted of one category type, e.g., birds, tools, or musical instruments. Forty-eight different categories were presented in 48 blocks. Within each block, half of the trials were category members and the remaining half, non-member foils. The average number of trials per block was 14.89 ($SD = 4.79$), with the shortest block containing 10 trials and the longest block containing 24 trials. All participants were presented with a total of 706 trials (353 members and 353 non-members) across all 48 blocks.

At the beginning of each experimental block, participants were given instructions to decide whether each word referred to a member of the category, e.g., "Is each of the following a bird?". Participants pressed the 'space bar' when they were ready to begin the experimental block. Following the display of the category type, a fixation sign appeared for 1000 ms followed by the target (a category member or non-member). The participant had to press the number one key for a 'yes' member response, and the number two key for a 'no' non-member response on a serial response box (Psychological Software Tools, Pittsburgh, PA). The target was displayed for 350 ms followed by a blank screen of 250 ms in which the participant could make a response. Following a response, the next trial was initiated after a blank screen of 500 ms. If the participant failed to make a response, or responded too slowly, a feedback screen displayed the message 'TOO LATE!". After 500 ms, the next trial was initiated.

To encourage participants to respond as quickly as possible, they were offered monetary bonuses throughout the experiment. In addition to the $10/hour earned for participation in the experiment, each participant earned an additional $.10 for each correct trial. They were also penalized for each incorrect trial (-$.10), and for each trial for which they did not respond in the allotted time (-$.20). Participants earned an average of $12.46 in addition to the $10/hour monetary compensation for their participation.

After the last trial of each category block, participants were presented with feedback on their performance. They were given information about their accuracy for the trials within a given category, the overall accuracy for all categories completed thus far, and the total amount of monetary bonuses they had earned. To reduce carry-over effects between blocks, participants were asked to determine the validity of ten simple arithmetic problems (e.g., $5 + 1 = 7$) by responding appropriately with 'true' or 'false'.

2.3. EEG/ERP recording and preprocessing

Before beginning the semantic categorization task, participants were fitted with an electrode cap. The scalp potentials were recorded from 128 sites using a Geodesic Sensor Net with Ag/AgCl electrodes (Electrical Geodesics, Eugene, OR). The EEG signals were recorded with a sampling rate of 1000 Hz and a hardware bandpass filter was set between 0.1 and 200 Hz. All impedances were kept under 40 KΩ. A vertex reference was used online during the recording; offline, the data were referenced against an average reference. Six eye channels allowed rejection of trials with eye movements and eye blink artifacts.

Data preprocessing was completed using Netstation software (Electrical Geodesics, Eugene, OR). ERPs were response-locked to the initiation of each response to a trial (member or non-member). For each response-locked event, ERPs were averaged over a 600 ms time segment; including 100 ms before the participants' responses and a 500 ms epoch after their responses. Segmented data were digitally filtered with a 30 Hz lowpass filter. For each segment, voltages on two separate pairs of electrodes that were greater than ± 75 mv were considered eye-movements, and voltages above ± 140 mv were considered eye blinks. Any electrode displaying more than ± 200 mv across the entire segment was considered a bad channel. All trial segments containing eye-movements, eye blinks, or more than 20% channel artifacts were rejected and not used in the final analyses. Bad channels were removed from the recordings and replaced by spherical spline interpolation using data from the remaining channels. Following trial rejection, segments were transformed using average reference. Finally, the ERP segments were corrected relative to a 100 ms baseline ending 50 ms before the response. Six participants were removed from the final analyses due to excessive artifacts or bad channels, yielding an $N = 22$ for final data analyses. After eliminating bad trials, the mean number of trials retained per trial type was 343.93 for members and 332.07 for non-member foils.

2.4. ERP data and statistical analysis

Electrodes used for analyses corresponded to the international 10–20 system electrode FCz (electrode 6) and a cluster of six electrodes that surround it, see Fig. 1. These fronto-central electrodes were selected a *priori* based on previous reports indicating maximal error-related brain activity at these sites (Ganushchak & Schiller, 2008a; 2008b; Masaki, Tanaka, Takasawa, & Yamazaki, 2001; Riesel, Weinberg, Endrass, Meyer, & Hajcak, 2013), including several experiments with linguistic tasks (Harris et al., 2014; Hauk et al., 2006). Data from this cluster were averaged for analysis. The ERN's peak is usually elicited at approximately 80 ms post-response (Falkenstein et al., 1991; Gehring et al. 1993). Visual inspection of our data indicated similar findings; the ERN can be observed from approximately 30 to 80 ms post-response, see Fig. 2. Thus, we tested the effects of our experimental conditions on the ERN mean amplitudes using a 0–100 ms time window.

