available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

# Multiple effects of sentential constraint on word processing

*Kara D. Federmeier*[a,b,c,*], *Edward W. Wlotko*[a], *Esmeralda De Ochoa-Dewald*[d], *Marta Kutas*[d,e]

[a]*Department of Psychology, University of Illinois, Urbana-Champaign, IL 61820, USA*
[b]*Neurosciences Program, University of Illinois, Urbana-Champaign, IL 61820, USA*
[c]*The Beckman Institute for Advanced Science and Technology, University of Illinois, Urbana-Champaign, IL 61820, USA*
[d]*Department of Cognitive Science, University of California, San Diego, CA 92093-0515, USA*
[e]*Neurosciences, University of California, San Diego, CA 92093-0515, USA*

ARTICLE INFO

*Article history:*
Accepted 28 June 2006
Available online 9 August 2006

*Keywords:*
Language
Sentence processing
Context effect
Sentential constraint
ERP
N400

ABSTRACT

Behavioral and electrophysiological studies have uncovered different patterns of constraint effects on the processing of words in sentences. Whereas response time measures have indicated a reduced scope of facilitation from strongly constraining contexts, event-related brain potential (ERP) measures have instead revealed enhanced facilitation for semantically related endings in such sentences. Given this disparity, and the concomitant possibility of functionally separable stages of context effects, the current study jointly examined expectancy (cloze probability) and constraint effects on the ERP response to words. Expected and unexpected (but plausible) words completed strongly and weakly constraining sentences; unexpected items were matched for contextual fit across the two levels of constraint and were semantically unrelated to the most expected endings. N400 amplitudes were graded by expectancy but unaffected by constraint and seemed to index the benefit of contextual information. However, a later effect, in the form of increased frontal positivity from 500 to 900 ms post-stimulus-onset, indicated a possible cost associated with the processing of unexpected words in strongly constraining contexts.



## 1. Introduction

A substantial body of literature using a variety of behavioral and neurophysiological measures attests to the impact of sentence context information on word processing. Congruent context information has been found to speed word naming and lexical decision (word/nonword) judgment times (Fischler and Bloom, 1979; Kleiman, 1980; Schuberth et al., 1981; Stanovich and West, 1979) and to enhance word perception in the Reicher–Wheeler task (Jordan and Thomas, 2002). During natural reading, contextually constrained words are more likely to be skipped and less likely to be regressed to, and, when fixated, are viewed for less time than less constrained words (Ehrlich and Rayner, 1981). A word in a supportive context also elicits a reduced N400 (Kutas and Hillyard, 1980), an event-related brain potential component that has been linked to the access and integration of meaning information (see review by Kutas and Federmeier, 2001).

Although it seems clear that context affects word processing, the mechanisms by which it does so remain controversial. A critical issue concerns when context has its effects, and – on the assumption that the effects of context may be multiple – whether there are different mechanisms involved at different points in time. For example, context

* *Corresponding author*. Department of Psychology, UIUC, 603 E. Daniel St. Champaign, IL 61820, USA.
E-mail address: kfederme@uiuc.edu (K.D. Federmeier).

0006-8993/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2006.06.101

effects might occur relatively late, at the point in which the word currently being analyzed is integrated with information stored in working memory, or much earlier in the processing stream (e.g., if the system anticipates likely upcoming words, features, or concepts and makes preparations to process them prior to their occurrence); they might involve enhancement (facilitation) or suppression (inhibition) or both; and they might be differentially impacted by matching and mismatching information. A wealth of empirical data has been brought to bear on such issues, yielding an interesting pattern of convergence and divergence across tasks and measures.

Whether measured in terms of response times, eye-movement patterns, or brain electrophysiology, facilitation has been observed for moderately or highly predictable words in congruent contexts, and this facilitation is graded by the degree of fit between a particular word and a particular sentence. Such fit is often defined empirically using a cloze procedure, in which participants are asked to complete a sentence fragment with the word that first comes to mind; the frequency with which a particular word is produced is its cloze probability for that context (Taylor, 1953). For all measures, benefits of context (i.e., advantages for words in predictive contexts relative to wholly unpredictive – sometimes called neutral – contexts) have generally been shown to be monotonically graded with cloze probability: the greatest benefit (or N400 reduction) is seen to items with high cloze probability, sometimes called "best completions", but some facilitation is observed even for items of only moderate predictability (e.g., Jordan and Thomas, 2002; Kutas and Hillyard, 1984; Rayner and Well, 1996; Schuberth et al., 1981).

The effects of context on the processing of unexpected words, however, have been less clear, and are of great theoretical import for determining the scope and nature of context's influence. Behavioral investigations, primarily using lexical decision tasks, have found that such effects are influenced by a number of factors, including the semantic similarity between the unexpected word and the best completion for that sentence and the context's constraint, e.g., the degree to which it narrows down the range of possible continuations. Schwanenflugel and her colleagues observed facilitation for the processing of unexpected endings related to an expected completion, but only when these items completed weakly constraining contexts (e.g., "She cleaned the dirt from her sandals," where "shoes" is the expected but low cloze probability ending) and not when they appeared in strongly constraining contexts (e.g., "On a hot summer's day, many people go to the lake," where "beach" is expected with high cloze probability) (Schwanenflugel and LaCount, 1988; Schwanenflugel and Shoben, 1985). These results have been interpreted as suggesting that strongly constraining contexts yield a narrower scope of activation than do weakly constraining contexts. In particular, Schwanenflugel and her colleagues (Schwanenflugel and LaCount, 1988; Schwanenflugel and Shoben, 1985) have hypothesized that more constraining contexts establish a larger set of featural restrictions for possible upcoming words. Since, on their proposal, facilitation occurs only when the semantics of an incoming word matches on all of the featural descriptions established by the context, the greater number of featural restrictions exacted by strongly constraining contexts increases the probability of a mismatch and thus makes facilitation less likely.

