Online Expectations for Verbal Arguments Conditional on Event Knowledge

Klinton Bicknell[1] **(kbicknell@ucsd.edu), Jeffrey L. Elman**[2] **(jelman@ucsd.edu),**
Mary Hare[3] **(mlhare@bgsu.edu), Ken McRae**[4] **(kenm@uwo.ca), and Marta Kutas**[2] **(mkutas@ucsd.edu)**

[1] Dept of Linguistics, University of California, San Diego; [2]Department of Cognitive Science, University of California, San Diego; [3]Department of Psychology, Bowling Green State University; [4]Department of Psychology, University of Western Ontario

Abstract

This paper provides support for the hypothesis that comprehenders form online expectations for upcoming verbal arguments using their knowledge of typical events. We test this hypothesis in a self-paced reading experiment and an experiment measuring event-related brain potentials. In both experiments, we use materials in which the likelihood of the verbal patient depends on event knowledge about the particular combination of agent and verb earlier in the sentence. By manipulating the agent for a given verb, we show that comprehenders experience more processing difficulty in sentences where the patient is less likely. Norming studies and a priming experiment provide evidence that this result is unlikely to have arisen from direct linguistic associations between patient and agent, suggesting that comprehenders use their event knowledge to form expectations.

Keywords: event knowledge; expectancy generation; thematic fit; self-paced reading; event-related potentials;

Introduction

Incremental language comprehension is a complex task – to understand an utterance, a comprehender has to recover structured meaning from a serial input rampant with local ambiguity. How do we do this? One strategy is to form expectations about upcoming material, updating them constantly as new words arrive (Elman, 1990). A number of sources of information have been shown to contribute to such expectations. Much of this information is syntactic, such as verb subcategorization biases (Trueswell, Tanenhaus, & Kello, 1993), but it was also noted that probabilistic semantic knowledge plays a significant role as well (Garnsey, Pearlmutter, Myers, & Lotocky, 1997). Here we focus on thematic fit, the likelihood of a noun to fill a particular argument slot of a verb, as one crucial source of semantic information.

In particular, we investigate the factors that comprehenders use to assess thematic fit online. In earlier work, McRae, Spivey-Knowlton, and Tanenhaus (1998) showed that a noun phrase's thematic fit to a given role is computed relative to specific verbs, not just verb-generally. In this paper, we go a step further and argue, based on evidence from self-paced reading and event-related potential (ERP) studies, that online estimates of thematic fit depend not on the verb alone, but on the relevant properties of the situation being described. If this is correct, it suggests that one important source of semantic knowledge in comprehension is generalized knowledge about typical events.

Much of the recent evidence for the use of event knowledge comes from priming studies, which have shown that verbs facilitate their typical agents, patients, instruments, and locations within 250 ms (Ferretti, McRae, & Hatherell, 2001; Ferretti, Kutas, & McRae, 2007) and that verbs themselves

are facilitated by their typical agents, patients, and instruments with a similar latency (McRae, Hare, Elman, & Ferretti, 2005). While these studies demonstrate the immediacy with which event knowledge appears to be automatically activated, the event knowledge which they show evidence for is limited to relations between verbs and their typical filler nouns. It would be even stronger evidence that comprehenders are using the wealth of information present in event knowledge to show that they make use of more complex relations. This is the goal of the present research.

Thus, the work presented in this paper makes use of complex dependencies between event participants that cannot be reduced to simple associations. Specifically, we use materials that manipulate the thematic fit between the direct object noun and the verb's patient (PA) role by changing the agent (AG) of the event being described. To create a counterbalanced design, we use direct objects that are likely patients for the verb in the event evoked with one agent but unlikely in the situation with the other, as in (1–2).

(1) a. The **journalist**$_{AG}$ checked the **spelling**$_{PA}$ of his latest report to the editor. (Congruent)

  b. The **mechanic**$_{AG}$ checked the **spelling**$_{PA}$ of his latest report about the engine. (Incongruent)

(2) a. The **mechanic**$_{AG}$ checked the **brakes**$_{PA}$ on the car he had gotten earlier that day. (Congruent)

  b. The **journalist**$_{AG}$ checked the **brakes**$_{PA}$ on the car he had rented earlier that day. (Incongruent)

We present data supporting the hypothesis that comprehenders make immediate use of complex dependencies encoded in event knowledge by showing that the patient in the Incongruent condition (1b,2b) induces a slow-down in a self-paced reading experiment and increases the amplitude of an immediate N400 response in an ERP study.

