Journal of Psycholinguistic Research, Vol. 24, No. 6, 1995

Eye Movements as a Window into Real-Time Spoken Language Comprehension in Natural Contexts

Kathleen M. Eberhard,[1,2] Michael J. Spivey-Knowlton,[1] Julie C. Sedivy,[1] and Michael K. Tanenhaus[1]

Accepted August 11, 1995

When listeners follow spoken instructions to manipulate real objects, their eye movements to the objects are closely time locked to the referring words. We review five experiments showing that this time-locked characteristic of eye movements provides a detailed profile of the processes that underlie real-time spoken language comprehension. Together, the first four experiments showed that listeners immediately integrated lexical, sublexical, and prosodic information in the spoken input with information from the visual context to reduce the set of referents to the intended one. The fifth experiment demonstrated that a visual referential context affected the initial structuring of the linguistic input, eliminating even strong syntactic preferences that result in clear garden paths when the referential context is introduced linguistically. We argue that context affected the earliest moments of language processing because it was highly accessible and relevant to the behavioral goals of the listener.

We thank D. Ballard and M. Hayhoe for the use of their laboratory (National Resource Laboratory for the Study of Brain and Behavior). We also thank J. Pelz for his assistance in learning how to use the equipment and K. Kobashi for assisting in the data collection. Finally, we thank Janet Nicol and an anonymous reviewer for their comments and suggestions. The research was supported by NIH resource grant 1-P41-RR09283; NIH HD27206 (M.K.T); an NSF graduate fellowship (M.J.S-K.); and a Canadian SSHRC fellowship (J.C.S.).

[1] University of Rochester, Rochester, New York 14627; K. M. Eberhard, M. J. Spivey-Knowlton; M. K. Tanenhaus, Department of Brain and Cognitive Sciences; J. C. Sedivy, Department of Linguistics.

[2] Address all correspondence concerning this article to Kathleen M. Eberhard, Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, New York 14627.

0090-6905/95/1100-0409$07.50/0

The interpretation of an utterance crucially depends on the discourse context in which it occurs. Consider, for example, the sentence *He is putting the ball in the box on the shelf.* Given just the knowledge of the English language, a listener would know that when the sentence was uttered, a male person (or animal) was moving a spherical object either from a box to a shelf or from some unknown place to a box which was located on a shelf. To arrive at the complete and intended interpretation of the sentence, the listener requires knowledge of the situational context of the sentence, i.e., knowledge of the referents and their spatial relations at the particular time of the utterance.

Although all models of language comprehension acknowledge the importance of discourse context in determining the interpretation of a sentence or an utterance, they differ in their assumptions about when context exerts its influence (for recent reviews, see MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Tanenhaus & Trueswell, 1995). For example, some models postulate an initial stage of processing in which the rapid and automatic initial processes that structure the linguistic input are encapsulated from the slower integrative processes that relate an utterance to its discourse context (e.g., Frazier, 1978, 1987; Swinney & Osterhout, 1990). In contrast, other models have emphasized that the linguistic input is immediately mapped onto a discourse representation, with context influencing even the earliest moments of linguistic processing (Altmann & Steedman, 1988; Bates & MacWhinney, 1989; Crain & Steedman, 1985; MacDonald et al., 1994; Marslen-Wilson & Tyler, 1987; Spivey-Knowlton et al., 1993; Taraban & McClelland, 1988, 1990).

Most of the research investigating the role of context in sentence processing has focused on the comprehension of written sentences in discourse contexts that typically consist of only a few sentences describing an imaginary situation. One reason for the focus on written comprehension is that testing subtle predictions about the time course of processing requires using methodologies that can provide immediate information about how each word is interpreted as the sentence unfolds. There are a variety of methodologies for studying reading that provide a continuous processing profile, chief among them being self-paced reading and eye-movement monitoring tasks which have the advantage of allowing the subject to process the input in a relatively natural way, i.e., without requiring any explicit decisions.

The situation is different for spoken language comprehension. Although there are a variety of on-line methodologies that provide insight into the temporal characteristics of spoken language comprehension, they do not allow for continuous monitoring. In addition, most of these methodologies measure comprehension indirectly via a superimposed secondary task, e.g., detection, cross-modal priming, or probe tasks. And methodologies that do

provide a more direct measurement, e.g., monitoring tasks, require listeners to consciously attend to the linguistic input in an unnatural way. As a result, these paradigms cannot easily be used to study immediate spoken language comprehension as it typically occurs in natural real-world situations, for example in interactive conversation where the objects and events being referred to are in the immediate environment and therefore are highly salient.

In this article, we present an overview of some work involving a new methodology that allows an in-depth investigation of incremental processing and contextual dependence in spoken language comprehension. The methodology involves monitoring listeners' eye movements as they follow short discourses instructing them to move or touch common objects in a display (e.g., "Put the candy above the fork. Now put it below the pencil"), thus providing an on-line, nonintrusive measure of spoken language comprehension as it occurs in natural situational contexts. Eye movements are monitored using a light-weight eye-tracking camera that is mounted on a helmet worn by the listeners. Also mounted on the helmet is a small video camera that records the visual scene from the listener's perspective. The visual scene image is displayed on a TV monitor along with a record of the listener's eye fixations superimposed as cross hairs.[3] Both the image and the experimenter's spoken instructions are synchronously recorded by a VCR which permits frame-by-frame playback.

Two essential features of this paradigm make it useful for studying spoken language comprehension in context. First, the visual context is *available* for the subject to interrogate as the spoken language message unfolds over time, and because the message directs the listener to interact with the context, the context is necessarily *relevant* to the comprehension process. This contrasts with a linguistically introduced context, which must be represented in memory and depending upon the experimental task, may or may not be immediately accessible or perceived as relevant by the reader or listener. Secondly, in all of the work we have conducted to date, we have found that subjects' eye movements to objects are closely time locked to the spoken words that refer to those objects (Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, & Carlson, 1995; Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy, 1995; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, in press). Thus the methodology provides a natural on-line measure of how comprehension unfolds over time, and how it is influenced by the

[3] Eye movements were monitored by an Applied Scientific Laboratories eyetracker which provides an infrared image of the eye at 60 Hz. This allows the tracking of the center of the pupil and the corneal reflection of one eye to be recorded every 16 msec. The accuracy of the tracking is about a degree over a range of $\pm 20°$. A short calibration routine is conducted before each experimental session to map the eye-in-head coordinates from the tracker to the visual scene image coordinates.

information provided by the visual context. The work that we have conducted demonstrates that the pattern and timing of these eye movements allows for strong inferences about component processes in comprehension, including referential processing, word recognition, and parsing.

We will discuss a total of five experiments. The first two experiments lay the foundation for the other three by clearly demonstrating the incremental nature of spoken language comprehension: Listeners interpreted each word of an utterance immediately with respect to the set of co-present visual referents. More specifically, they used information from each word to reduce the set of possible visual referents to the intended one. As a result, they established reference as soon as the utterance provided enough accumulative information for them to distinguish or disambiguate the intended referent from the alternatives. In all the experiments, the visual context was manipulated to vary the point in the utterance when information for uniquely identifying an intended referent occurred. This manipulation allowed us to examine the effects of the visual context on the time course of comprehension.

In the third experiment, we manipulated the prosody of the utterance as well as the visual context. We found that listeners made immediate use of disambiguating information provided by contrastive stress, as revealed by its facilitatory effect on the point in a complex noun phrase when listeners established reference. We also discuss the results of a fourth experiment showing that the point at which reference was established within a word (e.g., candle) was influenced by whether or not the relevant visual context contained an object with a similar name (e.g., both a candle and some candy).