On this view, then, strongly constraining contexts yield a robust but focused facilitation, which does not extend even to those unexpected words that share semantic features in common with the best completion (although at least one study (Schwantes, 1985), using a naming task, did find facilitation for unexpected words in strongly constrained sentences). Facilitation from weakly constraining contexts, in contrast, does seem to extend to other, related words/concepts. In Schwanenflugel and LaCount's (1988) study, there was no benefit for unexpected and unrelated words in any context, suggesting a limit to the scope of facilitation even for weakly constrained sentences. Indeed, for unexpected words that are anomalous in their contexts, inhibitory effects have sometimes been observed—most often in lexical decision tasks (Fischler and Bloom, 1979, 1980) but also in naming tasks under some conditions of stimulus degradation (Stanovich and West, 1979, 1983). It has been argued that these inhibitory effects of sentential context arise in a different (relatively late) processing stage.

However, N400 responses – which, as already mentioned, are graded with cloze probability in a manner that parallels behavioral data – have shown a pattern of sensitivity to contextual constraint and semantic relatedness that is intriguingly different from the behavioral effects. Initial reports found little influence of contextual constraint: Kutas and Hillyard (1984) examined ERP responses to low cloze probability words in strongly, moderately, and weakly constraining sentence contexts (as defined by the cloze probability of the best completion for that sentence), and found no difference in N400 amplitude. Thus, whereas in behavioral studies the amount of facilitation observed for an unexpected word also seemed to depend on the predictability of other words implied by the context (but never actually presented), N400 amplitudes seemed to pattern with cloze probability, independent of contextual constraint.

However, these initial studies did not systematically control for the degree of semantic relationship between the unexpected word and the best completion, a factor that had clear effects on the behavioral patterns. Federmeier and Kutas (1999b) examined the effects of relatedness and constraint by recording ERPs to unexpected sentence final words that had greater or lesser degrees of semantic overlap with the most expected completion. In an attempt to control for the expectancy of the two unexpected ending types, both were designed to be implausible completions for the sentence pairs. In sentence pairs such as, "They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of …" N400 responses were facilitated for expected ("palms") relative to unexpected completions, but among unexpected completions were smaller to those from the same semantic category ("pines") than those from a different category ("tulips"). Further, this pattern interacted with constraint: greater facilitation was found for these "within category violations" in strongly than in more weakly constraining sentences, an effect that went in the opposite direction from the rated plausibility of these words in their contexts. Federmeier and Kutas (1999b) interpreted their

findings as providing support for the hypothesis that listeners use context information to actively prepare for – i.e., to predict – semantic features of upcoming items. Facilitation is then a function of both the strength of the prediction (which varies with constraint) and the amount of semantic overlap between the predicted word and the one actually presented (greater for within-category than for between-category violations).

Note that the effects of constraint observed on N400 amplitude measures by Federmeier and Kutas (1999b) are opposite the pattern found for lexical decision times by Schwanenflugel and LaCount (1988). Rather than showing a narrowed scope of facilitation for strongly constraining sentences, the electrophysiological data suggest that, at least at some processing stages, strongly constraining sentences more strongly facilitate unexpected but semantically related words — including those that are actually implausible in their contexts. Such a pattern makes sense on the assumption that strongly constraining contexts can engender both greater benefits and greater costs for word processing. As emphasized by Federmeier and Kutas (1999b), increased constraint could mean increased pre-activation of some types of information, and thus increased benefit for words whose features overlap with those elicited by the context. However, as emphasized by Schwanenflugel and LaCount (1988), increased constraint could also increase the possibility of a mismatch and/or entail costs associated with altering the contextually-induced pattern of activation when an unpredictable item is encountered. In other words, the narrowed scope evident for strongly constraining contexts in behavioral measures may arise from processing that occurs downstream of the N400. Critically, the difference in the pattern seen across the two measures suggests that there may be multiple effects of context that differentially modulate processing at different times.

Given the apparent disparity between behavioral and electrophysiological effects of sentential constraint, and the concomitant possibility there are functionally separable stages of context effects on word processing, it makes sense to attempt to further disentangle the effects of cloze probability and constraint on the ERP response to words. In particular, while behavioral studies have examined the influence of contextual information on the response to unexpected and semantically unrelated words, this has not yet been done with ERPs in a systematic and controlled fashion. Such manipulations may yield important data, since mismatch effects and/or costs associated with switching between patterns of activation might be more prevalent for unexpected endings that are more semantically distant from the best completions. One recent study that manipulated lexical- and message-level sources of constraint (Hoeks et al., 2004), observed larger N400 responses to "poor fit" items in sentences with strong constraint ("The javelin was by the athletes *summarized*"; English translation of a Dutch stimulus sentence) than in sentences with weak constraint ("The javelin has the athletes *summarized*"). The cloze probability of both of these words was at floor and their rated plausibility did not differ, so the difference in the responses to them would seem to be due to the presence, in the strong constraint condition, of a preferred competitor. However, this study differed from previous studies in that constraint was manipulated syntactically; indeed, a strikingly divergent finding of this study was the lack of an effect of either cloze probability or constraint on the N400 response to "good fit" items (i.e., there was no difference in N400 amplitude between "The javelin was by the athletes *thrown*", which had a cloze probability of 77% and "The javelin has the athletes *thrown*", which had a cloze probability of 1%). Thus, a study that manipulated constraint semantically would provide a more straightforward comparison to the extant behavioral and ERP data.

In the current study, therefore, we examine ERPs to expected endings of strongly (e.g., "The children went outside to *play*") and weakly (e.g., "Joy was too frightened to *move*") constraining sentences, and also to plausible, but semantically distinct, unexpected endings in those same contexts (e.g., "look" completing either of the two prior example sentences). For expected endings, cloze probability and constraint cannot be dissociated, since, by definition, only strongly constraining sentences lead to high cloze probability completions (cloze probabilities were 91% and 35%, respectively, for the examples above). However, both types of sentences can accommodate low cloze probability items, and these items can be matched for their rated cloze probability (cloze probability of look was 3% in both of the example sentences) and controlled for their semantic relationship to the expected ending (e.g., *look* is not associated with or a close semantic relation of either *play* or *move*). Any difference in the processing of these unpredictable items can then be attributed to the constraint of the sentence context itself. Of special interest in this study will be whether constraint effects can be seen, when other variables are controlled for, in N400 amplitude patterns or in any other aspects of the ERP response to these semantically-unrelated, unexpected words.

