Psychophysiology, 39 (2002), 133–146. Cambridge University Press. Printed in the USA.

DOI: 10.1017.S004857720139203X

The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study

KARA D. FEDERMEIER,[a] DEVON B. MCLENNAN,[a] ESMERALDA DE OCHOA,[a]
AND MARTA KUTAS[a,b]

[a]Department of Cognitive Science, University of California, San Diego, California, USA
[b]Department of Neurosciences, University of California, San Diego, California, USA

Abstract

To examine changes in semantic memory organization and use during aging, we recorded event-related potentials as younger and older adults listened to sentences ending with the expected word, an unexpected word from the same semantic category, or an unexpected word from a different category. Half of the contexts were highly constraining. In both groups, expected words elicited less negativity 300–500 ms (N400) than unexpected ones, and unexpected words elicited smaller N400s when these were categorically related. Whereas younger adults showed the greatest N400 reduction to unexpected but related words in high constraint contexts, older adults showed the opposite tendency. Thus, unlike younger adults, older adults as a group do not seem to be using context predictively. Older adults with higher verbal fluency and larger vocabularies, however, showed the younger response pattern, suggesting resource availability may offset certain age-related changes.

Descriptors: Aging, Semantic memory, Sentence processing, N400, Verbal fluency

Our folk models send mixed messages about the changes in cognition expected to occur over the course of the adult lifespan. On the one hand, a familiar saying asserts that “you can’t teach an old dog new tricks,” perhaps reflecting that aging is associated with slowed motor processing and declines in the ability to explicitly recall new episodic memories. On the other hand, it is also said that “older is wiser.” Older adults seem to have relatively preserved, if not augmented, stores of world knowledge and largely maintain their abilities to communicate about that knowledge as well. In fact, these observations are generally borne out by research into cognitive aging, and the question of whether—and, if so, why—some cognitive processes remain stable with age has long been of interest.

Although, with age, low frequency words become more difficult to access in the absence of phonological or orthographic cues [e.g., increased tip of the tongue experiences (Bowles & Poon, 1985; Burke, MacKay, Worthley, & Wade, 1991)], information about word meaning actually seems to be well retained or even augmented with age. For example, older adults perform quite well on standard vocabulary measures, in some cases outscoring education-matched younger adults (e.g., review in Salthouse, 1993). Similarly, though their reaction times are slower, older adults are at least as accurate as younger ones at making speeded word/nonword judgments (lexical decision; e.g., Bowles & Poon, 1981, 1985; Howard, 1983) The nature and organization of this word-related information also seems stable across age. For example, when verbal abilities are matched, older adults generate word associations that are qualitatively similar to those of younger adults (Bowles, Williams, & Poon, 1983; Burke & Peters, 1986; Lovelace & Cooley, 1982; Scialfa & Margolis, 1986), with both groups producing primarily paradigmatic responses (words from the same grammatical class that share features in common) at approximately the same level of specificity and with approximately the same degree of variability. Younger and older adults also generate similar exemplars when given taxonomic category labels (Howard, 1980). Because such associations are generally taken to reflect the strength of connections between items in semantic memory, these findings suggest that the organization of semantic information is similar for younger and older adults.

Similar conclusions are also drawn when semantic memory organization is examined with more implicit tasks. As is true for younger adults, older adults’ on-line performance is facilitated in the presence of semantically related word information (semantic priming; e.g., Bowles, 1989; Burke, White, & Diaz, 1987; Laver & Burke, 1993), for both category coordinates and category–property relations (Howard, McAndrews, & Lasaga, 1981). This facilitation is similarly modulated by associative strength in both groups, with a tendency for strength to have greater effects on priming magnitude in the older sample (e.g., Balota & Duchek, 1988). In fact, semantic priming effects seem to be somewhat larger overall in older adults (e.g., Laver & Burke, 1993), perhaps because their

This study was supported by a Howard Hughes Predoctoral Fellowship to K.F. and by grants HD22614 and AG08313 to M. K.

Address reprint requests to: Kara D. Federmeier, Department of Cognitive Science, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA. E-mail: kfederme@cogsci.ucsd.edu.

responses are generally slower (such that there is more time for facilitation to "spread" to related items).

Overall, then, available evidence seems to suggest that the contents and structure of semantic memory do not change much with age. One question, however, is whether there might be age-related differences in how, and how effectively, this information is accessed/used during on-line processing. There have been suggestions, for example, of age-related changes in the time course with which information becomes active in semantic memory. In particular, some studies have reported that older adults fail to show semantic priming at very short stimulus-onset asynchronies where younger adults continue to do so (e.g., Howard, Shaw, & Heisey, 1986; but see also Balota, Black, & Cheney, 1992, and Balota & Duchek, 1988, for cases where no age-related differences were found). Electrophysiologically, older adults also elicit smaller, delayed N400s to nonassociated targets in priming tasks (Gunter, Jackson, & Mulder, 1998). Both results have been taken to suggest that the buildup and/or spread of activation in semantic memory may be slowed with age.

Rapid activation of semantic information is clearly important for normal, on-line language processing. If activation in semantic memory is weaker or delayed in older adults, then their sentence processing would be expected to differ. However, in many respects, older adults' language processing seems similar to younger adults'. For example, at least when working memory resources are not particularly taxed, older adults seem as likely as younger adults to use contextual information to draw inferences: for instance, instruments from action descriptions ["knife" from "The cook cut the meat" (Burke & Yee, 1984)], or exemplars from category terms ["bee" from "The insect in the clover stung the professor" (Light, Valencia-Laver, & Davis, 1991)]. Older adults also use contextual information to activate appropriate—and inhibit inappropriate—aspects of words' meanings [e.g., "organ" as instrument versus "organ" as body part (Balota & Duchek, 1991; see also Burke & Harrold, 1993; Hopkins, Kellas, & Paul, 1995)]. Thus, in many cases, older adults are apparently able to employ their world knowledge effectively for the purposes of sentence processing.

Older adults show facilitated word processing, as indexed behaviorally and electrophysiologically, in the presence of congruent contextual information (e.g., Cohen & Faulkner, 1983; Gunter, Jackson, & Mulder, 1992, 1995; Madden, 1989; Obler, Nicholas, Albert, & Woodward, 1985; Woodward, Ford, & Hammett, 1993). However, there are indications that they rely more heavily than younger adults on contextual cues for this facilitation (e.g., Madden, 1988; Tun & Wingfield, 1993), and, similar to ERP findings in semantic priming tasks, older adults' N400 responses to incongruent words in sentences are smaller and delayed (e.g., Gunter et al., 1992, for word-by-word reading in highly educated older adults, and Woodward et al., 1993, for auditory language comprehension). Hamberger, Friedman, Ritter, and Rosen (1995) also found differences in the pattern of N400 responses in younger versus older participants to different types of sentence-final words. They used sentences ending with the expected completion, an unexpected but semantically related completion (that was or was not congruent with the sentence), or an unexpected and unrelated completion; participants made a sense/nonsense judgment for each sentence. Whereas younger participants' N400 responses were graded by both congruity and semantic relatedness, older participants' N400 responses were facilitated only for the most expected completion. Hamberger et al. suggested that the younger and older adults may have employed different task-related strategies for reading the sentences. Thus, there are indications that, because of slowing or other factors, older adults may use relatively intact semantic information stores somewhat differently than younger adults in the face of on-line processing demands.

One way to try to characterize how semantic memory is used during language processing is to set semantic memory organization and language context information at odds. This is what we have done in a series of recent studies with young adults (Federmeier & Kutas, 1999a, 1999b, 2001). We examined the electrophysiological response to three types of target words in sentence contexts: (a) expected exemplars, the best completion for a particular sentence context, (b) within-category violations, contextually unexpected items from the same semantic category as the expected exemplar, and (c) between-category violations, unexpected items from a different semantic category. The results showed a clear influence of semantic memory organization on on-line language processing. Although expected exemplars were processed most easily (smallest N400 response), within-category violations were easier to process (generated smaller N400 responses) than between-category violations, even though both were implausible in the sentence contexts; this was true for both visually presented words (Federmeier & Kutas, 1999b) and for line drawings of the same concepts (Federmeier & Kutas, 2001). In fact, in the case of visually presented words, within-category violations were actually easiest to process in those contexts in which they were the *most* implausible—namely, highly constraining sentence contexts. These results suggest that language context information is used to preactivate the semantic features of upcoming concepts, such that items sharing those features (e.g., those from the same taxonomic category) also come to be facilitated even if they are otherwise implausible.

Here, we use the same paradigm to examine age-related changes in language processing and the use of semantic memory on-line. We reproduce everyday language processing as closely as possible by using natural, connected speech, allowing us to examine more directly whether older and younger adults' processing is similar under the temporal constraints of normal language comprehension. We compare results in younger adults for auditory language processing with those previous obtained for visual word (Federmeier & Kutas, 1999b) and picture processing (Federmeier & Kutas, 2001). More importantly, we examine the influence and interaction of (a) contextual congruity (b) contextual constraint, and (c) semantic memory organization on older adults' language comprehension. We expect that, like younger adults, older adults will show facilitation for contextually-expected items as compared with unexpected items. If they are using context predictively, older adults—like younger adults—should also show facilitation for within-category violations, especially in highly constraining contexts. On the other hand, if they are not predicting, then we should observe little difference between within- and between-category violations and little effect of contextual constraint [cf. the pattern of results in young adults for right-hemisphere-initiated processing (Federmeier & Kutas, 1999a)]. Finally, we examine individual differences in older adults' language processing as a function of their performance on standard neuropsychological tests to determine which effects are strictly due to aging and which may be modulated by the availability of various types of cognitive resources.