For the statistical analyses of the behavioral (reaction time and accuracy) and electrophysiological data, four different response types were of interest:

1. Correct responses to category members, or 'HITS'. These are trials on which participants responded 'YES' and were correct, e.g., to *canary* as a member of the category of **birds.**
2. Incorrect responses to category members, or 'MISSES'. These are trials on which participants responded 'NO', and were incorrect, e.g., to *canary* as a member of the category of **birds.**
3. Correct responses to non-category members, or 'CORRECT REJECTIONS'. These are trials on which participants responded 'NO' and were correct, e.g., to *beak* as a member of the category of **birds.**
4. Incorrect responses to non-category members, or 'FALSE ALARMS'. These are trials on which participants responded 'YES' and were incorrect, e.g., to *beak* as a member of the category of **birds.**

Using the data corresponding to each of these responses, three analyses were completed. First, the mean amplitudes of electrophysiological data from 0 to 100 ms for response type were analyzed using a 2 (category membership: member, non-

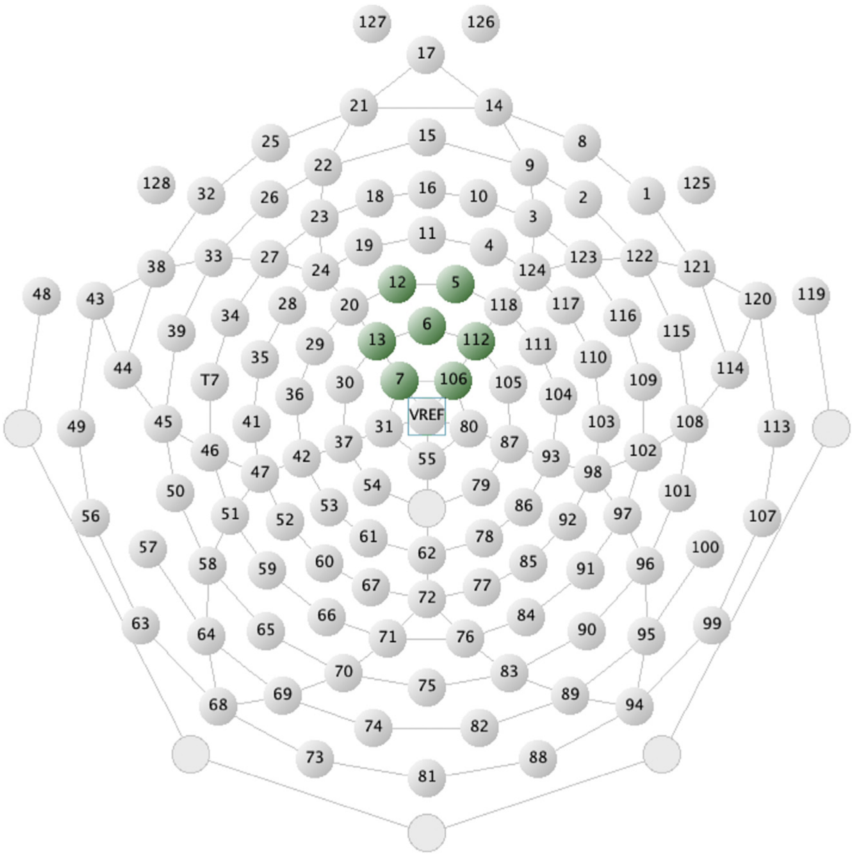

Fig. 1. An electrode map of the 128 electrodes of the Geodesic Sensor Net, indicating the seven front-central electrodes of interest.

member) × 2 (correctness: correct, incorrect) repeated measures analysis of variance (ANOVA). When appropriate, the Greenhouse-Geisser (Greenhouse-Geisser, 1959) correction for sphericity was applied, and ε is reported. Second, d-prime scores, computed by dividing the z-score of the 'hit' rate by the z-score of the 'false alarm rate', provided an online performance measure of participants' discriminability between category members and non-member foils. Third, correlational analyses (Pearson's r) were conducted to examine the relationship between the mean amplitudes for each of the four responses with offline measures of comprehension and lexical skill, and online measure of performance (d-prime).

3. Results

We tested three hypotheses about the ERN and semantic processing: First, that we would observe an ERN when decision errors were made, e.g., responding "yes" to a foil such as *BIRD*: *feather* and responding "no" to a positive instance of the category, e.g., BIRD: *canary*. We tested this hypothesis by examining the error-related brain activity associated with 'misses' and 'false alarms' trials, and compared them to 'hits' and 'correct rejections' trials. Second, we hypothesized that the size of the ERN would be related to individuals' task performance measured by accuracy. To support this prediction, we correlated the size of the ERN with participants' d-prime scores. Finally, we hypothesized that the size of the ERN would be related to offline assessments of participants' semantic knowledge. We tested this by correlating the measures of comprehension and lexical skill with the mean amplitudes of the ERN for all four trial types.

3.1. Behavioral data

3.1.1. Reaction time, error rates, and D-prime

Participants' decision times were analyzed for correct trials only (i.e., hits and correct rejections). Overall, participants were faster to respond "yes" to category members ($M = 341$ ms, $SD = 78.62$) than to respond "no" to non-category foils ($M = 373$ ms, $SD = 105.96$). A paired samples *t-test* confirmed the reliability of this difference, $t\,(21) = -4.24$, $p < .01$. Participants were more accurate in responding to category members (hits), $M = .76$ ($SD = .074$) than to foils (correct rejections), ($M = .50$, $SD = .15$), $t\,(21) = 9.68$, $p < .01$. More errors were made for non-category members (false alarms), $M = .25$ ($SD = .075$) than category members (misses), $M = .12$, ($SD = .038$), $t\,(21) = -9.40$, $p < .01$.