## 2. Results

### 2.1. *Behavior*

Participants correctly recognized an average of 38 of the 160 experimental words in the recognition test (24%) and false alarmed to an average of 4 of the 80 unseen words in the test (5%). They were thus able to discriminate between words they had and had not seen as sentence endings, indicating they were paying attention to the experimental stimuli. Recognition accuracy across the experimental conditions is shown in Table 1. These data were subjected to a repeated measures analysis of variance (ANOVA) with two levels of Constraint (word appeared in a strongly constraining or weakly constraining

**Table 1 – Number of words correctly identified as sentence endings as a function of sentential constraint and expectancy**

| | Expected | Unexpected |
|---|---|---|
| Strongly constraining | 5.9 (SD=3.2) | 12.1 (SD=3.9) |
| Weakly constraining | 9.1 (SD=3.6) | 11.1 (SD=4.2) |

sentence frame) and two levels of Expectancy (word was an expected or unexpected ending). There were main effects of both Constraint [$F(1,31)=4.65$; $p=0.04$] and Expectancy [$F(1,31)=77.79$; $p<0.001$], as well as a Constraint by Expectancy interaction [$F(1,31)=19.01$; $p<0.001$]. Overall, participants recognized unexpected words better than expected ones and recognized words from weakly constraining sentences better than those from strongly constraining sentences. Increased constraint was associated with slightly better memory for unexpected completions, but more noticeably was associated with reduced memory for expected completions. Recognition performance was thus consistent with the general finding that distinctive items (i.e., less expected completions, and perhaps particularly those that replaced a highly expected word) are recognized more accurately.

### 2.2. ERPs

Fig. 1 shows the grand average ERPs ($N=32$) to expected and unexpected endings in strongly and weakly constraining sentence frames. Sentence-final words in all conditions elicited the pattern characteristic of ERPs to visual stimuli. These components include, over occipital sites, an initial positivity (P1) peaking at about 80 ms, followed by a negativity (N1) at 160 ms, and a positivity (P2) around 275 ms; and, over frontal sites, a negativity (N1) peaking around 85 ms, followed by a positivity (P2) peaking around 200 ms. These responses were followed by a centro-posterior negativity between about 300 and 500 ms (N400). After the N400, the ERPs in all conditions became more positive again; over frontal sites, visual inspection suggested a greater positivity for the

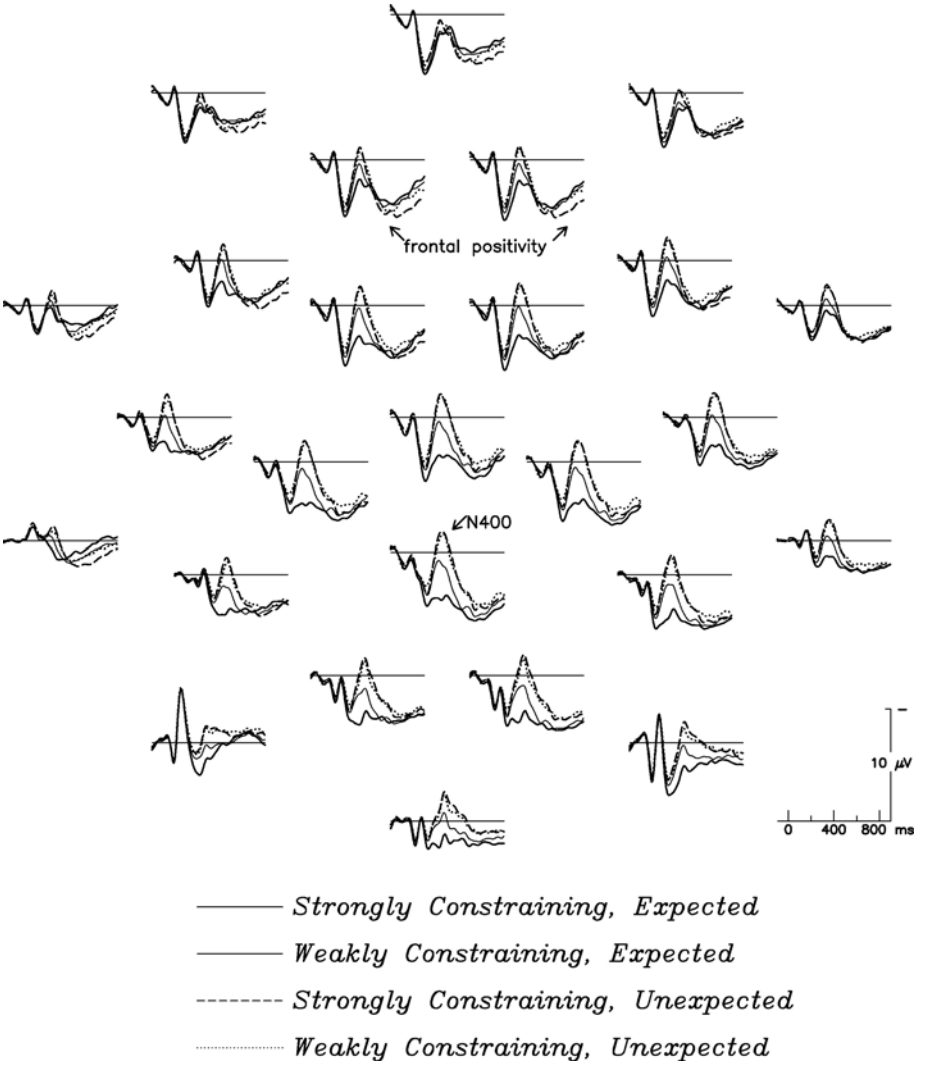


**Fig. 1 – Shown, at all 26 electrode sites, are grand average ERP waveforms for expected and unexpected sentence-final target words in strongly and weakly constraining contexts. The position of the sites on the diagram approximates their position on the head, with the front of the head at top. Negative is plotted up here and in all subsequent figures. Between 300 and 500 ms, a centro-posterior negativity (N400) is smaller for expected than for unexpected targets. The N400 response to expected targets is further graded by constraint, such that responses are smaller to expected words in strongly than in weakly constraining contexts (consistent with their cloze probabilities). Responses to unexpected targets did not differ as a function of constraint in this time window. However, unexpected targets in strongly constraining sentence contexts elicited an enhanced frontal positivity from 500 to 900 ms post-stimulus-onset.**

