fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity

Ellen F. Lau* and Anna Namyst

University of Maryland, Department of Linguistics, College Park, Maryland, United States of America; ellenlau@umd.edu and amnamyst@umd.edu

* Corresponding author
Abstract

Previous electrophysiological work argues that predictability and semantic incongruity rapidly impact comprehension, as indicated by modulation of the N400 component between ~300-500ms. An ongoing question is whether effects of predictability in fact reflect pre-activation in long-term memory as opposed to modulating the kind of integration processes triggered by incongruity. We used a functional localizer approach in fMRI to compare the impact of predictability and incongruity in comprehension of adjective-noun phrases. We found that predictability impacted activity in left posterior middle temporal gyrus (pMTG), while incongruity impacted activity in left middle frontal gyrus. Together with parallel data from ERP, these data support the hypothesis that left pMTG activity is a key contributor to N400 effects of predictability and that the mechanism driving this effect is reduced activation of stored lexical representations. We tentatively suggest that the left middle frontal gyrus may play a role in reanalysis when incongruity is encountered.

Keywords: N400, prediction, pre-activation, congruity, MTG, ERP, fMRI
Introduction

Much psycholinguistic research suggests that during language comprehension readers and listeners routinely predict the content, and perhaps the form, of the upcoming input, and that predicted input is correspondingly processed more easily (Federmeier & Kutas, 1999; Altmann & Kamide, 1999; Wicha, Moreno, & Kutas, 2004; and many others). However, from the early days of this work, a central challenge for studying prediction in language has been dissociating effects of prediction from other processes. In particular, predictability manipulations were often confounded with congruity, such that in one condition the critical word was both highly predictable and semantically congruous, and in the other condition the critical word was both unpredictable and semantically incongruous (e.g. He spread the warm bread with butter/socks; Kutas & Hillyard, 1980). If neural activity is greater in the latter condition than the former, the difference could be attributed either to facilitated lexical or conceptual access, which would reduce activity in the predictable condition, or to increased effort towards computing the sentence- or discourse-level meaning, which would increase activity in the incongruous condition.

This ambiguity has resulted in ongoing debate about the functional interpretation of the well-known N400 effect in ERP, and in uncertainty about the extent to which the N400 effect can be used to study predictive processes in language comprehension. The N400 response refers to a broad negative deflection in the ERP response to words and other meaningful stimuli which peaks around 400ms post-stimulus onset. Kutas and Hillyard (1984) and hundreds of studies since have observed that the amplitude of the N400 is smaller for words that are predicted by the context. This could be taken to indicate that the N400 indexes activation of lexical and conceptual networks, which can be activated less broadly when a predictive context allows the comprehender to narrow in on the correct candidate (Federmeier & Kutas, 1999). However, much of this data could also be taken to indicate that the N400 indexes activity associated with integrating incoming input into a larger sentence-level or discourse-level meaning representation, and that the unpredictable sentence stimuli tend to involve scenarios that are less common/plausible and thus require more integration work to construct the meaning (Van Berkum et al., 2005).

Recent ERP work from our group and others provides new evidence that N400 amplitude is modulated by predictive facilitation. Deconfounding predictability and semantic congruity in sentence contexts has been highly challenging because even large corpora are too sparse to provide precise estimates of the likelihood of particular sentence continuations, and human sentence continuation data provide precise estimates for highly predictable endings but less precise estimates for unpredictable endings. Lau et al. (2016) introduced a two-word adjective-noun paradigm, which allows better-supported estimates of predictability from corpus data. We compared the separate effects of predictability (runny nose vs. dainty nose) and semantic incongruity (yellow bag vs. innocent bag), motivated by the logic that if N400 effects of predictability were driven by semantic integration difficulty, large N400 differences should also be observed in the congruity comparison. However, when predictability was controlled, we found that congruity elicited only small, barely reliable N400 differences, while predictability drove a very large N400 reduction. These results suggest that predictability can modulate the N400
response independent of integration difficulty, by hypothesis through facilitated lexical and/or conceptual activation. Integration difficulty may separately and independently impact the ERP during this time-window, as suggested by recent sentence-level studies that more precisely quantify contextual predictability and show differences in latency and/or amplitude for N400 effects of congruity and predictability (DeLong et al., 2014; Brothers et al., 2015).