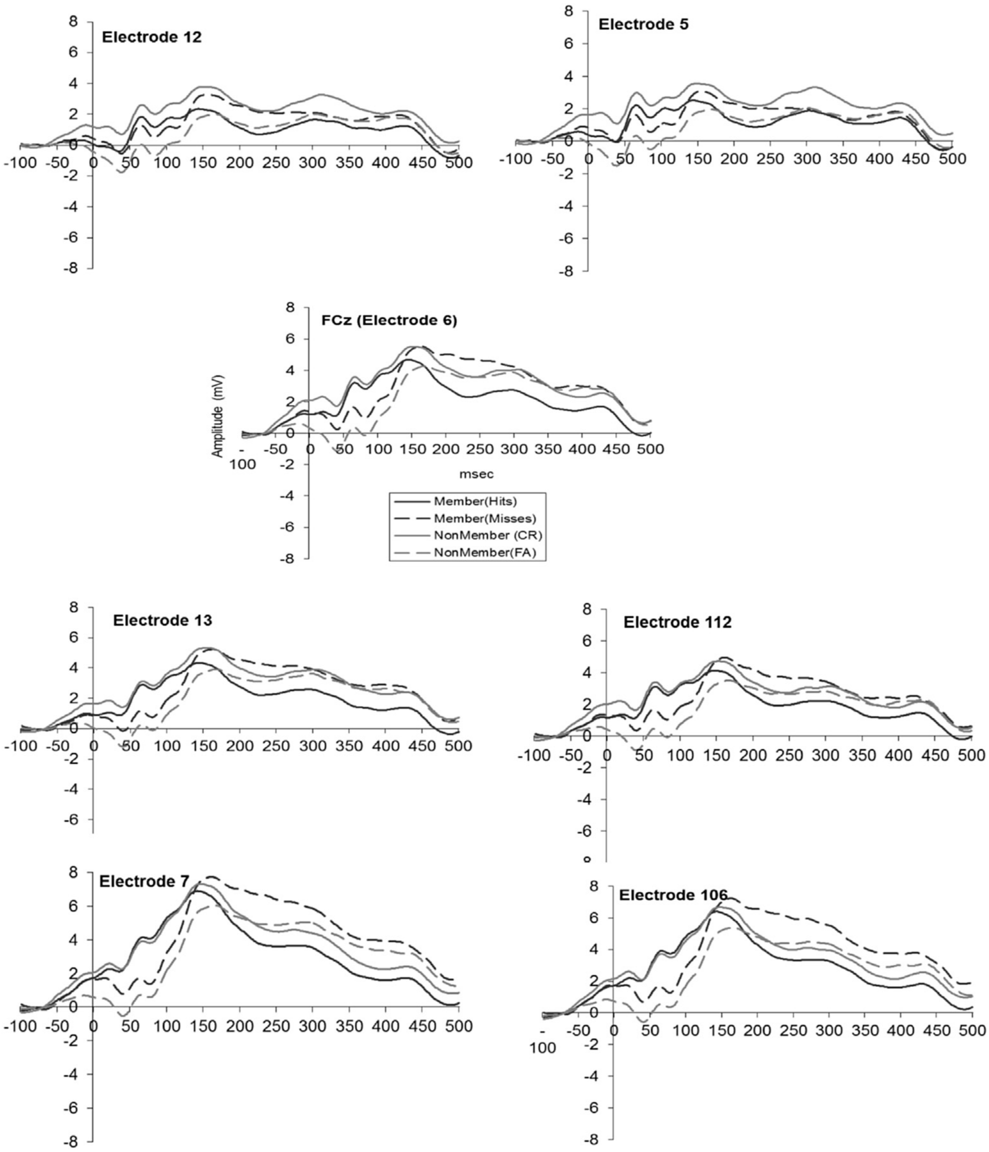

Fig. 2. Response-locked waveforms of each individual electrode of interest. The ERN is observed from approximately 30 ms–80 ms.

Participants' ability to discriminate category members from foils was measured by d-prime, $M = .65$, $SD = .53$, with a range from -0.05 to 1.91. This measure, in combination with the reaction times and error rate results, indicates a relatively low level of discrimination, reflecting the attractiveness of the foils in these speeded decision conditions.

3.2. ERP data

The grand average means for each of the four response types at FCz and the surrounding cluster of 6 electrodes are shown in Fig. 2. The ERN for the four different trial–response combinations can be seen from approximately 30 to 80 ms. The average number of trials retained for analysis to produce the grand average per response type were as follows: 'hits', $M = 260.22$ ($SD = 25.22$), 'misses', $M = 84.36$ ($SD = 26.01$), 'correct rejections', $M = 166.22$ ($SD = 49.14$), and 'false alarms', $M = 168.55$ ($SD = 51.11$). These averages indicated that more than a sufficient number of correct and error trials were retained for analysis, thus reducing the probability of distorted results due to too few error trials compared with correct trials (Luck, 2014).