Methods

Materials

Stimulus materials consisted of auditory versions of the sentence pairs used in Federmeier and Kutas (1999b), recorded as natural, connected speech by a female monolingual English speaker (16-bit resolution, digitized at 22,050 Hz). Sentence pairs ended with one of three types of target words: (a) expected exemplars, the highest cloze probability ending for a given sentence pair, (b) within-category violations, unexpected (cloze probability less than 0.05) items from the expected taxonomic category, and (c) between-category violations, unexpected (cloze probability less than 0.05) items from a different (unexpected) taxonomic category. The first sentence of each pair established the expectation for item and category, and these 132 context sentences were recorded in a separate session, after those containing the target words. Pair-final sentences contained no lexical associates of any of the possible endings and, independent of the context sentence, could plausibly be completed by any of the three target types. These target-containing sentences were recorded individually for each type of target, randomized across three recording sessions such that the same sentence context was not repeated in any given session.

Target items were derived from 66 different semantic categories, two items from each. With rare exceptions, these category coordinates were not lexically associated. Categories were chosen to be those at the lowest level of inclusion for which the average undergraduate student could be expected to readily differentiate several exemplars. For approximately half the categories used, this level was basic as determined by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) or by analogy. Other categories were based at the next highest level (a superordinate of the basic level) because it was unclear that the average participant could clearly and consistently differentiate below this level. Between-category targets for each sentence pair were chosen from a related category that shared key features (e.g., animacy, size, general function) with that from which the expected exemplar and within-category violation were derived. An example set of stimuli is given below, showing (in bold) the expected exemplar, within-category violation, and between-category violation completions, respectively:

They wanted to make the hotel look more like a tropical resort. So, along the driveway they planted rows of **palms/pines/tulips**.

The air smelled like a Christmas wreath and the ground was littered with needles.
The land in this part of the country was just covered with **pines/palms/roses**.

The tourist in Holland stared in awe at the rows and rows of color.
She wished she lived in a country where they grew **tulips/roses/pines**.

The gardener really impressed his wife on Valentine's Day.
To surprise her, he had secretly grown some **roses/tulips/palms**.

As can be seen from the example, across the stimulus set, target items appeared once as each kind of ending, so that conditions were perfectly controlled for word frequency, imageability, concreteness, and so forth. Neither target words nor target sentences differed in duration across conditions. Stimuli were organized in three lists, with no context or item repeated. Each list consisted of 44 of each type of target (expected exemplars, within-category violations, between-category violations) plus 44 plausible filler sentence pairs. Stimuli were randomized once within each list and then presented in the same order for each participant. More examples of the stimuli can be found in Appendix B of Federmeier and Kutas (1999b).

Constraint

Cloze probabilities were obtained from college-aged volunteers for the 132 sentence pair contexts (sentence pairs missing the final word of the second sentence), as previously described in Federmeier and Kutas (1999b). Cloze probability for a given word in a given context was calculated as the proportion of individuals choosing to complete that particular context with that particular word. Expected exemplars were always the item with the highest cloze probability for a given context. Mean cloze probability for the expected exemplars was 0.74 (standard deviation 0.20). Within-category violations and between-category violations always had cloze probabilities of less than 0.05. Mean cloze probability was 0.004 for the within-category violations and 0.001 for the between-category violations.

Cloze norms were also obtained from a sample of 20 older adults to confirm that there were no striking differences between the age groups in their expectations for the sentence-final words in these materials. As in previous studies comparing cloze probability judgments from younger and older adults (which have all reported no age-related changes; e.g., Hamberger, Friedman, & Rosen, 1996; Lovelace & Coon, 1991), we found that older and younger adults gave qualitatively and quantitatively similar patterns of responses to our materials. Older adults gave the same dominant response as the young on all but six of the items (and these differences were lexical rather than semantic in nature—e.g., a sentence that could be completed with either "rats" or "mice," with the younger adults tending to use "rats" and the older adults "mice"). Older adults also showed no tendency to be more variable than the young in their responses to the sentences, with a very similar mean cloze probability (0.78) and standard deviation (0.22). No older adult produced a between-category violation on the cloze task, and within-category violations were produced very infrequently (mean 0.007).

Although all expected exemplars were items with the highest cloze probability for their sentence contexts, the sentence contexts differed in their constraint, or the degree to which they led individuals to strongly expect one particular item versus a number of different items. To examine the effects of sentential constraint on the ERP response to target items, we divided the sentences into two groups, "high constraint" and "low constraint," by a median split on the cloze probability (college-age norms[1]) of the expected exemplar. An example of each type follows:

High Constraint

At the zoo, my sister asked if they painted the black and white stripes on the animal.
I explained to her that they were natural features of a **zebra/donkey/poodle**.

Low Constraint

By the end of the day, the hiker's feet were extremely cold and wet.
It was the last time he would ever buy a cheap pair of **boots/sandals/jeans.**

For the high constraint sentences, the cloze probability of the expected exemplars had a range of 0.78 to 1.0 and an average value of 0.896. For the low constraint sentences, the cloze prob-

[1]Given the lack of systematic differences between the cloze probabilities generated by the younger and older adults, we used the younger adults' data for the split by constraint, as it came from a much larger and thus more robust sample.

ability of the expected exemplars had a range of 0.17 to 0.78 and an average value of 0.588. High constraint sentences are thus those in which there is a single, highly preferred ending, and low constraint sentences are those that are compatible with a larger range of ending types and in which the expected exemplar has at least one, and generally several, close competitors. Word frequency and word length were controlled across all constraint and ending type conditions.

Plausibility Ratings

As detailed in Federmeier and Kutas (1999b), plausibility ratings were obtained from college-aged volunteers for all items in their sentence contexts. Expected exemplars had a mean plausibility rating of 95.6%, within-category violations had a mean plausibility rating of 28.3%, and between-category violations had a mean plausibility rating of 15.3%. Expected exemplars were thus considered more plausible than within-category violations, $t = 46.06$, $p < .001$, and within-category violations more plausible than between-category violations, $t = 15.75$, $p < .001$. These plausibility ratings were influenced by contextual constraint. Expected exemplars were considered more plausible in high (97.7%) than in low (93.5%) constraint sentences, $t = 5.00$, $p < .001$. In contrast, both violation types were rated as more plausible in low (within 30.2%; between 18.7%) than in high (within 23.6%; between 11.9%) constraint sentences, within $t = 3.54$, $p < .001$, between $t = 8.21$, $p < .001$. In other words, the pattern of plausibility ratings was congruent with claims (Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985) that more highly constraining contexts allow greater integration of best completions but reduced integration of improbable completions.

Participants

Twenty-one young adults were obtained from the population of University of California, San Diego (UCSD) undergraduates (10 men and 11 women, 18 to 27 years of age, mean age 20). Twenty-four older adults were recruited from the local San Diego population using a newspaper announcement (12 men and 12 women, 58 to 74 years of age, mean age 68). As a group, the older adults were more educated than the younger adults: all but 4 had at least two years of college education, 9 held a Bachelor's degree, and 5 held a Master's degree or Ph.D. Volunteers were compensated with cash and/or experimental credit hours. All participants were right-handed [as assessed by the Edinburgh Inventory (Oldfield, 1971)] native speakers of English who reported normal hearing. Participants were randomly assigned to the three stimulus lists.

Experimental Procedure

Volunteers were tested in a single experimental session conducted in a soundproof, electrically shielded chamber. They were seated in a comfortable chair and instructed to listen to the stimulus sentences for comprehension. They were informed at the start of the experiment that they would be given a recognition memory test over the stimuli at the conclusion of recording. The session began with a short practice trial designed to reiterate the experimental instructions and to acclimate volunteers to the experimental conditions and the task.

To insure that the stimulus materials were presented at approximately the same subjective volume for all participants (given the possibility of mild hearing loss, particularly for higher frequencies, among the elderly volunteers), we administered a brief audiometric analysis at the start of the experiment. Using the same speaker setup employed for the experimental sentences, each participant was played several sequences of 2000 Hz tones varying systematically (both ascending and descending) in intensity and asked to report the number of tones they experienced. The average intensity level at which the participant could detect the stimulus was calculated, and the experimental stimuli were then presented at 55 dB above this threshold.[2] Participants were asked to maintain visual fixation on a fixation cross and to avoid blinking during the presentation of each sentence pair. Each trial began with the presentation of the context sentence, followed by a 2-s pause, followed by the presentation of the target-containing sentence. A 5-s pause separated trials. Target words had an average duration of 635 ms, and the average duration of the target-containing sentence was 3,563 ms. Volunteers were given a short break after every 15 to 20 pairs of sentences.

At the conclusion of the recording session, participants were given a written recognition memory test consisting of 50 sets of sentence pairs: 10 new ones, 20 unchanged experimental pairs (of which 10 ended with expected exemplars, 5 ended with within-category violations, and 5 ended with between-category violations), and 20 modified sentence pairs in which the final word had been changed from that originally viewed by the volunteer (10 in which violations had been changed to expected exemplars and 10 in which expected exemplars had been changed to violations). Volunteers were instructed to classify the sentences as new, old, or similar (changed).

Neuropsychological Testing

Eighteen of the elderly subjects were able to return to the lab for a second experimental session, during which they were administered a battery of neuropsychological tests.[3] Tests were administered by a single individual in a private testing room free of distractions. Table 1 describes the tests used and the measures collected.