It is important to note that, because the verb is held constant within items, this pattern of results might, in principle, also be predicted by direct associations, if strong associations existed between the agents and patients in the Congruent condition. To rule out this explanation, we ensured that our agents and patients were not normatively associated in a norming study and explicitly tested for direct facilitation in a priming experiment.

The plan for this paper is as follows. We first describe two norming studies, the first to ensure that the materials strongly reflect comprehenders' event knowledge, and the second to measure direct associations between agents and patients. We

then present the results of a priming experiment testing for facilitation between agent-patient pairs in the Congruent condition. Next, we demonstrate that our manipulation is successful with a self-paced reading experiment. Finally, we further investigate the time course of this effect with ERP.

Norming Experiments

Norming Experiment 1: Patient Generation

The first norming experiment was conducted to directly probe people's intuitions about event knowledge. After candidate verbs and agents were selected, participants were asked to produce the patients that were likely for each combination of agent and verb in a questionnaire. A weighted sum of the responses was calculated and used as a basis for selecting patients to use in the self-paced reading and ERP experiments.

Participants Forty-seven students from UC-San Diego participated for course credit.

Materials and Procedure We began by selecting 50 verbs with 2 agent nouns per verb, such that the agent nouns altered the scenario described by the agent-verb combination, and thus the distribution of likely patients to the largest extent possible, in the judgment of the experimenters. We administered a web-based questionnaire that contained questions about these materials such as 'Who or what does a mechanic typically check?' and allowed for up to 5 answers to each question. Each participant saw only 1 question per verb.

Results Responses were weighted according to their position in the five answer slots per question and were then tallied to produce a ranked list of typical patients for each agent-verb pair. As expected, the responses indicate that the distribution of likely patients for a given event depends in part on the agent. In many cases, however, it does seem likely that there are direct links between the agents and their top patients. For example, the top response by to 'Who or what does a librarian typically arrange?' was *books*, which is obviously very closely associated to *librarian* (although this is due, of course, to people's knowledge of what libraries tend to do.) Since it would confound our materials to have such direct links between the agents and patients, we performed the second norming study to measure direct associations between our agents and the patients gathered in this study.

Norming Experiment 2: Word Association

We conducted single-response association norms using our agent nouns as stimuli and used the results to ensure that our agents and patients were not normatively associated.

Participants Twenty-two students from UC-San Diego participated for course credit. All were native speakers of English.

Materials and procedure A web-based questionnaire was constructed that displayed to each participant the agent nouns from Norming Experiment 1, with one text input box following each noun. Participants were asked to 'give the first word that comes to mind that is meaningfully related or strongly associated to' each presented word.

Results We used the results of the word association task to eliminate any patient from our potential materials that was normatively associated to its agent in the Congruent condition. We eliminated all patients from our potential materials that were provided by one or more participants as a response to the corresponding agent in the Congruent condition. Thus, we implemented an extremely strict criterion.

Of the remaining patients, two were then chosen for each verb based on the results of the first norming study. We selected patients for each agent-verb pair that had a high score from Norming Experiment 1 but were not given in response to that agent by any participant in Norming Experiment 2. In some cases where this was difficult, we used synonyms of the high-scoring patients. Finally, we removed 6 verbs for which the agents did not seem to reliably change the types of patient responses. This procedure resulted in 44 verbs, each of which had two agents and two patients associated with it.

Lexical Decision Experiment

In order to verify that our agents did not independently facilitate their corresponding patients in the Congruent condition, we performed a lexical decision experiment to probe for priming. Since the goal of this experiment was to confirm that no facilitation obtained between the agents and patients in our pairs, we also included a set of materials that had already shown robust facilitation in a lexical decision task from McRae and Boisvert's (1998) first experiment. We thus expected to see robust facilitation on the McRae and Boisvert items, and no facilitation on our items.

Methods

Participants Forty-four students from UC-San Diego participated for course credit. All were native speakers of English and had normal or corrected-to-normal visual acuity.

Materials The materials were created by using the 88 agent-patient pairs and 32 semantically similar pairs from McRae and Boisvert (e.g., *coconut* $\sim$ *pineapple*) for a total of 120 related pairs.

