COGNITION

Cognition 73 (1999) 89–134

www.elsevier.com/locate/cognit

The kindergarten-path effect: studying on-line sentence processing in young children

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

University of Pennsylvania, Philadelphia, PA, USA

Received 18 August 1998; received in revised form 29 January 1999; accepted 1 May 1999

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 context or situation (e.g. Altmann & Steedman, 1988; Britt, 1994; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995; Trueswell & Tanenhaus, 1991). In the study of language acquisition, however, surprisingly little is known about how children process language in real time and whether they coordinate multiple sources of information during interpretation. The lack of child research is due in part to the fact that most existing techniques for studying language processing have relied upon the skill of reading, an ability that young children do not have or are only beginning to acquire. We present here results from a new method for studying children's moment-by-moment language processing abilities, in which a head-mounted eye-tracking system was used to monitor eye movements as participants responded to spoken instructions. The results revealed systematic differences in how children and adults process spoken language: Five Year Olds did not take into account relevant discourse/pragmatic principles when resolving temporary syntactic ambiguities, and showed little or no ability to revise initial parsing commitments. Adults showed sensitivity to these discourse constraints at the earliest possible stages of processing, and were capable of revising incorrect parsing commitments. Implications for current models of sentence processing are discussed. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Syntactic ambiguity; Parsing; Sentence comprehension; Language acquisition; Development

* Corresponding author. Dr. John C. Trueswell, Department of Psychology, University of Pennsylvania, 3815 Walnut St., Philadelphia, PA 19104-6196, USA. Tel.: (215) 898-0911; fax: (215) 898-7301.
E-mail address: trueswel@psych.upenn.edu (J.C. Trueswell)

0010-0277/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0010-0277(99)00032-3

When adults interpret language, they do not wait to receive an entire sentence or phrase before making some determination of the sentence's intended meaning. Readers and listeners are much more impatient than this; they make partial commitments to interpretation ''on the fly'' as each sentence unfolds over time (e.g. Altmann & Steedman, 1988; Frazier & Rayner, 1982; Marslen-Wilson, 1973; Trueswell, Tanehhaus & Garnsey, 1994). A large body of psycholinguistic research has examined how adults incrementally process speech and text, and how and when various sources of information are coordinated to arrive at an interpretation (e.g. see MacDonald, Pearlmutter & Seidenberg, 1994, Tanenhaus & Trueswell, 1995, for recent reviews). As discussed below, the findings show that adults are capable of rapidly coordinating the detailed linguistic properties of the message with information from the context to incrementally determine the intended meaning.

In this paper, we present the results of a research program that is designed to examine how the rapid and relatively interactive adult processing system develops in young children. The work employs a new method for studying on-line sentence processing with children, in which a head-mounted eye-tracking system was used to monitor eye movements as participants responded to spoken instructions. Prior experiments have successfully used this method to examine referential and linguistic processing in adults (e.g. Allopenna, Magnuson & Tanehaus, 1998; Tanenhaus et al., 1995). To our knowledge, this is the first time that this method has been used to study the development of sentence processing mechanisms in children. The experiments focus on syntactic ambiguity and how information from the context is used to inform processing commitments. To introduce these issues, we first sketch recent descriptions of sentence processing in adults and the types of theories that have been proposed to account for this ability. We will then describe the limited research that has been done in the domain of on-line processing in children. Finally, we will introduce the eye movement method for studying listening, focusing on how it can be used to examine language development.

1. Adults' language comprehension is rapid and context-sensitive

During the last three decades, the study of on-line sentence processing has been one of the central topics of adult psycholinguistic research (e.g. Frazier & Fodor, 1978; Frazier & Rayner, 1982; Marslen-Wilson, 1973, 1987; Tanenhaus & Trueswell, 1995). The majority of findings have come from the task of reading, because fixation and eye movement patterns across text segments can be used to quantify moment-by-moment increases and decreases in processing difficulty. This work has documented that much of language comprehension takes place automatically, with readers and listeners making at least partial commitments to interpretation as each sentence unfolds over time (e.g. Altmann & Steedman, 1988; Frazier & Rayner, 1982; Marslen-Wilson & Tyler, 1987; Trueswell et al., 1994).

