UNIVERSITY OF ROCHESTER WORKING PAPERS IN THE LANGUAGE SCIENCES – VOL. 2, NO. 1 (SPRING, 2001)

Sussman, Rachel S., and Sedivy, Julie C. (2001). The time-course of processing syntactic dependencies: Evidence from eye movements during spoken narratives. In J. S. Magnuson and K. M. Crosswhite (Eds.), *University of Rochester Working Papers in the Language Sciences, 2 (1),* 52-70.

THE TIME-COURSE OF PROCESSING SYNTACTIC DEPENDENCIES: EVIDENCE FROM EYE MOVEMENTS DURING SPOKEN NARRATIVES

Rachel S. Sussman
University of Rochester
and
Julie C. Sedivy
Brown University

Abstract

The current study used head mounted eyetracking to examine the time course and nature of processing filler-gap relations in wh-questions, and the role of verb argument frame information. Subjects listened to a short narrative while viewing pictures of entities mentioned in the story and answered an auditorily presented question; eye movements in response to these questions were recorded. In Experiment 1, subjects heard wh-questions (e.g., "What did Jody squash the spider with?"), contrasted with yes/no question questions with no gap (e.g., "Did Jody squash a spider with her shoe?"). Results showed evidence of anticipatory gap projection at the verb only for wh-questions. In Experiment 2, wh-questions containing a strongly transitive verb used intransitively (e.g., "Who did the mother bear kill for?") were compared to questions containing an intransitive verb (e.g., "Who did the mother bear die for?"). Verb-argument information was used to constrain hypotheses of possible gap locations. These results provide converging evidence for active on-line resolution of syntactic dependencies mediated by verb argument structure, and introduce an important methodological tool for investigating processes that underlie syntactic dependencies.

1. Introduction

A substantial amount of research has been concerned with the on-line processing of syntactic dependencies present in wh-questions such as, "Who did Jesse remind August to invite?" In such a sentence, the wh-element *who* (the filler) must ultimately be associated with the direct object position of *invite* (the gap).[1] However, as discussed in Clifton and Frazier (1989) and Fodor (1978), the identification of the gap is potentially ambiguous. In the example above, for instance, the direct object of *remind* is a potential, but not actual, gap site, as this position is already occupied by the noun phrase "August". A number of studies have shown evidence that an overt filler in an utterance (e.g., "who," "which leopard," etc.) prompts the parser to actively search for sites of possible syntactic dependence for the filler in the unfolding utterance, generally interpreting the dependency as involving the first potential gap location after the verb (e.g., Stowe, 1986, Frazier and Flores d'Arcais, 1989). Gibson (1998) has argued that such a parsing strategy is

[1] It should be noted that we are using this terminology with agnosticism, and without committing to the representation of wh-dependencies as involving empty categories as opposed to direct association with subcategorizers as argued for by Pickering and Barry, (1991). None of the data presented in this paper either depends on or argues for a particular representational analysis.

driven by strong pressures from memory resource limitations to minimize the length of unresolved syntactic dependencies. In addition, it has been found that this process uses information based on a verb's syntactic and semantic constraints to predict or restrict possible gap locations (e.g., Stowe, Tanenhaus and Carlson, 1991, Clifton, Frazier and Connine, 1984, Boland, Tanenhaus, Garnsey and Carlson, 1995, but see deVincenzi, 2000). For instance, a verb that occurs most frequently without a direct object will be less likely to elicit expectation of a gap site immediately after the verb; similarly, the semantic content of a filler (e.g., *which movie*) will be considered against the semantic and real-world constraints on the nature of the direct object and other possible arguments, with a mismatch between the two diminishing the likelihood of positing a gap site in this location.

By and large, these phenomena are associated with structures known to contain some sort of syntactic gap. However, recent work by Altmann (1999) and Altmann and Kamide (1999) suggests that previous results with filler-gap constructions be viewed as a more general process of anaphoric interpretation of the verb's arguments. These recent studies have shown that listeners use context to project probable referents for a verb's arguments even in the absence of a filler. That is, even without the motivation of finding a corresponding syntactic position for a filler, listeners are using verb argument structures to predict referents for the verb's thematic roles. More specifically, Altmann (1999) shows that this integrative process appears to mimic many of the effects of verb-argument structure found with filler-gap constructions. This result brings into question the degree to which results from gap-filling studies are a result of a special parsing pressures that arise with structures involving an ambiguous syntactic dependency. One goal of the current study is to systematically evaluate whether the detection of a wh-dependency results in a more active and constrained search for the verb argument referent than that observed in sentences without such a dependency.

A second goal is to explore a new experimental methodology for investigating filler-gap processing, namely, measures of eye movements in response to spoken linguistic stimuli. To date, experiments that have reported anticipatory projection of gap sites have relied on one of two general experimental paradigms: (a) reading time studies in which the detection of an increase in reading times at a lexically overt NP for sentences containing wh-fillers is taken as an indication that a gap site was expected in that location (Stowe, 1986, Stowe et al., 1991, among others) or (b) anomaly detection measures, such as a word-by-word stop-making-sense task in which participants indicate noticing a semantic anomaly in a sentence based on anticipating a gap site in a certain location (e.g., Which child did the babysitter read…?" typically elicits a "doesn't make sense" judgment at this point, even though the gap site could perfectly sensibly be located later in the sentence(Boland et al., 1995). Additional anomaly detection measures such as the identification of N400 effects in ERP studies have led to the conclusion that gap sites are projected in an anticipatory fashion (Tanenhaus, Garnsey and Boland, 1991, Kluender and Kutas, 1993).

Each of these methods relies on measuring some behavioral response to surprise or unexpectedness; the anomaly detection tasks in particular depend upon the inclusion of genuinely nonsensical sentences in the experimental design. Thus, while these methods have clearly proven useful, they may not represent the bulk of naturalistic language processing in a real-world context where noticeable anomalies or difficulties are relatively rare. The current study relies on measures of eye movements to visual displays combined with a question-answer task. There are substantial arguments for applying this experimental paradigm to the study of language processing. First, this paradigm provides an extremely naturalistic language task for subjects, thus addressing concerns that previous behavioral measures such as stop-making-sense or judgment tasks introduced an

unnatural level of meta-linguistic analysis. Second, it provides an opportunity to observe what is *expected* rather than *unexpected*; it is a glimpse at everyday sentence processing rather than a measurement of breakdown due to a surprising occurrence. Third, it provides a continuous record of the listeners' eye movements/expectations as the utterance unfolds, yielding a more detailed picture of processing over time, as well as an indication of the time course of initial expectations versus readjustment of expectations and recovery from misanalyses. Fourth, it provides data that are referentially grounded; hence, while filled gap effects in reading time studies, for example, can tell us *whether* a gap was expected, eye movements can in principle indicate the referential content of the expected gap. Fifth, because of the simplicity of the task involved, it can be used to study processing in populations that are generally not amenable to more traditional experimental methods. The prospect of extending studies of the processing of filler-gap relations in this way is particularly attractive, given the attention this linguistic phenomenon has received in the language acquisition literature (e.g., see de Villiers, Roeper and Vainikka, 1990) and the discussion of complex syntactic dependencies in agrammatic populations (e.g., Caramazza and Zurif, 1976; Schwartz, Linebarger, Saffran and Pate, 1987).