Two lists were created, and each list contained half the items in each of these two sets in the related condition, and the other half in an unrelated condition formed by shuffling the target words. Each list thus had 44 (potentially) related agent-patient pairs, 16 related McRae and Boisvert items, 44 unrelated agent-patient pairs, and 16 unrelated McRae and Boisvert pairs. This made for a total of 60 pairs in the related condition, and 60 in the unrelated condition per list, so 120 unrelated pairs were added to have a relatedness proportion of 0.25 (if the agent-patient pairs were related.) Finally, 240 pairs with a nonword target were added so that there were the same number of trials in which the target was a nonword as those in which it was a word. Two pseudo-randomizations of each list were created, such that trials that included the test items (including both the agent-patient pairs and McRae

Table 1: Percent errors and decision latency (in milliseconds) by condition for each stimuli set, Lexical Decision Expt.

Dependent measure	McRae and Boisvert		Agent-patient	
	Mean	SE	Mean	SE
Percent errors				
Unrelated	11.2	1.4	5.1	0.8
Related	9.5	1.5	3.6	0.5
Facilitation	1.7		1.5	
Decision latency				
Unrelated	549	11.2	510	9.0
Related	531	12.4	510	9.0
Facilitation	18		0	

and Boisvert items) were always separated by at least one trial. Finally, in each pseudo-randomization, the McRae and Boisvert items occurred in the latter part of the experiment, after all the agent-patient items had been seen; this was done so that participants would not be biased to search for relations among the agent-patient pairs, given the strong relation for the McRae and Boisvert items.

Procedure Participants indicated whether each target was a word or non-word using two buttons on a CMU button box, which precisely measured decision latencies, accurate to the nearest millisecond. Each trial began with a fixation cross on a computer screen for 250 ms, followed by the prime for 200 ms, then by 50 ms of blank screen (for a total SOA of 250 ms), and finally the target, which was displayed until the participant responded. The inter-trial interval was 1500 ms. Participants were instructed to 'read and pay attention to the first word, then respond to the second word as quickly and accurately as possible.' The response latencies were recorded. Participants were given a break after every 40 trials, and the first item after a break was always a filler. An entire session lasted about 30 minutes.

Results and Discussion

Separate analyses were performed for the agent-patient pairs and the McRae and Boisvert pairs. In each analysis, response latencies more than 3 standard deviations above or below the grand mean were replaced with the cutoff value. This procedure affected 2.4% of the data for the McRae and Boisvert items and 1.6% of the data for the agent-patient pairs. Trials on which there were errors were excluded from the decision latency analysis. One-way ANOVAs were performed testing for an effect of relatedness on the error rates as well as decision latency.

The percent errors and decision latency are given by condition for each stimuli set in Table 1. There were no effects of relatedness on error rate, for either the McRae and Boisvert data or the agent-patient pairs. The McRae and Boisvert items showed a significant 18 ms of facilitation $[F_1(1, 42) = 8.67, p < 0.01; F_2(1, 30) = 4.47, p < 0.05]$). The agent-patient pairs, however, reveal no hint of an effect $[Fs < 1]$, and in fact have less than 1 ms mean difference.

This experiment thus confirmed that the agents did not independently facilitate their corresponding patients, suggesting that any effects obtained in the reading experiments are unlikely to be confounded.

Self-paced reading experiment

The first norming experiment provided materials that directly reflect event knowledge, and thus should be able to provide a strong manipulation of expectancies for patients. Furthermore, the second norming experiment allowed us to select materials in which the agent was not normatively associated with the patient, and the priming experiment confirmed that there was no significant facilitation between the agents and patients we chose. Using these materials in sentences such as (1–2), we performed a self-paced reading experiment to test whether comprehenders make use of event knowledge to form expectancies for patients, when the predictions made by event knowledge are crucially dependent on a combination of the verb and agent. Specifically, we predict that participants will be slower in the condition where the patient is Incongruent given the combination of agent and verb, versus the Congruent condition. We expect this effect to show up at the patient itself, or possibly in the post-patient region directly after.

Method

Participants Forty-eight students from UC-San Diego participated for course credit. All were native speakers of English and had normal or corrected-to-normal visual acuity.

Materials The materials were created from the 44 verbs from Norming Experiment 2, each of which had 2 agents and 2 patients. Each item was a unique combination of verb and patient noun (*e.g., checked the spelling*). Thus, two experimental items were formed by producing all four combinations of agent and patient for a given verb in a sentence-initial frame as in (3).

