INVITED TALK

Eye Movements and Spoken Language Comprehension

Michael K. Tanenhaus*
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY 14627
mtan@bcs.rochester.edu

Julie C. Sedivy
Department of Linguistics
University of Rochester
Rochester, NY 14627
sedivy@bcs.rochester.edu

Michael J. Spivey-Knowlton
Department of Psychology
Cornell University
Ithaca, NY 14583
mjsk@cornell.edu

Paul D. Allopenna
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY 14627
allopen@bcs.rochester.edu

Kathleen M. Eberhard
Department of Psychology
University of Notre Dame
Notre Dame, IN 46556
kathleen.m.eberhard.1@nd.edu

James S. Magnuson
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY 14627
magnuson@bcs.rochester.edu

Abstract

We present an overview of recent work in which eye movements are monitored as people follow spoken instructions to move objects or pictures in a visual workspace. Subjects naturally make saccadic eye-movements to objects that are closely time-locked to relevant information in the instruction. Thus the eye-movements provide a window into the rapid mental processes that underlie spoken language comprehension. We review studies of reference resolution, word recognition, and pragmatic effects on syntactic ambiguity resolution. Our studies show that people seek to establish reference with respect to their behavioral goals during the earliest moments of linguistic processing. Moreover, referentially relevant non-linguistic information immediately affects how the linguistic input is initially structured.

Introduction

Many important questions about language comprehension can only be answered by examining processes that are closely time-locked to the linguistic input. These processes take place quite rapidly and they are largely opaque to introspection. As a consequence, psycholinguists have increasingly turned to experimental methods designed to tap real-time language processing. These include a variety of reading time measures as well as paradigms in which subjects monitor the incoming speech for targets or respond to visually presented probes. The hope is that these "on-line" measures can provide information that can be used to inform and evaluate explicit computational models of language processing.

Although on-line measures have provided increasingly fine-grained information about the time-course of language processing, they are also limited in some important respects. Perhaps the most serious limitation is that they cannot be used to study language in natural tasks with real-world referents. This makes it difficult to study how interpretation develops. Moreover, the emphasis on processing "decontextualized" language may be underestimating the importance of interpretive processes in immediate language processing.

Recently, we have been exploring a new paradigm for studying spoken language comprehension. Participants in our experiments follow spoken instructions to touch or manipulate objects in a visual workspace while we monitor their eye-movements using a lightweight camera mounted on a headband. The camera, manufactured by Applied Scientific Laboratories, provides an infrared image of the eye at 60Hz. The center of the pupil and the corneal reflection are tracked to determine the orbit of the eye relative to the head. Accuracy is better than one degree of arc, with virtually unrestricted head and body movements [Ballard, Hayhoe, and Pelz, 1995]. Instructions are spoken into a microphone connected to a Hi-8 VCR. The VCR also records the participant's field of view from a "scene" camera mounted on the headband. The participant's gaze fixation is superimposed on the video image We analyze each frame of the instructions to determine the location and timing of eye movements with respect to critical words in the instruction.

We find that subjects make eye-movements to objects in the visual workspace that are closely time-locked to relevant information in the instruction. Thus the timing and patterns of the eye movements provide a window into comprehension processes as the speech unfolds. Unlike most of the on-line measures that have been used to study spoken language processing in the past, our procedure can be used to examine comprehension during natural tasks with real-world referents [Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C., 1996].

In the remainder of this paper, we review some of our recent work using the visual world paradigm. We will focus on three areas: (a) reference resolution; (b) word recognition, and (c) the interaction of referential context and syntactic ambiguity resolution.