EEG Recording Parameters

The electroencephalogram (EEG) was recorded from 26 geodesically spaced tin electrodes embedded in an Electro-cap. These sites included midline prefrontal (MiPf; equivalent to Fpz), left and right medial and lateral prefrontal (LMPf, RMPf, LLPf, RLPf), left and right medial, mediolateral, and lateral frontal (LMFr, RMFr, LDFr, RDFr, LLFr, RLFr), midline central (MiCe; equivalent to Cz), left and right medial and mediolateral central (LMCe, RMCe, LDCe, RDCe), midline parietal (MiPa; equivalent to Pz), left and right mediolateral parietal (LDPa, RDPa), left and right lateral temporal (LLTe, RLTe), midline occipital (MiOc; equivalent to Oz), and left and right medial and lateral occipital (LMOc, RMOc, LLOc, RLOc). All were referenced to the left mastoid during recording; the right mastoid was also recorded, referenced to the left. Blinks and eye movements were monitored via electrodes placed on the outer canthus (left electrode serving as reference) and infraorbital ridge of each eye (referenced to the left mastoid). Electrode impedances were kept below 5 KΩ. The EEG, processed through Grass amplifiers set at a bandpass of 0.01–

[2] As expected, older volunteers' thresholds were higher on average than those of the younger volunteers, but the difference was less than 10 dB (and, with one exception, under 20 dB), consistent with their self-report of normal hearing.

[3] We did not perform neuropsychological testing with our younger participants because, based on the results of our prior visual study (Federmeier & Kutas, 1999b), we did not expect to see enough variability in their response pattern to allow for an examination of individual differences with this group.

Table 1. *Neuropsychological Tests*

	Source
Language related	
Letter and category verbal fluency (FAS; animals, fruits and vegetables, and first names)	Benton and Hamsher (1978)
Peabody Picture Vocabulary Test III (PPVT-III; raw and age-standardized scores)	Dunn and Dunn (1997)
Semantic relations test from *Clinical Evaluation of Language Functions*, Third Edition (CELF-III)	Semel and Wiig (1994)
Reading comprehension (two essays with multiple choice questions; measured number correct and time to complete)	Lab material
Memory related	
Forwards and backwards digit span from *Wechsler Adult Intelligence Scale-Revised* (WAIS-R)	Wechsler (1981)
Reading span	Daneman and Carpenter (1980)
Executive function	
Modified Wisconsin card sorting test (measured number of categories, perseverative errors, and total errors)	Heaton, Chelune, Talley, Kay, and Curtiss (1993), modified as in Nelson (1976)

100 Hz, was continuously digitized at 250 Hz and stored on hard disk for later analysis.

Data Analysis

Data was re-referenced off-line to the algebraic mean of the left and right mastoids. Trials contaminated by eye movements, excessive muscle activity, or amplifier blocking were rejected off-line before averaging; less than 10% of trials in either group were lost due to such artifacts. In 6 subjects (1 from the younger group and 5 from the older group) with larger numbers of blink artifacts, blinks were corrected via a spatial filter algorithm devised by Dale (1994). Time-locking points for target word onset were determined manually (using both visual and acoustic cues), and ERPs were computed at each electrode location for epochs extending from 100 ms before word onset to 920 ms after. Averages of artifact-free ERP trials were calculated for each type of target word (expected exemplars, within-category violations, between-category violations) in each group after subtraction of the 100-ms prestimulus baseline. Data were bandpass filtered from 0.1 to 20 Hz prior to statistical analyses.

Results

Behavior

Younger adults correctly classified an average of 82% of the recognition sentences, whereas older adults were less accurate, classifying an average of 71% correctly, one-tailed $t(43) = 2.76$,[4] $p < .005$. For both groups, the most common type of error was a misclassification of “similar” sentences (those with an altered final word) as “old,” accounting for 31% of all errors in the younger adults and 41% of all errors in the older adults. There were few “false alarms” (misclassification of new sentences) in either group (7% in the younger adults and 11% in the older adults), and the rest of the errors were fairly evenly spread across misclassifications of similar sentences as new and genuinely old sentences as similar or new.

Perhaps because of the modality shift between study and test conditions, recognition accuracy for the younger adults in this experiment was a bit lower than that observed in previous experiments using the same materials [cf. 88% accuracy for central visual word presentation (Federmeier & Kutas, 1999b) and 93% for central picture presentation (Federmeier & Kutas, 2001)]. The pattern of errors was also somewhat different, as participants in previous studies were equally likely to mistake old for similar and similar for old. However, the performance of both younger and older participants in this experiment was well above chance, indicating that both groups were attending the experimental sentences and processing them for meaning.

ERPs: Younger Adults

Figure 1 shows grand average ERPs (across all 21 college-age volunteers) to sentence-final targets at all recording sites. Because natural speech is both fast and continuous, early components are obscured by the lack of a clear point of word onset and by habituation (see Naatanen & Picton, 1987). However, in all conditions, a negativity can be seen beginning around 250 ms and continuing until about 600 ms, with a peak around 400 ms (N400; e.g., Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Connolly, Phillips, Stewart, & Brake, 1992; Holcomb & Neville, 1991; McCallum, Farmer, & Pocock, 1984). This negativity appears smallest for expected exemplars (solid) and largest for between-category violations (dotted).

Peak latency analysis. The latency of the largest negative peak between 300 and 500 ms was measured for each ending type condition in each participant and subjected to an omnibus analysis of variance (ANOVA). Repeated measures included three levels of ending type (expected exemplars vs. within-category violations vs. between-category violations) and 26 levels of electrode. Note that *p* values in this and all subsequent analyses are reported after epsilon correction (Huynh–Felt) for repeated measures with greater than one degree of freedom. Mean peak latency was 385 ms for expected exemplars, 386 ms for within-category violations, and 406 ms for between-category violations. The analysis revealed a trend for slightly later peak responses to between-category violations as compared with the other two conditions, $F(2,40) = 2.48$, $p = .1$,; this did not interact with electrode, $F(50,1000) = 0.72$, n.s.

Mean amplitude analyses. Mean voltage measures were taken in a 200-ms window around 400 ms (i.e., 300–500 ms poststimulus onset). These measures were subjected to an omnibus ANOVA on two repeated measures: three levels of ending type (expected exemplar vs. within-category violation vs. between-category vio-

[4] Note that because of unequal sample sizes, all group comparison *t*-statistics used a pooled variance estimate.

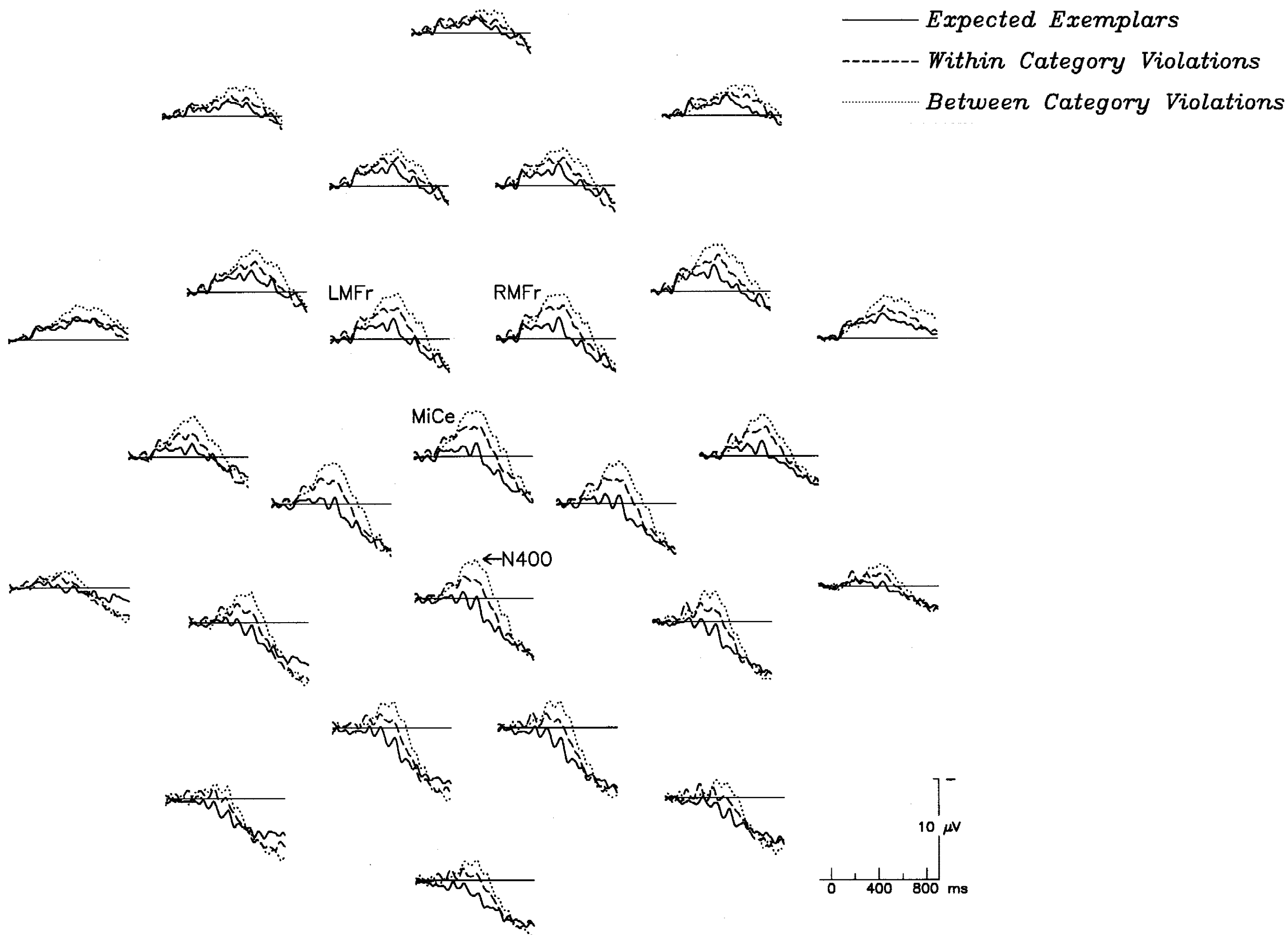

Figure 1. Overall pattern of response in younger adults. Grand average ($N = 21$) ERP waveforms are shown for the three ending types at all 26 electrode sites. The electrode sites are arranged in the figure to approximate their placement on the scalp, with the front of the head at top. Negative is plotted up. In all conditions, a negativity can be seen between 300 and 500 ms (N400). At all channels, this response is smallest for expected exemplars (solid line) and largest for between-category violations (dotted line), with an intermediate response to within-category violations (dashed line).

lation) and 26 levels of electrode. This analysis revealed a main effect of ending type, $F(2,40) = 18.84$, $p < .001$, and an Ending Type × Electrode interaction, $F(50,1000) = 4.39$, $p < .001$. Mean amplitudes of the N400 response were -0.32 uV, -1.66 uV, and -2.79 uV for expected exemplars, within-category violations, and between-category violations, respectively.