(3) The journalist checked the spelling of his latest . . .
     1   2   3   4   5   6 7 8   . . .

In this frame, word 1 was always *the*, word 2 was the agent noun, word 3 was the verb, word 4 was an article such as *a* or *the*, and word 5 was the patient noun. Across each item, words 6–8 remained constant and represented the post-patient region. After this sentence-initial frame, each sentence was completed by creating a plausible continuation. Finally, each sentence also had a follow-on sentence appended to form a short story, in an effort to encourage readers to pay closer attention to meaning. After this procedure was completed for all 44 verbs, the 32 verbs were selected that had the most natural sentences for all four agent-patient combinations, according to the judgment of the experimenters.[1] Each of the 32 verbs was used to produce two items, one for each of its two patients. This produced a total of 64 experimental items such

---

[1]The subset of data from the lexical decision experiment for these 64 agent-patient pairs also showed no trace of facilitation.

as (4), each of which had a Congruent condition (in which the patient matched the expectations created by agent and verb) and Incongruent condition:

(4) a. (*Congruent*) The **journalist** checked the **spelling** of his latest report to the editor. His editor was a stickler for correct spelling and made everyone else at the paper obsess about it too.

   b. (*Incongruent*) The **mechanic** checked the **spelling** of his latest report about the engine. Not having had much formal schooling, he strived to maintain professionalism at his shop.

Each agent was thus in the Congruent condition in one item for a verb, and in the Incongruent condition for the other item with for that verb. This counterbalancing controls for any associations that might exist between verb and agent or verb and patient. These 64 experimental items were rotated across 4 lists, so that each participant only read each verb once and encountered an equal number of sentences in the congruent and incongruent conditions. In addition to the 32 experimental items in each list, 48 filler sentence pairs were included in the experiment such that experimental items never occurred in sequence. Another 20 items were constructed for use during the practice session. Finally, a yes-no comprehension question was constructed for each sentence pair.

Procedure   Participants read the sentences on a computer screen using a word-by-word, non-cumulative, self-paced moving window reading paradigm (Just, Carpenter, & Woolley, 1982). Each trial began by displaying the entire target sentence on screen with dashes replacing all the letters. Participants pressed a button on the button box to reveal the next word, which simultaneously caused the previous word to change back to dashes. During this procedure, the time between the presentation of a word and the button press to show the next word was recorded. After reading the last word of the target sentence, a button push presented the follow-up sentence, also displayed with dashes. After participants made their way through the follow-up sentence, a comprehension question was displayed. Participants used two other buttons on the button box to indicate their answer. They were instructed to read as naturally as possible, and to slow down if they found themselves unsure of an answer to a comprehension question.

Before the main experiment, participants read a set of 20 practice sentence pairs to familiarize them with the task. They received a break prior to the start of the main experiment and received further breaks after every 21 sentence pairs. An entire session lasted about 30 minutes.

Analysis   In order to minimize the effects of word length differences across conditions and of participant reading rates, a regression was performed on the data for each participant, predicting reading time from word length in characters. This process utilized all words from both filler and target trials. The values predicted by the regressions were subtracted from

the actual reading times to produce residual reading times. Residual reading times that were more than 4 standard deviations from the mean for each combination of position and condition were removed. This process affected less than 1% of the data. One-way ANOVAs were performed on the question accuracy and the residual reading times at each of the critical regions, testing for an effect of Congruence.[2]

Results

Comprehension question accuracy   Overall, comprehension question accuracy on fillers and target items was 92.2% (SE: 0.5%), and no single participant was below 80%. There was no effect of Congruence on accuracy [Fs < 1].

Reading times   The mean residual reading times by word and condition are graphed in Figure 1. The significant effect of Congruence on reading times occurs on the first [$F_1(1, 47) = 9.93, p < .01; F_2(1, 62) = 8.01, p < .01$] and second [$F_1(1, 47) = 10.61, p < .01; F_2(1, 62) = 8.52, p < .01$] words of the post-patient region.[3]

Discussion

The results of the self-paced reading experiment support the hypothesis that comprehenders use event knowledge to form expectations about upcoming patients. Reading times on the two words immediately following the patient were elevated when it was not what would be predicted for the combination of verb and agent. Because the reading times on the patient itself showed no effect of Congruence, however, this result does not directly support the claim that this event knowledge is used immediately. To investigate this claim about time course, we performed an ERP experiment using the same materials.