The goal of the current study was to use the adjective-noun paradigm from Lau et al. (2016) to provide converging evidence from fMRI for the hypothesis that N400 effects of contextual predictability indeed reflect pre-activation of stored lexical and/or conceptual representations, as opposed to being solely driven by differences post-lexical integration difficulty. Just as in our ERP study, participants were presented with two-word adjective-noun phrases and attended to the materials in anticipation of a post-experiment memory recall test. If N400 effects of predictability reflect mechanisms that support basic access of stored lexical representations, we would expect to see effects of predictability in posterior temporal areas that respond to manipulations of lexical content. Alternatively, if N400 effects of predictability reflect mechanisms that integrate stored lexical or conceptual representations, we would expect to see effects of predictability in frontal or anterior temporal regions that respond to functional localizer manipulations of sentential structure.

Which brain regions generate N400 context effects has also been a matter of some debate. As reviewed by Lau et al. (2008; see also Van Petten & Luka, 2006), fMRI effects in mid-posterior temporal cortex appear to best track semantic priming manipulations that generate N400 effects, and MEG work using sentence contexts is also consistent with this. However, the results of fMRI studies using congruity/predictability manipulations in sentences have been much more variable. This likely reflects the fact that the temporally ‘sluggish’ fMRI response sums the response to most or all words in the sentence, resulting in increased variability that can mask the differential response to the single word that is manipulated. An additional virtue of the adjective-noun paradigm used here is that only the single context word and the target word contribute to the response.

Results

Overall accuracy on the post-run memory recall task was 69% for the predictability manipulation and 63% for the congruity manipulation.

Lexical/phrasal localizer

The results of the lexical/phrasal localizer were largely consistent with previous work. In the lexical contrast (scrambled sentences vs. consonant strings) we observed significant clusters of increased activity for scrambled sentences relative to consonant strings in left posterior MTG, left posterior STG, and left middle frontal gyrus. In the phrasal contrast (sentences vs. scrambled sentences) we observed a significant cluster of increased activity for sentences in left anterior STS (see table in Supplementary Materials for exact coordinates and extents). We evaluated effects of lexical predictability and semantic incongruity manipulations in these four ROIs.
We note that previous work using this kind of localizer has identified additional regions in similar lexical and phrasal contrasts, and indeed at a less conservative threshold we also observed phrasal effects in right hippocampus, right posterior STG, left IFG, and right anterior STS, and lexical effects in left anterior STG and right posterior STG, among other regions.

*Predictability effects*

The predictability manipulation had a reliable impact only in the left posterior MTG region of interest identified in the lexical localizer, such that highly predictable items demonstrated less activity than less predictable items (Figure 1). In the omnibus ANOVA we observed a significant interaction between condition and region \((F(1,23) = 2.7, p < .05)\), and in follow-up ANOVAs in each of the four regions, pMTG was the only region to demonstrate a significant effect of condition \((F(1,23) = 3.1, p = .05)\). Follow-up pairwise comparisons demonstrated a significant difference between high predictability and low predictability, low constraint conditions \((F(1,23) = 5.1, p < .05)\) and between high predictability and low predictability, high constraint conditions \((F(1,23) = 4.4, p < .05)\). In our previous ERP study using the same materials and paradigm (Lau et al., 2016) we observed the same pattern in N400 amplitudes to the critical noun, as illustrated in Figure 3.

*Congruity effects*

The semantic congruity manipulation had a reliable impact only in the left middle frontal region of interest identified in the lexical localizer, such that incongruous items demonstrated more activity than congruous items. In the omnibus ANOVA we observed a significant interaction between condition and region \((F(1,23) = 5.3, p < .05)\), and in follow-up ANOVAs in each of the four regions, the precentral gyrus was the only region to demonstrate a significant effect of condition \((F(1,23) = 5.1, p < .05)\).