Finally, we present evidence that a visually co-present referential context influenced initial syntactic commitments under conditions where linguistically introduced contexts have been shown to be ineffective. The referential effects we report are consistent with Crain and Steedman's (1985) and Altmann and Steedman's (1988) claims about the importance of referential pragmatic context in ambiguity resolution, and they provide strong evidence against modular models in which syntactic commitments are made during an encapsulated first stage of processing. More generally, we argue that these effects follow naturally from the behavioral goals and Gricean expectations of the listener. We also argue that the global pattern of results across visual and linguistic contexts is most naturally accommodated by constraint-based models of language processing.

THE INCREMENTAL NATURE OF ESTABLISHING REFERENCE

For communication to be successful, a listener must arrive at the intended referents of the words of a speaker's utterance. Intuitively, the es-

tablishment of reference seems to be incremental: As soon as a word is heard, its meaning becomes available and is interpreted with respect to the entities in the discourse model. However, from a linguistic perspective, we often talk as though the interpretation of words does not occur until the end of a phrasal constituent. Consider for example the definite noun phrase *the large beach ball.* We say that the adjectives *large* and *beach* modify the noun *ball* because they provide additional relevant information to the noun. This linguistic perspective implies that the referent of a phrase like *the large beach ball* cannot be understood until the noun *ball.* While this implication may be reasonable when considering decontextualized linguistic expressions, it clashes with our intuition about expressions spoken under normal circumstances, i.e., expressions spoken in context. Words uttered in context do not refer to or modify other words, they refer to or modify entities in the discourse model. Olson (1970) further elaborated this point in the following statement: "words do not 'mean' referents or stand for referents, they have a use—they specify perceived events relative to a set of alternatives; they provide information" (p. 263).

Olson illustrated his claim by pointing out that the expression that is used to refer to a particular object depends on the context in which the object occurs. For example, the definite expression *the ball* may be used to refer to one of the objects in Fig. 1a, but a different definite expression, e.g., *the beach ball,* must be used to refer to the same object when it occurs in the context of Fig. 1b. And yet another definite expression must be used when the object occurs in the context of Fig. 1c, e.g., *the large beach ball.* Speakers use different expressions to refer to the same object in different contexts in order to provide their listeners with the necessary information for distinguishing the intended referent from the set of alternatives. Listeners expect speakers to walk a middle line between providing too little information for distinguishing the referent and providing too much information (Grice, 1975). An important factor affecting a speaker's ability to walk this line is the extent to which his or her set of relevant discourse referents is the same as the listener's (Clark, 1992), i.e., the extent to which the information that the speaker intends the listener to consult in understanding the utterance is the information the listener actually consults. The speaker is apt to be most successful at providing the right information when the relevant set is visually co-present and therefore is highly accessible and salient to both the speaker and the listener (Clark, 1992).

From the listener's perspective, when the relevant set of referents is visible and co-present with the speaker's utterance, the listener should be able to immediately interpret each word of the utterance with respect to the

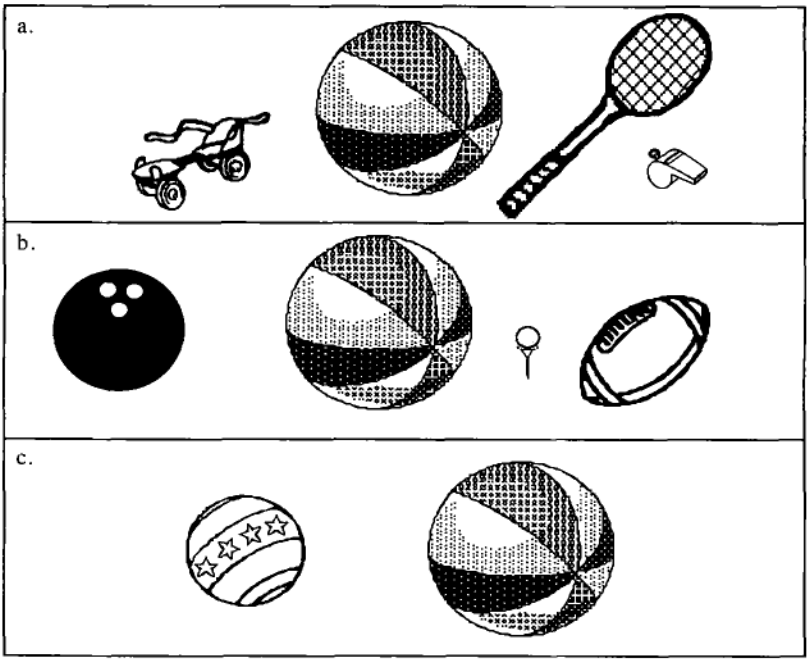


Fig. 1. The noun phrase, *the ball*, may be used to refer to one of the objects in Fig. 1a, but a different noun phrase, e.g., *the beach ball*, must be used to refer to the same object when it occurs in the context of Fig. 1b. And yet another noun phrase must be used when the object occurs in the context of Fig. 1c, e.g., *the large beach ball*.

set of referents.[4] For example, if a listener is told to *kick the large beach ball* in a co-present visual context depicted in Fig. 1c., where there are two beach balls but only one that is large, he or she may be able to identify the referent as soon as the adjective *large* is heard. This general prediction was investigated using the head-mounted eye-tracker paradigm (Tanenhaus et al., in press). We reasoned that because the paradigm allows us to observe the listeners' eye movements to the referent objects as the spoken input unfolds, it would provide important insight into not only how but also when listeners establish reference.

THE INCREMENTAL PROCESSING OF COMPLEX SPOKEN NOUN PHRASES

In the first experiment (see Tanenhaus et al., in press), subjects were given spoken instructions to touch various blocks arranged in simple dis-

[4] For a discussion of the relation between co-presence and mutual knowledge and their effect on the establishment of definite reference see Clark and Marshall (1992).

plays like those depicted in Fig. 2. The blocks differed along the dimensions of marking, color, and shape, and the spoken noun phrases that referred to the blocks specified each of those dimensions. For each display, the experimenter read aloud from a script a critical instruction like "Touch the starred yellow square" and several filler instructions (e.g., "Touch the plain red square. Now touch the plain blue square. Now touch it again"). Every display contained a centrally located fixation cross, and the subjects were instructed at the beginning of the experiment to look at the cross and rest their hands in their lap after they perform each requested action.

Critical instructions were given in three display conditions. The conditions determined the point in the instructions when disambiguating information occurred. Disambiguating information was provided by the marking adjective in the early condition, by the color adjective in the mid condition, and by the noun in the late condition. We predicted that if subjects interpreted the words in a noun phrase incrementally with respect to the relevant set of referents in the display, their eye movements to the target objects should occur shortly after the word that disambiguated the intended referent from the alternatives, rather than after the noun. In addition, we expected that the timing of the eye movements relative to the onset of the disambiguating words would reveal the speed with which nonlinguistic visual information was integrated with the spoken linguistic input.

The data were analyzed using frame-by-frame playback of the synchronized video and audio recordings of the experimental sessions. Eye-movement latencies to target objects were obtained by locating the frame on which a critical word in an instruction began and then counting the number of frames until the eye position, which was indicated by the crosshairs, moved from the central cross in the display to the target object. Each audio–video frame consisted of a 33-msec segment of time.

Figure 3 contains a graph of the mean eye-movement latencies to the target blocks measured from the onset of the spoken determiner *the* in each

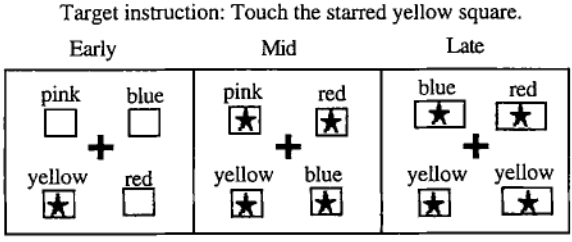


Fig. 2. Example displays representing the three point-of-disambiguation conditions in Experiment 1.

