The Development of On-Line Language Comprehension Abilities in Children

John C. Trueswell, Irina Sekerina, Nicole M. Hill, and Marian Logrip

(Department of Psychology and the Institute for Research in Cognitive Science,
University of Pennsylvania)

Abstract. A great deal of psycholinguistic research has focused on the question of how adults interpret language in real time. This work has revealed a complex and interactive language processing system capable of rapidly coordinating linguistic properties of the message with information from the situation or context [e.g., 1-7]. In the study of language acquisition, however, surprisingly little is known about how children interpret language in real time, in large part because most existing techniques have relied upon the adult skill of reading [8, 9]. We present here results from a new method for studying children's language processing abilities, in which a head-mounted eye-tracking system was used to monitor eye movements as subjects responded to spoken instructions. The results revealed systematic differences in how children and adults process spoken language: Five year olds relied heavily on the linguistic properties of the input, showing less ability or inclination than adults to coordinate these properties with information from the visual context. This finding suggests that an important component of human development is acquiring the capacity to rapidly coordinate information generated from multiple perceptual and cognitive systems.

Recent developments in miniaturized eye-tracking technology have permitted us to establish a new method for studying spoken language comprehension in children. Prior research with adults has shown that by monitoring eye movements during listening, much can be inferred about the processes underlying language interpretation [6, 10]. These studies have found that adults' eye movements are closely time-locked with speech — e.g., within a few hundred milliseconds of hearing a word that uniquely refers to an object in the world, adults can launch an eye movement to the intended referent. Moreover, consistent with the hypothesis that the language processing system is highly interactive, adults are capable of guiding their interpretation of grammatically ambiguous phrases based upon relevant information from the visual context. By using a new light-weight eye-tracking visor specifically designed for children (see Methods, Figure 1), we set out to examine the extent to which children process speech in real time and resolve local ambiguities based on cues from the visual context.

Two age groups participated in the study: sixteen Five Year Olds, and twelve Adults. Subjects' eye movements were monitored as they responded to spoken instructions to move stuffed animals on a tabletop. We compared instructions containing a temporary syntactic ambiguity, such as "*Put the frog on the napkin in the box*", with unambiguous versions, such as "*Put the frog that's on the napkin in the box*". In the ambiguous version, the phrase "*on the napkin*" could be interpreted as a destination of the putting event, indicating where to put the

frog, or as a modifier of the preceding noun phrase, indicating which frog. Each target sentence was heard in one of two visual contexts (Figure 1).

One context supported the modifier interpretation, consisting of two frogs, one of which was on a napkin, an empty napkin and an empty box (2-Referent Context). In this case, upon hearing "*the frog*", a listener would not know which frog is being referred to, and should thus interpret the phrase "*on the napkin*" as a modifier. The other context supported the destination interpretation and consisted of the same scene except the second frog was replaced with another animal such as a horse (1-Referent context). In this case, modification of "*the frog*" with "*on the napkin*" would be unnecessary because there is only one frog [1, 11]. Hence, we would expect listeners to interpret the prepositional phrase as a destination, referring to the empty napkin. In an earlier eye-tracking study, Tanenhaus et al. found that adults can use these referential constraints to inform syntactic commitments. Adults initially misinterpreted the ambiguous phrase as the destination in 1-Referent but not 2-Referent contexts, as shown by substantially more looks to the incorrect destination (the empty napkin) in I-Referent as compared to 2-Referent contexts [6].

Before examining how children deal with ambiguous phrases, it is important to note that their eye movements revealed that they were processing speech in real time, recognizing referential expressions with nearly the same speed as adults. When a definite noun phrase, such as "*the frog*", uniquely referred to a single object in the scene, subjects' direction of gaze quickly moved to the referent object — as measured from the onset of the noun, an average of 667 ms for Five Year Olds and 564 ms for Adults. This finding is well within the range of other eye movement studies [6, 10] and indicates that children's word recognition processes proceed at approximately the same speed as adults.

An examination of how subjects parsed the ambiguous prepositional phrase "*on the napkin*" revealed systematic differences between ages. As opposed to Adults, who showed sensitivity to context, the findings from Five Year Olds indicated they had consistently interpreted the ambiguous phrase as a destination and were unable to revise their initial commitment. This finding is of considerable interest because a preference for the destination