An important implication of the on-line nature of processing is that readers and listeners occasionally make incorrect commitments that require revision at a later point in the sentence. Readers and listeners can be led down the ''garden-path'' at a point of ambiguity and only later make their way back to the intended interpretation. The garden-path phenomenon is something of a gift from nature to the researcher interested in the machinery of sentence processing. This is because it can be used as a means for examining which kinds of evidence (syntactic, semantic, discourse context) inform initial commitments to interpretation at a point of ambiguity, helping to reveal the internal organization of the language comprehension system.

Determining which kinds of information take priority in comprehension so as to develop a theory of the parsing architecture is no simple matter, however, and has required a long-term experimental effort in which sources of evidence are parametrically manipulated. For example, in the sentence:

1. Anne hit the thief with the stick.

there are two ways that we can interpret the prepositional phrase ''with the stick''. We can interpret it as an Instrument with which the action is performed, in which case it is structurally linked to the verb (VP attachment). Or, we can interpret it as a Modifier of the direct object that is attached directly to that noun-phrase (NP attachment). It has been confirmed experimentally that readers have a strong tendency to commit to the Instrument interpretation when encountering the ambiguous preposition ''with'' in sentences like these (e.g. Rayner, Carlson & Frazier, 1983; Taraban & McClelland, 1988). This finding is consistent with our intuitions of a misinterpretation, or garden-path, when we encounter the final word in such sentences as:

2. Anne hit the thief with the wart.

The results of many initial studies in this area supported an encapsulated structure-based language processing system known as the Garden-path theory (e.g. Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner et al., 1983). This theory postulated an initial syntactic stage of parsing which at points of ambiguity selected the syntactically simplest alternative (the *Minimal Attachment Principle*, Frazier, 1987, 1989; Frazier & Fodor, 1978). For example, findings seemed to confirm that readers experienced garden-paths only when a temporary ambiguity resolved toward a syntactically more complex alternative (e.g. Frazier & Rayner, 1982). Manipulations of seemingly relevant non-syntactic cues such as semantically biasing information (e.g. the sticks and warts of sentences 1 and 2), or discourse-biasing information had little effect until a later, editing, stage at which readers revised their initial commitments (e.g. Ferreira & Clifton, 1986; Rayner et al., 1983). Moreover, because commitments appeared to be insensitive to manipulations even of subcategorical syntactic information, it was proposed that initial processing relied on broad (major) category information to structure input (Ferreira & Henderson, 1991; Frazier, 1989). So, for the examples above, the preference for VP attachment was predicted because this alternative was argued to be syntactically simpler than NP attachment (Rayner et al., 1983).

However, in the past 15 years, an accumulating body of evidence has shown that a wide range of syntactic and non-syntactic sources of information can mediate garden-path effects. For example, lexically specific syntactic information, semantic plausibility, frequency of lexical co-occurrence, and referential context have all been found to rapidly constrain adults' on-line commitments to interpretation, even in the face of syntactic ambiguity (e.g. Altmann, Garnham & Henstra, 1994; Altmann & Steedman, 1988; Britt, 1994; Garnsey, Pearlmutter, Myers & Lotocky, 1997; MacDonald, 1993; Pearlmutter & MacDonald, 1995; Taraban & McClelland, 1988; Tanenhaus et al., 1995; Trueswell, 1996; Trueswell et al., 1993, 1994). Earlier failures to demonstrate their effects may have been due to inadequate manipulation or control of these factors (see MacDonald et al., 1994; Trueswell, Tanenhaus & Kello, 1993, Trueswell et al., 1994). Findings that adults' initial parsing commitments were affected by lexically specific syntactic information (e.g. verb subcategorization preferences) have led many researchers to conclude that the processing system does not employ broad structural heuristics but rather is highly tuned to the statistical regularities pertaining to the syntax of the language. Findings that adults can also employ non-syntactic information to inform parsing commitments have also led some researchers to conclude that the language processing system does not include an encapsulated parsing system, but rather rapidly coordinates the linguistic properties of the message with information from the context to determine processing commitments.