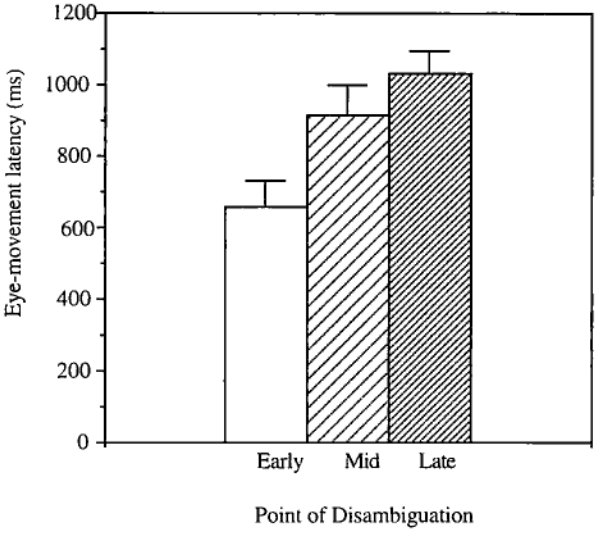


Fig. 3. Mean eye-movement latencies and standard errors, measured from the onset of the determiner *the* in each of the point-of-disambiguation conditions of Experiment 1.

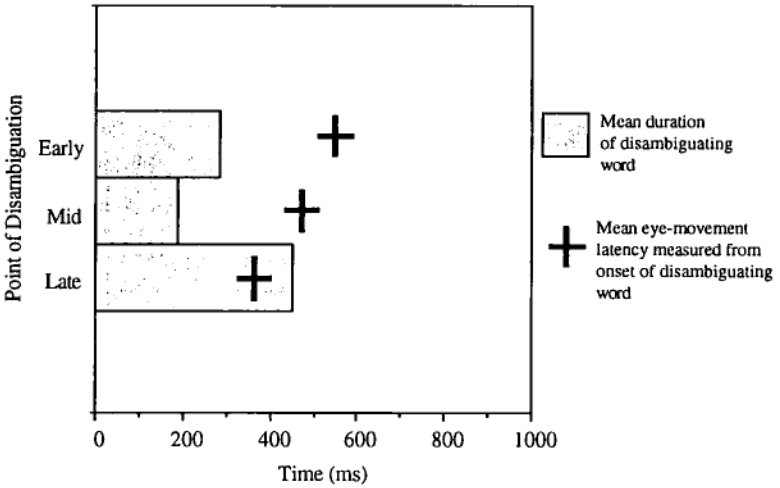


Fig. 4. Mean durations of the disambiguating words and mean eye-movement latencies to the target objects (measured from the onset of the disambiguating words) in Experiment 1.

of the three point-of-disambiguation conditions. The effect of point of disambiguation was reliable and provided evidence for rapid incremental processing: As shown in the figure, eye movements to target objects occurred sooner in the early condition than in the mid condition, which in turn, occurred sooner than in the late condition.

The graph in Fig. 4 highlights the time-locked nature of the eye movements to the disambiguating words. The bars represent the mean duration of the disambiguating words in the early, mid, and late conditions. The crosses indicate the mean eye-movement latencies to the target objects measured from the onset of the disambiguating words. If we take into account that the programming of a saccadic eye movement begins about 200 msec before the saccade is launched (Matin, Shao, & Boff, 1993), then subjects established reference on average 75 msec after the *offset* of the disambiguating words in the early and mid conditions and about 200 msec after the *onset* of the disambiguating words in the late condition. The faster latencies in the late condition suggest that, as the noun phrase unfolded, the information from each word was used to reduce the candidate set of blocks to just the two potential referents which were then distinguished by the last word of the noun phrase.

Thus, there were three important outcomes of the first experiment: First, it provided evidence for our intuition that we process spoken language incrementally. Second, the speed of the incremental processing demonstrated that a nonlinguistic visual context was rapidly integrated with the spoken linguistic input. And third, the experiment demonstrated the usefulness of the eye-movement methodology for investigating spoken language comprehension as it occurs in natural contexts.

However, there is an alternative account of the results of the first experiment that must be ruled out before the ramifications of these three outcomes can be further explored. Specifically, one could argue that, because the experiment involved very simple displays as well as instructions, the subjects developed a strategy that involved listening for specific words in the instructions rather than parsing and interpreting the entire instruction. To address this possibility, we examined whether incremental processing would be observed under more complex circumstances—circumstances where both the instructions and the displays were more complicated (Eberhard, Tanenhaus, Spivey-Knowlton, & Sedivy, 1995).

In this experiment subjects moved miniature playing cards according to spoken instructions like "Put the five of hearts that is below the eight of clubs above the three of diamonds." As in the preceding experiment, we manipulated the point in the instruction when disambiguating information occurred by giving it in three different display conditions depicted in Fig. 5.

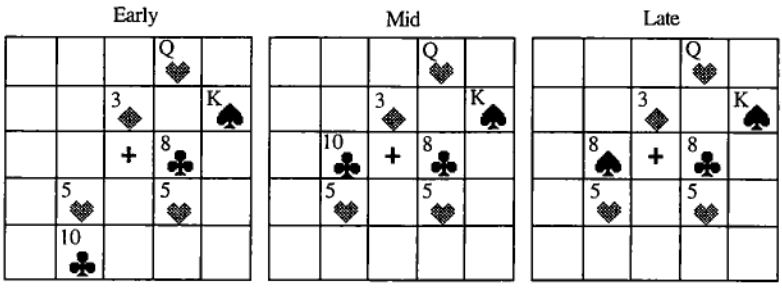


Fig. 5. Example displays for the three point-of-disambiguation conditions in Experiment 2.

All three conditions contained two potential target cards, e.g., two fives of hearts, and five other cards. In the early condition, the target card was disambiguated by the word *above* or *below* in the postmodifying clause because the other potential target card was not above or below any other card. (At the beginning of the experiment, subjects were told that the terms *above* and *below* would refer to a square that was immediately above or below a particular card.) In the mid condition, both target cards were below what will be referred to as "context" cards; therefore, the target card was disambiguated later in the postmodifying clause by the name of its context card, e.g., *eight*. In the late condition, the target was disambiguated by the last word of the complex noun phrase, e.g., *clubs* because the target's and potential target's context cards had the same name but different suit.

Seventeen displays were randomly presented to each subject in this experiment. Four displays represented each of the three point of disambiguation conditions. Subjects were permitted to watch the experimenter position the cards for each new display. The experimenter began every trial with the command "Look at the cross," which was located in the center of the display. The experimenter then read four or five instructions from a script using her natural speaking rate. Like the critical instructions, the filler instructions asked the subject to move a card to a location that was above or below another card (e.g., "Put the King of spades below the Queen of hearts. Now put the King of spades below the cross."). Although the five filler displays did not contain any identical cards, four of them were accompanied by an instruction that referred to a card with a complex noun phrase containing a postnominal clause (e.g., "Put the six of clubs that is above the ten of diamonds below the eight of spades."). The order of the filler and critical instructions varied across the displays to prevent subjects from predicting, before an instruction was given, which card they would be asked to move. Although the two identical cards in the critical displays were always

equidistant from the focus cross, their positions as well as the positions of the miscellaneous cards varied across the critical displays.

Figure 6 contains a graph of the mean eye-movement latencies in each of the three point-of-disambiguation conditions measured from the onset of the first content word of the critical noun phrase, e.g., from the word *five*, until the eye movement to the target card that preceded the reaching for the card. As in the first experiment, the effect of point of disambiguation was reliable and provides evidence that, even under these more complex circumstances, spoken language comprehension was incremental.