**Discussion**

The current study aimed to bring fMRI evidence to bear on the hypothesis that N400 effects of contextual predictability reflect lexical or conceptual pre-activation, as opposed to being solely driven by differences in post-lexical integration difficulty. In the short two-word materials used here, a manipulation of predictability robustly modulates N400 amplitude in ERP and a manipulation of semantic congruity does not (Lau et al., 2016). The corresponding fMRI results show that activity in a region of left posterior MTG picked out by a ‘lexical’ localizer tracks the N400 predictability pattern, and at the same time does not differentiate semantically congruous and incongruous items. Activity in ‘phrasal’ regions of left anterior STS and left IFG were modulated by neither predictability nor congruity. However, we found that activity in a ‘lexical’ region of left middle frontal gyrus was increased in response to incongruity.

Like our prior ERP study (Lau et al. 2016), these data argue against a view in which N400 predictability effects can simply be reduced to semantic integration difficulty. If semantic integration does occur when processing such short phrases, then a large
Difference in integration difficulty would be observed for incongruous phrases such as *innocent bag*. However, Lau et al. (2016) observed very little difference in the ERP response to incongruous phrases when compared to unpredictable congruous phrases, and in the current study we find that semantic congruity modulates a different brain region than lexical predictability. These data are consistent with the view that N400 effects of predictability reflect the pre-activation of stored lexical and/or conceptual representations in temporal cortex (Lau et al., 2008). Although this hypothesis could be implemented in various ways, one idea is that when a word/concept is not predicted by the context, a relatively large neighborhood of the lexical and conceptual network is activated by the bottom-up input and remains active until a given candidate ‘wins’ the competition for selection. But when a word is pre-activated by the context, it can be selected almost immediately after the bottom-up input is encountered, without requiring activation of broader regions of the network, resulting in a brief but massive reduction in neural activity.

**N400 effects of lexical and conceptual prediction: a posterior-to-anterior gradient**

As reviewed in the Introduction, previous work has most frequently localized N400 effects to temporal cortex, but the position of those effects has varied from anterior to posterior temporal cortex. In a review of most extant semantic priming fMRI studies at that time (Lau et al., 2008) we found that the predominant location of effects was in posterior middle temporal gyrus. However, in subsequent studies using semantic priming in a within-subjects EEG-MEG/fMRI design, we found that N400 effects localized instead to a more anterior position in STG/STS (Lau et al., 2013a; 2014). We suggested that the explanation lay in the emphasis of the experiments and tasks. The majority of the early semantic priming studies used lexical decision tasks, so that the emphasis would be on identification of a specific lexical item (e.g. Gold et al., 2006). This could result in a higher baseline activation of the lexical network to each word, and thus when an item was predicted by the context, the greatest gains (in reduced neural activity) would be observed in the lexical network in posterior temporal cortex. On the other hand, the experiments in Lau et al. (2013a, 2014) used a conceptual probe task (*Press the button when you see an animal word*). This could result in a higher baseline activation of the conceptual network to each word, and thus when an item was predicted by the context, the greatest gains would be observed in the conceptual network in anterior temporal cortex (see Lau et al., 2014, for further discussion).

In the current study we used a functional ROI approach to focus specifically on regions associated with lexical processing on the one hand and phrasal processing on the other. We observed reliable effects of predictability in our posterior temporal ROI, but found no reliable effects of predictability in our anterior temporal phrasal ROI. We believe that our observation of posterior temporal reductions in the conditions that elicit N400 effects is consistent with the task emphasis explanation described above. In the current study, participants performed a memory probe task at the end of the block requiring them to distinguish complete adjective-noun phrases they had seen from permuted catch trials. This might have resulted in greater attention to the particular words that were observed on each trial, and corresponding facilitation in a posterior temporal lexical network. Similar logic might explain the absence of facilitation in the anterior temporal ROI.
through the reduced emphasis on conceptual processing in this task (although it is worth noting that the anterior MTG/STS region picked out by the localizer is slightly inferior to the anterior STG region that showed priming effects in our prior work).