2. Experiment 1

2.1 Rationale

The purpose of Experiment 1 was to determine whether the presence of a wh-filler, and hence, the detection of a syntactic dependency, results in an active search for the gap referent over and above any mechanism involved in general anaphoric mapping of a verb's arguments to an available discourse context. As shown by Altmann (1999) and Altmann and Kamide (1999), a certain degree of anticipation of a verb's upcoming arguments can be observed even in the absence of a displaced wh-filler. To compare the effects seen in gapped constructions with those of their non-gapped counterparts, the current study included narratives followed by questions containing overt WH-fillers as well as similar questions presented in yes/no form (e.g., "What did Jody squash the spider with?" vs. "Did Jody squash the spider with a shoe?"). The measure of interest was the pattern of participants' eye movements to a visual display depicting characters and objects referred to in the narrative. Cooper (1974) found that subjects generally looked at line drawings of objects depicting referents or objects related to referents shortly after the relevant referential expression was uttered. More specifically, Altmann and Kamide (1999; henceforth Altmann and Kamide) found that anticipatory eye movements were made to visual objects even before the onset of the referential expression in cases where the semantics of the verb allowed for the identification of a sole object in the display that could serve as a possible complement of the verb (e.g., *eat*). Verbs that could take any object in the display as a complement (e.g., *move*) did not show anticipatory eye movements. In addition, analyzing the relationship between eye movements to a visual display in response to spoken stimuli has proven to be highly sensitive in the investigation of a wide range of processing phenomena, in experimental situations involving both narrative and non-narrative stimuli (e.g., Allopenna, Magnuson and Tanenhaus, 1998; Arnold, Eisenband, Brown-Schmidt and Trueswell, 2000; Sedivy, Tanenhaus, Chambers and Carlson, 1999; Tanenhaus, Spivey-Knowlton, Eberhard and Sedivy, 1995).

We expected to find differences between the patterns of eye movements for wh-questions versus questions presented in yes/no form. Yes/No questions contain no overt fillers or corresponding gaps and therefore should involve no added pressure to project the referents of upcoming verb arguments. If wh-dependencies do trigger an active search for the syntactic position where the dependency could be resolved, then the Wh-condition should show greater evidence of

anticipatory eye movements to referents corresponding to potential upcoming arguments of the verb than the Yes/No condition.

2.2 Participants, materials, and procedure

16 members of the Brown University community were recruited and paid for their participation in the study. All participants were non-bilingual, native speakers of English. All had normal uncorrected vision or wore soft contact lenses.

Subjects were presented with a short narrative followed by a question directly related to the story they had just heard. The questions involved highly transitive verbs and were either yes/no questions ("Did Jody squash the spider with her shoe?") or wh-questions where the direct object argument of the transitive verb was filled by an overt NP ("What did Jody squash the spider with?"). The stories provided contexts which unambiguously defined the participants and their roles with regard the action described by the main verb of the question. Subject were instructed to give verbal answers at the end of each question. Since all of the critical wh-questions had a gap in oblique position (i.e. as object of the preposition), an equivalent number of filler trials consisting of wh-questions containing an unambiguous direct object gap was also included in the experiment. In addition, as all of the critical yes/no questions required a positive response, an equivalent number of fillers eliciting a negative response was included for counterbalancing. In total, Experiment 1 contained 10 critical items, (5 of which appeared in wh-form and 5 of which appeared in yes-no form), 5 wh-question filler trials with unambiguous direct object gaps, and 5 yes-no questions for which the correct answer was "no." In order that each critical item should appear in both yes/no and wh form, two separate lists of materials were constructed. For any given item, half the subjects heard the question in its wh form, while the other half heard the yes/no version. A complete sample item is shown below, with the accompanying visual display in Figure 1.

> Jody was eating breakfast one morning when she saw a big hairy spider creeping across the table towards her. Jody, whose terrible arachnophobia had caused her to seek therapy a few years ago, drew on the techniques of relaxation and anxiety management that her psychologist had taught her. Instead of screaming or freaking out, she calmly took off her shoe and slammed it down on top of the spider. She ate the rest of her Froot Loops in peace.
>
> Did Jody squash the spider with her shoe?
> Or
> What did Jody squash the spider with?

While hearing both the narrative and the question, subjects viewed a display containing characters and objects mentioned in the story, as illustrated above. The display contained both the object that had been defined in the story as the direct object of the main verb of the question (e.g., the spider), as well as an object that had played another role in the same action and that would eventually turn out to be the correct response to the wh-question (e.g., the shoe). Stories and displays were presented on a G3 Apple computer and 20" ColorSync monitor, using RSVP, a Macintosh-based program designed for the presentation of timed visual and auditory stimuli. Subjects were instructed to keep their eyes on the computer screen during the experiment, but were otherwise given no instruction of where to look.

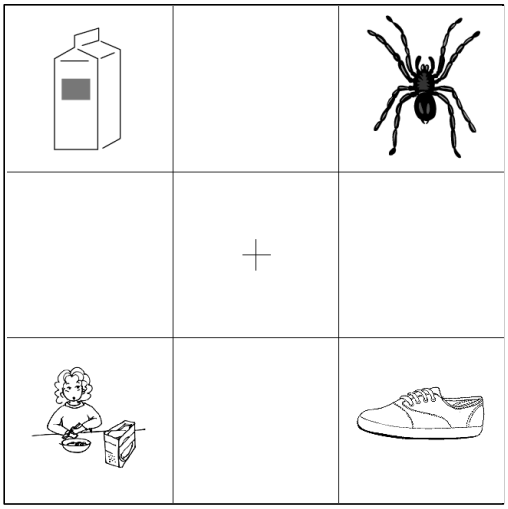


FIGURE 1**:** Visual display for Experiment 1.