The theory that appears to best capture the existing evidence is the constraint-based lexicalist theory (see MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). This theory assumes a constraint-satisfaction approach to ambiguity resolution (Marslen-Wilson & Tyler, 1987; McClelland, 1987), in which multiple sources of information can be used to converge on an interpretation. One component of this theory is a grammatical processing system that is highly tuned to the structural preferences of individual lexical items; hence ''lexicalist.'' For example, the preference for VP attachment in the examples above is explained as arising from a system that is sensitive to the grammatical preferences of the verb ''hit'', which include the use of an Instrument role typically introduced by the preposition ''with'' (e.g. see Taraban & McClelland, 1988, for evidence in favor of this explanation).

Some of the more intriguing evidence in favor of this approach to sentence processing has come from studies finding that readers' initial commitments to syntactically ambiguous phrases can be affected by the referential context of ambiguous expressions (e.g. Altmann, Garnham & Dennis, 1992; Altmann & Steedman, 1988; Britt, 1994; Trueswell & Tanenhaus, 1991). In the PP-attachment ambiguity, for example, it was hypothesized that the preference for NP attachment for an ambiguous phrase like ''with the stick'' in sentence 1 would be affected by contexts which demand modification of the preceding NP ''the thief'' (i.e., the *Referential Principle*, Crain & Steedman, 1985). In particular, consider a context in which there are two thieves in discourse focus, one of which is holding a stick (2-Referent context). Upon hearing ''Ann hit the thief...'', the definite NP ''the thief'' requires a referent, but since a unique referent cannot be found, additional linguistic information is expected to uniquely specify the referent, i.e., through post-nominal modification. So, ''with the stick'' is then expected to be interpreted as a Modifier rather than an Instrument.

Altmann and Steedman (1988) found support for the Referential Principle in a reading study, in which no signs of garden-pathing were found in PP-attachment ambiguities when the sentences appeared in 2-Referent contexts (see also Altmann et al., 1994; Pynte & Kennedy, 1993; Spivey-Knowlton & Sedivy, 1995). Subsequently, some researchers have integrated this referential component of processing into the constraint-based lexicalist position, by indicating that relevant sources of information from the context can affect processing commitments, especially when lexical properties of the stimulus are relatively neutral (Spivey-Knowlton & Sedivy, 1995, Spivey-Knowlton & Tanenhaus, 1994). This position is supported by the finding that, in reading studies at least, effects of referential context are less immediate and less salient when the lexical properties of the input strongly bias one alternative (Altmann, van Nice, Garnham & Henstra, 1998; Britt, 1994; Spivey-Knowlton & Sedivy, 1995; see also MacDonald, 1994; Trueswell, 1996). Previous failures to find effects of referential context are explained as arising from the use of experimental stimuli that had strong lexical biases in favor of one syntactic alternative.

In sum, research on sentence processing, which for the most part has focused on the task of reading, has found that adults incrementally process language by rapidly coordinating detailed linguistic properties of the message with information from the context. In fact, it can be argued that the adult system gives equal status to linguistic information from the message and relevant contextual information (see MacDonald et al., 1994; though see Boland & Cutler, 1996; Trueswell, 1996). Independent of any debates about the priority of different classes of information, it is apparent from the now large body of experimentation that both linguistic and contextual information are coordinated with impressive speed by the adult comprehension system. Contextually dependent expressions are quickly evaluated with respect to the context and inform a variety of on-line decision processes, including those pertaining to the structural analyses of a sentence.

2. Children's on-line language processing abilities

In this paper, we examine the on-line nature of children's sentence processing abilities. We focus on the ability of children to resolve the PP-attachment ambiguity, and whether, like adults, children use the Referential Principle to help determine on-line parsing decisions. As in the adult studies described above, we will compare contexts that support NP modification (2-Referent contexts) with contexts that support VP modification (1-Referent contexts) and observe whether these contexts affect children's on-line processing commitments to the ambiguous and unambiguous phrases.