Reference Resolution

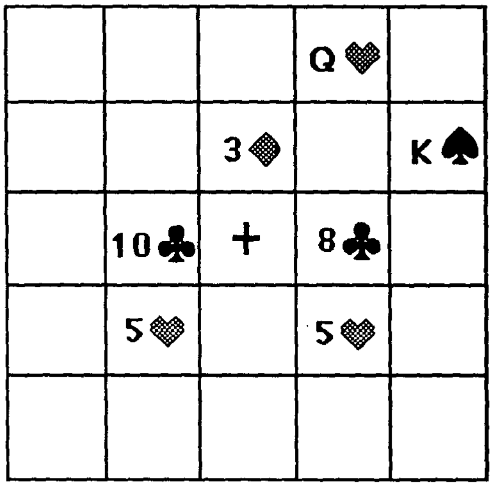
Evidence for Incremental Interpretation

In order to investigate the time course with which people establish reference we use different displays to manipulate where in an instruction the referent of a definite noun phrase becomes unique. The timing and patterns of the eye-movements clearly show that people establish reference incrementally by continuously evaluating the information in the instruction against the alternatives in the visual workspace. For example, in one experiment [Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995], participants were told to touch one of four blocks. The blocks varied along three dimensions: marking (plain or starred), color (pink, yellow, blue and red) and shape (square or rectangle). The instructions referred to the block using a definite noun phrase with adjectives (e.g., "Touch the starred yellow square."). The display determined which word in the noun phrase disambiguated the target block with respect to the visual alternatives For example, the earliest point of disambiguation would be after "starred" if only one of the blocks was starred, after "yellow" if only one of the starred blocks was yellow, and after "square" if there were two starred yellow blocks, only one of which was a square (Instructions with definite noun phrases always had a unique referent).

An instruction began with subjects looking at a fixation cross. We then measured the latency from the beginning of the noun phrase until the onset of the eye-movement to the target object. Subjects made eye-movements before touching the target block on about 75% of the trials.

Eye-movement latencies increased monotonically as the point of disambiguation shifted from the marking adjective to the color adjective to the head noun. Moreover, eye-movements were launched within 300 milliseconds of the end of the disambiguating word. It takes about 200 milliseconds from the point that an eye-movement is programmed until when the eye actually begins to move. On average then, participants began programming an eye-movement to the target block once they had heard the disambiguating word and before they had finished hearing the next word in the instruction.

We used the same logic in an experiment with displays containing more objects and syntactically more complex instructions [Eberhard et al, 1995]. Participants were instructed to move miniature playing cards placed on slots on a 5X5 vertical board. Seven cards were displayed on each trial. A trial consisted of a sequence of three instructions. On the instructions of interest, there were two cards of the same suit and denomination in the display. The target card was disambiguated using a restrictive relative clause, e.g. "Put the five of hearts that is below the eight of clubs above the three of diamonds." Figure 1 shows one of the displays for this instruction.



"Put the five of hearts that is below the eight of clubs above the three of diamonds."

Figure 1: Display of cards in which their are two fives of hearts. As each five of heart is below a different numbered card, the above instruction becomes unambiguous at "eight".

The display determined the point of disambiguation in the instruction. For the display in Figure 1, the point of disambiguation occurs after the word "eight" because only one of the fives is below an eight. We also used an early point of disambiguation display in which only one of the potential target cards was immediately below a" "context" card and a late point of disambiguation display in which the denomination of the "context" card disambiguated the target (i.e., one five was below an eight of spades and the other was below and eight of clubs).

Participants always made an eye-movement to the target card before reaching for it. We again found a clear point of disambiguation effect. The mean latency of the eye-movement that preceded the hand movement to the target card (measured from a common point in the instruction) increased monotonically with the point of disambiguation.

In addition, participants made sequences of eye-movements which made it clear that interpretation was taking place continuously. We quantified this by examining the probability that the subject would be looking at (fixating on) particular classes of cards during segments of the instruction. For example, during the noun phrase that introduced the potential targets, "the five of hearts", nearly all of the fixations were on one of the potential target cards.

During the beginning of the relative clause "...that is below the..." , most of the fixations were to one of the context cards (i.e. the card that was above or below a potential target card). Shortly after the disambiguating word, the fixations shifted to the target card.

Contrastive focus

The presence of a circumscribed set of referents in a visual model makes it possible to use eye-movements to examine how presuppositional information associated with intonation is used in on-line comprehension. [Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard & Carlson, 1995] For example, semantic analyses of contrast have converged on a representation of contrastive focus which involves the integration of presupposed and asserted information [e.g., Rooth, 1992; Kratzer, 1991; Krifka, 1991]. Thus a speaker uttering "Computational linguists give good talks" is making an assertion about computational linguists. However, a speaker who says "COMPUTATIONAL linguists give good talks." is both complimenting the community of computational linguists and making a derogatory comparison with a presupposed set of contrasting entities (perhaps the community of non-computationally oriented linguists).

We explored whether contrast sets are computed on-line by asking whether contrastive focus could be used to disambiguate among potential referents, using a variation on the point of disambiguation manipulation described earlier. We used displays with objects that could differ along three dimensions: size (large or small), color (red, blue and yellow), and shape (circles, triangles and squares). Each display contained four objects [see Sedivy et al., 1995 for details].