Figure 7 contains a graph of the mean eye-movement latencies to the target cards measured from the onset of the disambiguating words. The bars indicate the average duration of the disambiguating words in the early, mid, and late conditions. Again, if we take into account that the programming of an eye movement begins about 200 msec before the initiation of the eye movement, then reference was established on average about 500 msec *after* the offset of the disambiguating words in the early and mid conditions and about 100 msec *before* the offset of the disambiguating word in the late condition. Although the overall latencies to establish reference in this experiment differed quantitatively from the latencies in the first experiment with simple blocks, they were nonetheless qualitatively the same. The longer

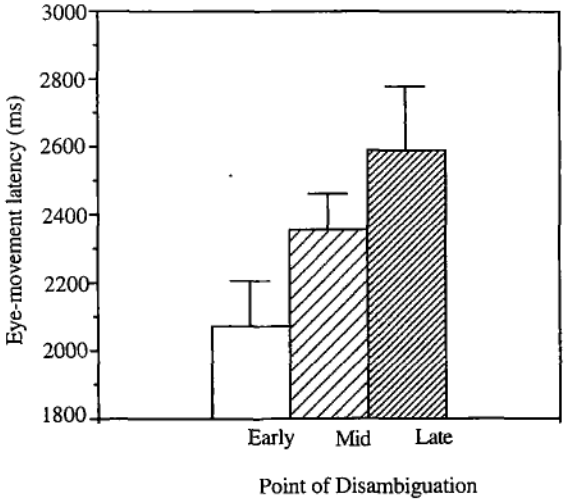


Fig. 6. Mean eye-movement latencies and standard errors, measured from the onset of the first content word of the complex noun phrase in each of the point-of-disambiguation conditions of Experiment 2.

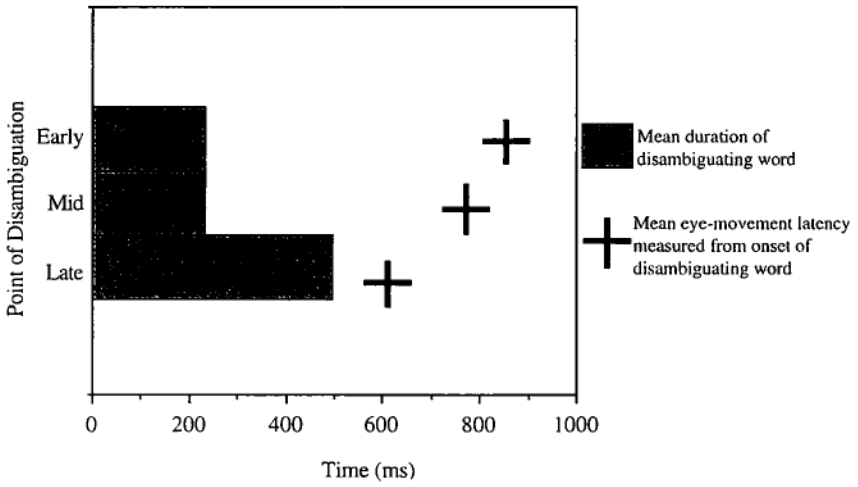


Fig. 7. Mean durations of the disambiguating words and mean eye-movement latencies to the target objects (measured from the onset of the disambiguating words) in Experiment 2.

latencies in this experiment can be attributed to the larger displays and less discriminable objects.

Unlike in the simpler blocks experiment, subjects in this experiment typically made several eye movements to various objects in the display as the complex noun phrase unfolded. Consider, for example, the sequence of eye movements that was made by a subject who received the instruction "Put the five of hearts that is below the eight of clubs above the three of diamonds" in the mid display condition (see Fig. 5). Upon hearing the word *hearts*, the subject moved her eyes to the five of hearts on the left of the display. She then looked at the other five before coming back to the first one. As she heard the word *below* she looked at the context card that the five was below, which was the ten of clubs. Shortly after hearing the disambiguating word *eight* she looked over to the other context card, which was the eight of clubs. She then made an eye movement to the target five below the eight, which indicated that she had established reference. This was verified by the fact that she soon grasped the five. While she grasped the five, she looked to the three of diamonds in the middle of the spoken word *diamonds* demonstrating that she was now attempting to establish where the card was to be put.

The graph in Fig. 8 shows the probability that subjects looked at the target card, the wrong target card (e.g., the other five of hearts), and one of the irrelevant or miscellaneous cards in the display during five temporal

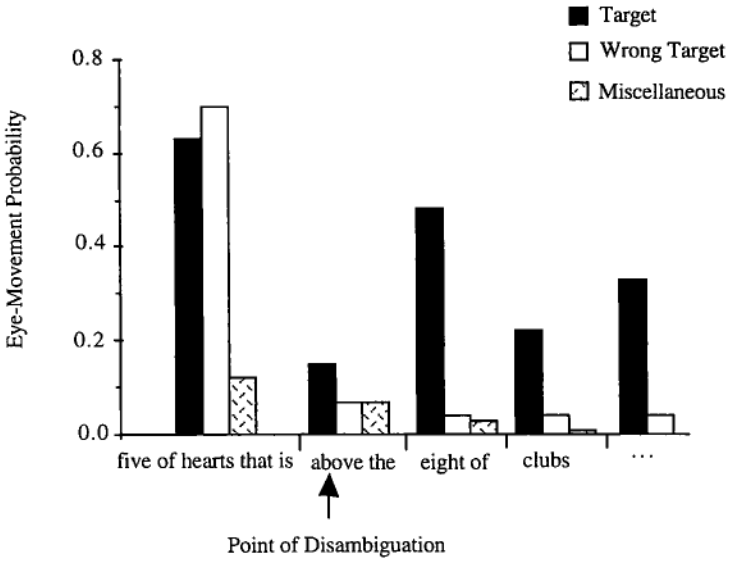


Fig. 8. Probabilities of eye movements to the cards in the early condition of Experiment 2. The probabilities that listeners looked at the target card, the wrong target card, or one of the irrelevant cards in the display are given for five segments of the complex noun phrase. The probabilities do not sum to one in each segment because some trials involved multiple eye movements or no eye movements during a segment of the speech stream.

segments of the complex noun phrase when it was given in the early condition. Probabilities were calculated by dividing the total number of looks to a particular card across all subjects by the total number of looks that occurred during the particular segment of the speech stream.

The first region, which corresponds to *five of hearts that is*, is ambiguous as to which five of hearts is the target card, so there is a high probability of looks to both the target and wrong target cards in this segment. Both probabilities are much higher than the probability of a look to an irrelevant or miscellaneous card. Disambiguating information arrives in the next segment, and there is a decrease in all three probabilities. However, in the segment immediately following, there is a sharp increase in the probability of a look to the target card (.48). This increase or "peak" in probability indicates the establishment of reference on the basis of the preceding disambiguating information.

A similar pattern of probabilities was also observed in the mid and late conditions. In the mid condition the probability of a look to the target card peaked at .61 in the segment immediately following the disambiguating

word *eight*, and in the late condition the probability of a look to the target card peaked at .59 in the segment immediately after the disambiguating word *clubs*. The peak probabilities in all three conditions occurred about 400 msec after the offset of the disambiguating words. This estimate of when reference was established is less conservative than the measurements given in Fig. 6. According to this estimate, the programming of an eye movement to the target card occurred about 200 msec after the offset of the disambiguating words, once again demonstrating that listeners rapidly integrate the visual input with the spoken input and establish reference as soon as they receive disambiguating information.