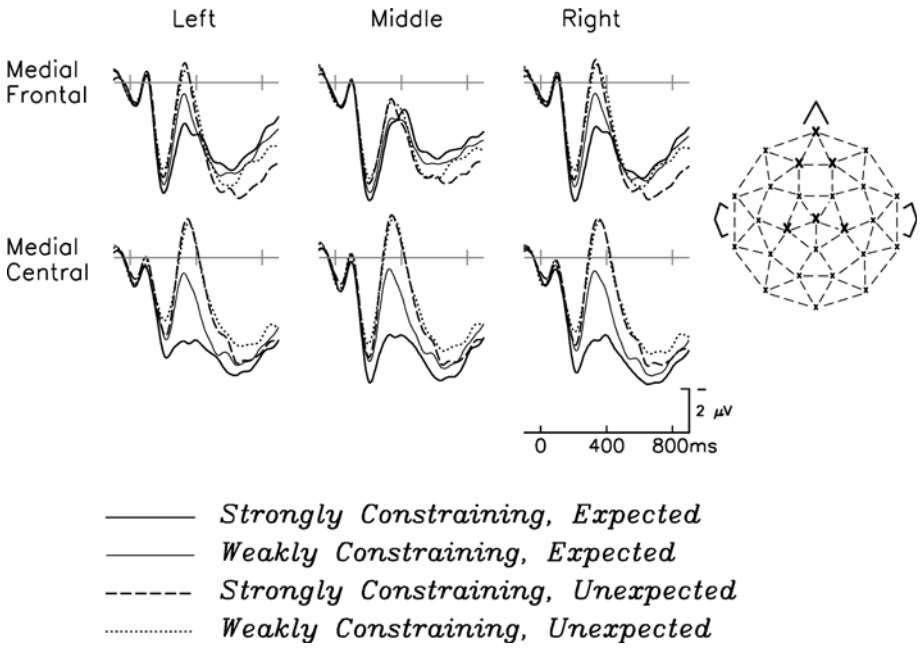


**Fig. 2 – Close-up of the effects at six electrode sites, 3 over the central part of the head (showing the N400 effect pattern) and 3 over the front of the head (showing the frontal positivity to unexpected items in strongly constraining contexts). The small head diagram at right shows the positions (with X's) of the electrode sites.**

unexpected endings in the strongly constraining sentences. Mean amplitudes were measured for both the N400 (300–500 ms) and the later positivity (500–900 ms), and condition effects (constraint and expectancy) were assessed by way of a repeated measures ANOVA across all electrode sites. Interactions with electrode are reported only when of theoretical significance; in those cases, follow-up analyses of distribution were performed using 16 representative scalp sites divided into 2 levels of Hemisphere (right hemisphere and left hemisphere scalp sites), 2 levels of Laterality (lateral and medial scalp sites) and 4 levels of Anteriority (prefrontal, frontal, central, and occipital scalp sites). For each ANOVA the Huynh–Feldt adjustment to the degrees of freedom was applied to correct for violations of sphericity associated with repeated measures. Accordingly, for all F tests with more than 1 df in the numerator, the corrected $p$-value and the Huynh–Feldt epsilon value are reported.

#### 2.2.1. *N400 (300–500 ms)*

The bottom half of Fig. 2 shows the responses to the four critical ending type conditions at a set of medial–central electrodes where N400 effects are typically largest. A repeated measures ANOVA with two levels of Constraint (strongly and weakly constrained), two levels of Expectancy (expected and unexpected), and twenty-six levels of Electrode revealed main effects of both Constraint [$F(1,31)=10.29$, $p=0.003$] and Expectancy [$F(1,31)=95.22$, $p<0.001$] as well as a Constraint by Expectancy interaction [$F(1,31)=17.60$, $p<0.001$]. Consistent with past findings, expected endings elicited N400s of smaller amplitude than unexpected endings; planned comparisons revealed that this was true both within strongly [$F(1,31)=82.89$; $p<0.001$] and within weakly [$F(1,31)=39.26$, $p<0.001$] constraining sentence frames. Also consistent with their cloze probabilities, expected endings elicited smaller N400s within strongly than within weakly constraining sentences (2.7 μV versus 1.5 μV) [$F(1,31)=39.01$, $p<0.001$].

Of critical interest for the present study was whether unexpected endings, matched for cloze probability, would elicit N400s of different amplitudes as a function of constraint. Planned comparisons confirmed the pattern suggested by the Constraint by Expectancy interaction: there was no difference in N400 amplitude to the unexpected endings when these completed strongly as compared with weakly constraining sentences (0.1 versus 0.2 μV, respectively) [$F(1,31)=0.16$, $p=0.695$]. To ensure that we were not missing a small, more temporally-restricted effect, we conducted the same analysis within a 50 ms window (335–385 ms) around the peak of the N400, with the same result [$F(1,31)=0.32$, $p=0.574$][1].