Consider now the display illustrated in Figure 2 which contains a small yellow triangle, a large blue circle and two red squares, one large and one small. With the instruction "Touch the large red square." the point of disambiguation comes after "red" . After "large" there are still two possible referents: the large red square and the large blue circle. After "red" only the large red square is a possible referent. . However, with the instruction "Touch the LARGE red square", contrastive focus on "large" restricts felicitous reference to objects that have a contrast member differing along the dimension indicated by the contrast (size). In the display in Figure 2, the small red square contrasts with the large red square. However, the display does not contain a contrast element for the large blue circle. Thus, if people use contrastive stress to compute a contrast set on-line, then they should have sufficient information to determine the target object after hearing the size adjective Thus eye-movements to the target object should be faster with contrastive stress. That is, in fact, what we found. Latencies to launch a saccade to the target were faster with contrastive stress than with neutral stress.

However, there is a possible objection to an interpretation invoking contrasts sets. One could argue that stress simply focused participants' attention on the size dimension, allowing them to restrict attention to the large objects. To rule out this alternative, we also included displays with two contrast sets: e.g., two red squares, one large and one small, and two blue circles, one large and one small. With a two contrast display, contrastive focus is still felicitous. However, the point of disambiguation now does not come until after the color adjective for instructions with contrastive stress and with neutral stress. Under these conditions, we found no effect of contrast. The interaction between type of display and stress provides clear evidence that participants were computing contrast sets rapidly enough to select among potential referents.

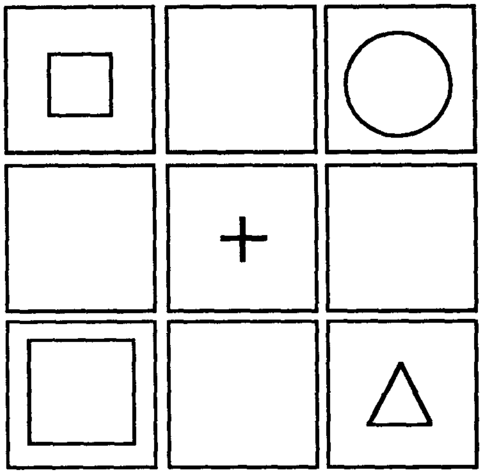


Figure 2: Display with one large and one small red square. The large circle is blue; the small triangle is yellow.

Word Recognition

The time course of spoken word recognition is strongly influenced by both the properties of the word itself (e.g., its frequency) and the set of words to which it is phonetically similar. Recognition of a spoken word occurs shortly after the auditory input uniquely specifies a lexical candidate [Marslen-Wilson, 1987]. For polysyllabic words, this is often prior to the end of the word. For example, the word "elephant" would be recognized shortly after the "phoneme" /f/. Prior to that, the auditory input would be consistent with the beginnings of several words, including "elephant", "elegant", "eloquent" and "elevator".

Most models of spoken word recognition account for these data by proposing that multiple lexical candidates are activated as the speech stream unfolds. Recognition then takes place with respect to the set of competing activated candidates. However, models differ in how the candidate set is defined. In some models, such as Marslen-Wilson's classic Cohort model, competition takes place in a strictly "left-to right" fashion. [Marslen-Wilson, 1987]. Thus the competitor set for "paddle" would contain "padlock", which has the same initial phonemes as "paddle", but would not include a phonetically similar word that did not overlap in its initial phonemes, such as a rhyming word like "saddle". In contrast, activation models such as TRACE [McClelland & Elman, 1986] assume that competition can occur throughout the word and thus rhyming words would also compete for activation.

Our initial experiments used real objects and instructions such as "Pick up the candy". We manipulated whether or not the display contained an object with a name that began with the same phonetic sequence as the target object [Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Spivey-Knowlton, Tanenhaus, Eberhard & Sedivy, 1995]. Examples of objects with overlapping initial phonemes were "candy" and "candle", and "doll" and "dolphin". An eye-movement to the target object typically began shortly after the word ended, indicating that programming of the eye-movement often began before the end of the word. The presence of a competitor increased the latency of eye-movements to the target and induced frequent false launches to the competitor. The timing of these eye-movements indicated that they were programmed during the "ambiguous" segment of the target word. These results demonstrated that the two objects with similar names were, in fact, competing as the target word unfolded. Moreover, they highlight the sensitivity of the eye-movement paradigm.