To test the first hypothesis, an ANOVA 2 (category membership: member, non-member) by 2 (correctness: correct, incorrect) was used to examine the differences in ERN amplitudes for 'hits', 'misses', 'correct rejections', and 'false alarms'. A main effect for category membership, $F\ (1, 21) = 9.03$, $p < .01$, indicated that mean amplitudes for category members ($M = 1.63$, $SD = 1.27$) were more positive than for non-category members foils ($M = 1.29$, $SD = 1.27$). The main effect of correctness was also significant, F (1, 21) = 42.87, $p < .01$, with greater positive amplitudes for correct trials ($M = 2.38$, $SD = 1.60$) than for incorrect trials ($M = 0.54$, $SD = 1.19$), thus confirming a basic ERN effect. As illustrated by Fig. 3, the interaction of category membership and correctness was also significant, F (1, 21) = 15.92, $p < .01$. The interaction reflects a larger correctness effect for non-category members ('correct rejections', $M = 2.65$, $SD = 1.86$ versus 'false alarms', $M = -0.06$, $SD = 1.33$) than for category members ('hits', $M = 2.11$, $SD = 1.55$ versus 'misses' $M = 1.14$, $SD = 1.29$). Greater mean amplitude positivity for 'correct rejections' was unexpected, as we predicted that the greatest positive amplitudes would be observed for 'hits'. We revisit this issue in the next section. Overall, these results confirm our first hypothesis that ERNs would be elicited for trials that produced semantic categorization decision errors.

3.3. Individual differences

Our second and third hypotheses suggest that ERNs can reflect individual differences in two ways: 1) with on-line performance in the semantic judgment task; 2) with off-line assessments of word knowledge or reading ability. For the second hypothesis, the question is whether better performance on the semantic decision task is associated with the magnitude of the ERN. We defined the ERN magnitude as the difference in mean amplitude between 'hits' and 'false alarms'; a measure that is based on "YES" responses in both cases. Considering first the on-line task performance relationship, we found a very strong correlation, $r = .85$, between participants' d-prime scores and the magnitude of their ERN ($p < .001$). Participants who made relatively few errors produced larger ERN magnitudes than did participants who made relatively more errors.

To examine this effect in more detail, we carried out an ANOVA modeled after the main ANOVA, with performance accuracy (d-prime) added as a between-participants factor, defined by a median split. The high accuracy group (N = 11) had a mean d-prime of 1.07 and the low accuracy group (N = 11) had a mean d-prime of .24. The analysis showed a significant interaction of correctness (correct, incorrect) with d-prime (high, low), F (1, 20) = 19.19, $p < .01$. Fig. 4 illustrates these results, showing that the ERN mean amplitude of the high d-prime group was more sensitive to the correctness of response than was the ERN of the low-d-prime group. The high d-prime group showed a larger difference in ERN mean amplitude between 'hits' and 'false alarms' and between 'correct rejections' and 'misses'. The difference between hits and false alarms for this group is visibly greater.

The other aspect of individual differences (our third hypothesis) is whether the magnitude of the ERN is associated with measures of relevant abilities (especially word knowledge) outside the experimental environment. For this we correlated the magnitude of the ERN (i.e., the difference in mean amplitude between 'hits' and 'false alarms') with all individual differences measures (vocabulary, comprehension, spelling, phonological awareness, ART, and Raven's). No significant correlations were found for measures of lexical and comprehension skills and the ERN magnitude, with correlation coefficient values ranging from 0.02 to 0.26 and no p-value less than 0.23. However, we also conducted a post-hoc analysis examining the mean amplitudes for each condition separately with the individual differences measures. We found that the mean amplitude for 'hits' was significantly correlated with vocabulary knowledge ($r = .428$, $p < .05$). The mean amplitude for 'correct rejections' was significantly correlated with phonological knowledge ($r = .439$, $p < .05$). Thus, while the ERN magnitude as a measure of participants' discrimination of correct from incorrect responses was not related to off-line measures of skill, the positivities evoked by correct responses were.

4. Discussion

We tested three hypotheses regarding the ERN: 1) that an ERN would be observed when participants made errors in simple semantic category decisions, demonstrating that ERNs are sensitive to semantic processes that produce errors; 2) that

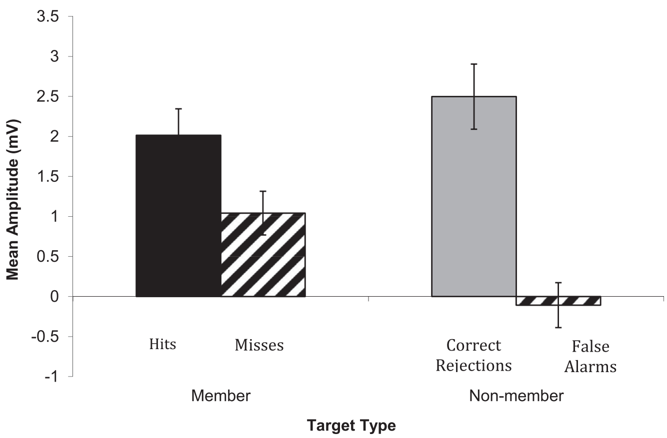

Fig. 3. Interaction effect of member type and correctness. Mean amplitude represents the average ERN amplitude of the seven electrodes from 0 to 100 ms for each target type.