If the N400 response indeed reflects the summed contribution of both lexical and conceptual networks, and if the emphasis of different paradigms modulates their relative contributions, then we might expect to see corresponding differences in the scalp distribution of the N400 effect. If systematic, such differences could be a useful diagnostic of the functional interpretation of an N400 effect in a particular study. Although we can observe slight differences in our own previous ERP work with these paradigms (e.g. Lau et al. 2013b; 2016), given the widespread distribution of the standard N400 effect, they are subtle at best. However, such differences could be usefully examined in future EEG/MEG work using within-subjects comparisons.

**Why are hemodynamic correlates of N400 prediction effects relatively weak?**

Although the ROI analysis revealed reliable effects of predictability in left posterior MTG/STS that corresponded to patterns of N400 amplitude, no effects of predictability were robust enough to survive the whole-brain analysis even at a relatively high threshold. This is fairly consistent with prior hemodynamic work: in our own previous fMRI semantic priming work we observed relatively small and circumscribed effects (Lau et al., 2013a; 2014; Weber et al., 2016), and this is also true in many previous studies (e.g. Gold et al., 2006). In fMRI experiments that used sentence-level manipulations that typically lead to N400 effects, the results were even more variable, as reviewed in Lau et al. (2008). However, against the ERP literature this might seem surprising: compared to other ERP responses N400 predictability effects are relatively large in magnitude, long in duration (several hundred milliseconds) and can often be clearly seen in datasets from individual participants.

Why is such a strong and reliable effect in the ERP measure so much less robust in fMRI? The question has not been resolved, but there are many possibilities. One is that it results from different properties of the averaging in ERP and fMRI. Some ERP differences can arise from increased synchronization of oscillations rather than a net difference in neural activity; if this were the mechanism driving N400 effects, it would be unlikely to lead to a net change in hemodynamic activity (see Luka & Van Petten, 2006 for discussion). Because the N400 response is smeared across the scalp in EEG, slightly different distributions from individual to individual still sum together to give a net effect, but if the response is actually generated in somewhat different regions of temporal cortex from person to person, then spatially more precise fMRI would not end up summing these effects across individual participants (see Fedorenko et al. 2010 for broader discussion of this issue). Another possibility is that even responses that last a few hundred milliseconds in ERP are too short to generate reliable differences in measurable hemodynamic activity, and that standard fMRI effects of language manipulations tend to reflect differential activity across many seconds. We believe that these are important questions for future research.

**Semantic incongruity effects in dorsolateral prefrontal cortex**
Although we did not predict that lexicality or semantic incongruity would modulate activity in left middle frontal / precentral gyrus, it is also not wholly surprising. Dorsolateral prefrontal cortex has been implicated in a range of high-level executive functions, from working memory to monitoring to decision-making. In the scrambled sentence vs. consonant string localizer contrast, this activity may reflect an effort to relate the unstructured words. Similarly, incongruous phrases may have triggered the engagement of executive routines, whether to examine if an analysis mistake was made, to search for an alternative word meaning, or to generate a metaphorical interpretation. Increased activity in this region for incongruous sentences has been observed in previous fMRI studies, although not consistently (e.g. Kuperberg et al., 2003). An interesting question for future investigation is whether this differential dorsolateral prefrontal activity is the source of the minor differences in N400 amplitude that we observed for the same congruity manipulation in ERP.