In ongoing work, we are exploring more fine-grained questions about the time-course of lexical activation. For example, in an experiment in progress [Allopenna, Magnuson & Tanenhaus, 1996], the stimuli are line drawings of objects presented on a computer screen (see Figure 3). On each trial, participants are shown a set of four objects and asked to "pick up" one of the objects with the mouse and move it to a specified location on the grid. The paddle was the target object for the trial shown in Figure 3. The display includes a "cohort" competitor sharing initial phonemes with the target (padlock) a rhyme competitor (saddle) and an unrelated object (castle).

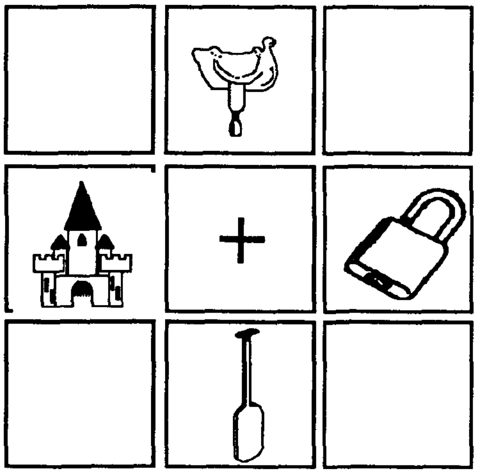


Figure 3: Sample Display for the Instruction: "Pick up the paddle."

Figure 4 shows the probability that the eye is fixating on the target and the cohort competitor as the spoken target word unfolds. Early on in the speech stream, the eye is on the fixation cross, where subjects are told to look at the beginning of the trial. The probability of a fixation to the target word and the cohort competitor then increases. As the target word unfolds, the probability that the eye is fixated on the target increases compared to the cohort competitor. These data replicate our initial experiments and show how eye-movements can be used to trace the time course of spoken word recognition. Our preliminary data also make it clear that rhyme competitors attract fixations, as predicted by activation models.

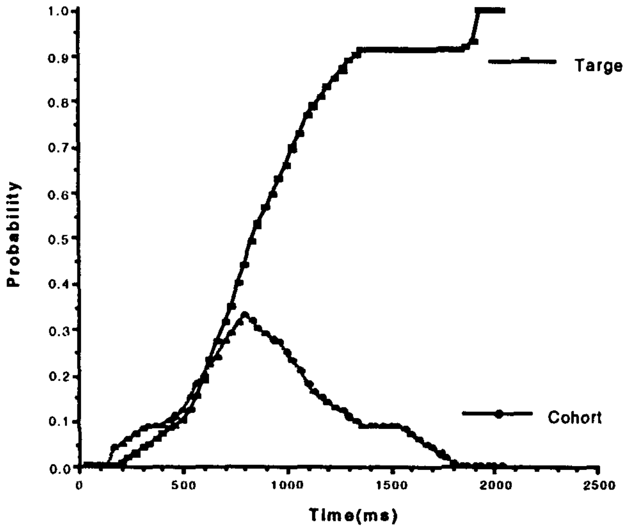


Figure 4: Probabilities of eye-fixations in a competitor trial.

Reference and Syntactic Ambiguity Resolution

There has been an unresolved debate in the language processing community about whether there are initial stages in syntactic processing that are strictly encapsulated from influences of referential and pragmatic context. The strongest evidence for encapsulated processing modules has come from studies using sentences with brief syntactic "attachment" ambiguities in which readers have clear preferences for interpretations, associated with particular syntactic configurations. For example, in the instruction "put the apple on the towel...," people prefer to attach the prepositional phrase "on the towel" to the verb "put", rather than the noun phrase "the apple", thus interpreting it as the argument of the verb (encoding the thematic relation of Goal), rather than as a modifier of the noun.

If the instruction continues "Put the apple on the towel into the box", the initial preference for a verb-phrase attachment is revealed by clear "garden-path" effects when "into" is encountered. Encapsulated models account for this preference in terms of principles such as pursue the simplest attachment first, or initially attach a phrase as an argument rather than as an adjunct. In contrast, constraint-based models attribute these preferences to the strength of multiple interacting constraints, including those provided by discourse context. [For a recent review, see Tanenhaus and Trueswell, 1995]

An influential proposal, most closely associated with Crain and Steedman [1985], is that pragmatically driven expectations about reference are an important source of discourse constraint. For example, a listener hearing "put the apple..." might reasonably assume that there is a single apple and thus expect to be told where to put the apple (the verb-phrase attachment). However, in a context in which there was more than one apple, the listener might expect to be told which of the apples is the intended referent and thus prefer the noun phrase attachment.