As mentioned earlier, surprisingly little is known about how children process language in real-time and, in particular, how they resolve temporary syntactic ambiguities. Only a handful of studies in the literature have used real-time processing techniques to study children's language comprehension (Holcomb, Coffey & Neville, 1992; McKee, Nicol & McDaniel, 1993; Swinney & Prather, 1989; Tyler, 1983; Tyler & Marslen-Wilson, 1981; see also McKee, 1996). Of these, only two have addressed issues related to syntactic processing (McKee et al., 1993; Tyler & Marslen-Wilson, 1981), and none have specifically examined the resolution of temporary syntactic ambiguities. Explicit theorizing about children's sentence processing abilities has also been limited to a relatively small set of researchers (e.g. Bever, 1970, 1982; Crain, Ni & Conway, 1994; Goodluck, 1990; Goodluck & Tavakolian, 1982; Hamburger & Crain, 1982, 1984; Frank, 1998; Mazuka, 1990). Indeed, processing issues have typically been set apart in developmental psycholinguistic theorizing as though they were mere contaminants or ''performance factors'' that mask the competence that lies below (e.g. Roeper, 1981).

Despite the current state of affairs, there are some expectations from the adult psycholinguistic literature and from the language acquisition literature about how children might employ lexical, syntactic and referential factors during syntactic ambiguity resolution. First, highly interactive models of language processing (e.g. Marslen-Wilson & Tyler, 1987; McClelland, 1987) would expect that if a child has knowledge of the Referential Principle, i.e., in the form of a probabilistic constraint on syntax, then he or she would use this constraint to inform on-line processing decisions. However, theories that emphasize a greater role for structural processes (e.g. Frazier, 1989) or theories that emphasize a greater role for lexically based syntactic processes (e.g. Boland & Cutler, 1996; MacDonald et al., 1994; Trueswell & Tanenhaus, 1994) might expect a developmental distinction between the use of linguistic and referential factors in children's on-line resolution of syntactic ambiguity. From the perspective of the language acquisition literature, there are also some expectations about the role of referential and syntactic information in ambiguity resolution. In particular, there is some indication that, for the age group of interest in our current study (four and Five Year Olds), pragmatic and referential factors play an equal or greater role than structural factors in children's interpretation of a sentence (Hamburger & Crain, 1982, 1984; Holcomb, 1992; Tyler, 1983; Tyler & Marslen-Wilson, 1981, but c.f. Goodluck 1990; Goodluck & Tavakolian, 1982). We briefly consider the on-line and off-line evidence related to these issues.

In one of the earliest attempts to study children's language processing in real-time, Tyler and Marslen-Wilson (1981) examined children's reliance on syntactic, semantic and pragmatic information during interpretation. Children (ages 5, 7 and 10) performed a word-monitoring task as they listened to either normal speech, semantically anomalous speech, or scrambled (syntactically anomalous) speech. All age groups showed patterns similar to those reported for adults, with the ability to detect a word becoming slower and less accurate as one compared normal speech to semantically anomalous speech to syntactically anomalous speech. This finding suggested that children were developing essentially the same types of analyses of the input as adults, and that the time course of the construction of these syntactic and interpretative analyses did not differ across ages. Using a similar monitoring technique (the detection of a mispronunciation), Tyler (1983) focused specifically on children's processing of referential expressions such as definite NPs and pronouns. These experiments found an important developmental difference. While anaphoric mapping processes appeared to be well-mastered in all age groups (5, 7, 10 year olds), developmental differences were observed for pronouns, with Five Year Olds relying more on pragmatic plausibility than lexical factors (i.e., gender-marking) in their assignment of pronominal co-reference.

Although the work of Tyler and Marslen-Wilson has made a valuable contribution to the study of language processing in children, we believe the results should be taken with some caution. This is because techniques like word monitoring provide a relatively indirect measure of the processes related to language interpretation, and therefore are often difficult to evaluate. For instance, if one were to assume that children and adults are highly sensitive to the word-distribution patterns of their language, this information alone could explain the observed differences between the three classes of speech used in the Tyler and Marslen-Wilson (1981) study. This is because each class of sentences contains decreasingly less distributional information (normal, semantically anomalous, syntactically anomalous speech). Thus, it is possible that the corresponding decreases in performance on the word-monitoring task arose not from the disruption of interpretative and/or syntactic processes, but rather from differences in the distributional cues found in the input signal.