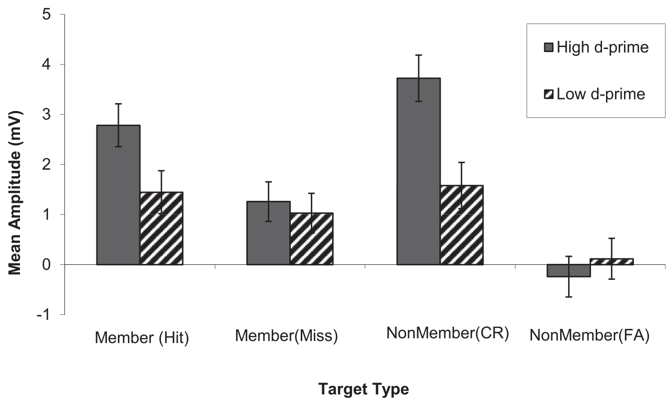

Fig. 4. Interaction of target type and d-prime. Mean amplitude represents the average ERN amplitude of the seven electrodes from 0 to 100 ms by d-prime group and target type.

the magnitude of the ERN would be related to performance accuracy during the semantic decision task, demonstrating the role of transient semantic performance factors in the ERN; 3) that the magnitude of the ERN would be related to general semantic knowledge assessed off-line. The results provide strong evidence for the first two hypotheses but only partial evidence for the third.

On the sensitivity of the ERN to semantic category decision, we found greater negativities at approximately 50 ms post-response for error trials relative to correct trials. Participants' false alarms (e.g., “yes” to ‘*beak*’ for kind of **bird**) and misses (e.g., “no” to *canary* for kind of **bird**) both showed this pattern. As suggested by the condition comparisons in Fig. 3, the ERN was especially large for a false positive, e.g., when a participant responded incorrectly to ‘*beak*’ as a member of the bird category. Errors that were misses, e.g., failing to categorize ‘*canary*’ as a bird, produced a somewhat less pronounced ERN. This suggests that with ‘*beak*’, the participant experiences more conflicting information—there is semantic overlap with the category—and this shows up in a strong ERN. A failure on ‘*canary*’ may sometimes reflect a conflict; this is because of the large feature overlap of the foils with category members (a beak is something birds have). There is task-induced conflict even for well-known category members (such as ‘*canary*’) when the participant strives for accuracy under speeded conditions. However, on some portion of positive trials, the errors may represent something other than conflict, e.g., an attention lapse, or an incorrect button press. Overall, we have clear evidence that semantic decision tasks produce ERN effects, and that such effect are especially visible when errors are made to category foils that share semantic features with the category members. Although, to further understand the link between semantic processing and the ERN, a follow-up experiment that compares errors for related and unrelated foils is necessary.

To examine task performance in relation to the ERN (the second hypothesis), we correlated d-primes, a measure of the discrimination of category members from category foils, with ERN magnitude. Overall, participants' d-prime scores were low, reflecting that the decision task was made difficult by the combination of time pressure (only 350 ms for a decision) and the attractiveness of the category foils. In particular, the time pressure may interrupt a verification stage that checks the defining features of the category and verifies that a category criterion is met by the word, beyond its high level of semantic overlap (Smith, Shoben, & Rips, 1974). Thus, false alarms are rather probable without enough time to execute a verification stage, and the overall discrimination performance is rather low. This may be exactly the conditions under which the ERN emerges. The verification stage is not always executed in time to control the behavioral response, but it is in time to send an error message, producing an ERN.

The individual differences results showed that better performance on the task, as measured by d-prime, was associated with larger ERN magnitudes. The largest difference in performance was indicated for non-category members (i.e., ‘correct rejections’); high performers on the task showed significantly larger amplitude positivities than low performers. High performers also showed greater positivities for category members (i.e., ‘hits’), whereas positive amplitudes for category and non-category members were not significantly different in the low performers group. These findings are consistent with the results of Pailing and Segalowitz (2004), who found that ERNs for correct trials were correlated with the participant's level of decision certainty in letter and tone discriminations tasks. The stronger positivity observed during a correct semantic decision may reflect a high certainty about the relevant knowledge state (a ‘canary’ IS a **bird**, a ‘beak’ is NOT a **bird**), which is related to the participant's overall accuracy.

Especially noticeable (in Fig. 4) is that the low d-prime participants showed little difference between ERNs on trials when they correctly responded “YES” to a category member (‘hits’) and when they incorrectly responded “NO” to a category member (‘misses’). However, the source of this indifference, relative to the difference shown by high-performers, is in the ‘hits’, not the ‘false alarms’. Low-performing participants did not show the degree of positivity for category members that high performers did. This may suggest that low performing subjects have less confidence in their category knowledge in this task; either the defining attributes of the category or whether given exemplars have these attributes.