Numerous experiments have investigated whether or not the referential context established by a discourse context can modify attachment preferences. These studies typically introduce the context in a short paragraph and examine eye-movements to the disambiguating words in a target sentences containing the temporary ambiguity. While some studies have shown effects of discourse context, others have not. In particular, strong syntactic preferences persist momentarily, even when the referential context introduced by the discourse supports the normally less-preferred attachment. For example, the preference to initially attach a prepositional phrase to a verb requiring a goal argument (e.g., "put") cannot be overridden by linguistic context. These results have been taken as strong evidence for an encapsulated syntactic processing system.

However, typical psycholinguistic experiments may be strongly biased against finding pragmatic effects on syntactic processing. For example, the context may not be immediately accessible because it has to be represented in memory. Moreover, readers may not consider the context to be relevant when the ambiguous region of the sentence is being processed.

We reasoned that a relevant visual context that was available for the listener to interrogate as the linguistic input unfolded might influence initial syntactic analysis even though the same information might not be effective when introduced linguistically.

Sample instructions are illustrated by the examples in (1).

1. a. Put the apple on the towel in the box.
   b. Put the apple that's on the towel in the box.

In sentence (1a), the first prepositional phrase "on the towel", is ambiguous as to whether it modifies the noun phrase ("the apple") thus specifying the location of the object to be picked up, or whether it modifies the verb, thus introducing the goal location. In example (1b) the word "that's" disambiguates the phrase as a modifier, serving as an unambiguous control condition.

These instructions were paired with three types of display contexts. Each context contained four sets of real objects placed on a horizontal board. Sample displays for the instructions presented in (1) are illustrated in Figures 5, 6, and 7 Three of the objects were the same across all of the displays. Each display contained the target object (an apple on a towel) the correct goal, (a box) and an incorrect goal (another towel). In the one referent display (Figure 4) there was only one possible referent for the definite noun phrase "the apple", the apple on the towel. Upon hearing the phrase "the apple", participants can immediately identify the object to be moved because there is only one apple and thus they are likely to assume that "on the towel" is specifying the goal. In the two-referent display (Figure 5), there was a second possible referent (an apple on a napkin). Thus, "the apple", could refer to either of the two apples and the phrase "on the towel" provides modifying information that specifies which apple is the correct referent. Under these conditions a listener seeking to establish reference should interpret the prepositional phrase "on the towel" as providing disambiguating information about the location of the apple. In the three and one display, we added an apple cluster. The uniqueness presupposition associated with the definite noun phrase should bias the listener to assume that the single apple (the apple on the towel) is the intended referent for the theme argument. However, it is more felicitous to use a modifier with this instruction. This display was used to test if even a relatively subtle pragmatic effects will influence syntactic processing

Strikingly different fixation patterns among the visual contexts revealed that the ambiguous phrase "on the towel" was initially interpreted as the goal in the one-referent context but as a modifier in the two-referent contexts and the three-and-one contexts [for details see Spivey-Knowlton et al, 1995; Tanenhaus et al., 1995] In the one-referent context, subjects looked at the incorrect goal (e.g., the irrelevant towel) on 55% of the trials shortly after hearing the ambiguous prepositional phrase, whereas they never looked at the incorrect goal with the unambiguous instruction. In contrast, when the context contained two possible referents, subjects rarely looked at the incorrect goal, and there were no differences between the ambiguous and unambiguous instructions. Similar results obtained for the three-and-one context.

Figures 5 and 6 summarize the most typical sequences of eye-movements and their timing in relation to words in the ambiguous instructions for the one-referent and the two-referent contexts, respectively. In the one-referent context, subjects first looked at the target object (the apple) 500 ms after hearing "apple" then looked at the incorrect goal (the towel) 484 ms after hearing "towel". In contrast, with the unambiguous instruction, the first look to a goal did not occur until 500 ms after the subject heard the word "box".