Planned comparisons were then conducted via an omnibus ANOVA on two levels of ending type (expected exemplar vs. within-category violation and within-category violation vs. between-category violation) and 26 levels of electrode. Within-category violations were significantly more negative than expected exemplars, $F(1,20) = 7.73$, $p < .01$; this effect interacted with electrode, $F(25,500) = 3.54$, $p < .005$, suggesting that their scalp distributions are not the same. Between-category violations were also more negative than within-category violation, $F(1,20) = 8.85$, $p < .001$. These effects did not seem to differ in distribution, Ending Type × Electrode interaction $F = 1.29$, n.s. This pattern of ending type effect can be seen in Figure 2; it replicates the findings of Federmeier and Kutas (1999b).

Because the analysis over all electrode sites indicated a possible distributional difference between the response to expected exemplars and that to violations, a follow-up distributional analysis was conducted. Data were normalized according to the procedure described in McCarthy and Wood (1985) and then subjected to an ANOVA on four repeated measures: two levels of ending

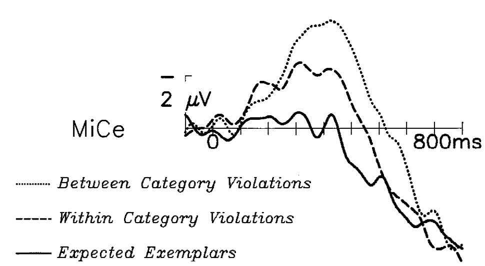

Figure 2. Effect of ending type, younger adults. The effect of ending type is shown here at the middle central site (MiCe, equivalent to Cz). N400s were smaller to the congruent expected exemplars (solid line) than to either violation type. Furthermore, the response to within-category violations (dashed line), those violations that shared many semantic features with the items expected in the context, was reduced relative to the response to between-category violations (dotted line).

type (expected exemplar vs. within-category violation), two levels of hemisphere (left vs. right), two levels of laterality (lateral vs. medial) and four levels of anteriority (prefrontal vs. frontal vs. parietal vs. occipital). There was a significant interaction of Ending Type with Anteriority, $F(3,60) = 4.86$, $p < .05$, and a marginal Ending Type × Laterality interaction, $F(1,20) = 3.41$, $p = .08$. Differences between the conditions were more pronounced over the back of the head and tended to be bigger over medial electrode sites, reflective of the typical N400 distribution (e.g., Holcomb & Neville, 1990, 1991).

Effects of constraint. Effects of ending type between 300 and 500 ms were examined as a function of constraint using the 14 mediocentral electrode sites where N400 responses are typically largest (LDFr, LMFr, RMFr, RDFr, LDCe, LMCe, MiCe, RMCe, RDCe, LDPa, LMOc, MiPa, RMOc, RDPa). An omnibus ANOVA on two levels of constraint (high vs. low), three levels of ending type (expected exemplars vs. within-category violations vs. between-category violations), and 14 levels of electrode revealed no main effect of constraint, $F(1,20) = 2.23$, $p = .15$, but a main effect of ending type, $F(2,40) = 20.88$, $p < .001$, and a Constraint × Ending Type interaction that just missed significance, $F(2,40) = 3.00$, $p = .057$. The interaction becomes significant in the narrower time window of 350–450 ms, $F(2,40) = 3.54$, $p < .05$, and replicates the findings of Federmeier and Kutas (1999b) for presentation in the visual modality.

To look at effects of ending type for each constraint level separately, planned comparisons were then conducted via an omnibus ANOVA on two levels of ending type and 14 levels of electrode. At both constraint levels, within-category violations elicited larger N400 responses than expected exemplars, though in high constraint contexts, this difference was only significant in the early half of the time window [High (300–400 ms): $F(1,20) = 4.16$, $p = .05$; Low (300–500 ms): $F(1,20) = 14.24$, $p < .01$]. At both constraint levels, between-category violations also elicited larger N400 responses than within-category violations, though in low constraint contexts, this effect was only significant in the narrower time window of 350–450 ms [High (300–500 ms): $F(1,20) = 5.00$, $p < .05$; Low (350–450 ms): $F(1,20) = 5.49$, $p < .05$].

The two constraint levels (high vs. low) were also compared directly for each ending type (mean amplitude 300–500 ms) at the same 14 channels. Constraint did not affect the response to either expected exemplars, $F(1,20) = 1.23$, n.s., or between-category violations, $F(1,20) = 1.28$, n.s. It did, however, affect the response to within-category violations, $F(1,20) = 4.63$, $p < .05$, with smaller N400s observed to within-category violations in high (−1.12 uV) than in low (−2.89 uV) constraint contexts. The interaction of Constraint and Ending Type can be seen in Figure 3 and is the same as that observed in young adults for visual presentation of the same stimuli (Federmeier & Kutas, 1999b).

ERPs: Older Adults

Figure 4 shows grand average ERPs (across all 24 elderly volunteers) to sentence-final targets at all recording sites. As was true for younger adults' ERPs, in all conditions, a negativity is visible beginning around 250 ms and continuing until about 600 ms, with a peak around 400 ms (N400). This negativity appears smallest for expected exemplars (solid) and largest for between-category violations (dotted).

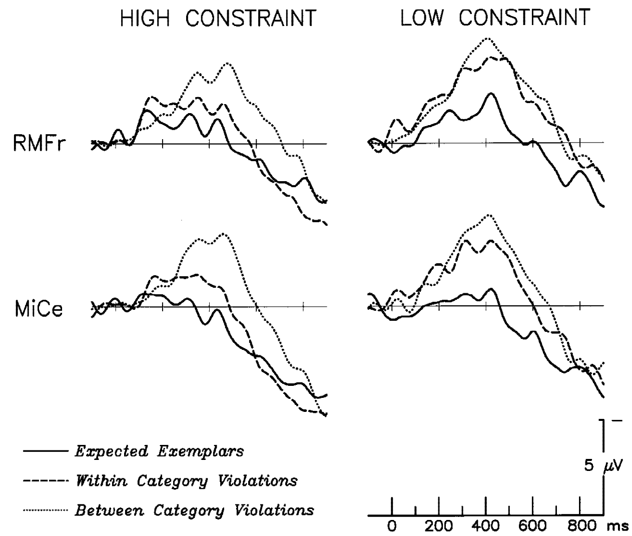

Figure 3. Effect of constraint, younger adults. The effect of sentential constraint on the N400 response is shown at right frontal site RMFr (top) and at MiCe (bottom). All three ending types were different from one another at both constraint levels. When compared directly, constraint did not significantly affect the response to expected exemplars (solid line) or between-category violations (dotted line). Within-category violations (dashed line) in high constraint sentences (left) elicited smaller amplitude N400s than within-category violations in low constraint sentences (right).

Peak latency analysis. The latency of the largest negative peak between 300 and 500 ms was measured for each ending type condition in each participant and subjected to an ANOVA on three levels of ending type (expected exemplars vs. within-category violations vs. between-category violations) and 26 levels of electrode. Mean peak latency was 372 ms for expected exemplars, 397 ms for within-category violations, and 401 ms for between-category violations. Responses to expected exemplars peaked earlier than responses to either type of violation, $F(2,46) = 4.29$, $p < .05$; this did not interact with electrode, $F(50,1150) = 1.22$, n.s. Average N400 peak latency (collapsed across ending type condition and electrode site) did not differ between younger and older adults, one tailed $t(43) = 0.12$, n.s. (see Figure 5).

Mean amplitude analyses. Mean voltage measures, taken in a 200-ms window around 400 ms, were 0.43 uV, −0.72 uV, and −1.00 uV for expected exemplars, within-category violations, and between-category violations, respectively. Overall, as can be seen in Figure 5, these N400 amplitudes (collapsed across ending type and electrode) were significantly smaller in older than in younger adults, one-tailed $t(43) = 1.76$, $p < .05$; this result held for all three ending types examined individually [expected exemplars: one-tailed $t(43) = 1.70$, $p < .05$; within-category violations: one-tailed $t(43) = 2.14$, $p < .05$; between-category violations: one-tailed $t(43) = 4.48$, $p < .001$].[5] The mean amplitude measures were subjected to an omnibus ANOVA on three levels of ending type (expected exemplar vs. within-category violation vs. between-category violation) and 26 levels of electrode. The analysis revealed a main effect of ending type, $F(2,46) = 14.56$, $p < .001$,

[5] Note, however, that responses within the first 200 ms are of similar amplitude in older and younger adults.