interpretation over the modifier interpretation is precisely what is expected from a processing system that is relying on the linguistic properties of the message and not the context — i.e., the verb "*put*" has a strong syntactic expectation for a second argument in the form of a prepositional phrase [3, 5]. This preference was most apparent in Five Year Olds' responses to ambiguous instructions, in which 58% of the trials resulted in an incorrect action (Figure 2). An action was considered correct if the intended object (e.g., the frog on the napkin) was moved directly to the intended destination (e.g., the empty box). An action was also considered correct if both the object and the modifying object were moved to the intended destination (e.g., both the frog and the napkin underneath the frog were placed in the box), consisting of less than 3.3% of responses overall.
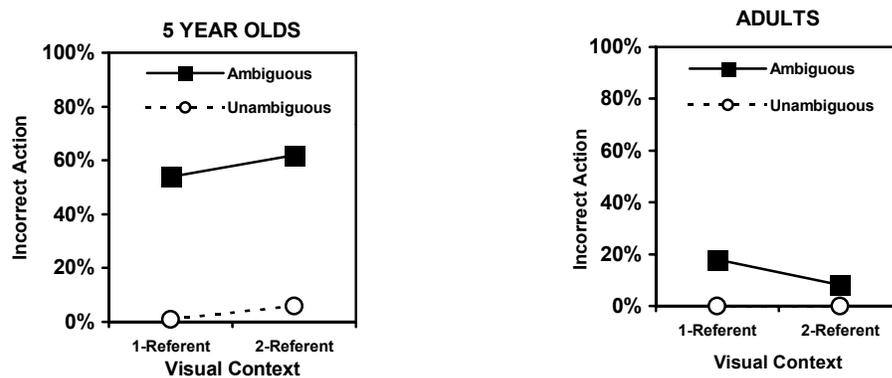
Figure 2. Percentage of trials in which objects were moved incorrectly.

All of these errors involved moving the animal to the Incorrect Destination, i.e. the empty napkin. The two most common errors were to place the animal onto the Incorrect Destination, or to hop the animal to the Incorrect Destination and then to the Correct Destination. Subjects showed significantly fewer errors (4%) on the unambiguous instructions, showing that children are sensitive to the linguistic cues to syntactic structure, $F(1,12)=42.75$, $p<.001$. This difference was unaffected by the type of visual context, $F<1$. In addition, the eye movement record further supported the hypothesis that Five Year Olds considered the phrase "*on the napkin*" as the destination. From the onset of the preposition, we calculated the percentage of trials in which subjects directed their gaze to the Incorrect Destination (Figure 3), e.g., the empty napkin, as measured from the onset of ambiguous phrase, e.g., "*on the napkin*".

5 YEAR OLDS

Look to Incorr. Destin.

100%
80%
60%
40%
20%
0%

■ Ambiguous
○ Unambiguous

1-Referent   2-Referent
Visual Context

ADULTS

Looks to Incorr. Destin.

100%
80%
60%
40%
20%
0%

■ Ambiguous
○ Unambiguous

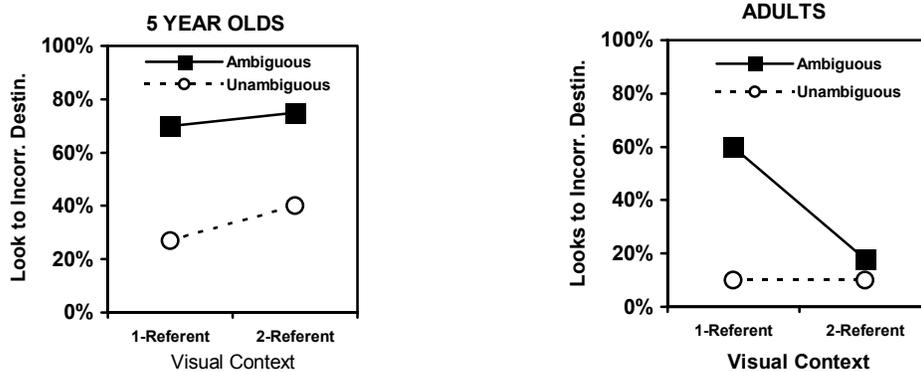1-Referent   2-Referent
Visual Context

Figure 3. Percentage of trials in which there was a look to the Incorrect Destination.

Approximately 70% of these eye movements were the first eye movement made after fixating on an animal, and typically occurred soon after hearing "*napkin*". Subjects looked at the Incorrect Destination much more often in the ambiguous trials (73%) than in the unambiguous trials (36%), $F_{(1,12)}=12.62$, $p<.005$. Again, this difference was unaffected by visual context, $F<1$.

Adult performance was strikingly different, showing evidence that adults used the visual context to inform processing commitments. Adults initially misinterpreted the ambiguous phrase as the destination in the 1-Referent Context, but not in the 2-Referent Context. In particular, adults showed very few incorrect actions, most of which occurred for ambiguous instructions in the 1-Referent context (Figure 2). The percentage of looks to the Incorrect Destination on ambiguous trials was greater in 1-Referent Contexts (62%) than in 2-Referent Contexts (18%, Figure 3), indicating that listeners temporarily considered the Incorrect Destination only in the 1-Referent context. By comparison, 9% of the unambiguous trials contained looks to the Incorrect Destination, resulting in a significant interaction between Context and Ambiguity, $F_{(1,8)}=9.89$, $p<.05$. A direct comparison between the eye movement patterns of the Adults and Five Year Olds revealed a triple interaction between Context, Ambiguity and Age, $F_{(1,26)}=4.57$; $p<.05$, supporting the hypothesis that Adults are highly sensitive to the context whereas Five Year Olds are not.