Finally, although isolated two-word phrases provide critical methodological advantages for determining the hemodynamic correlates to brief ERP effects, results from the comprehension of isolated two-word phrases can in no way be taken to provide a full picture of the processes involved in comprehension of sentences and discourses. Many integration processes are not engaged by two-word phrases, and recent fMRI work investigating lexical surprisal (an information-theoretic measure of expectedness) in more naturalistic discourses in a much broader set of regions (Willems et al., 2016; Henderson et al., 2016; Brennan et al. 2016). Similarly, our functional localizer, which involved passive reading with no explicit task, is unlikely to provide a comprehensive picture of the brain network that supports more naturalistic sentence comprehension. Previous work using sentences has reported modulation of left IFG by semantic congruity (Hagoort et al., 2004), but the absence of robust IFG modulation in our functional localizer prevented us from evaluating semantic congruity effects in isolated phrases. Interestingly, an exploratory analysis using a BA45/47 cluster derived from the localizer at a higher threshold (p = .003) resulted in a marginally significant congruity effect in the later part of the hemodynamic response (8-12s); given the greater functional-anatomical variability associated with IFG (Amunts et al., 1999), it is likely that individual functional localizer data will be needed to better evaluate this response. However, we believe the key point going forward is that simpler manipulations targeting a subset of processes make it easier to relate ERP and fMRI effects.

Conclusion

We have presented the results of an fMRI study that replicated a recent ERP manipulation of predictability and congruity. This multimodal approach allowed us to provide novel support for the hypothesis that N400 effects of contextual predictability do in fact reflect the impact of predictive context on activation of lexical or conceptual representations, and not only later supra-lexical processes such as semantic integration. Although the current contribution is modest, we hope this type of multimodal ERP-fMRI mapping work will ultimately lay the foundation for a more integrated cognitive neuroscience of language, where electrophysiological results could be roughly translated into cortical space based on prior knowledge rather than a time-consuming and error-prone source localization procedure.
Methods

Participants

Participants were 24 right-handed (Oldfield, 1971) University of Maryland students (15 female; mean age 22.6, range 19-29) who participated in the study for monetary compensation and gave informed consent in accordance with the Institutional Review Board of the University of Maryland. One additional participant was excluded from analysis for excessive movement.

Lexical/phrasal Localizer

The lexical/phrasal localizer was composed of three conditions: sentences, word lists, and consonant string lists. All stimuli were nine words long. Forty sentences with varied but relatively simple structures were selected (20 from Rogalsky and Hickok, 2008) and divided across two presentation lists. Scrambled sentences for each list were created by scrambling words across all the sentence items from the other list. The 20 nonword stimuli in each list consisted of a sequence of random consonant strings, where the length of each consonant string was matched to the length of the words in the sentences. Each participant saw only one list. All stimuli are available in Supplementary Materials.

Each trial consisted of 9 words (300ms on, 100ms off), preceded by a fixation cross (200ms on, 200ms off), for a total time of 4s per trial. 20 trials were randomly intermixed from each of 3 conditions: sentences, scrambled sentences, and consonant strings. In order to optimize deconvolution of the fMRI signal, the 60 4-s trials were intermixed with 80 seconds of rest (fixation) using the optseq algorithm (http://surfer.nmr.mgh.harvard.edu/optseq), and 10 seconds of rest was added to the end of the run for a total run-time of 330 seconds.

Predictability and congruity manipulations

Predictability and congruity manipulations were designed as sub-experiments conducted in two separate runs, following the design of the initial two ERP experiments in Lau et al. (2016). Experiment 3 in Lau et al. (2016) demonstrated that the ERP effects of predictability and congruity were relatively unchanged when the materials were fully intermixed.

The predictability manipulation consisted of three conditions: predictable (runny nose), unpredictable, low-constraint (dainty nose), and unpredictable, high-constraint (runny yogurt). Lau et al. (2016) collected ERPs from all three conditions in their Experiment 1, but focused their report on the contrast between the first two conditions only and its relation to effects of congruity. We included the additional high-constraint condition here in order to evaluate any BOLD responses associated with inhibiting the originally predicted noun (note that much prior work suggests that any such inhibition does not impact the N400 response; e.g. Federmeier et al. 2007). Lau et al. (2016) describes materials creation in greater detail. We derived predictability counts from the Corpus of Contemporary American English (COCA; Davies, 2009). For predictable items,
p(noun|adjective) was larger than .5, mean .65; for unpredictable items, p(noun|adjective) was less than .02. The low-constraint adjectives were selected such that there existed no noun for which p(noun|adjective) was larger than .15. 120 3-item sets were created and distributed across two lists in a Latin Square design, such that each participant saw exactly one version of each item (no words were repeated within a given participant), and 40 items from each condition.