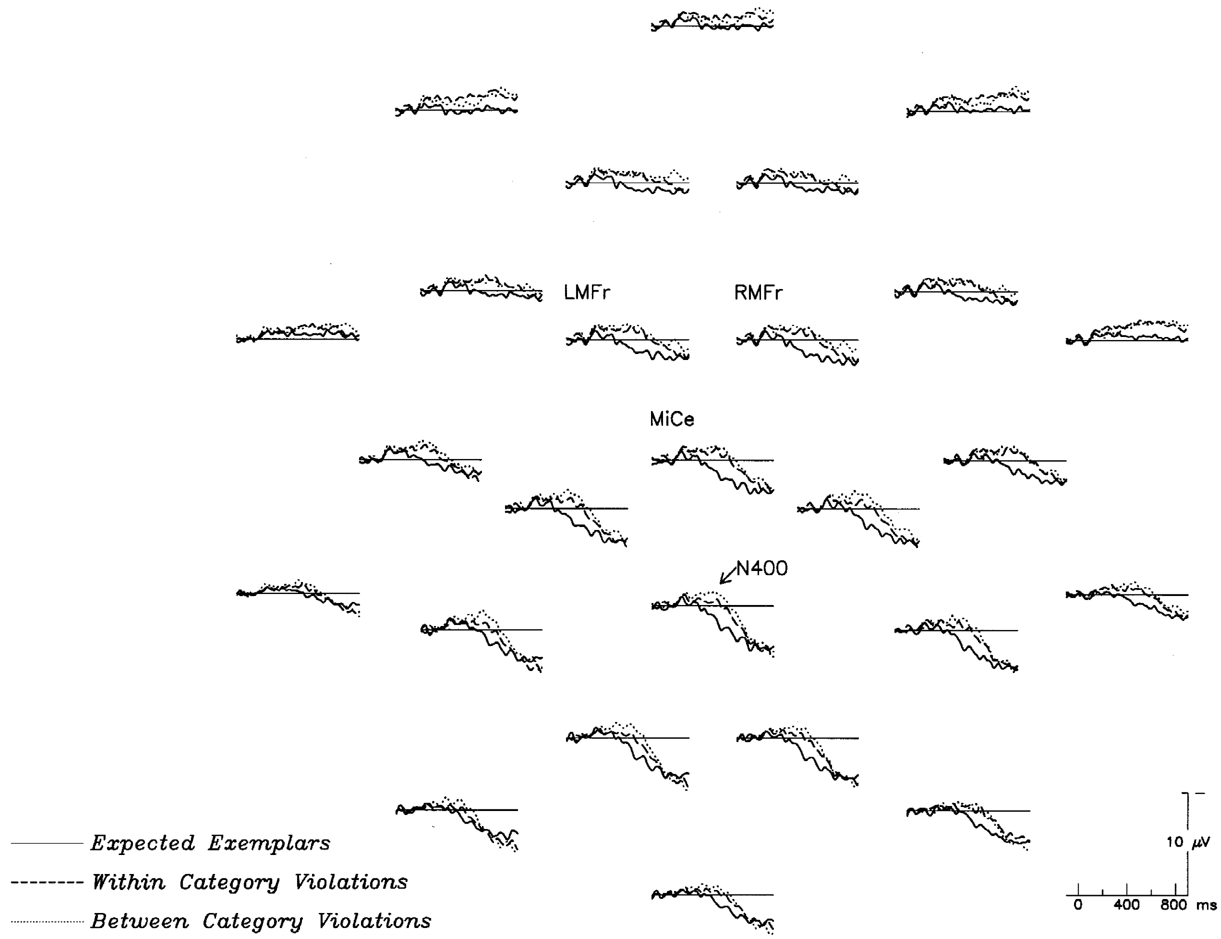

Figure 4. Overall pattern of response in older adults. Grand average ($N = 24$) ERP waveforms are shown for the three ending types at all 26 electrode sites. As was true for younger adults, N400 responses can be seen between 300 and 500 ms (N400). These are smallest for expected exemplars (solid line) and, especially over medial, posterior sites, smaller for within-category violations (dashed line) than for between-category violations (dotted line).

and an Ending Type × Electrode interaction, $F(50,1150) = 3.40$, $p < .001$.

Planned comparisons were conducted via an omnibus ANOVA on two levels of ending type (expected exemplar vs. within-category violation and within-category violation vs. between-category violation) and 26 levels of electrode. Within-category violations were significantly more negative than expected exemplars, $F(1,23) = 13.06$, $p < .005$; in contrast to responses in younger adults, this effect did not interact with electrode, $F(25,575) = 1.69$, n.s. There was no significant main effect for the comparison of within- and between-category violations, $F(1,23) = 1.41$, n.s. However, there was an interaction of Ending Type with Electrode, $F(25,575) = 2.93$, $p = .01$, suggesting that the two conditions might differ over some electrode sites. In fact, between-category violations are significantly more negative than within-category violations, $F(1,23) = 5.25$, $p < .05$, when analyses are restricted to the 10 central-posterior sites (MiCe, LDCe, LMCe, RDCe, RMCe, MiPa, LDPa, LMOc, RDPa, RMOc). The pattern of ending type effects can be seen in Figure 6.

Effects of constraint. Effects of ending type between 300 and 500 ms were again examined as a function of constraint using the 14 mediocentral electrode sites where N400 responses are typically largest (LDFr, LMFr, RMFr, RDFr, LDCe, LMCe, MiCe, RMCe, RDCe, LDPa, LMOc, MiPa, RMOc, RDPa). An omnibus ANOVA on two levels of constraint (high versus low), three levels of ending type (expected exemplars vs. within-category violations vs. between-category violations), and 14 levels of electrode revealed only a main effect of ending type, $F(2,46) = 15.93$, $p < .001$. There was no main effect of constraint, $F(1,23) = 0.08$, n.s., and no Constraint × Ending Type interaction, $F(2,46) = 0.70$, n.s., in this or any subset of this time window.

To look at effects of ending type for each constraint level separately, planned comparisons were then conducted via an omnibus ANOVA on two levels of ending type and 14 levels of electrode. At both constraint levels, within-category violations elicited larger N400 responses than expected exemplars, though in low constraint contexts, this difference was only significant in the late half of the time window [High (300–500 ms): $F(1,23) = 14.65$, $p < .01$; Low (400–500 ms): $F(1,23) = 4.82$, $p < .05$]. For older adults, however, between-category violations elicited larger N400 responses than within-category violations only in low constraint contexts; in high constraint contexts, the violation types did not differ, even in more restricted time windows [High (300–

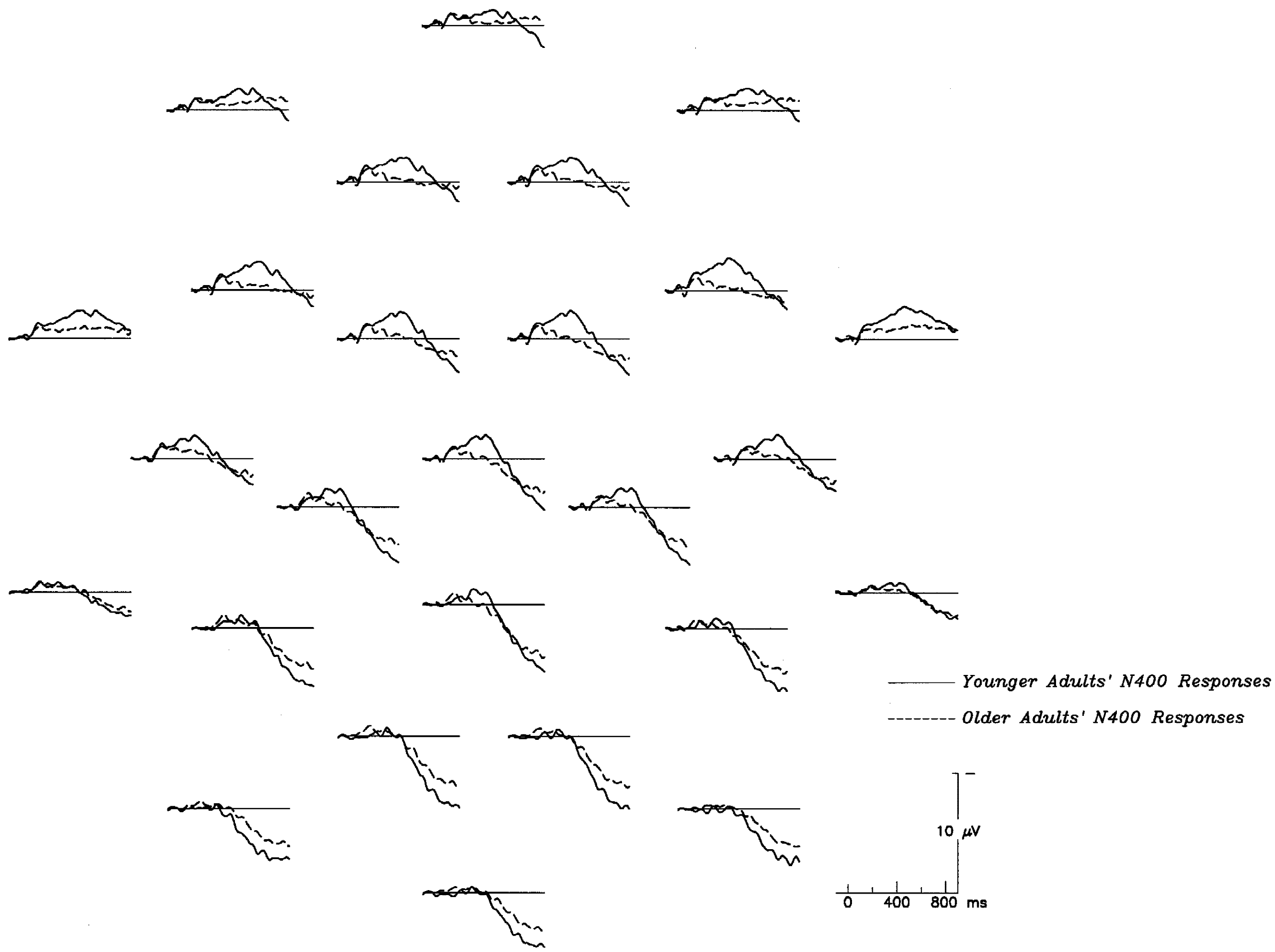

Figure 5. Comparison of N400 response in younger and older adults. N400 responses (averaged across ending type condition and constraint) are shown at all channels for younger (solid line) and older (dashed line) adults. The timing and amplitude of the ERP response is similar across groups until about 250 ms. N400 responses (300–500 ms) are significantly reduced in amplitude in older as compared with younger adults. Peak latency of the N400, however, is similar in both groups.