#### 2.2.2. *Late positivity (500–900 ms)*

Visual inspection suggested a condition-related difference following the N400, especially over frontal electrode sites (top half of Fig. 2). A repeated measures ANOVA with two levels of Constraint (strongly and weakly constrained), two levels of Expectancy (expected and unexpected), and twenty-six levels of Electrode revealed a three-way interaction of Constraint, Expectancy, and Electrode [$F(25,775)=3.84$, $p=0.003$, $\varepsilon=0.188$]. A distributional analysis was conducted to follow up on this interaction. A repeated measures ANOVA on two levels of Constraint, two levels of Expectancy,

[1] Analyses restricted to occipital sites, where there was the largest visual difference between the two unexpected conditions in the grand average plot, also failed to yield any significant effects in either the broad or the narrow time window.

two levels of Hemisphere, two levels of Laterality, and four levels of Anteriority revealed a five-way interaction of Constraint × Expectancy × Hemisphere × Laterality × Anteriority [$F(3,93)=3.40$, $p=0.021$, $\varepsilon=1.0$]. Unexpected items elicited enhanced positivity, but only when these items were embedded in strongly constraining sentence contexts; this effect was most prominent at frontal electrode sites and was bigger laterally than medially and bigger over left than right hemisphere scalp sites.

Pairwise comparisons were conducted between 600 and 900 ms[2] (the slightly later window was used to avoid overlap with N400 effects) over the 11 frontal-most channels. Unexpected endings in strongly constraining sentences (SC–UE) were more positive than the same words embedded in weakly constraining sentences (3.6 versus 2.8 μV for WC–UE) [$F(1,31)=4.94$, $p=0.03$]. The other three conditions did not differ from one another (SC–EE versus WC–EE (both 2.9 μV): [$F(1,31)=0.00$, $p=0.95$]; WC-EE versus WC-UE: [$F(1,31)=0.05$, $p=0.82$]).

#### 2.2.3. *Summary*

N400 amplitudes were graded by cloze probability (SC–EE < WC–EE < UE), but showed no additional effects of constraint—in particular, no difference was observed for unexpected items embedded in strongly as compared with weakly constraining sentence frames. However, an effect of constraint did emerge later, between 500 and 900 ms over frontal electrode sites, where there was a selective enhancement of positivity for unexpected items embedded in strongly constraining sentences, relative to the other three conditions (which did not differ).

## 3. Discussion

The goal of this study was to disentangle and examine the influences of expectancy/plausibility (as indexed by cloze probability) and sentential constraint on the unfolding of the brain's response to words embedded in sentence contexts. Prior work indicated intriguing – and still poorly understood – differences between the pattern of constraint effects obtained with behavioral measures and those seen on electrophysiological measures that manifest prior to the behavioral response. Both have shown a robust, graded, facilitative influence of expectancy, with speeded behavioral responses (e.g., lexical decisions) and reduced N400 amplitudes to words with higher cloze probabilities in their contexts (Jordan and Thomas, 2002; Kutas and Hillyard, 1984; Rayner and Well, 1996; Schuberth et al., 1981). Behavioral measures have revealed an additional effect of sentential constraint, such that unexpected words related to an expected completion show facilitation only when embedded in weakly constraining contexts (Schwanenflugel and LaCount, 1988; Schwanenflugel and Shoben, 1985). Constraint effects for these related, unexpected items on the N400 component of the ERP, in contrast, show the opposite pattern: N400 amplitudes are smaller for these items in strongly, as compared with weakly, constraining contexts (Federmeier and Kutas, 1999b). This discrepancy across measures points to the possibility that context has multiple effects on word processing, unfolding at different points in time.

To follow up on these results, in this study strongly and weakly constraining sentence frames were completed with expected and unexpected sentence final words. For expected completions, constraint and cloze probability are confounded, since only strongly constraining sentences yield endings with high cloze probabilities. However, both types of sentences can be completed with plausible but unexpected (low cloze probability) words. In this experiment, unexpected endings with notable feature overlap or lexical association with the expected completion were excluded, to allow us to better examine the brain response associated with revising an initial bias for a given concept and/or lexical item. Careful control was exerted over the unexpected words in the two constraint conditions: these words were identical lexical items and were matched for cloze probability in the two types of contexts. The question, then, is whether sentence-level constraint has any effect on processing when expectancy for a particular word within its context is controlled.

Consistent with prior work (Kutas and Hillyard, 1984), N400 amplitudes were graded by cloze probability. For both constraint conditions, expected words elicited smaller N400s than unexpected words, and expected words in strongly constraining contexts elicited N400s of reduced amplitude relative to expected words in weakly constraining contexts, consistent with the average cloze probability of these items. There was, however, no indication of an additional effect of sentence-level constraint on the amplitude of the N400 response to unexpected words. The cloze probability of the unexpected completions was matched across strongly and weakly constraining sentences, and these items elicited N400 responses of equivalent size.

Federmeier and Kutas (1999b) have reported one circumstance in which N400 amplitudes are not monotonically related to cloze probability (see also Kutas et al., 1984). When unexpected (in this case, also implausible) sentence endings share semantic features with the most expected completion, N400 amplitudes are reduced, and this facilitation of the N400 response is actually greater for violations in strongly constraining than in less constraining sentences, counter to their expectancy/plausibility in the sentences. These results suggest that sentence context information can be used to predict (preactivate) semantic features of likely upcoming words, such that words that share those features obtain at least a temporary processing benefit, even if they are not expected or even plausible in the context. The results of the present experiment show that such facilitation does not extend to unexpected items that contain no notable semantic feature overlap with the anticipated word/concept.

Taken together, the studies suggest that the processes reflected in the N400 are primarily – or exclusively – sensitive to the match between the information gleaned from a sentence context (including predictions about the semantic features of likely upcoming words) and the semantic feature information associated with a word presented in that context. Cloze probability measures provide information about the

[2] The pattern of results for both the general and distributional ANOVAs reported in the prior paragraph were identical for 500–900 and 600–900 ms timewindows.

contextually-induced expectancy for particular words and concepts and are therefore highly correlated with N400 amplitude under most circumstances. However, facilitation can also be observed for low cloze probability words that contain expected (and presumably therefore preactivated) semantic features, and this facilitation is graded by the overall level of expectancy for those features. The present experiment revealed no evidence to suggest that the process(es) indexed by the N400 is/are sensitive to sentential constraint as such. Neither the fact that strongly constraining contexts provide a richer basis for mismatches between a word and its context nor the fact that an unexpected word in a strongly constraining context must compete with (and/or entail the suppression of) a strongly-expected competitor seems to affect processing at this time (300–500 ms post-word onset). N400 amplitude thus seems to reflect the net benefit that contextual information provides for particular words, semantic features, and/or concepts.