Studies that have used a more direct measure of a child's interpretation of a sentence, such as acting-out a sentence or instruction, have provided a more off-line measure of comprehension processes. Of most relevance to the current study are experiments which suggest that children have an understanding of the relevant contextual factors associated with the Referential Principle and employ this information in their ultimate comprehension of a sentence (e.g. Hamburger & Crain, 1982; but c.f. Goodluck, 1990). In particular, Hamburger and Crain (1982), argued that the general failure of children (ages 4 to 6) to correctly act out sentences like ''The camel kicked the tiger that bumped the zebra'' (e.g. Sheldon, 1974; Tavakolian, 1981) was not due to a lack of knowledge of the relative clause structure but rather a failure on the part of experimenters to meet discourse-related requirements for the use of restrictive modifiers like the relative clause (see also Crain, McKee & Emiliani, 1990). Specifically, as mentioned above, restrictive modifiers help pick out a member of set. It was shown that when a set of tigers were present (e.g. three toy tigers), children showed a much more accurate understanding of the restrictive relative clause in the act-out task. These studies suggest that children have knowledge of the contextual factors associated with the Referential Principle, and, in keeping with the general notion that this information plays a greater role than syntactic cues, this knowledge is able to interfere with the understanding of the restrictive relative clause. It is important to note, however, that some studies have found weaker or no effects of referential factors on relative clause comprehension (e.g. Goodluck, 1990; Lee, 1992), a point which we will return to at the end of the paper.

In sum, studies that have managed to examine children's language processing abilities in real-time have found equal or perhaps greater reliance on pragmatic and discourse cues by children as compared to adults (Tyler & Marslen-Wilson, 1981;

Tyler, 1983). However, these studies have relied on somewhat difficult-to-interpret secondary tasks (e.g. monitoring for a particular word, or monitoring for a mispronunciation). Studies that rely on more direct measures of interpretation, such as the act-out task, have also tended to find a greater impact of pragmatic and contextual factors on the ultimate action taken by the child. However, these studies did not employ real-time measures of sentence comprehension, and there is some debate about the consistency of these effects. Moreover, the studies in both of these areas have not examined the conditions of temporary syntactic ambiguity that are typically used in adult studies of comprehension.

3. Studying children's eye movements during listening

In the experiments reported below, we employed a new experimental technique for studying spoken language comprehension in real time, in which participants' eye movements were monitored as they responded to spoken instructions to move objects about on a table. This technique provides a new means of examining the moment-by-moment processes of children's spoken language comprehension, in the relatively natural situation of acting out spoken instructions. It was only recently that this technique was developed for use with adults (Tanenhaus et al., 1995). These studies showed that by monitoring the eye movements of adults during listening, much can be inferred about the processes underlying language interpretation. Adults' eye movements were found to be 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.

The current experiments used a highly miniaturized head-mounted eye-tracking system specifically designed for use with children (see Section 4.1, Fig. 1). We compared instructions containing a temporary syntactic ambiguity involving PP-attachment, 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 (Fig. 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 that 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