ERP Experiment

To examine the time course of the effect of event knowledge, we conducted an ERP experiment using the same materials as in our self-paced reading experiment. ERPs can be a more sensitive measure and reveal subtle effects that are not evidenced in response latencies. For this reason, they could provide evidence for an effect of event knowledge directly at the patient noun. We predicted that the N400 response to the patient noun would have a larger negative amplitude in the Incongruent condition, in which the patient noun would be less expected.

Methods

Participants   Twelve students from UC-San Diego participated for course credit. All were right-handed, native monolingual speakers of English and had normal or corrected-to-normal visual acuity. None had a history of mental illness.

---

[2]The pattern of results described is also found in raw reading times.

[3]One item's comprehension question was never answered correctly in one condition, meaning that it could not be included in the by-items ANOVA. The *df* for all F_2s for this experiment reflect this: (1,62) instead of (1,63).
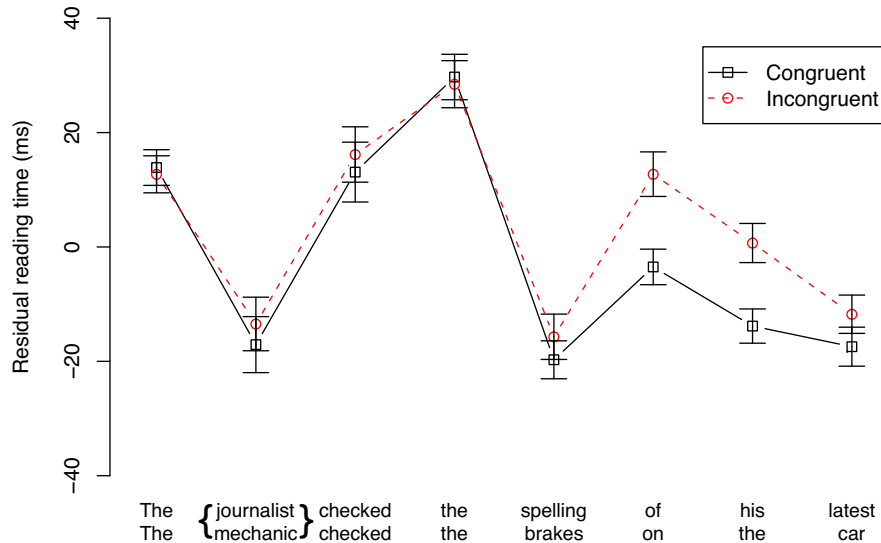
Figure 1: Mean residual reading times. Error bars show one standard error above and below the mean, calculated by participants.

Materials   The materials used for the ERP experiment were identical to those used in the self-paced reading experiment.

Procedure   Each pair of sentences was presented with rapid serial visual presentation (RSVP). Each word was presented along with a fixation dot for 200 ms followed by just the fixation dot for 100 ms, making for a 300 ms SOA. It is important to note that, in contrast to many N400 studies, this is faster than the mean self-paced reading speed per word (around 350 ms), meaning that participants have no more time to read the words in the ERP experiment than in the self-paced reading. Between the two sentences in a pair was a 1000 ms break, and there was a comprehension question after each pair. The task was divided into four blocks and participants were given a break between each block. The entire reading task took 20–30 minutes to complete.

The EEG was recorded from 26 electrodes distributed evenly over the scalp, referenced to the left mastoid. Eye movements and blinks were monitored via electrodes placed on the outer canthus and infraorbital ridge of each eye. Electrode impedances were kept below 5 KΩ . The EEG was amplified with Nicolet amplifiers with a bandpass of .016–100 Hz and was digitized at 250 samples per second.

Results and Discussion

The grand average ERP at electrode Pz in response to the patient noun is presented in Figure 2. We computed the average ERP to the patient nouns using epochs extending from 500 ms before the onset of the noun to 1548 ms after. Any trials contaminated with blinks, eye movements, excessive muscle activity, or amplifier blocking were rejected offline before averaging.

To analyze for differences in N400 amplitude, we entered the mean amplitude between 300–500 ms post-stimulus, relative to a pre-stimulus baseline of 500 ms, for each participant in each condition into a repeated measures ANOVA. We found an effect of Congruence in the N400 response to the patient noun such that the Incongruent condition was more negative [-1.63 *vs.* 0.18 μV; $F(1,11) = 6.44; p < .05$]. This result provides support for the immediacy of comprehenders' use of event knowledge in sentence processing.