However, a very different pattern of effects emerged later in processing. Between 500 and 900 ms, unexpected words completing strongly constraining sentence frames elicited an enhanced, slow positivity over frontal electrode sites. This positivity reflected a clear interaction of expectancy and constraint: it was not observed for unexpected items when these same lexical items completed weakly constraining sentences, and it was not observed in the same strongly constraining contexts for words that were expected. This processing stage thus seems to be sensitive to the greater degree of mismatch between the rich information provided by a strongly constraining sentence and an unrelated (though plausible) unexpected word, leading to the possibility of surprise and/or increased resource demands entailed by the need to override or suppress a strong prediction for a different word or concept. Although further work is needed to clearly link this electrophysiological pattern with results seen in behavioral measures, a likely hypothesis is that this effect could be related to the narrowed scope of behavioral facilitation (on lexical decision times) observed for strongly constraining sentences (Schwanenflugel and LaCount, 1988). If this response is indeed related to prediction, then it should be reduced or absent under conditions where predictive processing is less likely, as has been reported for elderly individuals (Federmeier et al., 2002) and for right-hemisphere-biased processing (Federmeier and Kutas, 1999a); we are currently examining this issue further. A similar frontal positivity (650 to 850 ms) was observed in Spanish–English bilinguals in response to low cloze probability words completing English sentence fragments and idioms (e.g., The driver of the speeding car was given a *citation*. The truth hit me like a ton of *stones*.) as well as to unexpected switches into Spanish (e.g., "multa (ticket)" completing the first sentence above or "ladrillos (bricks)" completing the second) (Moreno et al., 2002). Another perhaps similar effect – increased anterior positivity between 700 and 900 ms – was described by Coulson and Wu (2005) in response to probe words that were unrelated in meaning to a previously presented one-sentence joke (e.g., Everyone had so much fun diving from the tree into the swimming pool that we decided to put in a little platform. CRAZY); when these probe words were lateralized, only those initially projected to the left hemisphere showed this effect. Finally, frontal positivities (sometimes, but not always, accompanied by posterior positivities) in this latency range have also been observed in syntactically anomalous and syntactically ambiguous sentences (Carreiras et al., 2004; Friederici et al., 2001, 2002; Hagoort and Brown, 2000; Kaan and Swaab, 2003); however, the functional similarity between these "frontal P600" effects and the effect reported here is not obvious. If these frontal positivities do reflect related processing, the functional identification of the frontal P600 with syntactic integration difficulty (Friederici et al., 2002) or discourse complexity (Kaan and Swaab, 2003) would need to be reconsidered.

In conclusion, the current study provides clear evidence that context effects on word processing unfold over multiple processing stages that are functionally, temporally, and neurally distinct. The first stage, indexed by the N400, seems to be sensitive only to the match between information contained in or implied by a sentence context and that associated with the word currently being processed. The degree of match (given by cloze probability as well as semantic feature overlap) has a graded effect on N400 amplitudes, presumably reflecting the facilitated processing of the critical word. The fact that a different word might have been predicted or preferred in the context does not seem to impact processing at this time. Instead, such "costs" – that is, those associated with processing an unexpected word in a context that leads to robust expectations for a different item – emerged about 100 ms later in processing, in the form of a slowly-developing frontal positivity. This later stage of processing thus seems to reflect the appreciation of mismatch and/or the allocation of resources necessary to revise a prediction. Electrophysiological measures thus reveal both the benefits and the costs of contextual information for word processing, which are summed in behavioral measures but seem to arise at different time points in the language comprehension stream.

## 4. Experimental procedures

### 4.1. *Participants*

Thirty-two right-handed native speakers of English at the University of California, San Diego participated in the study in exchange for course credit or cash. Sixteen of the participants were women and sixteen were men. The mean age was 20 years, with a range of 18 to 28 years. All participants were right-handed as assessed by the Edinburgh inventory (Oldfield, 1971); 7 reported having left-handed or ambidextrous family members. All participants reported normal vision and none had a history of neurological or psychiatric disorders.

### 4.2. *Materials*

The experimental stimuli consisted of 282 sentence frames, half of which were strongly constraining and half of which were weakly constraining (as determined by cloze probability norming, described below). Each sentence was completed with both its most expected ending (the word

with the highest cloze probability for that sentence) and an unexpected but plausible ending (with a cloze value near 0). The endings of the sentences served as the critical words for the experiment, leading to four conditions: expected endings in strongly constrained sentence frames (SC–EE), expected endings in weakly constrained sentence frames (WC–EE), unexpected endings in strongly constrained sentence frames (SC–UE), and unexpected endings in weakly constrained sentence frames (WC–UE). Table 2 shows examples.

To determine the cloze probability of the endings in their sentence frames, a norming procedure was conducted with native English speakers at the University of California, San Diego (none of whom participated in the main ERP experiment). Sentence frames (368 total) were divided into four lists of 92 each; three of the four lists were completed by 18 participants and one list was completed by 19 participants. In accordance with standard cloze norming procedures, participants were asked to read each sentence frame and to write down the word they "would generally expect to find completing the sentence fragment." In an extension of the standard procedure, the instructions directed participants to give two additional plausible completions. We thus could compute cloze probabilities not only for the best completion of the sentence frame but also for a larger set of "next best" endings. Such information is especially important when trying to control for the fit of unexpected endings across constraint, because with traditional cloze measures the presence of a single, highly preferred ending in strongly constraining sentences means that one necessarily gets less information about what other words/concepts might be afforded by the sentence frames and actively considered by the participants.