During both the story and the question, eye movements were recorded using an ISCAN ETL-410 lightweight video-based head-mounted tracking system. The eye camera provided an infrared image of the eye at 60 Hz, and determined monocular eye position by monitoring the locations of the center of the pupil and the corneal reflection. A scene camera was mounted on the side of the helmet, providing an image of the subject's field of view. Each subject underwent a brief, 5-point calibration procedure prior to the experimental task. Calibration was carefully monitored throughout each trial, and minor adjustments were occasionally made between trials to maintain calibration. A frame-accurate SONY DSR-30 digital VCR was used to make a videotaped record for each experimental trial, consisting of the spoken stimuli as recorded via microphone from audio speakers interfacing directly with the computer, as well as the subjects' verbal response to the stimulus question and their moment-by-moment gaze fixation superimposed over the scene camera image. Because the scene camera was mounted onto the helmet itself, and moved with the subject's head, the VCR record took into account any head movements made by the subject, allowing for unrestricted head and body movements throughout the experiment. The entire experimental procedure, including familiarization, calibration, experiment and debriefing, took approximately 40 minutes.

2.3 Results and discussion

Of interest were the patterns of eye movements in the interval between the onset of the verb in each type of question and the onset of the subject's verbal answer.

Figure 2 shows the eye movement patterns for each type of object in the display as the question unfolded over time. Within the confines of the display, subjects had the choice of fixating the potential direct object referent (the competitor), the target referent (the ultimately correct referent to the WH-question), or either of two objects that were not possible answers to the question at hand but that were related to the story (distractor items). The proportion of trials for which subjects looked at each of these objects is charted as a single line within the graph. It should be noted that the proportions on the graphs do not sum to one, as would normally be expected. This is because in order to better reflect the true proportion of looking to any one distractor item, we

plotted the average proportion of looking to a distractor item. Also, since the task did not directly require the subject to interact with the display on any level, there were a fair number of fixations not relevant to the task (such as looks to the subject's watch, blank regions on the display, the fixation cross, etc.) that are not displayed on the graphs here.

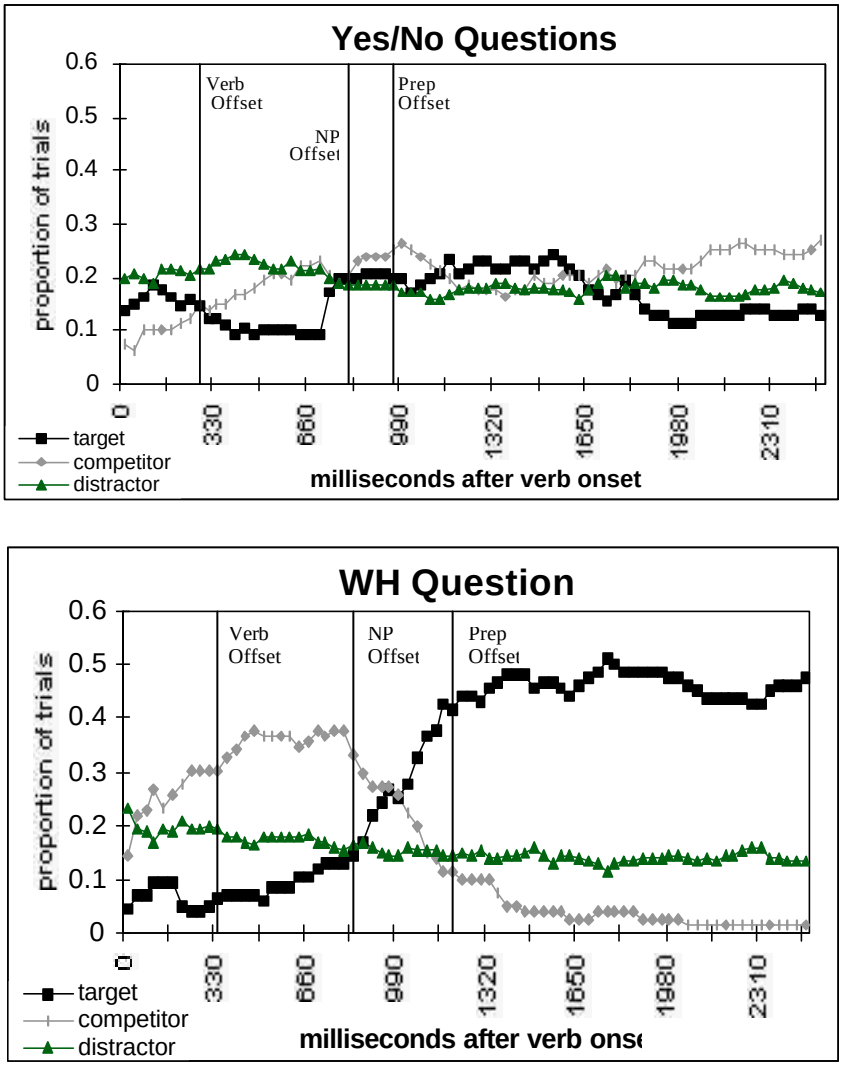


FIGURE 2: Fixation proportions to display items in Experiment 1.

Qualitatively, the data seem to bear out the predictions regarding patterns of looks. In the WH-condition, we see the expected increase of looks to the competitor object around the region of the verb, suggesting that participants are expecting the wh-filler to correspond with the upcoming direct object position. After encountering the overt NP and preposition, looks to the competitor decrease while looks to the target item increase as participants begin projecting the next likely (and correct) gap site and referent. By contrast, in the Yes/No condition, no such pattern is immediately evident. Looks to the three types of items remain at approximately the same level throughout the duration of the trial.

For the purposes of statistical analysis, we divided the time course information into scoring regions of special interest, and compared the amount of time within that region spent fixating each object type (see Arnold et al. [2000] for a similar analysis). Three scoring regions of interest were defined. The first region corresponded to the duration of the verb, the second to the overt noun phrase, and the third to the duration of the preposition. Each of these regions was displaced 200 ms from the actual onset and offset of the constituent in question so as to take into account the minimum time necessary to program and launch an eye movement, as described in Matin, Shao and Boff (1993). To insure that differences in constituent length across items were taken into account, regions were individually calculated for each item.

Fixation durations in milliseconds within a region were calculated for both the competitor item and the target. We treated the fixation durations for distractor objects as a baseline, and computed an adjusted score for the competitor item and the target by subtracting the average fixation duration for distractors within that scoring region from the total fixation times for competitors and targets. Table 1 shows the means for adjusted scores for targets and competitors across conditions and regions.

Verb Region		**Object**	
	Condition	target	competitor
	Yes/No	-24	-13
	WH	-41	31
Overt NP Region		**Object**	
	Condition	target	competitor
	Yes/No	-31	14
	WH	-19	77
Preposition Region		**Object**	
	Condition	target	competitor
	Yes/No	2	3
	WH	91	0

TABLE 1: Baseline-adjusted fixation durations (in ms) across regions for Experiment 1.