500 ms): $F(1,23) = 0.68$, n.s.; Low (300–500 ms): $F(1,23) = 4.51$, $p < .05$]. This, then, is opposite from the pattern observed in younger adults, where the violation types were most similar in low constraint contexts.

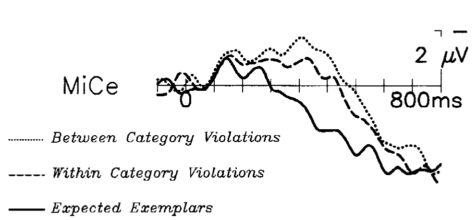

Figure 6. Effect of ending type, older adults. The effect of ending type is shown here at the middle central site (MiCe, equivalent to Cz). As for younger adults, N400s were smaller to the congruent expected exemplars (solid line) than to either violation type. Over some channels, the response to within-category violations (dashed line) was also reduced relative to the response to between-category violations (dotted line).

The two constraint levels (high vs. low) were also compared directly for each ending type (mean amplitude 300–500 ms) at the same 14 channels. As was true for younger adults, constraint did not affect the response to either expected exemplars, $F(1,23) = 0.35$, n.s., or between-category violations, $F(1,23) = 0.00$, n.s. For the analysis on the expected exemplars, however, there was a significant interaction of Constraint with Electrode, $F(13,299) = 2.31$, $p < .05$, reflecting a tendency for greater positivity in high than in low constraint contexts over the more frontocentral electrode sites.[6] In contrast to the results for younger adults in this study and in Federmeier and Kutas (1999b), for older adults there was no effect of constraint on the response to within-category violations, $F(1,23) = 1.05$, n.s. (Figure 7). In fact, a Constraint × Electrode interaction, $F(13,299) = 2.38$, $p < .05$, revealed a tendency, over frontocentral electrode sites, for the opposite effect—namely, more negative responses to within-category violations in high than in low constraint sentences.

[6]The main effect of constraint, however, does not become significant even when analyses are restricted to frontal electrode sites.

Individual variability. Although, on average, older adults' responses to within-category violations were not affected by constraint, there was individual variability, with some of the participants showing differences between high and low constraint within-category items in the same direction and of approximately the same magnitude as that seen on average for younger adults. We therefore conducted a linear regression analysis to see if the magnitude and direction of the constraint effect (high constraint within-category violation N400 amplitude minus low constraint within-category N400 amplitude, collapsed across the same 14 channels) was predicted by subject characteristics and/or neuropsychological measures.

The constraint effect was not correlated with age, $r^2 = .03$, $F(1,22) = 0.59$, n.s., or with performance on the recognition memory test, $r^2 = .05$, $F(1,22) = 1.12$, n.s. There was, however, a small but significant positive correlation with years of education, $r^2 = .16$, $F(1,22) = 4.35$, $p < .05$. For those 18 subjects for whom neuropsychological data were available, a step-wise multiple linear regression analysis was performed using all tests. Included were total verbal fluency (letter and category combined), PPVT-III raw scores, reading comprehension scores (number correct), CELF-III semantic relations scores, digit span (forward and backward combined), reading span, and number of categories on the Wisconsin card sorting test. As a set, the measures were highly correlated with the constraint effect, overall $R^2 = .86$, $F(7,10) = 8.92$, $p = .001$. However, significant independent contributions to the prediction of the constraint effect were made by only two tests: total verbal fluency, beta $= 0.62$, $t(10) = 3.31$, $p < .01$, and PPVT-III raw scores, beta $= 0.49$; $t(10) = 2.32$, $p < .05$. These two variables had a .57 correlation with one another. All other tests—reading span (beta $= -0.26$), digit span (beta $= -0.07$), the reading comprehension test (beta $= 0.13$), the semantic relations test (beta $= 0.13$), and the modified Wisconsin card sorting test (beta $= 0.27$)—did not significantly contribute independently.[7]

The mean total verbal fluency score was 116 (range 83 to 146). Mean total letter fluency (FAS) was 52 (range 35 to 77) and mean total category fluency (animals, fruits and vegetables, first names) was 64 (range 46 to 81). These scores are somewhat higher than but generally comparable to previously published averages for educated older adults using similar methods (e.g., Bolla, Gray, Resnick, Galante, & Kawas, 1998; Kozora & Cullum, 1995; Tombaugh, Kozak, & Rees, 1999). Mean raw score on the Peabody Picture Naming Vocabulary test was 194 (range 167 to 203); mean age-standardized score (Dunn & Dunn, 1997) was 121 (range 85 to 152). Our participants did quite well on this test as a group, with an average percentile rank of 92 (Dunn & Dunn, 1997).

Discussion

The pattern of ERP results observed for young adults in this study replicates that seen previously for visual presentation of the same materials (Federmeier & Kutas, 1999b). The anticipated effect of contextual congruency was observed: Expected exemplars elicited smaller N400s than violations of either type. In addition, we found that within-category violations, those contextually unexpected items that had greater semantic feature overlap with the expected completion, elicited smaller N400 responses than did the between-category violations.[8] This was true despite the fact that neither type of violation was a good completion for the sentence context. Thus, the overall pattern of ending type responses is not simply a function of contextual plausibility—that is, not driven only by how well the semantic features of the presented item fit the feature constraints provided by the sentence context. Instead, the structure of semantic memory—that is, the context-independent semantic similarity between the within-category violation and the expected exemplar—is clearly affecting processing, resulting in a relative facilitation for the within-category as compared with the similarly implausible between-category violation.

That contextual plausibility alone is not driving the N400 response pattern in younger adults is seen even more clearly in the interaction of contextual constraint with ending type. The amplitude of the response to the within-category violations (and only to this target type) was modulated by the strength of the sentence context. N400 responses to within-category violations were smaller in high as opposed to low constraint sentences, an effect that goes in the *opposite* direction from their rated plausibility. As we have discussed in detail previously (Federmeier & Kutas, 1999b; Kutas & Federmeier, 2001), this overall pattern of ending type effects as a function of constraint indicates that sentence processing in

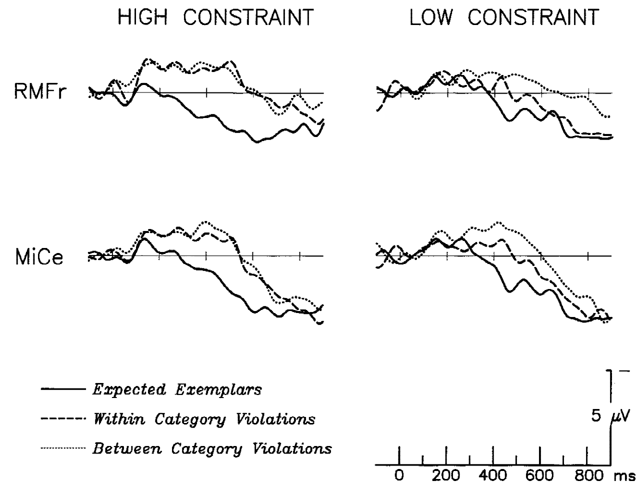

Figure 7. Effect of constraint, older adults. The effect of sentential constraint on the N400 response is shown at right frontal site RMFr (top) and at MiCe (bottom). At both constraint levels, expected exemplars elicit smaller N400 responses than violations of either type. Within-category violations elicit smaller N400s than between-category violations only in low constraint sentences, a pattern that goes in the opposite direction from that seen with young participants. When compared directly, constraint did not significantly affect the response to any of the ending types in this group of participants.

[7]The outcome of the partial correlation analysis was not altered in substance by including either age or education as an additional predictor, and neither variable made a significant independent contribution.

[8]Preliminary comparisons suggest that although the pattern of mean amplitude responses across ending types is similar for visual and auditory presentation of these materials, there may be modality-related differences in effect onsets. In particular, whereas in this experiment the overall difference between expected items and violations is evident prior to the onset of the difference between the two violation types, in the visual experiment, all conditions appear to differentiate from one another at about the same time.

younger adults has a predictive component.[9] Younger adults seem to use sentence context information to preactivate the semantic features of likely upcoming items, and this prediction then mediates their processing of the word that is actually presented. Ease of processing is thus not a function of the fit between the presented word and the context itself, but of the fit between the features of the presented word and the predicted one. Because within-category violations share significant semantic feature overlap with the item that is predicted (but, in this case, not actually presented), their processing is facilitated—especially in highly constraining contexts where the prediction is stronger and more consistent. This is true despite the fact that, in general, participants rate violations of all types as more implausible in highly constraining contexts, where the preferred completion is so well defined.

This study demonstrates that such predictive processing is not restricted to word-by-word reading, but also occurs under the more typical, and more temporally demanding, conditions of comprehending connected speech. Participants here had only about two-thirds the amount of time to process the target-containing sentences as compared with the reading study (average auditory sentence duration was 3.5 s as opposed to an average of about 5–6 s in Federmeier and Kutas, 1999b), yet the pattern of results was identical. In fact, more generally, the results from this study support the idea that the semantic processing of words is similar, independent of the initial modality of presentation [whereas semantic processing seems to differ in certain respects for pictures (Federmeier & Kutas, 2001)].