**Table 2 – Examples of the stimuli: SC=strongly constraining, WC=weakly constraining**

| | Sentence Frame | Expected | Unexpected |
|---|---|---|---|
| SC | He bought her a pearl necklace for her | birthday | collection |
| WC | He looked worried because he might have broken his | arm | collection |
| SC | There were brightly colored pictures on every | wall | card |
| WC | He was so busy and overwhelmed that he had forgotten to respond to the | question | card |
| SC | The construction worker had developed very powerful arms from unloading bags of cement from the | truck | dock |
| WC | The county decided to tear down the gas station by the | park | dock |
| SC | Sam could not believe her story was | true | published |
| WC | I was impressed by how much he | knew | published |
| SC | It's hard to admit when one is | wrong | scared |
| WC | Keith read a story to his little girl, but after only a short time she started to get | sleepy | scared |
| SC | The children went outside to | play | look |
| WC | Joy was too frightened to | move | look |
| SC | He was cold most of the night and finally got up to get another | blanket | log |
| WC | He fell on the floor after tripping on the | crack | log |
| SC | His skin was red from spending the day at the | beach | farm |
| WC | Ken built his new house on a quiet | hill | farm |

**Table 3 – Cloze probabilities of unexpected completions for both constraint conditions**

| | Overall cloze | Cloze as best completion | Cloze as alternate completion |
|---|---|---|---|
| Strongly constraining | 3.1% | 0.6% | 2.6% |
| Weakly constraining | 3.1% | 1.5% | 1.6% |

In accord with standard cloze norming procedures, "Cloze as best completion" reflects the percentage of people who completed a given sentence fragment with a given item as that they "would generally expect to find completing the sentence fragment". In an extension of standard procedures, participants were also asked to provide two additional, plausible completions. The probability of use of a given item as one of these additional endings for a given sentence frame is reported in the "Cloze as alternate completion" column. "Overall cloze" gives the probability of use of a given item as either of these two types of completions for a given sentence frame.

From the resulting database, we selected 141 strongly constraining sentence frames, for which the best completion had a cloze value of 67% or greater (mean 85.3%; mean use as a "next best" completion 4.9%), and 141 weakly constraining sentence frames, for which the best completion had a cloze value of 42% or lower (mean 26.9%; mean use as a "next best" completion 9.3%). Sentence frames of the two types were matched for length (average of 10 words per sentence in each type) and the two types of expected items were also matched for word frequency (122 and 141 for strongly and weakly constraining sentences, respectively (Francis and Kucera, 1982)) and word length (average 5 letters for both). The experimenters then chose unexpected endings (average word frequency 74; average word length 6), each of which could be paired with both a strongly constraining and a weakly constraining sentence frame. Across constraint, then, lexical properties of the unexpected endings were perfectly controlled[3]. Cloze probabilities for these unexpected completions are shown in Table 3.

Several experimenters judged each of the unexpected endings to be both plausible in its sentence frame and to come from a different semantic category from (and thus share relatively little feature overlap with) the corresponding expected ending for that sentence frame. Mean association strength (as assessed by the Edinburgh Associative Thesaurus; approximately 90% of the experimental stimuli were in the database) between the expected and unexpected endings for each sentence frame was less than 0.005 for both constraint conditions (and did not differ). To assess association between the sentence ending and the words in the sentence frame, we counted the number of sentences containing at least one

[3] In a small set of cases, there were minor changes in the inflection of lexical items (e.g., one singular, one plural) across their uses in the two constraint conditions.

moderate to strong associate (0.2 or greater) of the sentence final word. Overall, there were few words in the sentence contexts associated with the critical words: 4% for SC–EE and 1% for the other three conditions. Mean association strength between the sentence ending and all of the other content words in the sentence frame was less than 0.005 for all four conditions (SC–EE, WC–EE, SC–UE, WC–UE).

Stimuli were divided into two lists, such that each participant saw each sentence frame only once; within each list, half of the frames for each constraint condition were completed by the expected ending and half were completed by the unexpected ending (yoked so that the matched unexpected endings did not appear in the same list). Stimulus characteristics were matched across conditions within each list. The order of sentence frames was randomized once for each list and then presented in the same order to each participant.

### 4.3. *Procedure*

Participants were seated 100 cm in front of a 21″ CRT computer monitor. Each trial began with a warning sign (several pluses on the screen) presented for 500 ms; the blank screen between the warning sign and the first word of the sentence varied randomly from 500 to 1200 ms (to prevent the consistent buildup of anticipatory slow-wave activity). Sentences were then presented word by word in the center of the screen. Each word was presented for 200 ms with an interstimulus interval of 300 ms. A 3-s pause separated each sentence.

Participants were asked to minimize blinks, eye movements, and muscle movement while reading. They were instructed to read the sentences for comprehension while keeping in mind that they would be asked questions about what they had read at the conclusion of the recording session. The recording session began with a short set of practice sentences to acclimate the participants to the task situation. The main experimental session was divided into four blocks of sentences, with participants taking a short rest between each block; recording time was approximately 1 h.

After the recording session ended, participants completed a recognition test. A list of 240 words was selected such that, for each participant, 80 of the words were never seen as sentence-final words during the experiment, and, of the remaining 160 words, 40 sentence-final words came from each experimental condition. Participants were asked to circle all the words that they remembered seeing as a final word of one of the sentences in the experiment.

After the experiment, participants also completed a short set of neuropsychological measures, which included verbal fluency (category and letter), reading span (Daneman and Carpenter, 1980), and author and magazine recognition questionnaires (Stanovich and West, 1989). These data were not used for the analyses presented here and are thus not reported.