For example, for the category **fish**, participants' overall accuracy for the category member 'salmon' was M = 0.93 and for 'tuna', M = .83, compared with a lower accuracy for 'cod', and 'trout', M = .59, and M = .62, respectively. More generally, it may be that when performance depends on semantic memory (or other "permanent" knowledge), the ERN includes a component of knowledge feedback. When that knowledge feedback is strong, there is a correspondingly strong correctness response in the ERN time window. Such a response may not occur in perceptual tasks that minimize the role of semantic memory.

Finally, for the third hypothesis, that ERN effect would reflect relevant semantic knowledge assessed off-line, the results report mixed support. Individual ERN measures for 'hits' (separated from 'false alarms') were correlated with vocabulary knowledge; greater positive ERN amplitudes for correct category members ('hits') were observed for participants with higher vocabulary knowledge scores. This association with correct rather than incorrect responses is the same pattern we observed for the association of performance measures with the ERN. This association with greater positive amplitudes for correct responses ('hits' and 'correct rejections') is well established, although with alternative explanations (e.g., Scheffers & Coles, 2000), and associated with participants' confidence (Pailing & Segalowitz, 2004). In light of these preliminary results, the question that needs to be addressed with further studies is whether the increased positivities in the ERN to correct responses might be more diagnostic of knowledge states than are the incorrect responses.

The larger picture of how the ERN might reflect a person's stable knowledge states remains to be determined. However, there is evidence that the ERN does correlate with skill when very low-skill is compared with high-skill (Horowitz-Kraus, 2011; Horowitz-Kraus & Breznitz, 2008). In the current study with college adults without reading disability, the range of skill differences was smaller. Nevertheless, evidence that skill differences in this population are detectable in ERNs comes from experiments by Harris et al., (2014) who found correlations of .88 and .66 between the ERN effect in spelling decision and off-line spelling knowledge. Thus, even within a relatively narrow range, the ERN may expose differences in lexical knowledge, including knowledge of both word meaning and word form.

Acknowledgments

This research has been supported by National Institute of Health Grant 115260 award to Charles A. Perfetti. We thank Benjamin Rickles for his assistance with data processing.

Appendix A

Semantic Category	Member	Non-member
A bird	blue jay	beak
	cardinal	chirp
	crow	claw
	eagle	feather
	hawk	gobble
	parrot	nest
	pigeon	quack
	robin	song
	sparrow	wing
A building for religious services	cathedral	alter
	church	choir
	mosque	pew
	synagogue	rosary
	temple	steeple
A carpenter's tool	drill	beaker
	hammer	bulb
	nails	lens
	saw	scalpel
	screwdriver	shears
	wrench	socket
A cooking tool	knife	cabinet
	pan	faucet
	pot	lid
	spatula	napkin
	spoon	shelf
	strainer	sponge
A family member	aunt	accountant

(*continued on next page*)

(*continued*)

Semantic Category	Member	Non-member
	brother	babysitter
	cousin	barber
	dad	boss
	grandma	doctor
	grandpa	friend
	mom	girlfriend
	nephew	landlord
	niece	lawyer
	sister	mailman
	uncle	teacher
A fish	bass	gill
	cod	net
	salmon	school
	trout	spawn
	tuna	stream
A four-footed animal	cat	ant
	cow	crab
	deer	duck
	dog	gorilla
	elephant	kangaroo
	goat	lobster
	horse	ostrich
	lion	scorpion
	mouse	spider
	pig	wasp
	tiger	whale
A fruit	apple	jelly
	banana	juice
	grapes	nectar
	kiwi	peel
	orange	pie
	peach	punch
	pear	rind
	pineapple	seed
	plum	stem
	strawberry	tart
	watermelon	zest
A gardener's tool	glove	dirt
	hoe	grass
	hose	lawn
	rake	terrace
	shovel	yard
A kind of money	cent	account
	dime	balance
	dollar	bank
	euro	charge
	franc	debt
	nickel	deposit
	penny	loan
	peso	purse
	pound	savings
	quarter	teller
	yen	wallet
A member of the clergy	bishop	convent
	cardinal	eucharist
	nun	monastery
	pope	parisher
	priest	vatican
A military title	colonel	causality
	general	draft
	lieutenant	navy
	private	platoon
	sergeant	salute
A musical instrument	cello	band
	clarinet	bow
	drums	key
	flute	mute
	guitar	notes
	piano	orchestra
	saxophone	pick
	trombone	pitch

(*continued*)