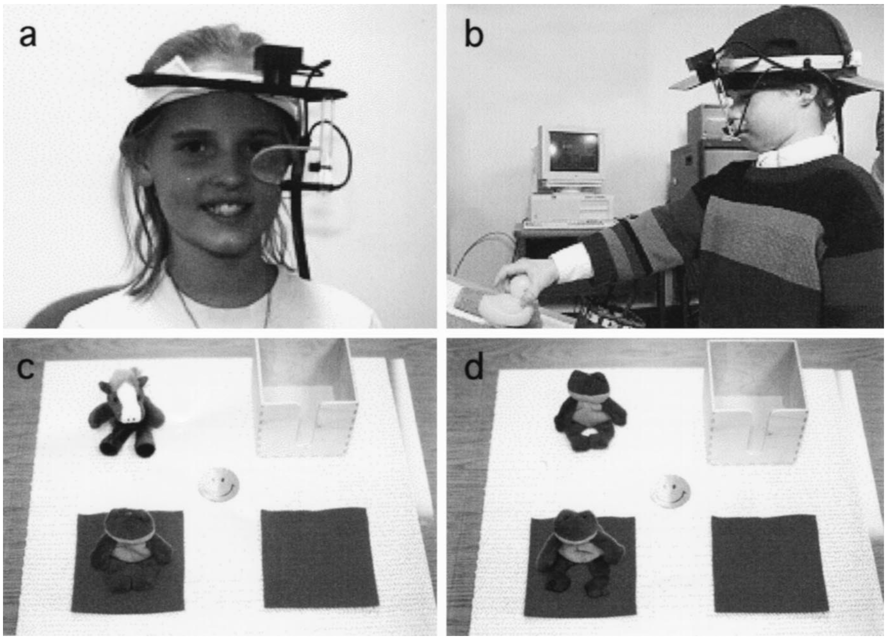

Fig. 1. (a) An eight year old girl wearing the eye-tracking visor. (b) A Five Year Old boy performing the task. (c) The 1-Referent context for ''Put the frog on the napkin in the box''. (d) The 2-Referent context for this instruction.

frog. 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. (1995) found that adult listeners could use the Referential Principle to inform syntactic commitments in the way described above (see also Spivey & Tanenhaus, 1999). Adults initially misinterpreted the ambiguous phrase as a Destination in the 1-Referent context but not in the 2-Referent context, as shown by substantially more looks to the Incorrect Destination (the empty napkin) in the 1-Referent as compared to the 2-Referent context. In fact, the number of looks to the Incorrect Destination in the 2-Referent context was essentially identical to the number observed for unambiguous sentences (''Put the frog that's on the…''), suggesting that adults' initial interpretation of the ambiguous phrase in the 2-Referent contexts was as a Modifier.

A number of possible outcomes could be expected from our study with children. If Five Year Olds rely heavily on pragmatic and referential factors, we would expect to observe a large effect of referential context on interpretation, perhaps in both the ambiguous and unambiguous instructions. This is because our unambiguous controls were relative clause constructions, whose interpretation has been found to be affected by contextual factors associated with the Referential Principle (e.g. Hamburger & Crain, 1982). If Five Year Olds rely less on the Referential Principle and more on relevant linguistic constraints to inform syntactic ambiguity resolution (e.g. as determined by structural parsing heuristics, or lexically specific syntactic preferences), we might expect signs of strong VP-attachment preference in the ambiguous trials, strong NP-attachment in the unambiguous trials, and no effect of the referential context. Finally, it is possible that Five Year Old children have essentially the same processing abilities as adults, predicting no difference in performance between children and adults.

4. Experiment 1: Five Year Old children

4.1. Method

4.1.1. Participants

Sixteen children (nine males, seven females) participated in the study. The children ranged in age from 4 years 8 months to 5 years 10 months. All participants were raised in English-speaking households. They received a small stuffed animal for their involvement in the study.

4.1.2. Equipment

Eye movements were recorded using a light-weight ISCAN eye-tracking visor, specifically designed for children. The visor was worn like a baseball cap, and consisted of a monocle and two miniature cameras (see Fig. 1). One camera recorded the visual environment from the perspective of the participant's left eye, and the other camera recorded a close-up image of the left eye. A computer analyzed the eye image in real-time, superimposing the horizontal and vertical eye position on the scene image. The scene image and the superimposed eye position, along with all auditory stimuli, were recorded to tape using a frame-accurate digital video recorder (a SONY DSR-30).

Because the scene and eye cameras were attached to the eye-tracking visor, movement of the subject's head had no impact on the accuracy of the eye position signal. Head movements resulted in the panning of the scene image. Eye position continued to be plotted on the scene image throughout any movement of the head, and appeared as a cross-hair on the scene image. Thus, for any given video frame of the scene image, the object being fixated appeared behind the superimposed cross-hair.