These adjusted scores for competitor and target objects were submitted to a repeated measures 2x2 ANOVA with condition (wh- vs. yes/no) and object type (competitor vs. target) as independent factors, and subjects and items as random variables. For region 1, the region corresponding to the verb (offset by 200 ms), the analysis yielded a significant main effect of object by subjects ($F_1(1,15)= 14.74$, $p<0.01$) and by items ($F_2(1,9)=23.48$, $p=.001$), with a greater fixation proportion for the competitor object than the target. The critical interaction of condition by

object was marginally significant by subjects ($F_1(1,15)= 3.98$, $p=0.06$) and significant by items ($F_2(1,9)=9.29$, $p<0.05$). Pair-wise comparisons revealed that for the wh-conditions, there were more looks to the competitor objects versus the target objects ($F_1(1,15)=15.39$, $p=.001$; $F_2(1,9)=29.08$, $p<.001$). In the yes/no condition, on the other hand, this difference did not approach significance ($F_1(1,15)=0.31$, $p=0.59$; $F_2(1,9)=.78$, $p=0.40$). For region 2, (the complement NP, offset by 200 ms), there was again a main effect of object ($F_1(1,15)=15.39$, $p=.001$; $F_2(1,9)=29.08$, $p<.001$). There was no main effect of condition, and the interaction of condition by object was not significant ($F_1(1,15)=1.33$, $p=0.27$; $F_2(1,9)=.1.80$, $p=0.21$), although numerically, the difference between the objects was greater for the wh-condition than the yes/no condition. For region 3 (the preposition, offset by 200 ms), there was a significant main effect of condition ($F_1(1,15)=9.02$, $p<0.01$; $F_2(1,9)= 5.92$, $p<.01$), probably reflecting the longer duration of the preposition in the wh-condition, where this was the sentence final word. The main effect of object was marginally significant by subjects ($F_1(1,15)=3.51$, $p=0.08$) and significant by items $F_2(1,9)= 14.58$, $p<.01$). The longer fixation times on the target reflected in this effect were entirely driven by the wh-condition; numerically, the reverse pattern was found for the yes/no condition. The crucial interaction of condition by object was robust both by subjects and items ($F_1(1,15)=10.28$, $p<0.01$; $F_2(1,9)= 11.18$, $p<.01$). Pair-wise comparisons showed there to be significantly longer fixation times for the target versus the competitor in the wh-condition : $F_1(1,15)=13.55$, $p<0.01$; $F_2(1,9)= 9.00$, $p=.05$), while the difference between the objects did not approach significance in the yes/no condition ($F_1(1,15)=0.004$, $p=0.95$; $F_2(1,9)= 0.08$, $p=0.78$).

These results indicate a strikingly different pattern of eye movements for the wh- questions as compared to the yes/no questions, which do not involve a filler-gap dependency. Generally, the eye movement data for the wh-questions are considerably more constrained by the information flow available from the speech stream, with marked differences between target and competitor objects evident in regions 1 and 3, and a dramatic reversal in the eye movement patterns occurring across time. By contrast, eye movements to the various objects in the display for the yes/no questions seem to cluster more closely together. These data suggest that the process of interpreting wh-questions does indeed involve a heightened anticipation of potential referents for gap sites, and does not appear to be reducible to expectations driven by anaphoric mappings of verb argument structure to contextually supplied referents, as suggested by Altmann (1999). There appears to be at best marginal evidence from these data for the general anticipatory processing in the absence of wh-fillers as was documented by Altmann and Kamide. Their study revealed that subjects began anticipating potential verb complements as early as during the presentation of the verb itself. In these data, looks to the direct object referent (i.e., the competitor object) never significantly exceeded the target object; the only (non-significant) trend towards increased looks to this object occur during the NP itself, where the overt mention of this referent would in any case be expected to yield increased looks.

In order to more directly compare these data to Altmann and Kamide's results, additional analyses were performed by computing the cumulative saccades for each object over time, following procedures reported in that paper. The cumulative responses were computed by including only fixations that occurred following the verb onset, as looks prior to this point do not reflect responses made based on pertinent information. These numbers give a measure over time of the proportion of trials which included a look to each of the relevant objects, as depicted graphically in Figure 3.

This method of analysis allows for the identification of a principled point in time at which it is predicted that differences will exist between pertinent objects or stimulus conditions in terms

of the cumulative probability of having made an eye movement to a specific object. A qualitative assessment of these data reveal that the difference in cumulative responses between target and competitor objects appears to be greater for the wh-condition than the yes/no condition. For the purposes of statistical analysis, we identified the point corresponding to 200 ms after verbs offset, and compared scores for the competitor and target objects minus the baseline (i.e., one half of the cumulative proportions of trials containing a look to one of the distractor objects).

These data were submitted to a repeated measures 2x2 ANOVA with object and condition as independent factors, and subjects and items as random variables. The analysis revealed a main effect of object ($F_1(1,15)=8.81$, $p=.01$; $F_2(1,9)= 8.88$, $p<.05$), with more looks to the competitor than target overall, and an effect of condition that approached significance ($F_1(1,15)=4.55$, $p=.05$; $F_2(1,9)= 4.04$, $p<.08$). The crucial interaction of subject by condition was not significant ($F_1(1,15)=1.41$, $p>.25$; $F_2(1,9)= 1.55$, $p>.2$). Thus, while the general pattern of data was consistent across the two measures, the analysis of cumulative responses did not reflect the differences in patterns for looks to objects across conditions as robustly as did the fixation time measure. This may be due to the fact that, in a task such as this where a good deal of scanning behavior was elicited, the cumulative response measure treated fleeting glimpses due to scanning as equivalent to prolonged looks to objects that reflected attempts to establish reference. The fixation time measure, however, would not have treated these equivalently, and therefore may have been a more sensitive measure in this case.

In addition, paired t-tests were computed for each condition separately, and revealed that cumulative looks to target and competitor items differed significantly in the wh-condition ($t_1(15)=2.77$, $p<.05$, $t_2(9)=3.23$, $p<.05$) but not in the yes/no condition. While the results of the t-tests need to be interpreted with caution, given the marginal significance of the interaction, they are quite consistent with the results of the fixation times analyses reported above.

Further confirmation for the differences across wh- and yes/no conditions as measured by cumulative responses can be gleaned by determining the earliest point at which looks to the target item and looks to the competitor item diverge significantly from baseline. We defined this as the earliest point where a t-test shows that cumulative proportions of trials containing a look to the item in question (target or competitor) versus looks to the distractor item become significantly different from each other. We also stipulated that this difference persist for at least 100 ms to be considered stable enough to represent a real divergence. As these values were bounded by 0 and 1, this analysis was performed on the arcsine transformed version of the cumulative proportions.