In the mid and late conditions the target and wrong target cards were above or below what were referred to as context cards. In the example depicted in Fig. 5, the target five was below the eight of clubs in both conditions and the wrong target five was below a ten of clubs in the mid condition and the eight of spades in the late condition. We found that in the segments of the spoken noun phrase where information made these context cards relevant (i.e., *above the* and *eight of*) the probabilities of looking at each of these two cards exceeded the probabilities of looking at any of the other cards in the display, including the target and wrong target cards. This finding shows that subjects looked at objects in the visual discourse model that were made relevant by the current spoken input. In sum, evidence from both the pattern of eye-movement probabilities and eye-movement latencies clearly shows that subjects in this experiment did not simply listen for specific words in the speech stream; rather, they interpreted each word of the spoken input with respect to the visual discourse model and established reference as soon as distinguishing information was received.

THE EFFECTS OF CONTRASTIVE STRESS ON THE INCREMENTAL PROCESSING OF SPOKEN NOUN PHRASES

Although much of the information for establishing reference comes from the words of a spoken expression, additional information may be conveyed by the pitch, amplitude, and timing variations of those words, i.e., by their prosody. Specifically, our focus was on how prosody can be used to direct listeners attention to relevant entities and their properties. This next experiment examined whether disambiguating information that is conveyed by contrastive stress would be used on-line by listeners to establish reference (Sedivy, Carlson, Tanenhaus, Spivey-Knowlton, & Eberhard, 1994).

Contrastive stress signals that a specific set of referents is relevant to the discourse, namely a "contrastive" set whose members share many properties but differ on one particular property. For example, consider the effect

of contrastive stress (represented by capital letters) in the instruction "Touch the LARGE blue square" when it is given in the context of Fig. 9a. The stress makes relevant the two objects that differ only on the dimension of size (i.e., the two blue squares), and it selects the one that is positive on the stressed dimension (e.g., "Touch the LARGE blue square, not the SMALL one") (Jackendoff, 1972; Kanerva, 1990; Krifka 1991; Rooth, 1985, 1992). Thus, although there are two large objects in the display, contrastive stress on the adjective *large* provides disambiguating information because only one of the large objects is a member of a contrastive set. When the same instruction is uttered with neutral stress the disambiguating information does not occur until the following word *blue*.

We examined whether listeners can make immediate use of the disambiguating information conveyed by contrastive stress by giving instructions like "Touch the large blue square" uttered with contrastive or neutral stress on *large* in displays like the one depicted in Fig. 9a (see Sedivy et al., 1994; Sedivy et al. 1995; Tanenhaus et al., in press). We predicted that, if the disambiguating information from contrastive stress can be used online, then eye movements to the target objects should occur sooner in the stressed condition than in the neutral condition.

However, there is an alternative account for any facilitatory effects that may be observed in the stressed condition. Because stressing a word increases its duration, listeners will have more time in this condition to use the information from the word *large* to reduce the set of objects to just the two large ones and this may facilitate the processing of the disambiguating information from the following word *blue*, irrespective of any contrastive set. Therefore to ensure that any facilitatory effects observed in the stressed condition were due to the disambiguating information provided by contrastive stress, both the stressed and neutral instructions were also presented in displays like the one depicted in Fig. 9b. Because these displays contain two contrastive sets, the stress account predicts that stressing *large* will not pro-

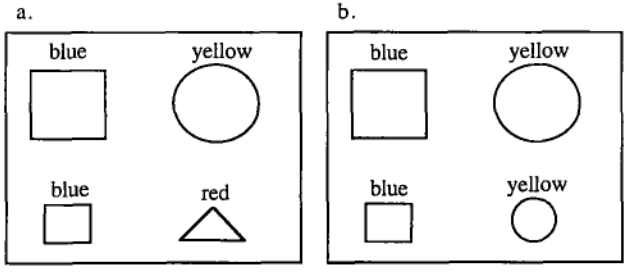


Fig. 9. Example displays representing the one- and two-contrast set conditions in Experiment 3.

vide disambiguating information, and therefore no facilitatory effects should be observed with this display. However, the alternative duration-based account predicts facilitatory effects from stress with these displays as well because, as in the display represented in Fig. 9a, there are only two large objects.

In all conditions, eye-movement latencies to the target objects were measured from the onset of the color adjective. As shown in Fig. 10, eye-movement latencies were faster with stressed instructions than with neutral instructions, but only when the displays contained a single contrast set, thus providing evidence that information conveyed by contrastive stress was used on-line to reduce the set of visual referents to the intended one.

THE EFFECTS OF A VISUALLY-PRESENT COHORT COMPETITOR ON THE INCREMENTAL PROCESSING OF SUBLEXICAL INFORMATION

Evidence from studies of spoken word recognition show that recognition occurs shortly after the spoken input uniquely specifies a lexical item from the mental lexicon. For example, the word *elephant* would be recognized shortly after the "phoneme" /f/. Prior to that point, the spoken input is consistent with other words such as *eloquent*, *elevator*, and *elegant*. Thus, the recognition of a word is influenced by the words it is phonetically similar to. The set of words that is phonetically similar to a target word is referred to as its "cohort" (Marslen-Wilson, 1987).

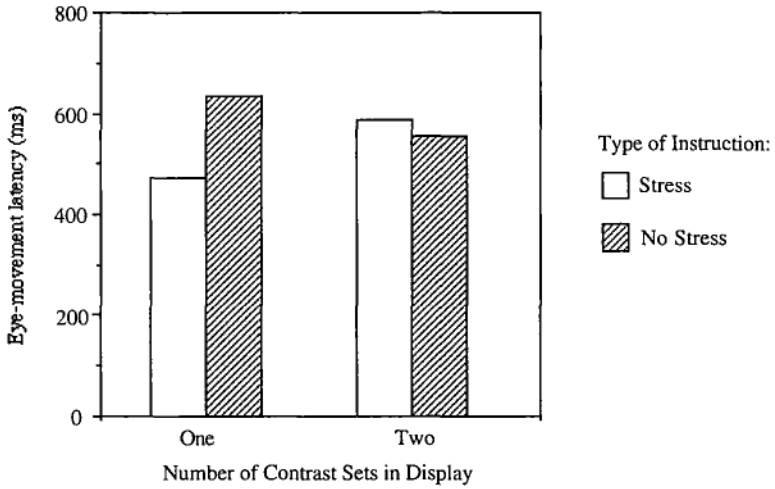


Fig. 10. Mean eye-movement latencies to the target objects measured from the onset of the color adjective in the four conditions of Experiment 3.

Using the head-mounted eye-tracker paradigm, we examined whether the amount of spoken input that a listener would need to establish reference would be influenced by the *names* of the referents in the visual context (Spivey-Knowlton, Sedivy, Eberhard, & Tanenhaus, 1994; Tanenhaus et al., in press). In particular, we hypothesized that eye movements to a target object (e.g., candle) would be slower when the context contained a "competitor" object with a similar name to the target (e.g., candy) compared to when it did not. Although the first three experiments provided strong evidence for the rapid effects of the visual context on spoken language processing, findings of a visual cohort "competitor" effect would demonstrate that the relevant visual context affected even the earliest moments of processing the spoken input. In addition, it would further demonstrate the usefulness of this methodology as an on-line measure that can be applied to issues in spoken word recognition.

In this experiment, we instructed subjects to move common objects in displays like the one depicted in Fig. 11. On some trials, the displays contained a target object whose name shared initial sounds with the name of another "competitor" object in the display (e.g., candle and candy). On other trials, the target appeared in the display without its competitor. The same critical instruction, e.g., "Pick up the candle," was given in both the competitor-present and -absent displays. Additional filler instructions directing the subject to move nontarget objects were also given (e.g., "Pick up the mouse. Now put it below the box").

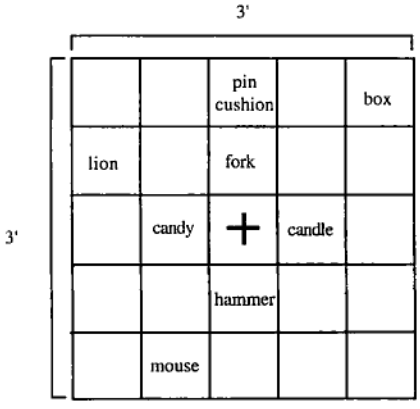


Fig. 11. Example display representing the competitor-present condition in Experiment 4. The critical instruction was "Pick up the candle."