Semantic Category	Member	Non-member
	trumpet	slide
	tuba	tune
	violin	valve
A natural earth formation	canyon	avalanche
	hill	canal
	lake	dock
	mountain	fort
	ocean	hut
	river	lodge
	rock	mine
	valley	pier
	volcano	tent
A non-alcoholic beverage	coffee	burp
	coke	chocolate
	juice	honey
	lemonade	mug
	milk	pitcher
	pepsi	pulp
	soda	stein
	sprite	straw
	tea	sugar
	water	syrup
A part of a building	door	airport
	floor	barn
	roof	duplex
	stairs	hospital
	wall	mall
	window	skyscraper
A part of speech	adjective	comma
	adverb	hyphen
	noun	suffix
	pronoun	tense
	verb	vowel
A part of the human body	arm	bandage
	ears	bruise
	eyes	fin
	fingers	hide
	foot	horn
	hands	rash
	head	scale
	leg	shell
	mouth	splinter
	nose	tear
	toes	tusk
A precious stone	diamond	granite
	emerald	limestone
	pearl	marble
	ruby	salt
	sapphire	slate
A science	astronomy	business
	biology	design
	chemistry	english
	physics	history
	psychology	philosophy
A sport	baseball	bleacher
	basketball	catch
	football	dribble
	golf	dugout
	hockey	goal
	lacrosse	helmet
	soccer	jersey
	swimming	puck
	tennis	putt
	track	racquet
	volleyball	roster
A state in the US	Alabama	Albany
	California	Atlanta
	Colorado	Boston
	Delaware	Charleston
	Florida	Cincinnati

(*continued on next page*)

(*continued*)

Semantic Category	Member	Non-member
	Georgia	Dallas
	Indiana	Denver
	Maryland	Miami
	Ohio	Raleigh
	Pennsylvania	Richmond
	Texas	Seattle
	Virginia	Trenton
A thing made of wood	chair	bag
	desk	bottle
	house	fire
	pencil	pen
	table	syringe
	toothpick	tupperware
A thing that flies	airplane	blade
	bird	breeze
	bug	cloud
	fly	sky
	helicopter	tail
	kite	wind
A thing that makes noise	car	antenna
	computer	monitor
	horn	screen
	people	sight
	radio	trash
	television	wheel
A transportation vehicle	airplane	beep
	bike	cockpit
	boat	highway
	bus	ignition
	car	pedal
	motorcycle	throttle
	train	tire
	truck	track
	van	trunk
A type of dance	ballet	club
	jazz	glide
	salsa	palace
	swing	partner
	tango	rhythm
	tap	romance
	waltz	tutu
A type of fabric	cotton	coat
	nylon	scarf
	polyester	stitch
	silk	strand
	wool	thread
A type of flower	carnation	bud
	daffodil	bush
	daisy	garden
	dandelion	grass
	rose	pollen
	sunflower	thorn
	tulip	vine
A type of footwear	boot	cloth
	clog	lint
	loafer	robe
	sandal	scruff
	shoe	sole
	slipper	strap
	sneaker	suede
A type of fuel	coal	heat
	diesel	neon
	gas	silicon
	oil	smog
	unleaded	sulfur
A type of human dwelling	apartment	den
	condo	hall
	house	mill
	mansion	sewer
	shack	yard
A type of jewelry	bracelet	bead

(*continued*)

Semantic Category	Member	Non-member
	locket	beauty
	necklace	clock
	pendant	fastener
	ring	glitter
	tiara	hook
	watch	treasure
A type of metal	aluminum	brick
	copper	cement
	gold	clay
	iron	dirt
	silver	foil
	steel	leather
A type of music	alternative	beat
	classical	chorus
	country	harmony
	jazz	melody
	pop	studio
	rap	tempo
	reggae	voice
A type of reading material	book	author
	journal	editor
	magazine	page
	newspaper	reference
	novel	subscription
A type of snake	anaconda	charmer
	cobra	coil
	python	fang
	rattle	hiss
	viper	slither
A type of tree	aspen	acorn
	maple	cone
	oak	leaf
	pine	needle
	redwood	paper
	willow	trunk
A unit of time	century	alarm
	day	beep
	decade	buzz
	hour	calendar
	millennium	ding
	millisecond	inch
	minute	limit
	month	mile
	second	schedule
	week	snooze
	year	space
A vegetable	broccoli	bulb
	carrot	butter
	celery	cob
	corn	flower
	cucumber	kernel
	lettuce	leaf
	onion	pod
	peas	seed
	potato	soup
	tomato	stalk
A water vessel	canoe	anchor
	rowboat	hammock
	sailboat	hull
	ship	mast
	speedboat	oar
	yacht	propeller
A weather phenomenon	flood	electricity
	hail	mitten
	hurricane	mud
	lightening	parka
	rain	plow
	snow	sleigh
	tornado	umbrella
Alcoholic beverage	beer	chai

(*continued on next page*)

(*continued*)

Semantic Category	Member	Non-member
	gin	espresso
	rum	gatorade
	tequila	latte
	vodka	lipton
	whiskey	mocha
	wine	snapple
An article of clothing	bra	brim
	hat	buckle
	jacket	button
	jeans	cuff
	pants	hem
	shirt	knit
	shoes	lace
	shorts	logo
	skirt	pocket
	socks	sleeve
	sweater	slit
	underwear	zipper
An article of furniture	bed	blanket
	chair	candle
	couch	drawer
	desk	laptop
	dresser	plant
	loveseat	remote
	sofa	stereo
	table	window
An elective office	governor	advocate
	mayor	ambassador
	president	bureaucrat
	representative	diplomat
	secretary	liaison
	senator	lobbyist
	treasurer	
An herb	basil	chicken
	oregano	flake
	parsley	pasta
	rosemary	pizza
	thyme	sauce
An insect	ant	bite
	bee	bumble
	beetle	hive
	butterfly	stinger
	grasshopper	swatter
	mosquito	venom

References

Acheson, D. J., Wells, J. B., & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*(1), 278–289.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: a replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80*, 1–46.