For wh-questions, the competitor item diverged significantly from baseline at 33 ms from verb onset, both in the subjects and items analyses. In contrast, for Yes/No questions, the competitor item differed significantly from baseline 924 ms after verb onset in the subjects analysis, but never diverged from baseline in the items analysis. The target item diverged from baseline 1320 ms after verb onset in the subjects analysis (924 ms by items) in wh-questions, which was considerably earlier than in the yes/no conditions, where the target item diverged from baseline at 2409 ms and 1881 ms by subjects and items respectively.

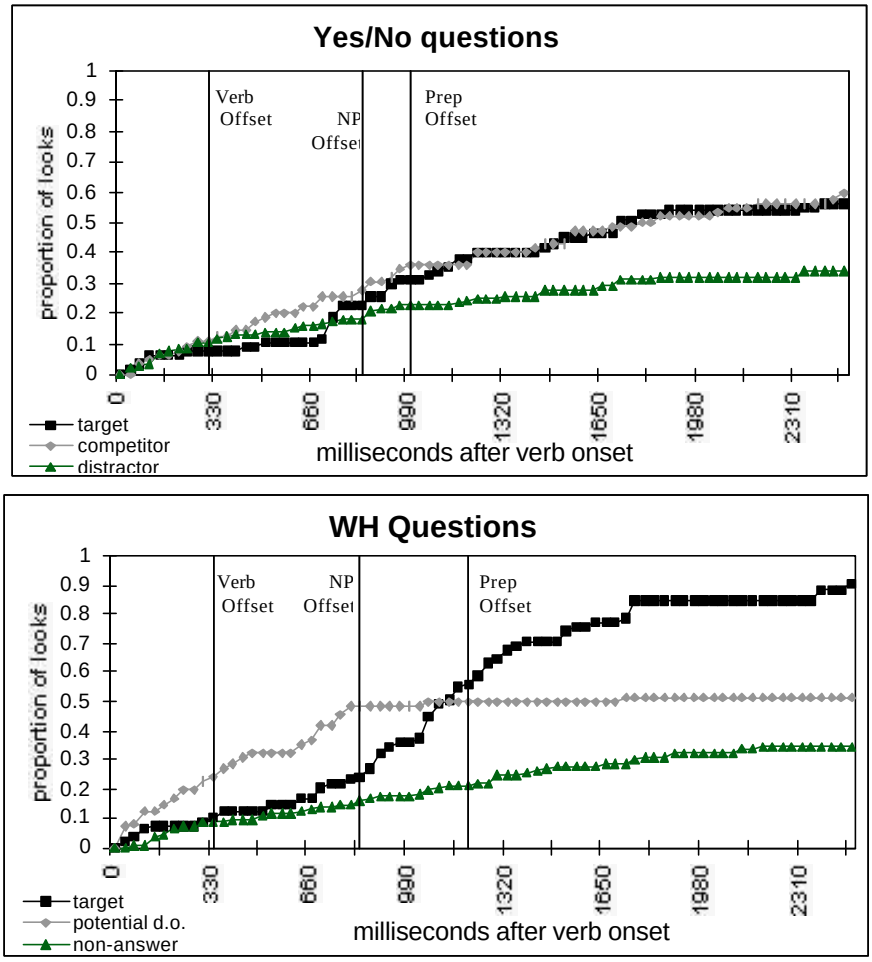


FIGURE 3: Cumulative fixations to display items in Experiment 1.

The analyses show consistently that wh-questions prompt looks to their potential direct objects as soon as the verb is encountered. Later on, as conflicting evidence mounts, looks to the target item increase while looks to the competitor item decrease. This is supported by the differences in relative amounts of looks seen both in the region analysis and by the difference seen in the cumulative proportions of trials containing looks to target and competitor items at verb offset. The points at which target and competitor items respectively diverge from baseline also indicates that participants were projecting the potential direct object as the referent of the filler until the incoming information no longer supported that hypothesis. With this disconfirmation, participants begin to project a new referent: the target item.

Yes/no questions, on the other hand, do not show this pattern. Analyses show that target and competitor items receive the same proportions of looks throughout all three regions as well as at verb offset, and points of first divergence in this condition tended to be somewhat later and not particularly stable across the subjects and items analyses. Interestingly, the anticipatory nature of eye movements in the absence of wh-dependencies that was observed by Altmann and Kamide was not found here. It should be noted, however, that the number of subjects and items in the Altmann and Kamide study was considerably larger. In addition, the cumulative proportions in their study

ultimately rose to considerably higher levels than was the case here. It is possible that either the complexity of the preceding narrative or the increased familiarity with the display over the course of the narrative resulted in a generally noisier pattern of eye movements. Therefore, while the difference between the wh- and yes/no conditions is apparent from these data, it would be premature to conclude that these present counterevidence against anticipatory mapping of verbal arguments onto contextual referents in the absence of wh-dependencies.

3. Experiment 2

3.1 Rationale, materials, participants, and procedure

Experiment 1 provides evidence from eye movements of an active search for a gap referent upon identification of a wh-dependency. In Experiment 2, we set out to determine whether eye movements would reveal this search to be constrained by verb-based information, as suggested by previous experiments using other experimental methods. In addition to methodology, the current study differs from its predecessors in that it uses a bare wh-phrase (such as "who") to test the contribution of verb argument structure. Previous studies relied upon wh-phrases including lexical content (e.g., "*which architect*") as this type of filler can produce implausibility effects when associated with an inappropriate verb. The presence of an implausibility effect can thus be taken as a measure of when and if a gap has been posited, providing a way to examine the process in the presence of verbs with varying argument structure. However, De Vincenzi (2000) has suggested that the effects observed in these studies are instead the result of a principle of interpretation associated with the discourse linking properties of WHICH-N type phrases. Under this view, WHICH-N type phrases are more sensitive to lexical properties such as verb argument structure; it is expected that questions involving non-discourse linked wh-phrases such as the ones used here will not exhibit the same effects.

The experiment employed the same sort of narrative and question/answer task as used in Experiment 1, but was restricted to wh-questions. In one condition, subjects heard a WH-question containing an intransitive verb such as "die". In another condition subjects heard a corresponding question with a typically transitive verb used intransitively. Stories were carefully constructed such that the situation contained a potential direct object for the verb in the transitive condition, as well as a salient entity that would eventually be the correct referent for the wh-filler. If argument structure is used to project gap sites, we predict more looks to the potential direct object (i.e., competitor item) as the verb is processed in cases where the question involves a transitive-biased verb. For intransitive verbs, we predict no difference in looks to competitor items and looks to target items.