In addition to demonstrating that it is now possible to study children's language processing in real time, the results reveal that children who otherwise have a strong command of language have yet to fully develop contextually-dependent processing mechanisms for language interpretation. The presence of consistent grammatical preferences on the part of the youngest

children indicates a partial independence of the encoding of linguistic and non-linguistic regularities in the human language system. The findings from adults suggest that this independent encoding is coupled with an ability to rapidly integrate these multiple sources of information [e.g., 4, 12, 13]. More generally, the results contribute to an emerging picture of cognitive development, in which the various perceptual and cognitive subsystems of the human brain, which have been established through the course of evolution, operate in relative isolation in the young child [e.g., 14, 15, 16, 17]. By limiting learning to a subset of information, each of these systems may succeed by the "divide and conquer" approach [see 18, 19, 20 for proposals of improved learning with limited information]. A critical step in later development is the establishment of interactive mechanisms that coordinate the operations of each subsystem, thereby creating the complex but relatively unified cognitive system of the adult [e.g., 15, 21, 22].

Methods

Subjects. Age range of subject groups were as follows: Five Year Olds (4.5 to 5.5 years), and Adults (18 to 23 years). All subjects were raised in English-speaking households and had no reported language or cognitive impairments. Parents and adult subjects gave informed consent prior to experimentation.

Materials and Design. Sixteen different target items were generated and randomly assigned to one of four conditions. Items were rotated through each condition generating four different presentation lists. Subjects were assigned to one presentation list, resulting in no subject seeing the same target item twice. Trials began with an instruction to look at a central fixation, i.e., "Look at the happy face", followed by the target instruction, and one to four additional instructions. A small number of distractor trials (8 to 16) were used with children. More distractor trials (26) were used with adults because of concerns that they would detect the experimental design. Significant differences reported above used subjects as the random factor. These differences were also significant when using items as the random factor.

Equipment and Procedure. Eye movements were recorded using an ISCAN eye-tracking visor, consisting of a monocle and two miniature cameras (Figure 1). One camera recorded the visual environment from the perspective of the left eye, and the other camera recorded a close-up image of the eye. A computer analyzed the eye in real-time, superimposing a horizontal and

vertical eye position on the scene image. Horizontal and vertical resolution is less than one degree of visual angle. Objects in this experiment filled three to four degrees of visual angle when viewed at arm's length. The eye-tracker was precalibrated on an adult's eye before using it with the youngest of children. Calibration accuracy was checked by having the child track a moving stuffed animal and by instructing the child to fixate on various objects. The scene image and predicted eye position were inspected. For all subjects reported here, eye position was accurate enough to indicate clearly which object was being fixated.

References

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

[2] Altmann, G., Garnham, A., & Dennis, Y. (1992). Avoiding the garden-path: Eye movements in context. *Journal of Memory and Language* 31, 685-712.

[3] 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.

[4] MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101, 676-703 .

[5] Spivey-Knowlton, M.J., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227-267.

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

[7] Trueswell, J. C. & Tanenhaus, M. K. (1991). Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes* 6, 303-338.

[8] Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14, 178 - 210.

[9] Rayner, K., & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.) *Attention and Performance XII*. Erlbaum, London.

[10] Allopenna, P.D., Magnuson, J.S., Tanenhaus, M.K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language* 38, 4 19-439.

[11] Crain, S. & Steedman, M. I. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*. Cambridge University Press, Cambridge.

[12] Boland, J. & Cutler, A. (1996). Interaction with autonomy: Multiple output models and the inadequacy of the Great Divide. *Cognition* 58, 309-320.

[13] Trueswell, J. C. & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, K. Rayner & L. Frazier (Eds.) *Perspectives on Sentence Processing*. Erlbaum, Hillsdale, NJ.

[14] Hermer L and Spelke E. (1994). A geometric process for spatial reorientation in young children. *Nature* 370, 57-59.

[15] Hermer, L. & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition* 61, 195-232.

[16] Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA.

[17] Pinker, S. (1994) *The Language Instinct*. William Morrow and Company, New York.

[18] Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition* 48, 71-99.

[19] Newport E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Science* 10, 147-172.

[20] Newport E. L. (1990). Maturational Constraints on Language Learning. *Cognitive Science* 14, 11-28.

[21] Karmiloff-Smith, A. (1992). *Beyond Modularity*. MIT Press, Cambridge, MA.

[22] Rozin, P. (1976). The Evolution of Intelligence and Access to the Cognitive Unconscious. *Progress in Psybiological and Physiological Psychology* 6, 245-280.