The results, which are summarized in Fig. 12, showed a reliable competitor effect. The mean eye-movement latency to the target object measured from the onset of the target's spoken name was slower when a competitor object was present in the display relative to when it was not. The average duration of the spoken target names was 300 msec. Taking into account that it takes 200 msec to program an eye movement, listeners established reference on average 55 msec *before* the offset of the target's name in the competitor-absent condition, and 30 msec *after* the offset of the name in the competitor-present condition. This finding of incremental interpretation within words provided striking evidence for the rapidity with which spoken information can be integrated with the visual input.

THE INCREMENTAL PROCESSING OF PP-ATTACHMENT AMBIGUITIES: EVIDENCE FOR THE IMMEDIATE EFFECTS OF A VISUAL REFERENTIAL CONTEXT ON SYNTACTIC PROCESSING

Given the evidence we have just reviewed for immediate effects of a real-world visual context on the interpretation of spoken input, and the sensitivity of the eye movements to the immediate comprehension processes, we decided to examine whether a real-world visual context could influence listeners' initial syntactic decisions (Spivey-Knowlton et al., 1995; Tanen-

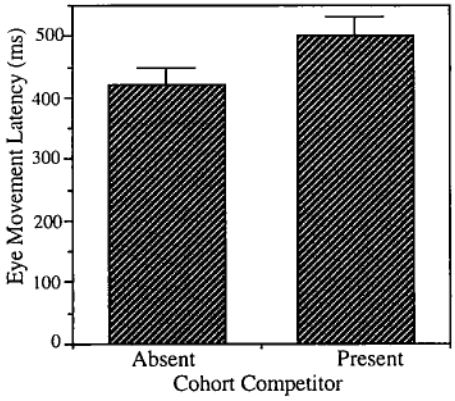


Fig. 12. Mean eye-movement latencies and standard errors to the target objects measured from the onset of the targets' names in the competitor-present and -absent conditions of Experiment 5.

haus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). We manipulated referential contexts for instructions with prepositional phrase (PP) ambiguities using the verb *put* in double prepositional phrase constructions such as "Put the saltshaker on the envelope in the bowl." As we discussed in the introduction to this article, the first prepositional phrase (*on the envelope*) is ambiguous as to whether the speaker intends it to be the goal of the putting event, i.e., the place where the saltshaker is to be moved, or a modification of the theme of the putting event further specifying the properties of the saltshaker, i.e., it is currently on the envelope.

The double-PP construction has two important properties. First, all of the parsing models that posit an encapsulated first stage predict that the first PP will be initially interpreted as the goal rather than as a modifier of the theme: The verb phrase attachment is the minimal attachment (Frazier, 1987), it assigns the PP as an argument rather than as an adjunct (Abney, 1989), and the goal argument is an obligatory argument for the verb *put* (Britt, 1994). Second, studies investigating the written comprehension of sentences with double-PP constructions using verbs with obligatory goal arguments, such as *put*, have shown that readers initially interpret the first PP as specifying the goal even when the prior linguistic context supports the modifier interpretation by introducing two possible referents for the theme (Britt, 1994; Ferreira & Clifton, 1986). These results have provided some of the strongest evidence for the initial encapsulation of syntactic processing.

However, as we noted earlier, in reading studies the context must be maintained in memory and therefore may not be immediately accessible. A stronger test of the encapsulation hypothesis would come from examining the on-line processing of ambiguous sentences when they are uttered in real-world discourse situations. Thus, in this last experiment we presented spoken instructions like (1) that contained a PP-attachment ambiguity as well as

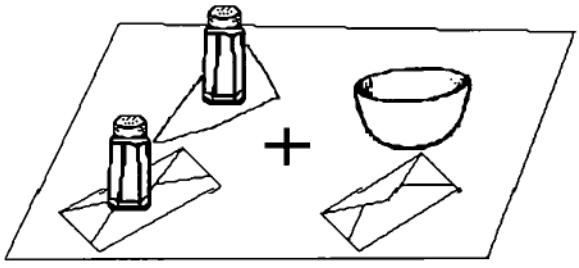


Fig. 13. Example display for the two-referent context condition in Experiment 5.

unambiguous control instructions like (2) in contexts like the one depicted in Fig. 13 that supported the less-preferred modifier interpretation:

(1) Put the saltshaker on the envelope in the bowl.
(2) Put the saltshaker that's on the envelope in the bowl.

The visual contexts contained two possible referents for the definite noun phrase *the saltshaker*. The two saltshakers were distinguished by an attribute, namely which object they were on. Thus, the contexts created conditions where modification was needed for felicitous reference (cf. Altmann, 1987; Altmann & Steedman, 1988; Crain & Steedman, 1985).

We reasoned that if the syntactic processing of spoken input were initially structured independently of context (e.g., Frazier, 1987), then listeners should show evidence of misinterpreting the PP *on the envelope* as specifying the location of where the object is to be put (i.e., a goal interpretation) in these "two-referent" contexts. If, however, listeners interpreted the spoken input very rapidly with respect to the discourse model, as our previous experiments have demonstrated, they would know at the point of hearing *saltshaker* that they did not yet have enough information to identify the intended theme referent; therefore, they might immediately and correctly interpret the first PP *on the envelope* as providing the necessary information for distinguishing which saltshaker was the intended theme. If so, this would provide definitive evidence against parsing models that postulate an initial stage of encapsulated syntactic processing.

In addition to presenting ambiguous and unambiguous instructions like (1) and (2) in a "two-referent" context, we presented them in a "one-referent" context like the one depicted in Fig. 14. Although the modifier interpretation of the ambiguous instruction was the correct interpretation in this context as well, the context did not support that interpretation. Both the incremental processing account and the encapsulated account predict that listeners will initially misinterpret the first ambiguous PP *on the envelope*

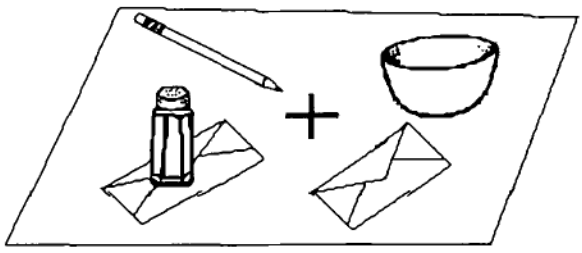


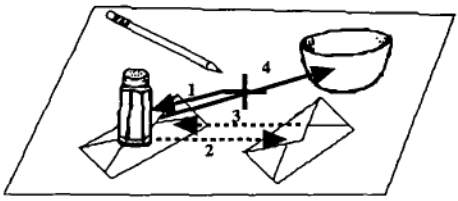
Fig. 14. Example display for the one-referent context condition in Experiment 5.

as the goal in this context; however, they predict misinterpretation for different reasons. Again, the encapsulated account predicts misinterpretation because, regardless of the context, it assumes that the sentence will initially be assigned the simpler syntactic construction, which corresponds to a goal interpretation of the first PP. The incremental account predicts misinterpretation because, in this context, listeners will be able to unambiguously establish the theme referent upon hearing *saltshaker*; therefore, they should move on to establish the goal referent. Although it will ultimately be incorrect, the visual context allows listeners to interpret the ambiguous PP *on the envelope* as the goal (i.e., there is an envelope available for the saltshaker to be placed upon). Thus, they cannot realize that the goal interpretation is incorrect until after they receive the second PP.