A complete sample item is provided below, with an example display in Figure 4:

A hunter out in the Alaskan wilderness came across a mother bear and her cub foraging for berries. Once he had gotten the big bear in his sights, the hunter fired a shot and missed. Though normally the mother bear would have turned and run from such a threat, she knew her cub was too little to be able to run fast enough or far enough to escape, and the safety of her young was her first concern. Instead of running, she charged the hunter, who by this time had realized his mistake and began firing desperate shots. This time he hit the bear, giving her the wound that would eventually cause her death. But before she died, the bear managed to maul the hunter and snap his neck, killing him instantly. The bear died several minutes later, knowing that her baby was safe and out of danger.

Who did the mother bear die for?
Or
Who did the mother bear kill for?

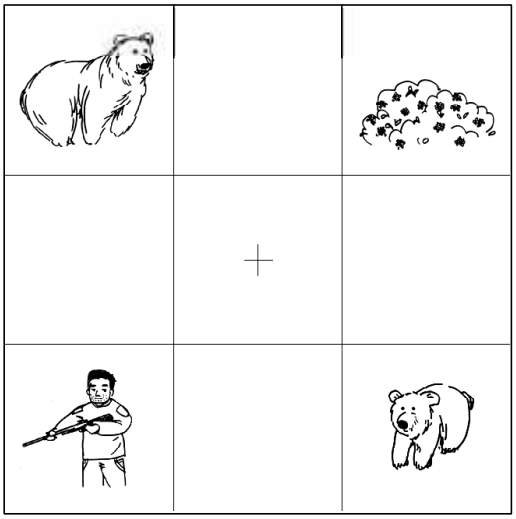


FIGURE 4: Visual display for Experiment 2.

Since all critical questions eventually resulted in a non-direct object response, a set of filler trials eliciting answers that corresponded to the direct object of the main verb used in the question was also included. In total, Experiment 2 consisted of 10 critical items, five of which were presented using a transitive preference verb and five of which were presented using intransitive preference verbs, plus 10 filler trials containing a direct object gap.

Experiments 1 and 2 were run simultaneously and thus involved the same set of 16 subjects. Method of presentation for displays and stories as well as method of recording eye movements were identical for the two experiments. As in Experiment 1, subjects viewed a display containing entities from the story which always included the potential direct object as well as the object corresponding to the correct answer to the WH-question, as seen in the example above. Patterns of eye movements in the interval between the onset of the verb and the onset of the subject's verbal answers were analyzed.

3.2 Results and Discussion

Figure 5 shows the time course of eye movement data for Experiment 2. As before, the averaging of looks to distractor items and the exclusion of stray fixations makes it so the proportions do not sum to one.

A qualitative assessment of these data shows general support for the hypothesis that information about argument structure plays an immediate role in determining possible referents for the WH-filler. In the transitive-preference condition, there is a sharp increase in looks to the potential direct object as the verb is heard, followed by an increase in looks to the target item with a corresponding decrease in looks to the potential direct object as the preposition is encountered. For intransitive preference verbs, all item types showed roughly the same proportion of looks until the appearance of the preposition, after which the target item showed a higher proportion of looks.

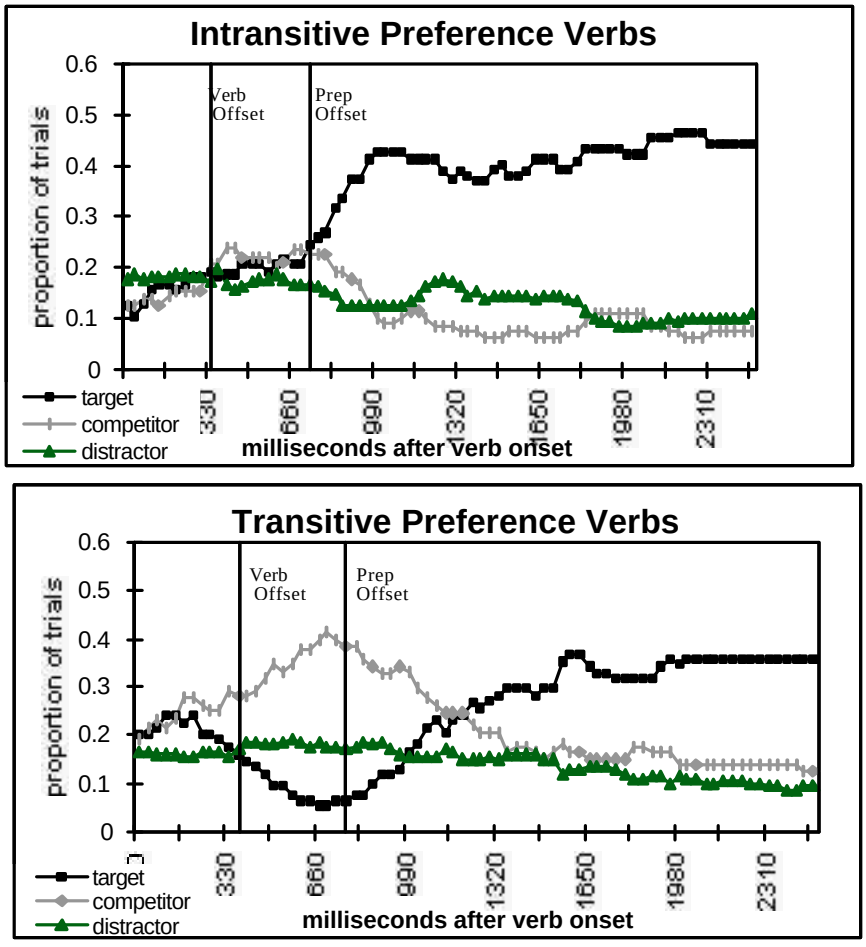


FIGURE 5: Proportions of fixations to display items in Experiment 2.