Conclusions

This paper presents evidence from self-paced reading and ERP studies supporting the hypothesis that comprehenders can make immediate use of complex contingencies encoded in their event knowledge to form linguistic expectancies. The norming studies ensured that our items reflected comprehenders' event knowledge and that the agent nouns did not independently facilitate the patient nouns. Thus, the results from the self-paced reading experiment demonstrate that comprehenders are using their knowledge about the specific combinations of an agent with a verb to form expectations for likely patients. Finally, the ERP results show that the use of this knowledge is immediate, reflected in the processing of the earliest possible word. This apparent difference in where the effect occurs between the self-paced reading and ERP experiments is intriguing, and worthy of further study.

Of course, with these data, it is impossible to rule out an alternative interpretation that the reading time slowdown and larger N400 in the Incongruent condition are not the result of prediction, but merely of more difficult integration. While this debate is not central to the focus of this paper, which only argues for the immediate use of event knowledge, we be-
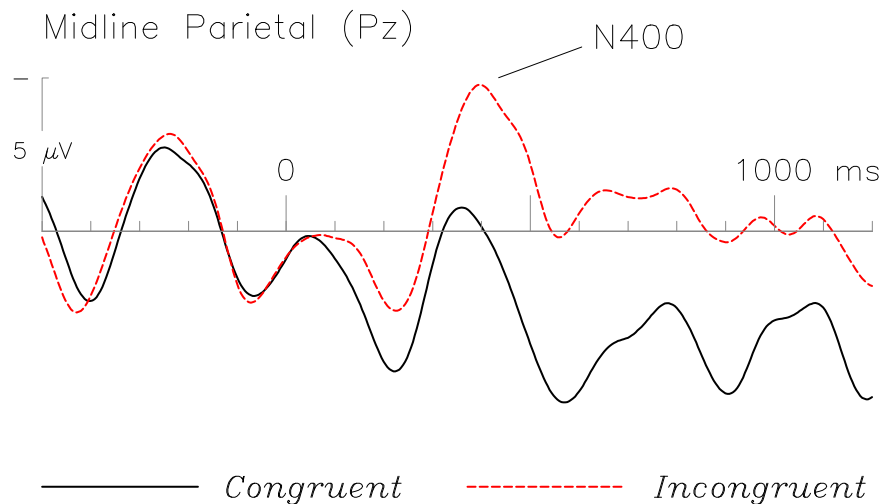
Figure 2: Grand average ERP at Pz in response to the patient noun (word 5), low pass filtered at 7 Hz.

lieve there is evidence to interpret these effects as arising from expectations (e.g., Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; DeLong, Urbach, & Kutas, 2005).

The hypothesis that comprehenders make use of complex contingencies in event knowledge to form expectancies for upcoming event participants (such as the patient) is of course not limited to agent-verb-patient contingencies. Other key information about an event, such as its location or the instrument that was used should also affect expectancies. As mentioned above, verbs prime their typical instruments (Ferretti et al., 2001), and instruments prime verbs for which they are typically used (McRae et al., 2005). Recent work by Matsuki, McRae, Tabaczynski, Elman, and Hare (2008) tested this prediction with self-paced reading using instrument-verb-patient contingencies as in (5).

(5) a. Mary used the saw to cut the expensive wood that she ... (Cong.)

   b. Mary used the saw to cut the expensive paper that she ... (Incong.)

Similar to the results reported in the present study, Matsuki et al. (2008) found a main effect of Congruence on the patient, such that it was read more slowly in the Incongruent condition.

This line of work raises the question of what other kinds of event knowledge are used, and how complex these immediately-accessed contingencies can be. For example, it could be the case that comprehenders can immediately calculate a complex contingency between agent, location, verb, and patient. Future work will explore this possibility.

Acknowledgments

References

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Ferretti, T. R., Kutas, M., & McRae, K. (2007). Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 182–196.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*, 516–547.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238.

Matsuki, K., McRae, K., Tabaczynski, T., Elman, J. L., & Hare, M. (2008). *Conceptual and structural event-based expectancy generation in language comprehension.* Poster presentation at the 2008 CUNY Conference on Human Sentence Processing, Chapel Hill, NC, 13–15 March 2008.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 558–572.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*(7), 1174–1184.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283–312.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 528–553.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.