Four lists each containing 12 different critical displays were constructed for this experiment. Each list was presented to two subjects. Half of the displays on each list represented the one-referent context condition and half represented the two-referent context condition. An equal number of ambiguous and unambiguous instructions were given in the two display conditions. Across all four lists each of the 12 displays appeared once representing each of the four conditions. The critical instructions were always preceded by the instruction "Look at the cross." Several filler instructions were also given in each of the critical displays. In addition, 18 filler displays were presented to prevent subjects from anticipating the nature of the instructions.

The results showed very different eye-movement patterns when the ambiguous instruction was given in the one- versus two-referent display conditions. Consistent with the predictions of the incremental account, the ambiguous phrase *on the envelope* was initially interpreted as the goal of the putting event in the one-referent context but as the modifier of the theme referent in the two-referent context. When the ambiguous instructions were given in the one-referent context, subjects looked at the incorrect goal (e.g., the irrelevant envelope) on 55% of the trials shortly after hearing the first ambiguous PP. They never looked at the incorrect goal when the unambiguous control instructions were given in this context. In contrast, when the ambiguous instruction was given in the two-referent context, subjects looked at the incorrect goal on only 17% of the trials, and this percentage did not differ reliably from the percentage of trials that subjects looked at the incorrect goal when the unambiguous instruction was given in this condition (11%).

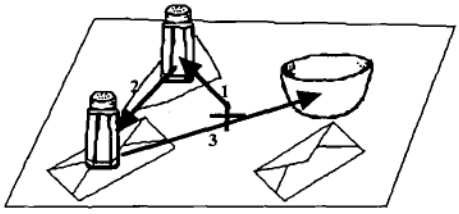
The sequences and timing of the eye movements that occurred when the ambiguous and unambiguous instructions were given in the one- and two-referent context conditions are summarized in Figs. 15 and 16. In the one-referent condition (Fig. 15), subjects first looked at the target object (the saltshaker) 500 msec after hearing *saltshaker*. They then looked at the in-



Ambiguous: Put the saltshaker on the **1** envelope in **2** the bowl. **3** **4**

Unambiguous: Put the saltshaker that's **1** on the envelope in the bowl. **4**

Fig. 15. Typical sequence and timing of eye movements to the objects in the one-referent display condition for both the ambiguous and unambiguous instructions. The solid arrows indicate eye movements that occurred during both instructions. The dashed arrows indicate the additional eye movements that occurred during the ambiguous instruction. The numbers in the instructions indicate the point at which the eye movements typically occurred.



Ambiguous: Put the saltshaker on the **1** envelope **2** in the bowl. **3**

Unambiguous: Put the saltshaker that's **1** on the envelope **2** in the bowl. **3**

Fig. 16. Typical sequence and timing of eye movements to the objects in the two-referent display condition for both the ambiguous and unambiguous instructions. The solid arrows indicate eye movements that occurred during both instructions. The numbers in the instructions indicate the point at which the eye movements typically occurred.

correct goal referent (the envelope on the right of the display) 484 msec after hearing *envelope*, thus providing evidence of a misinterpretation. When the unambiguous instruction was given in this condition, the subjects also first looked at the target object shortly after hearing *saltshaker*. However, they never looked at the incorrect goal referent; instead, they looked to the correct goal referent (the bowl) shortly after hearing *bowl*.

When the ambiguous instruction was given in the two-referent condition (Fig. 16), the subjects often looked at both saltshakers during the segment of the speech stream *the saltshaker on the envelope*, reflecting the fact that the intended theme referent could not be unambiguously identified until the word *envelope*. However, as stated above, they rarely looked at the

incorrect goal referent (the envelope on the right); instead, they looked at the correct goal referent (the bowl) shortly after hearing *bowl*. In fact, the timing and pattern of eye movements to the objects in the two-referent displays for the ambiguous instructions were nearly identical to the timing and pattern for the unambiguous instructions, indicating that, in both cases, the subjects initially interpreted the first PP *on the envelope* correctly as a modifier of the theme. Thus, the results of this experiment provided clear evidence that subjects interpreted the spoken input rapidly and incrementally with respect to the discourse model and, as a result, the visual referential context had an immediate effect on the syntactic structure that was assigned to the linguistic input.

These results are clearly inconsistent with modular theories of the language processing system in which initial syntactic commitments are made by an encapsulated syntactic processing module, i.e., when faced with several possible structural assignments, the syntactic processor makes an initial commitment based on processing principles that make reference only to grammatical information. The garden-path model of Frazier and colleagues, which postulates that all initial syntactic commitments are made using structural principles, is the best-known example of this type of model (cf. Frazier, 1987). In addition, a number of other researchers have proposed models in which initial commitments for the type of structure that we manipulated here are made without reference to contextual constraints (e.g., Britt, 1994; Mitchell, Corley, & Garnham, 1992; Perfetti, 1990; Pritchett, 1992).

However, one could argue that our results are fully compatible with modular models that assume that an encapsulated syntactic processor proposes alternative structures in parallel and then context is used to select from among these alternatives during an immediately subsequent integration stage. A detailed comparison of these two-stage "propose and dispose" models and nonmodular constraint-based models would take us well beyond the scope of this article, so we will limit ourselves to a few brief remarks (more detailed discussions of our views on these issues can be found in Boland, Tanenhaus, Garnsey & Carlson, in press; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton et al., 1993; Spivey-Knowlton & Tanenhaus, 1994, 1995; Tanenhaus & Trueswell, 1995; also see MacDonald et al., 1994, for related discussions).

The best-known example of a propose and dispose model is the referential theory developed by Crain, Steedman, Altmann, and colleagues (e.g., Altmann, 1987; Altmann & Steedman, 1988; Crain & Steedman, 1985; Steedman, 1987) in which all ambiguity resolution is accomplished by discourse-based principles. The specific proposals made by these researchers are, in fact, compatible with the results of the above experiment. This should come as no surprise because our visual contexts manipulated referential con-

straints. What makes our results so striking is that prior studies using linguistic contexts with sentences similar to ours have often reported only weak or delayed effects of referential context (Britt, 1994; Ferreira & Clifton, 1986), leading many researchers to conclude that discourse context is used only after an initial syntactic commitment to a preferred structure. The reason why discourse constraints have relatively weak effects in these studies is that discourse constraints operate in conjunction with other constraints. Thus, the effects of linguistic referential contexts are strongest when other constraints are weakest (Britt, 1994; Spivey-Knowlton et al., 1993; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton & Tanenhaus, 1994). Similar results hold for other types of nonsyntactic contextual constraints (for a recent review see Tanenhaus & Trueswell, 1995). Crucially, there does not appear to be a temporal window in which syntactic processing is not sensitive to relevant nonsyntactic constraints. Because a visual context and spoken instructions to manipulate it maximize the strength, relevance, and availability of referential constraints, we observe clear and immediate referential effects, even when they oppose other local constraints.

Of course, one could still argue that the parser is encapsulated, but that all of the effects that we can observe with our psychophysical measures are integration effects, i.e., the syntactic processing system is encapsulated and experimentally impenetrable. This view of encapsulated processing reduces to the following empirical claims: (1) The language processing system exhibits "bottom-up priority" (Marslen-Wilson, 1987); (2) the language processing system is sensitive to syntactic constraints. The same claims are embodied in constraint-based models which assume that grammatical constraints are integrated into processing systems that coordinate linguistic and nonlinguistic information as the input is processed. Thus, in its reduced form, the encapsulation hypothesis no longer does any work, and "propose and dispose" models become notational variants of constraint-based models.

Furthermore, because syntactic processing is tied to lexical information, which is processed continuously (i.e., as it unfolds over time), syntactic ambiguity arises at virtually every point in an utterance, and, as our experiments have demonstrated, is resolved almost immediately. Thus, viewed from the perspective of real-time processing, it is not clear what the theoretical import is of labeling processing as "encapsulated" when the encapsulation holds only up to the point of ambiguity (i.e., when processing is encapsulated but ambiguity resolution is constraint based).