These effects were confirmed by statistical analysis. Again, we used the method of dividing time course information into theoretically relevant regions and comparing the fixation durations for each object of interest within this region. Two regions of interest were identified: the region corresponding to the duration of the verb, and the region corresponding to the duration of the final preposition. These two regions were calculated individually for each item to take into

account differences in the individual item lengths. Regions were displaced by 200 ms to account for average programming time of eye movements. Baseline-adjusted scores within each scoring region were calculated as in Experiment 1 for each item (as shown in Table 2), and submitted to a repeated measures ANOVA with condition and object as independent factors. In region 1 (corresponding to the verb, offset by 200 ms), the only effect to approach significance was the main effect of object in the subjects analysis (F_1 (1,15)=4.02, p=0.06; F_2 (1,9)= 1.86, p=0.21), with a slight overall bias in favor of the competitor object. In region 2 (corresponding to the preposition, offset by 200 ms), there was a main effect of object (F_1 (1,15)=10.15, p<0.01; F_2 (1,9)= 6.08, p<0.05), with competitor objects overall eliciting longer fixation durations. This effect was entirely due to the transitive condition, with values virtually identical (and in the opposite direction) for the intransitive verbs. The interaction of object and condition was significant (F_1(1,15)=5.23, p<0.05; F_2(1,9)= 5.00, p=0.05). Pair-wise comparisons revealed there to be a significantly greater proportion of fixations to the competitor object in the transitive condition (F_1(1,15)=14.78, p<0.01; F_2(1,9)= 15.27, p=0.01), but no difference in the intransitive condition (F_1(1,15)=0.01, p=0.94; F_2(1,9)= 0.02, p=0.90).

Verb Region

	Object	
Preference	target	competitor
Intransitive	-2	10
Transitive	0	37

Preposition Region

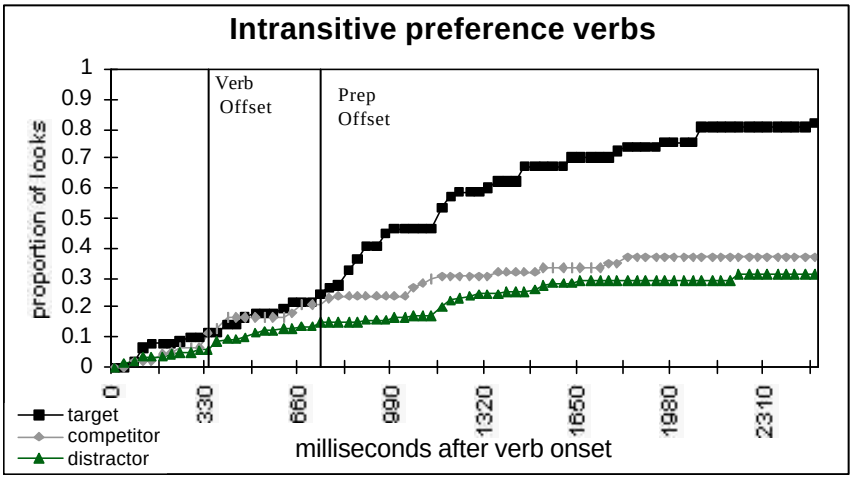
	Object	
Preference	target	competitor
Intransitive	43	41
Transitive	-33	84

TABLE 2: Baseline-adjusted fixation durations (in ms) across regions for Experiment 2.

As in Experiment 1, cumulative responses were computed and analyzed at the verb offset plus 200 ms (see Figure 6). Cumulative proportions of trials looks to competitor and target were adjusted to baseline and submitted to a repeated measures ANOVA. There was no main effect of condition (Fs<1). The main effect of object was significant in the subjects analysis only (F_1(1,15)=4.53, p=0.05; F_2(1,9)= 2.27, p>.15). The interaction of object by condition approached significance in both analyses (F_1(1,15)=3.31, p<0.09; F_2(1,9)= 3.39, p<.1).

Separate paired t-tests were computed for each condition. For transitive preference verbs, this analysis revealed significantly higher proportions of looks to the competitor item than the target item (t_1(15)=2.44, p<.05, t_2(9)=2.04, p<.05). For intransitive preference verbs, however, no such difference was found. Using the procedure and guidelines described in experiment 1, we calculated the point of divergence from baseline on the arcsine transformed cumulative proportions of looks for each item in both conditions. For wh-questions involving highly transitive verbs, cumulative looks to competitor items became significantly higher than baseline 462 ms after

verb onset by subjects, (660 ms by items), For questions involving highly intransitive verbs, there was no divergence between competitor and baseline items. These data reflect a temporary bias towards the competitor object in the transitive condition, with no such expectations in the intransitive conditions, suggesting that subjects initially take the wh-dependency to involve the direct object position. This temporary misanalysis results in an ultimate slowing in the identification of the correct target, where cumulative looks to target item rose above baseline 1551ms. after verb onset in both the subjects and items analyses as compared to 627 ms (and 792 ms by in the items analysis) for the intransitive condition.



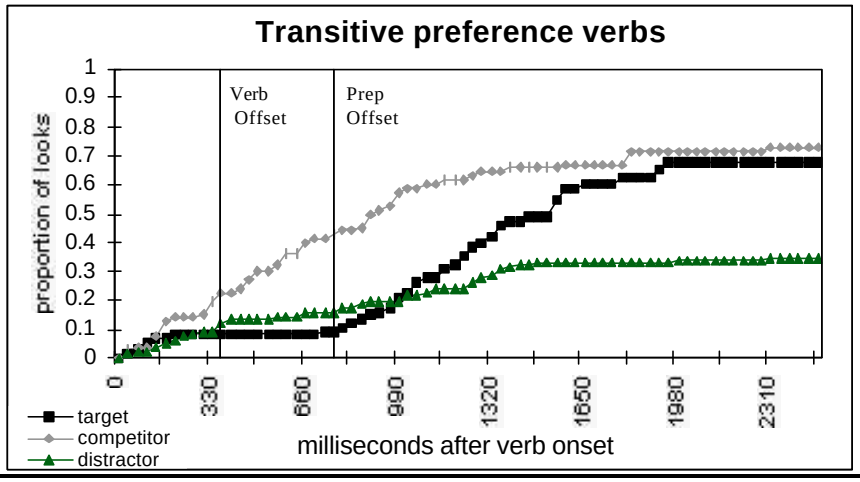


FIGURE 6: Cumulative fixations to display items in Experiment 2.

The results of the above analyses indicate that the projection of the gap site is moderated by verb argument structure and provides converging evidence with other studies arriving at similar conclusions. In wh-questions where the verb used is typically seen in transitive form, participants assume that the potential direct object of the verb will be the referent of the filler, as indicated by the larger proportions of looks to the competitor item than to the target item, and by the early divergence of proportions of looks to the competitor from baseline. When the verb used in the wh-question is typically found in intransitive form, the potential direct object is not regarded as a

potential referent for the filler. Instead, it is treated as a distractor item, in that it never diverges from baseline within the course of the trial. Nor does it ever receive more attention than the target item. These effects occurred in the absence of a discourse-linked wh-phrase (WHICH N), suggesting that bare wh-phrases are in fact sensitive to lexical information such as verb argument structure.