Bors, D. A., & Stokes, T. L. (1998). Raven's advanced progressive matrices: norms for first year university students and the development of a short form. *Education and Psychological Measurement, 58*(3), 383–398.

Botvinick, M., Braver, T., Barch, D., Carter, C., & Cohen, J. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624–652.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Cain, K., Oakhill, J., & Bryant, P. E. (2004). Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*, 31–42.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of cross-modal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurospsychology, 78*, 447–455.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology, 51*(2–3), 87–107.

Frishkoff, G. A., Perfetti, C. A., & Westbury, C. (2009). ERP measures of partial semantic knowledge: left temporal indices of skill differences and lexical quality. *Biological Psychology, 80*(1), 130–147.

Ganushchak, L., & Schiller, N. (2006). Effects of time pressure on verbal self-monitoring: an ERP study. *Brain Research, 1125*, 104–115.

Ganushchak, L., & Schiller, N. O. (2008a). Brain error-monitoring activity is affected by semantic relatedness: an event-related brain potentials study. *Journal of Cognitive Neuroscience, 20*(5), 927–940.

Ganushchak, L. Y., & Schiller, N. O. (2008b). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. *NeuroImage, 39*, 395–405.

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Dochin, E. (1993). The error-related negativity: an event-related brain potential accompanying errors. *Psychophysiology, 27*, S34.

Gernsbacher, M. A. (1993). Less skilled readers have less efficient suppression mechanisms. *Psychological Science, 4*(5), 294–298.
Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95–112.
Harris, L. N., Perfetti, C. A., & Rickles, B. (2014). Error-related negativities during spelling judgments expose orthographic knowledge. *Neuropsychologia, 54*, 112–128.
Hauk, O., Patterson, K., Woollams, A., Watling, L., Pulvermüller, F., & Rogers, T. T. (2006). When would you prefer a SOSSAGE to a SAUSAGE? At about 100 msec. ERP correlates of orthographic typicality and lexicality in written word recognition. *Journal of Cognitive Neuroscience, 18*(5), 818–832.
Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*, 679–709.
Horowitz-Kraus, T. (2011). Does development affect error-related negativity of impaired and skilled readers? An ERP study. *Developmental Neuropsychology, 36*(7), 914–932.
Horowitz-Kraus, T., & Breznitz, Z. (2008). An error-detection mechanism in reading among dyslexic and regular readers – an ERP study. *Clinical Neurophysiology, 119*(10), 2238–2246.
Ito, J., & Kitagawa, J. (2006). Performance monitoring and error processing during lexical decision task in patients with parkinson's disease. *Journal of Geriatric Psychiatry and Neurology, 19*(1), 46–54.
Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.
Landi, N., & Perfetti, C. A. (2007). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain and Language, 102*, 30–45.
Luck, S. J. (2014). *An introduction to the event-related potential technique* (second ed.). Cambridge, MA: MIT Press.
Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *NeuroReport, 12*, 1851–1855.
Meyers, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.
Nation, K., Snowling, M. J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech-Language Pathology, 9*, 131–139.
Nelson, M. J., & Denny, E. C. (1973). *The Nelson-Denny Reading test*. Houghton Mifflin Company.
Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*(3), 228–242.
Pailing, P. E., & Segalowitz, S. J. (2004). The effects of uncertainty in error monitoring on associated ERPs. *Brain and Cognition, 56*, 215–233.
Perfetti, C. A. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383.
Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.
Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam: John Benjamins.
Psychological Software Tools, Inc. (2000). *E-Prime [Computer software]*. Pittsburgh, PA: Psychological Software Tools, Inc.
Quinn, W. M., & Kinoshita, S. (2008). Category congruence effect in semantic categorization with masked primes with narrow and broad categories. *Journal of Memory and Language, 58*, 286–306.
Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN ? Convergent validity of error-related brain activity across different tasks. *Biological Psychology, 93*, 377–385.
Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 141–151.
Smith, E., Shoben, E., & Rips, L. (1974). Structure and process in semantic memory: featural model for semantic decisions. *Psychological Review, 81*, 214–241.
Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: an updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*, 289–335.
Vidal, F., Hasbrouq, T., Grapperson, J., & Bonnet, M. (2000). Is the 'error' negativity specific to errors? *Biological Psychology, 51*(2–3), 109–128.
Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2005). Less skilled comprehenders' ERPs show sluggish word-to-text integration process. *Written Language & Literacy, 8*(2), 233–257.
Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 55–89.
Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931–959.