In the two-referent context, subjects often looked at both apples, reflecting the fact that the referent of "the apple" was temporarily ambiguous. Subjects looked at the incorrect object on 42% of the unambiguous trials and on 61% of the ambiguous trials. In contrast, in the one-referent context, subjects rarely looked at the incorrect object (0% and 6% of the trials for the ambiguous and unambiguous instructions, respectively). In the two-referent context, subjects selected the correct referent as quickly for the ambiguous instruction as for the unambiguous instruction providing additional evidence that the first prepositional phrase was immediately interpreted as a modifier.

The three-and-one context provided additional information. Typical sequences of eye-movements for this context are presented in Figure 7. Participants rarely looked at the apple cluster, making their initial eye-movement to the apple on the towel. The next eye-movement was to the box for both the ambiguous and unambiguous instruction. These data also rule out a possible objection to the results from the two referent condition. One could argue that participants were, in fact, temporarily misparsing the prepositional phrase as the goal. However, this misanalysis might not be reflected in eye-movements to the towel because the eye was already in transit, moving between the two apples. However, in the three-and-one condition, the eye remains on the referent throughout the prepositional phrase. Given the sensitivity of eye-movements to probabilistic information, e.g., false launches to cohort and rhyme competitors, it is difficult to argue that the participants experienced a temporary garden-path that was too brief to influence eye-movements.

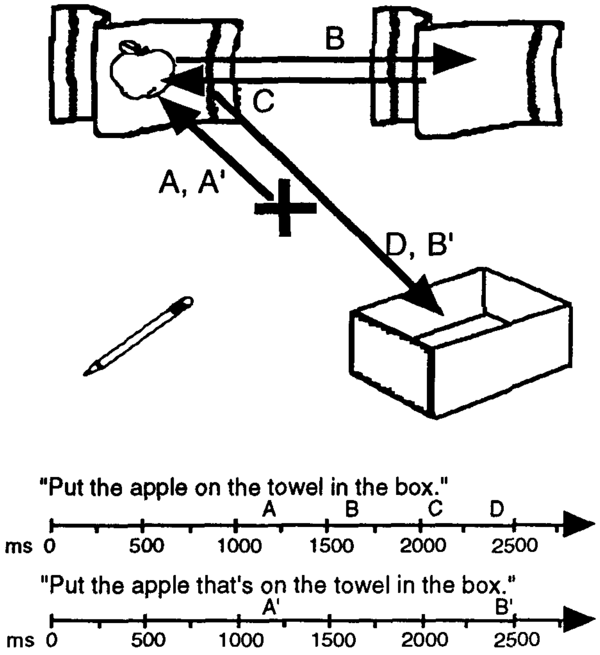

B
C
A, A'
D, B'
"Put the apple on the towel in the box."
A B C D
ms 0 500 1000 1500 2000 2500
"Put the apple that's on the towel in the box."
A' B'
ms 0 500 1000 1500 2000 2500

Figure 5: Typical sequence of eye movements in the one-referent context for the ambiguous and unambiguous instructions. Letters on the timeline show when in the instruction each eye movement occurred, as determined by mean latency of that type of eye movement (A' and B' correspond to the unambiguous instruction).

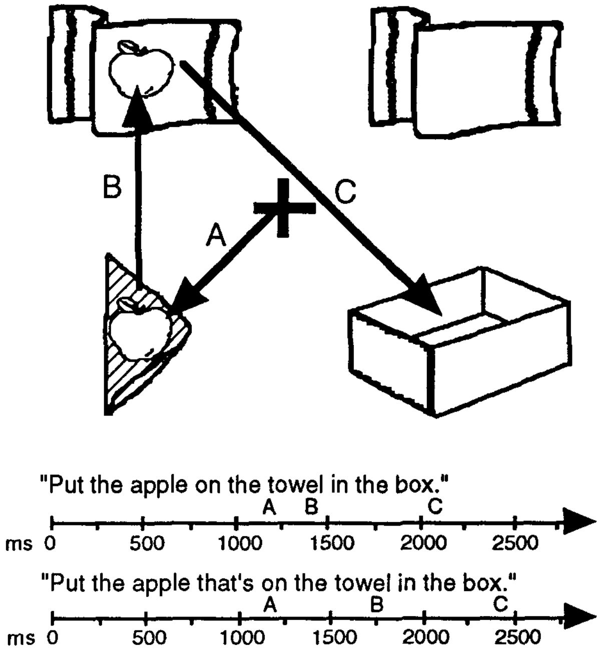

B
C
A
"Put the apple on the towel in the box."
A B C
ms 0 500 1000 1500 2000 2500
"Put the apple that's on the towel in the box."
A B C
ms 0 500 1000 1500 2000 2500

Figure 6: Typical sequence of eye movements in the two-referent context. Note that the sequence and the timing of eye movements, relative to the nouns in the speech stream, did not differ for the ambiguous and unambiguous instructions.