The congruity manipulation consisted of two conditions: congruous, unpredictable (yellow bag) and incongruous, unpredictable (innocent bag). All items had a p(noun|adjective) less than .005, and all adjectives were relatively unconstraining. Eighty 2-item sets were created and distributed across two lists in a Latin Square design. Each participant saw 40 items from each condition and exactly one version of each item (no words were repeated within a given participant).

In each trial, the adjective was presented for 500ms, followed by a 100ms blank screen, and then the noun was presented for 900ms, followed by at least 100ms of fixation, for a total time of 2s per trial. In the predictability block, the 120 2s trials were intermixed with 240s of rest, and 10s of rest was added to the end of the run to fully capture the responses to the last trials, for a total run time of 490s. In the congruity block, the 80 2s trials in were intermixed with 160s of rest, and 10s of rest was added to the end of the run, for a total run time of 330s.

In order to maintain attention to the materials, participants were asked to complete a memory recall test after each run. This test consisted of 20 bigrams, of which 10 had appeared in the preceding block and 10 were mismatched adjective-noun pairs from the stimulus set. Participants were asked to make a binary button-press decision to indicate whether they had seen the pair in the previous run or not.

Analysis

AFNI software (Cox, 1996) was used to analyze the functional MRI data. After slice-time correction, functional images were motion corrected to the second time point of the predictability manipulation run using a 6-parameter rigid-body transformation (Cox & Jesmanowicz, 1999), aligned to the anatomical image, and warped to Talairach space. Functional images were spatially smoothed using a Gaussian 6 mm FWHM kernel. For the predictability and congruity manipulations, first-level General Linear Model (GLM) analyses were carried out on the functional images with a finite impulse response (FIR) model that gave estimates of the hemodynamic response at each TR (every 2s) between 0-14s using AFNI’s 3dDeconvolve function with the ‘tentzero’ parameter, which sets activity at the first and last TRs to zero. Activity estimates were computed for each condition at each voxel in each participant, and the estimated values at 2, 4, and 6s post-stimulus onset were summed to create an estimate for each condition and each voxel to carry forward to the second-level analysis (Ashby, 2011). We focused on the first three of the six time points estimated for the BOLD response because our goal was to measure the fMRI response that corresponds to the relatively rapid and short-lived N400 response in ERP. Group-level analyses were conducted on these summed activity estimates using AFNI’s 3dANOVA2 function. We used the canonical HRF approach for analysis of the
localizer because these stimuli were longer in duration (4s) and the activity of interest was not time-locked to a particular point in the stimulus.

Functional ROIs for evaluating the contrasts of interest were extracted from the lexical contrast (scrambled sentences > consonant strings) and phrasal contrast (sentences > scrambled sentences) using an uncorrected threshold of \( p < .0005 \) and a cluster size threshold of 20 contiguous voxels.

Acknowledgments

We thank Allison Fogel, Tania Delgado, Caitlin Richter, Jennifer Stark, Wang Zhan, and William Matchin for assistance with materials creation, data collection, and data analysis.

Figure 1. Estimated hemodynamic timecourses for the predictability manipulation in regions of interest identified by the lexical and phrasal localizers.
Figure 2. Estimated hemodynamic timecourses for the congruity manipulation in regions of interest identified by the lexical and phrasal localizers.

Figure 3. Side-by-side comparison of the effect of the predictability manipulation on the estimated hemodynamic response (across 0-6s) in the posterior MTG ROI in the current study and in the ERP study reported by Lau et al. (2016).
References