4. General Conclusions

The results of Experiment 1 provide support for the claim that the identification of a wh-filler results in an active search for the referent corresponding to the filler. This search is measurably more urgent and constrained than attempts to identify relevant referents on the basis of discourse context and verb-argument information in the absence of a wh-dependency. Furthermore, Experiment 2 provides evidence for the use of verb-based information in constraining possible referents for fillers, and shows that the use of such information is not restricted to fillers with lexical content, as suggested by DeVincenzi (2000).

Much of the literature on wh-dependencies has focused on the relationship between the wh-element and the syntactic/thematic properties of the verb, and it has been widely argued that constraints originating from the verb's thematic/argument structure provide much of the impetus for predictive identification of the gap site. It has been found, for instance, that while subjects typically attempt to resolve the dependency at the earliest potential post-verbal location, similar expectations for the subject location, another possible gap location, have been difficult to document (e.g., Stowe, 1986). However, there is some suggestion in the data from the present study that attempts to identify the referent corresponding to the wh-word may occur prior to the verb. Specifically, note that in Experiment 1, looks to the competitor object began to diverge from looks to other objects essentially at the onset of the verb, indicating that eye movements to this object were beginning to be programmed in some cases prior to any information about the verb itself. It is possible that predictive processing is occurring based on expectations regarding the subject noun and its agenthood in certain kinds of events. Interestingly, Experiment 2 showed looks to competitor objects diverging from others at some later point, presumably in response to information available from the verb. It may be the case that the materials in Experiment 1 permitted for more constrained hypotheses regarding likely events prior to the verb than the materials in Experiment 2. This seems plausible, as the stories in Experiment 2 were carefully constructed such that the situation contained a salient entity that would eventually be the correct referent for the wh-filler as well as an entity that served as a potential direct object for the verb in the transitive condition. The stories in Experiment 1 may have favored events focusing on the potential direct object to a greater extent. Some support for this comes from an informal post-hoc questionnaire, in which 16 new subjects who had not participated in the previous experiment were recruited and asked to read the stimulus stories, and supply a reasonable question querying the content of the story, based on a fragment up to and including the subject noun (e.g., "What did Jody _____?"). Subjects provided questions for which the competitor object was the correct answer 64% of the time for Experiment 1 material, versus 51% for Experiment 2 materials. This difference, while not large, was significant ($F(1,15)=6.36$, $p<0.05$),and is therefore suggestive. More research is required to better identify the nature of the information that can be used to predict likely thematic roles for wh-words, and to determine whether this experimental paradigm has the potential to be more sensitive at probing constraints that may be more subtle than those originating from verb-based representations.

This study also raises some significant methodological implications. The naturalness and time-course sensitivity of the task allays some concerns over results from methodologies such as stop-making-sense tasks as tapping into post-processing stages only. In addition, the naturalness and simplicity of the tasks makes it possible to probe not only the processing mechanisms, but also representational capacities of populations for whom administering tasks that involve overt judgments or a high degree of literacy can be problematic. For instance, this paradigm might be used to investigate the nature of verb argument representations in children, or to search for evidence of control over grammatical constraints on wh-dependencies, as well as to evaluate models of information integration and processing under various complexity demands. Research in domains such as this could potentially serve as a catalyst for the closer integration of understanding of language processing, learning and language disorders.

References

Allopenna, P.D., Magnuson, J.S. and Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419-439.

Altmann, G. (1999). Thematic role assignment in context. *Journal of Memory and Language, 41*, 124-145.

Altmann, G. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247-264.

Arnold, J.E., Eisenband, J.G., Brown-Schmidt, S. and Trueswell, J.C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 75, 1-14.

Boland, J., Tanenhaus, M., Garnsey, S. and Carlson, G. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*, 774-806.

Clifton, C., Frazier, L. and Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behaviour, 23*, 696-708.

Clifton, C. and Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. Carlson and M. Tanenhaus (Eds.), *Linguistic Structure in Language Processing* (pp. 273-317). Dordrecht: Reidel.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory and language processing. *Cognitive Psychology*, 6, 84-107.

De Villiers, J., Roeper, T. and Vainnika, A. (1990). The acquisition of long-distance rules. In L. Frazier and J. De Villiers (Eds.). *Language Processing and Language Acquisition*. Dordrecht: Kluwer.

De Vincenzi, M. (2000). Cross-linguistic Psycholinguistics. In Crocker, M., Pickering, M. and Clifton, C. (Eds.), *Architectures and Mechanisms for Language Processing* (pp. 282-300). Cambridge, UK: Cambridge University Press.

Fodor, J.D. (1978) Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9*, 427-474.

Frazier, L. and Flores d'Arcais, G. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language, 28*, 331-344.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition 68(1)*, 1-76.

Kluender, R. and Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5(2)*, 196-214.

Matin, E., Shao, K., and Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perceptual Psychophysics, 53*, 372-380.

Pickering, M. and Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6(3)*, 229-259.

Saffran, E.M., Schwartz, M.F., and Linebarger, M.C. (1998). Semantic influences on thematic role assignment: evidence from normals and aphasics. *Brain and Language*, 62, 255-297.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109-147.

Stowe, L. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, 1, 227-245.

Stowe, L., Tanenhaus, M., and Carlson, G. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech, 34(4),* 319-340.

Tanenhaus, M., Garnsey, S., and Boland, J. (1991). Combinatory lexical information and language comprehension. In G. Altmann, (Ed.), *Cognitive Models of Speech Processing.* Cambridge, MA: MIT Press.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., and Sedivy, J.C. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science,* 268, 1632-1634.

Trueswell, J. Sekerina, I., and Hill, N. (1999). The kindergarten path effect: Studying on-line processing in young children. *Cognition, 73,* 89-134.

Authors' notes

Preliminary reports of these results were delivered at the 13th CUNY Conference on Human Sentence Processing, San Diego, CA. The authors are grateful for comments from this audience, and also from Mike Tanenhaus, Eiling Yee, Aparna Nadig. We thank A-Mi Kim for creating the drawings accompanying the linguistic stimuli, and John Mertus for technical assistance with recordings and speech sampling. This research was partially supported by NIH grant #HD27206 to Michael K. Tanenhaus and Gregory N. Carlson and NSF grant # BCS-0079497 to Julie C. Sedivy. Correspondence may be sent to either author: Rachel Sussman, Brain and Cognitive Sciences Dept., Rochester, NY 14627; Julie C. Sedivy, Dept. of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912.

UNIVERSITY OF ROCHESTER WORKING PAPERS IN THE LANGUAGE SCIENCES – VOL. 2, NO. 1 (SPRING, 2001)

James S. Magnuson and Katherine M. Crosswhite, Editors
Joyce Mary McDonough, Series Editor