The more central question for this study, however, concerned the performance of the older group of adults. As is typically observed, this group did not perform as well as the younger adults on the recognition test, indicating age-related decrements in explicit memory processes (cf. Burke et al., 1987; Howard et al., 1981). Nevertheless, it is clear from the pattern of ERP responses that the older adults were processing the meaning of the sentences on-line. Although older adults elicited overall smaller N400 responses than younger adults (as has been observed previously—e.g., Gunter et al., 1992, 1995; Woodward et al., 1993), they show the same congruency effect with similar timing: larger N400s to violations than to expected sentence completions. Older adults thus seem to be able to use context information to make immediate, rapid judgments about the fit of an item to the ongoing discourse. Further, they are able to do so with sufficient granularity to differentiate the expected item from a close semantic neighbor (the within-category violation). Our ERP findings thus cohere with the body of behavioral data suggesting that sentence context information has a facilitating effect on word processing for older, as for younger, adults (e.g., Balota & Duchek, 1991; Burke & Harrold, 1993; Burke & Yee, 1984; Hopkins et al., 1995; Light et al., 1991).

But is this contextual facilitation in older adults arising from the same processing mechanism(s)? To get at this issue, we specifically examined how responses to the three ending types were modulated by the constraint of the sentence contexts. Like younger adults, older adults show an overall effect of semantic similarity on the response to the unexpected items, with reduced N400 responses to within-category as compared to between-category violations. This difference between the violation types, however, is smaller in older adults than in younger adults and seen only when analyses are restricted to the more limited set of medial posterior electrode sites. More importantly, this main effect of ending type is modulated by a very different interaction with constraint than that observed in the younger adults. Whereas in younger adults the facilitation for within-category violations was driven by the response in high constraint sentences, in older adults, there is no difference between the violation types when these are embedded in highly constraining contexts. Instead, the slight facilitation for the within-category violations is coming from responses in *low* constraint sentences, consistent with rated plausibility [cf. behavioral patterns in younger adults reported by Schwanenflugel & LaCount (1988) and Schwanenflugel & Shoben (1985)]. More generally, older adults' responses were less influenced by constraint, as responses to all of the ending types were similar in amplitude as a function of constraint when compared directly.

Consistent with prior behavioral and electrophysiological work, therefore, we find that older adults can rapidly use contextual information to facilitate word processing (e.g. Cohen & Faulkner, 1983; Gunter et al., 1992, 1995; Madden, 1989; Obler et al., 1985; Woodward et al., 1993). However, they seem to use context *differently* from younger adults, as can be seen in particular in their responses to contextually unexpected items as a function of contextual constraint. Younger adults show facilitation for unexpected items when these share semantic features with the *expected* item—and the strength of this facilitation is directly related to the predictive strength of the context. Older adults, in contrast, show less facilitation for semantically similar violations—and do so only when contextual information is weaker and thus where such unexpected items are actually more plausible in the sentence contexts. In other words, the pattern of N400 amplitudes we observe for younger adults [both here and for word-by-word reading (Federmeier & Kutas, 1999b)] cannot be explained by plausibility alone, suggesting that contextual information is used to actively predict semantic features of upcoming words. N400 amplitudes in older adults, however, pattern directly with plausibility (i.e., the fit between the item actually presented and the feature constraints of the sentence context), with no evidence for prediction.

The age-related difference we observe is similar in some ways to that observed by Hamberger et al. (1996). With young adults they observed the well-established pattern that N400 amplitudes are graded by both their predictability in the context and their relationship with the most expected word (e.g., Kutas & Hillyard, 1984). Thus, an unexpected but semantically related completion (e.g., "They left dirty dishes in the faucet") elicits a smaller N400 than an unexpected completion without a semantic relationship to the expected word (e.g., "He mailed the letter without a lace"). Normal elderly participants, however, seemed to show an N400 reduction *only* to the best completion and not to the semantically related endings. This pattern of results again suggests that although older adult's sentence processing is affected by the fit of a given word to the context, it is less affected by the relationship between that word and a contextually *predicted* item. Thus, older adults' seem to be using context differently—specifically making less (or less efficient) use of context as a means to prepare for the processing of likely upcoming stimuli.

Although this pattern of results holds on the average, there is a subset of older adults that do seem to use context similarly to younger adults, showing increased facilitation for within-category violations in high as opposed to low constraint contexts. These older adults thus seem to be able to more effectively use the predictive information available in the sentence contexts to prepare for the processing of upcoming stimuli. The question then is why

[9] When we use the term "predictive" we do not mean that participants are guessing at a conscious or strategic level; rather, we use the term to mean that the nature of the processing is such that features of likely upcoming items become active prior to their actual occurrence.

the sentence processing of some, but not all, older adults resembles that of younger individuals.

One possibility is resource availability. As a group, older adults seem to have reduced working memory capacities relative to younger adults (e.g., Wingfield, Stine, Lahar, & Aberdeen, 1988) and differences in working memory span have been shown to be predictive of some patterns of language performance (e.g., Cohen, 1981; Light & Anderson, 1985; Light & Capps, 1986). One hypothesis, then, is that some older adults with reduced memory spans must allocate all available resources to dealing with the memory demands of the language comprehension task, leaving little left over for predictive processing. In this experiment, however, neither digit span nor reading span was correlated with the tendency of older adults to show the predictive pattern of ERP responses (and the older adults' average reading span was in the same range as is typically observed for younger adults). Thus, the tendency to use predictive context information does not seem to be any simple function of working memory resources.

Processing speed might be another variable that differentiates the group of older adults. As already discussed, there have been suggestions that older adults' semantic activation is slower than that of younger adults, though not all studies find this result (Balota et al., 1992; Balota & Duchek, 1988; Howard et al., 1986). Concordant with this idea, previous ERP work looking at sentence processing has found delays in the peak latency of the N400 response in older relative to younger adults (Gunter et al., 1992, 1995; Woodward et al., 1993). However, in this study, we found no differences in the overall peak latency of the N400 response between younger and older adults. There was only a difference in the *pattern* of latencies across conditions. For younger adults, the response to both expected exemplars and within-category violations was slightly earlier than the response to between-category violations. In older adults, instead, the response to expected exemplars was faster than the response to violations of either type (which did not differ). Again, therefore, we find a stronger coupling between the response to expected exemplars and within-category violations in younger as opposed to older adults, consistent with the idea that younger adults are making predictions. Averaged across conditions, however, N400 latencies were the same for older and younger adults in this study. Our experiment differs from previous sentence-processing studies that do find latency shifts in that here material was presented as natural speech. Prior studies have either been in the visual modality (Gunter et al., 1992, 1995) or, with auditory material, have imposed a delay between the context and the final, target word (Woodward et al., 1993). It may be that natural connected speech, with its coarticulatory cues, provides the extra time and information needed for elderly adults' word processing to keep pace with younger adults'. Regardless of the underlying reasons, however, we do not find evidence for slowing, at least as indexed by N400 latency.[10]

The individual variability in the older adults also did not seem to be based on the ability to switch sets (Wisconsin card sorting test), to comprehend text passages in general (semantic relations test and reading comprehension test), or to explicitly remember the test sentences. The individual differences, however, *were* predicted by performance on the verbal fluency test and by raw scores on the Peabody Picture Vocabulary Test. Older adults who were able to generate more lexical items of the appropriate type (i.e., beginning with a given letter or belonging to a particular semantic category) in 1 min showed larger N400 reductions to within-category violations in high as compared with low constraint contexts. The tendency to show the young pattern of brainwave responses was also predicted by a larger (auditory) receptive vocabulary. Thus, the tendency to show a predictive pattern of ERP responses seems to be related both to the ability to assign meaning to a wide range of words when these are presented auditorily and to the ability to generate lexical items quickly and appropriately, on demand.

Among the neuropsychological measures we collected, the best predictor of the tendency for older adults to show a response pattern similar to young adults was not a comprehension measure but rather a production one: verbal fluency. Verbal fluency has been linked to frontal and temporal lobe functions (reviewed in Stuss et al., 1998) and declines in (at least some subcomponents of) verbal fluency measures have been reported with normal aging (Kozora & Cullum, 1995; Tombaugh et al., 1999). As a group, our older adults performed quite well on the fluency test, surpassing the typical means reported for individuals in this age and education range (Bolla et al., 1998; Kozora & Cullum, 1995). Although we were not able to obtain neuropsychological data from the specific younger adults in this study, a comparable group of 32 UCSD undergraduates (tested for a different study in our lab) had a mean total fluency score of 117. Thus, with a mean score of 116, the older adults in our study do not seem to differ from younger adults tested under comparable conditions; this may be because age-related changes are offset by the higher average education of our older relative to our younger participants [verbal fluency performance is predicted by both (Tombaugh, Kozak, & Rees, 1999)]. Nevertheless, there was variability in the older adults' performance on this test, and those who obtained higher fluency scores were more likely to elicit brainwave responses that varied with ending type and constraint in the manner seen here (and previously) for younger adults.

This finding lends support to our claim that the young response pattern is related to predictive processing. It seems that older adults who are able to rapidly generate lexical items on demand can take advantage of these abilities during on-line language processing to preactivate the likely semantic (and perhaps in some cases even lexical) candidates for the upcoming item. Much of the time, such preactivation could be expected to make language comprehension more efficient (though of course there will be times when predictions are incorrect, requiring reanalysis). The difference in the use of predictive context information across older individuals might not be very apparent under nontaxing language processing conditions. However, the ability to use context predictively could be expected to significantly improve language comprehension under more difficult situations—for example, when input is speeded or accompanied by noise or competing stimuli, or under dual task or otherwise stressful processing conditions.