GENERAL DISCUSSION AND CONCLUSION

We have reviewed research showing that eye movements that accompany performance of natural tasks can provide a detailed profile of the rapid

mental processes that underlie real-time spoken language comprehension. When the context is available and salient, listeners process the input incrementally, establishing reference with respect to their behavioral goals during the earliest moments of linguistic processing. Referentially relevant pragmatic information (e.g., contrast sets) is immediately computed and used in resolving reference. Moreover, relevant nonlinguistic information affects the speed with which individual words are recognized as well as how the linguistic input is initially structured.

We believe that these results have important methodological and theoretical implications for research in language comprehension. As Clark (1992) pointed out, there have been two main theoretical approaches to research in language processing: a language-as-action approach and a language-as-product approach. The language-as-action approach emphasizes the intentional nature of language use: Utterances are spoken with a communicative purpose and the primary goal of the comprehension system is to recognize that purpose. The realization of that goal crucially depends on the situational context in which the utterance occurs, i.e., the particular speaker, time, place, and circumstance of the utterance. Thus, research in this tradition typically employs methodologies that examine goal-directed conversational interactions between speakers and addressees.

In contrast, much of the research in the language-as-product approach has focused on how the comprehension system recovers a sentence's linguistic structure. Implicit in this research is the assumption that the assignment of a structure occurs at a stage prior to interpretation and proceeds largely independently of the situational context. As a result, much of the research in this tradition has examined the comprehension of sentences under circumstances that strip them of their illocutionary force, e.g., circumstances in which listeners are seated in a room by themselves and are asked to comprehend sentences presented on a computer screen or over headphones while various sorts of their reactions to those sentences are measured. Research in this tradition has emphasized the use of methodologies that provide fine-grained information about the time course of comprehension. However, these methodologies have not been applicable to studying *spoken* language comprehension as it occurs in natural interactive situations where the context is highly constraining.

Although no behavioral measurement provides a direct window onto the parser, the eye-movement methodology does have important advantages over other methodologies used to investigate spoken language comprehension. In particular, it provides an opportunity for investigating the time course of comprehension in well-defined interactive situations. This is important because although studying language in a "vacuum" can provide information about some aspects of comprehension, a complete and accurate

theory of comprehension requires knowing how those aspects are affected when language is processed in situations at the other end of the continuum, i.e., in highly structured situational contexts where the communicative intent of the sentences is recognizable and relevant to the listener's behavioral goals. When listeners know the general purpose of the communicative exchange, as they typically do in natural situations, they have strong expectations about how the speaker will behave linguistically. Specifically, according to Grice (1975), listeners expect speakers to make their utterances as informative as is required for the purpose of the exchange and to not make their utterances more informative than is required. From this perspective, the ambiguous double-PP instructions in our fifth experiment were felicitous in the two-referent condition, because they provided enough information for the listener to identify which of the two possible referents was the intended theme referent, but they were infelicitous in the one-referent condition because they provided too much information for the listener to identify the intended theme. The listeners' on-line comprehension performance in these two conditions was consistent with this felicity difference.

We are not suggesting that research conducted in such well-defined situations provides the only definitive answer about how language is processed, but we would like to point out that studying language in impoverished situations encourages, rather than challenges, a modular information-encapsulation view of the system. An adequate model must be able to accommodate the range of results that come from both approaches to studying language processing. We believe that constraint-based models (e.g., MacDonald et al., 1994; Spivey-Knowlton et al., 1993, Tanenhaus & Trueswell, 1995) which emphasize the incremental and integrative nature of language processing are better equipped for handling the range of results than models that assign a central role to encapsulated subsystems.

REFERENCES

Abney, S. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18*, 129–144.

Altmann, G. (1987). Modularity and interaction in sentence processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp 249–257). Cambridge: MIT Press.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing, *Cognition, 30*, 191–238.

Bates, E. & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Boland, J., Tanenhaus, M. K., Garnsey, S., & Carlson, G. (in press). Argument structure and filler-gap assignment. *Journal of Memory and Language*.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251–283.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clark, H. H., & Marshall, C. R. (1992). Definite reference and mutual knowledge. In H. H. Clark, (Ed.), *Arenas of language use* (pp. 9–59). Chicago: University of Chicago Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Kartunnen, & H. Zwicky (Eds.), *Natural language parsing*. Cambridge, England: Cambridge University Press.

Eberhard, K., Tanenhaus, M., Spivey-Knowlton, M., & Sedivy, J. (1995). *Investigating the time course of establishing reference: Evidence for rapid incremental processing of spoken language*. Unpublished manuscript.

Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII*. Hove, England: Erlbaum.

Grice (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts*. New York: Academic Press.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Kanerva (1990). Focusing on phonological phrases in Chichewa. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection* (pp. 145–161). Chicago: University of Chicago Press.

Krifka, M. (1991). A compositional semantics for multiple focus constructions. *Proceedings of Semantics and Linguistic Theory (SALT) 1* (Cornell University Working Papers 11.)

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Marslen-Wilson, W., & Tyler, K. (1987). Against modularity. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp. 37–62). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71–102.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics, 53*, 372–380.

Mitchell, D. C., Corley, M. M. B., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 69–88.

Olson, D. (1970). Language and thought: Aspects of cognitive theory of semantics. *Psychological Review, 77*, 143–184.

Perfetti, C. A. (1990). The cooperative language processors: Semantic influences in an autonomous syntax. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Rooth, M. (1985). *Association with focus*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75–116.

Sedivy, J., Carlson, G., Tanenhaus, M., Spivey-Knowlton, M., & Eberhard, K. (1994). The cognitive function of contrast sets in processing focus constructions. In *Working Papers of the IBM Institute for Logic and Linguistics*.

Sedivy, J., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 375–380). Mahwah, NJ: LEA Associates.

Spivey-Knowlton, M., & Sedivy, J. (1995). Parsing attachment ambiguities with multiple constraints. *Cognition, 55*, 227–267.

Spivey-Knowlton, M., Sedivy, J., Eberhard, K., & Tanenhaus, M. (1994). Psycholinguistic study of the interaction between language and vision. In *Proceedings of the 12th National Conference on Artificial Intelligence: Workshop on the Integration of Natural Language and Vision Processing*.

Spivey-Knowlton, M., & Tanenhaus, M. K. (1994). Referential context and syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: Erlbaum.

Spivey-Knowlton, M., & Tanenhaus, M. (1995). *Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency within an integration-competition framework*. Manuscript submitted for publication.

Spivey-Knowlton, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (1995). Eye-movements accompanying language and action in a visual context: Evidence against modularity. (Eds.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 25–30). Mahwah, NJ: LEA Associates.

Spivey-Knowlton, M., Trueswell, J., & Tanenhaus, M. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 37*, 276–309.

Steedman, M. (1987). Combinatory grammars and human language processing. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural language understanding* (pp. 187–205). Cambridge, MA: MIT Press.

Swinney, D., & Osterhout, L. (1990). Interference generation during auditory language comprehension. *The Psychology of Learning and Motivation, 25*, 17–33.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). The interaction of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (in press). Using eye-movements to study spoken language comprehension: Evidence for visually-mediated incremental interpretation. In T. Inui & J. McClelland (Eds.), *Attention & Performance XVI: Integration in Perception and Communication*.

Tanenhaus, M., & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception: Vol. 11. Speech and Language*. San Diego, CA: Academic Press.

Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role expectations. *Journal of Memory and Language, 27*, 597–632.

Taraban, R., & McClelland, J. (1990). Sentence comprehension: A multiple constraints view. In D. Balota, K. Rayner, & G. Flores d'Arcais (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.