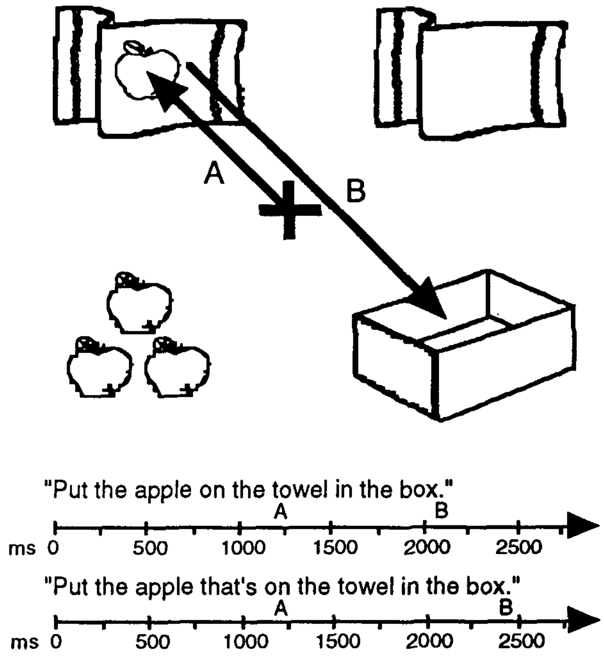

A
B
"Put the apple on the towel in the box."
A B
ms 0 500 1000 1500 2000 2500
"Put the apple that's on the towel in the box."
A B
ms 0 500 1000 1500 2000 2500

Figure 7: Typical sequence of eye movements in the three-and-one context. Note that the sequence and the timing of eye movements, relative to the nouns in the speech stream, did not differ for the ambiguous and unambiguous instructions.

Conclusion

We have reviewed results establishing that, with well-defined tasks, eye-movements can be used to observe under natural conditions the rapid mental processes that underlie spoken language comprehension. We believe that this paradigm will prove valuable for addressing questions on a full spectrum of topics in spoken language comprehension, ranging from the uptake of acoustic information during word recognition to conversational interactions during cooperative problem solving.

Our results demonstrate that in natural contexts people interpret spoken language continuously, seeking to establish reference with respect to their behavioral goals during the earliest moments of linguistic processing. Thus our results provide strong support for models that support continuous interpretation. Our experiments also show that referentially relevant non-linguistic information immediately affects how the linguistic input is initially structured. Given these results, approaches to language comprehension that emphasize fully encapsulated processing modules are unlikely to prove fruitful. More promising are approaches in which grammatical constraints are integrated into processing systems that coordinate linguistic and non-linguistic information as the linguistic input is processed.

Acknowledgments

* This paper summarizes work that the invited talk by the first author (MKT) was based upon. Supported by NIH resource grant 1-P41-RR09283; NIH HD27206 to MKT; NIH F32DC00210 to PDA, NSF Graduate Research Fellowships to MJS-K and JSM and a Canadian Social Science Research Fellowship to JCS.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1996). Watching spoken language perception: Using eye-movements to track lexical access. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Ballard, D., Hayhoe, M. & Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7, 68-82.

Crain, S. & Steedman, M. (1985). On not being led up the garden path. In Dowty, Kartunnen & Zwicky (eds.), *Natural Language Parsing*. Cambridge, MA: Cambridge U. Press.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J. & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409-436.

Kratzer, J. (1991). Representation of focus. In A. von Stechow & D. Wunderlich (Eds.), *Semantik: Ein Internationales Hundbuch der Zeitgenossichen Forschung*. Berlin: Walter de Guyter.

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

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

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

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

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). Hillsdale, NJ: Erlbaum.

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

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye-movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In T Inui & J.L. McClelland (eds.). *Attention and Performance XVI: Information integration in perception and communication.*, 457-478. Cambridge Mass: MIT Press.

Tanenhaus, M. & Trueswell, J. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.). *Handbook of Perception and Cognition: Volume 11: Speech, Language and Communication.* Academic Press., 217-262. New York: Academic Press.