A secondary, but also significant, predictor of the young response pattern was auditory receptive vocabulary. All but one of our older adults actually scored above average for their age on this test, with more than half scoring above the 90th percentile; thus, it is unlikely that our older participants had any general difficulties comprehending the experimental sentences (which as a whole did not contain extremely low frequency words). Nevertheless, those older adults with more elaborate semantic memories may be in a better position to take advantage of the more semantically specific information provided by the highly constraining contexts. To the extent that such individuals use context predictively, therefore,

[10] There was also no correlation on an individual basis between N400 latency and the constraint effect, $r^2 = .05$, $F = 1.27$, n.s.

their predictions are likely to be better guided by contextual information and thus more sensitive to constraint. Additionally, vocabulary size has been shown to be positively correlated with reading frequency/exposure to print (e.g., West, Stanovich, & Mitchell, 1993), which is, in turn, correlated with various measures of reading skill (e.g., Stanovich & Cunningham, 1992). It may be, therefore, that underlying the correlation with vocabulary we observe is a tendency for older adults who are more skilled readers to show the young (predictive) response pattern. Further research will be needed to tease apart the contributions of vocabulary, reading practice, and reading skill to these effects.

Overall, then, our results support previous work showing (a) that older adults are able to use contextual information on-line to facilitate the semantic processing of congruent items and (b) that the organization of semantic memory (here, in terms of category structure) remains relatively intact with age. We further show that, for older adults as for younger, semantic memory organization has a direct, early (within 400 ms) impact on on-line language processing, such that implausible items with expected semantic features are facilitated relative to similarly implausible items without such feature overlap. Our results suggest, however, that on average, older adults differ from younger adults in how they use context on-line. Whereas younger adults seem to use context to prepare for the processing of likely upcoming items by preactivating appropriate semantic features, older adults do not show the pattern of brainwaves associated with such predictive processing. This indicates that either older adults do not engage in predictive processing (using instead an integrative mechanism that is driven by plausibility) or that the output of predictive processes in older individuals does not come in time to affect the stages of language processing indexed by the N400.

People age differently, however, and we in fact found a subset of older adults whose brainwaves patterned like those of younger adults, and who thus seem to have retained the ability to make use of context information to actively prepare for the processing of likely upcoming stimuli. These are individuals who tend to have larger receptive vocabularies and who are able to more fluently generate appropriate lexical items on demand. Our results do not seem to support the idea that it is declines in either vocabulary or fluency per se that are responsible for the age-related changes we observe, as our older adults as a group perform comparably to younger adults on both measures. However, our results suggest that a large vocabulary and high verbal fluency can help to offset other effects of aging—whatever their underlying cause—and thereby allow older adults, like younger ones, to use context predictively during on-line sentence processing.

REFERENCES

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain & Language*, *39*, 187–205.

Balota, D. A., Black, S. R., & Cheney, M. (1992). Automatic and attentional priming in young and older adults: Reevaluation of the two-process model. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 485–502.

Balota, D. A., & Duchek, J. M. (1988). Age-related differences in lexical access, spreading activation, and simple pronunciation. *Psychology & Aging*, *3*, 84–93.

Balota, D. A., & Duchek, J. M. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain & Language*, *40*, 181–201.

Benton, A. L., & Hamsher, K. (1978). *Multilingual aphasia examination manual*. Iowa City, IA: University of Iowa.

Bolla, K. I., Gray, S., Resnick, S. M., Galante, R., & Kawas, C. (1998). Category and letter fluency in highly educated older adults. *Clinical Neuropsychologist*, *12*, 330–338.

Bowles, N. L. (1989). Age and semantic inhibition in word retrieval. *Journal of Gerontology*, *44*, P88–P90.

Bowles, N. L., & Poon, L. W. (1981). The effect of age on speed of lexical access. *Experimental Aging Research*, *7*, 417–425.

Bowles, N. L., & Poon, L. W. (1985). Aging and retrieval of words in semantic memory. *Journal of Gerontology*, *40*, 71–77.

Bowles, N. L., Williams, D., & Poon, L. W. (1983). On the use of word association norms in aging research. *Experimental Aging Research*, *9*, 175–177.

Burke, D. M., & Harrold, R. M. (1993). Automatic and effortful semantic processes in old age: Experimental and naturalistic approaches. In L. L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 100–116). New York: Cambridge University Press.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory & Language*, *30*, 542–579.

Burke, D. M., & Peters, L. (1986). Word associations in old age: Evidence for consistency in semantic encoding during adulthood. *Psychology & Aging*, *1*, 283–291.

Burke, D. M., White, H., & Diaz, D. L. (1987). Semantic priming in young and older adults: Evidence for age constancy in automatic and attentional processes. *Journal of Experimental Psychology: Human Perception & Performance*, *13*, 79–88.

Burke, D. M., & Yee, P. L. (1984). Semantic priming during sentence processing by young and older adults. *Developmental Psychology*, *20*, 903–910.

Cohen, G. (1981). Inferential reasoning in old age. *Cognition*, *9*, 59–72.

Cohen, G., & Faulkner, D. (1983). Word recognition: Age differences in contextual facilitation effects. *British Journal of Psychology*, *74*, 239–251.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain & Language*, *43*, 1–18.

Dale, A. M. (1994). *Source localization and spatial discriminant analysis of event-related potentials: Linear approaches*, University of California San Diego.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450–466.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test*, 3rd ed. Circle Pines, MN: American Guidance Service.

Federmeier, K. D., & Kutas, M. (1999a). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*, 373–392.

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

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 202–224.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle-aged academics. *Psychophysiology*, *29*, 38–54.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1995). Language, memory, and aging: A electrophysiological exploration of the N400 during reading of memory-demanding sentences. *Psychophysiology*, *32*, 215–229.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1998). Priming and aging: An electrophysiological investigation of N400 and recall. *Brain & Language*, *65*, 333–355.

Hamberger, M. J., Friedman, D., Ritter, W., & Rosen, J. (1995). Event-related potential and behavioral correlates of semantic processing in Alzheimer's patients and normal controls. *Brain & Language*, *48*, 33–68.

Hamberger, M. J., Friedman, D., & Rosen, J. (1996). Completion norms collected from younger and older adults for 198 sentence contexts. *Behavior Research Methods, Instruments & Computers*, *28*, 102–108.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test, revised and expanded*. Odessa, FL: Psychological Assessment Resources, Inc.

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

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

Hopkins, K. A., Kellas, G., & Paul, S. T. (1995). Scope of word meaning activation during sentence processing by young and older adults. *Experimental Aging Research*, *21*, 123–142.

Howard, D. V. (1980). Category norms: A comparison of the Battig and Montague (1969) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology*, *35*, 225–231.

Howard, D. V. (1983). The effects of aging and degree of association on the semantic priming of lexical decisions. *Experimental Aging Research*, *9*, 145–151.

Howard, D. V., McAndrews, M. P., & Lasaga, M. I. (1981). Semantic priming of lexical decisions in young and old adults. *Journal of Gerontology*, *36*, 707–714.

Howard, D. V., Shaw, R. J., & Heisey, J. G. (1986). Aging and the time course of semantic activation. *Journal of Gerontology*, *41*, 195–203.

Kozora, E., & Cullum, C. M. (1995). Generative naming in normal aging: Total output and qualitative changes using phonemic and semantic constraints. *The Clinical Neuropsychologist*, *9*, 313–320.

Kutas, M., & Federmeier, K. D. (2001). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*, 463–470.

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

Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology & Aging*, *8*, 34–43.

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology*, *40*, 737–747.

Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Developmental Psychology*, *22*, 580–585.

Light, L. L., Valencia-Laver, D., & Davis, D. (1991). Instantiation of general terms in young and older adults. *Psychology & Aging*, *6*, 337–351.

Lovelace, E. A., & Cooley, S. (1982). Free associations of older adults to single words and conceptually related word triads. *Journal of Gerontology*, *37*, 432–437.

Lovelace, E. A., & Coon, V. E. (1991). Aging and word finding: Reverse vocabulary and Cloze tests. *Bulletin of the Psychonomic Society*, *29*, 33–35.

Madden, D. J. (1988). Adult age differences in the effects of sentence context and stimulus degradation during visual word recognition. *Psychology & Aging*, *3*, 167–172.

Madden, D. J. (1989). Visual word identification and age-related slowing. *Cognitive Development*, *4*, 1–29.

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

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, *12*, 313–324.

Obler, L. K., Nicholas, M., Albert, M. L., & Woodward, S. (1985). On comprehension across the adult lifespan. *Cortex*, *21*, 273–280.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Salthouse, T. A. (1993). Effects of aging on verbal abilities: Examination of the psychometric literature. In L. L. Light, & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 17–35). New York: Cambridge University Press.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *14*, 344–354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232–252.

Scialfa, C. T., & Margolis, R. B. (1986). Age differences in the commonality of free associations. *Experimental Aging Research*, *12*, 95–98.

Semel, E., & Wiig, E. (1994). *The Clinical Evaluation of Language Function*, 3rd ed. San Antonio: The Psychological Corporation.

Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory, & Cognition*, *20*, 51–68.

Stuss, D. T., Alexander, M. P., Hamer, L., Palumbo, C., Dempster, R., Binns, M., Levine, B., & Izukawa, D. (1998). The effects of focal anterior and posterior brain lesions on verbal fluency. *Journal of the International Neuropsychological Society*, *4*, 265–278.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*, 167–177.

Tun, P. A., & Wingfield, A. (1993). Is speech special? Perception and recall of spoken language in complex environments. In J. Cerella, J. M. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 425–457). San Diego: Academic Press.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised manual*. New York: Psychological Corporation.

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly*, *28*, 35–50.

Wingfield, A., Stine, E. A. L., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research*, *14*, 103–107.

Woodward, S. H., Ford, J. M., & Hammett, S. C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography, & Clinical Neurophysiology*, *87*, 306–320.

(Received September 22, 2000; Accepted August 7, 2001)