### 4.4. *EEG recording and processing*

EEG was recorded from twenty-six geodesically arranged sites on the scalp using tin electrodes embedded in an Electro-cap. The sites are Midline Prefrontal (MiPf), Left and Right Medial Prefrontal (LMPf and RMPf), Lateral Prefrontal (LLPf and RLPf), Medial Frontal (LMFr and RMFr), Mediolateral Frontal (LDFr and RDFr ), Lateral Frontal (LLFr and RLFr), Midline Central (MiCe), Medial Central (LMCe and RMCe), Mediolateral Central (LDCe and RDCe), Midline Parietal (MiPa), Mediolateral Parietal (LDPa and RDPa), Lateral Temporal (LLTe and RLTe), Midline Occipital (MiOc), Medial Occipital (LMOc and RMOc), and Lateral Occipital (LLOc and RLOc); the head icon in Fig. 2 shows the arrangement. These electrodes were referenced online to the left mastoid and later referenced offline to the average of the left and right mastoids. Eye movements were monitored using a bipolar recording of EOG with electrode places on the outer canthus of each eye. Blinks were monitored with an electrode placed over the infraorbital ridge of the left eye, referenced to the left mastoid. Electrode impedances were kept below 5 kΩ and signals were amplified with Grass amplifiers set at a bandpass of 0.01 to 100 Hz. EEG was sampled at 250 Hz and saved on a hard drive.

EEG records were examined and marked for EOG, EMG, or other artifactual contamination. These trials (average 10.5%) were excluded from further analysis, with the exception of trials containing eye blinks for four subjects, which were corrected (Dale, 1994) and added back into the EEG record. ERPs were computed from 100 ms before the onset of critical words to 920 ms after, and averages of artifact-free ERPs were calculated for each type of critical word (SC–EE, WC–EE, SC–UE, WC–UE) after subtraction of the 100 ms pre-stimulus baseline. Measurements were taken after a digital bandpass filter of 0.2 to 20 Hz was applied.

## Acknowledgments

We thank David Groppe for determining lexical association in the stimulus set, Marcus Lauer for assistance with cloze norming, and Matt Rambert for help with stimulus development. Support for this research came from grants AG26308 to K. D. F. and from HD22614 and AG08313 to M. K.

## REFERENCES

Carreiras, M., et al., 2004. Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. Cogn. Brain Res. 20, 98–105.

Coulson, S., Wu, Y.C., 2005. Right hemisphere activation of joke-related information: an event-related brain potential study. J. Cogn. Neurosci. 17, 494–506.

Dale, A.M., 1994. Source Localization and Spatial Discriminant Analysis of Event-Related Potentials: Linear Approaches. Department of Cognitive Science. University of California San Diego, La Jolla, CA.

Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. J. Verbal Learn. Verbal Behav. 19, 450–466.

Ehrlich, S.F., Rayner, K., 1981. Contextual effects on word perception and eye movements during reading. J. Verbal Learn. Verbal Behav. 20, 641–655.

Federmeier, K.D., Kutas, M., 1999a. Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. Cogn. Brain Res. 8, 373–392.

Federmeier, K.D., Kutas, M., 1999b. A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. 41, 469–495.

Federmeier, K.D., et al., 2002. The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. Psychophysiology 39, 133–146.
Fischler, I., Bloom, P.A., 1979. Automatic and attentional processes in the effects of sentence contexts on word recognition. J. Verbal Learn. Verbal Behav. 18, 1–20.
Fischler, I., Bloom, P.A., 1980. Rapid processing of the meaning of sentences. Mem. Cogn. 8, 216–225.
Francis, W.N., Kucera, H., 1982. Frequency Analysis of English Usage, Vol. Houghton Mifflin Company, Boston.
Friederici, A.D., 2001. Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. Cogn. Brain Res. 11, 305–323.
Friederici, A.D., et al., 2002. Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. J. Psycholinguist. Res. 31, 45–63.
Hagoort, P., Brown, C.M., 2000. ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. Neuropsychologia 38, 1531–1549.
Hoeks, J.C.J., et al., 2004. Seeing words in context: the interaction of lexical and sentence level information during reading. Cogn. Brain Res. 19, 59–73.
Jordan, T.R., Thomas, S.M., 2002. In search of perceptual influences of sentence context on word recognition. J. Exper. Psychol., Learn., Mem., Cogn. 28, 34–45.
Kaan, E., Swaab, T.Y., 2003. Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. J. Cogn. Neurosci. 15, 98–110.
Kleiman, G.M., 1980. Sentence frame contexts and lexical decisions: sentence-acceptability and word-relatedness effects. Mem. Cogn. 8, 336–344.
Kutas, M., Federmeier, K.D., 2001. Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn. Sci. 4, 463–470.
Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.
Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307, 161–163.
Kutas, M., Lindamood, T.E., Hillyard, S.A., 1984. Word expectancy and event-related brain potentials during sentence processing. In: Kornblum, S., Requin, J. (Eds.), Preparatory States and Processes. Lawrence Erlbaum Associates, Hillsdale NJ, pp. 217–237.
Moreno, E.M., et al., 2002. Switching languages, switching Palabras (words): an electrophysiological study of code switching. Brain Lang. 80, 188–207.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Rayner, K., Well, A.D., 1996. Effects of contextual constraint on eye movements in reading: a further examination. Psychon. Bull. Rev. 3, 504–509.
Schuberth, R.E., et al., 1981. Effects of stimulus and contextual information on the lexical decision process. Mem. Cogn. 9, 68–77.
Schwanenflugel, P.J., LaCount, K.L., 1988. Semantic relatedness and the scope of facilitation for upcoming words in sentences. J. Exper. Psychol. Learn., Mem., Cogn. 14, 344–354.
Schwanenflugel, P.J., Shoben, E.J., 1985. The influence of sentence constraint on the scope of facilitation for upcoming words. J. Mem. Lang. 24, 232–252.
Schwantes, F.M., 1985. Expectancy, integration, and interactional processes: age differences in the nature of words affected by sentence context. J. Exp. Child Psychol. 39, 212–229.
Stanovich, K.E., West, R.F., 1979. Mechanisms of sentence context effects in reading: automatic activation and conscious attention. Mem. Cogn. 7, 77–85.
Stanovich, K.E., West, R.F., 1983. On priming by a sentence context. J. Exp. Psychol. Gen. 112, 1–36.
Stanovich, K.E., West, R.F., 1989. Exposure to print and orthographic processing. Read. Res. Q. 24, 402–433.
Taylor, W.L., 1953. "Cloze procedure": a new tool for measuring readability. Journal. Q. 30, 415–433.