The ISCAN tracker determines eye position by continuously tracking both the center of the pupil and the corneal surface reflection (seen as the ''gleam'' in a person's eye). The center of the pupil is tracked relative to the corneal reflection so as to cancel out eye position error that can result from unwanted movement/slippage of the eye-tracking visor on the subject's head. This is because small movements of the visor on the subject's head result in an equal translation of both the pupil and corneal reflection in the eye video image. Thus, a movement of the visor on the head results in no change in the relative distance between the pupil and corneal reflection.

4.1.3. Procedure

Calibration of the eye-tracker requires participants to remain still while holding gaze on a series of five spatial positions. During this time, the computer is provided with alignment information from the computer operator. Because the calibration procedure requires a relatively fixed head position, the procedure was not feasible for most Five Year Old children. Instead, calibration was first done on a female adult who had relatively small eyes. The visor was then placed on the child, and the calibration accuracy was checked by having the child follow a stuffed animal around with his/her eyes (i.e., with smooth-pursuit eye movements), and by instructing the child to look at various objects. The scene image and predicted eye position were then inspected by the computer operator. If eye position was judged to be off, minor adjustments were made to the angle of the eye-tracking visor until an accurate track could be obtained. With a few exceptions, eye position remained accurate enough to indicate clearly which object was being fixated.[1]

Participants were run individually, and seated in front of an angled tabletop. Verbal instructions were provided to both parent and child, and parental consent was obtained. Each trial consisted of the child moving a set of objects around on a table based on verbal instructions. Objects included small stuffed animals (e.g. cows, pigs, frogs, ducks, and bears) and everyday objects (plates, napkins, towels, trays, bowls, boxes, and pots).

At the beginning of each trial, a small array of objects was placed on the table, and objects were identified verbally by one of the experimenters, to ensure that the child understood the names of the animals and objects. A female experimenter gave the verbal instructions. The same female voice was used for all participants. The instructions always began with a request to look at a central location (i.e., ''Look at the happy face''), followed by 2 to 4 additional instructions to move objects around. For instance, a trial might consist of the following instructions: ''Look at the happy face. Now put the cow in the box. Now put it back. Now put the pig next to the cow.'' Regardless of whether their actions were correct or not, participants were given encouraging feedback (i.e., ''Very good''). The first instruction in every trial was planned in advance, but because we were uncertain about how a child might execute the first command (and thus uncertain about what the new configuration of objects might be), the remainder of the instructions for a trial were spontaneously generated by the experimenter based on the configuration of the objects on the table. The experiment included 24 trials and lasted about 15–20 min, excluding instructions and debriefing.

One experimenter was dedicated to watching the scene image throughout the course of the experiment. This person made sure that all objects on the tabletop

[1] The reader may be concerned that using an adult for the calibration procedure would generate an inaccurate eye position signal. However, we note that objects typically subtended 3° to 4° of visual angle, making it relatively easy to determine eye fixation on an object. In addition, we have recently acquired a new ISCAN eye-tracking device that allows for the automatic calibration of Five Year Old subjects (via a ''point-of-light'' calibration procedure). Preliminary results using this new calibration technique replicate the central findings of Experiment 1 (see Section 6). We therefore conclude that the calibration procedure of Experiment 1, though more difficult, resulted in an accurate position signal.

were captured in the scene image during target trials. He also evaluated the accuracy of the eye position before each trial. This was done by checking whether the crosshair aligned with objects as they were being introduced verbally by a second experimenter, and by checking whether the crosshair aligned with the happy face during the initial instruction ''Look at the happy face.'' If the eye position was judged to be poor, a more thorough calibration check was conducted, by asking the subject to fixate objects and smoothly track a moving object. During this check, the angle of the eye-tracking monocle would be adjusted until eye position was judged to be accurate.

4.1.4. Materials

The first sentence of every target trial involved the verb ''put'', and was one of two sentence types, as shown in the example item below.

3a. Put the frog on the napkin in the box. (Ambiguous)
3b. Put the frog that's on the napkin in the box. (Unambiguous)

For ambiguous sentences, the prepositional phrase ''on the napkin'' is temporarily ambiguous. It could be indicating the Destination (i.e., where the frog is to be put) or it could be a Modifier phrase (i.e., indicating a frog which is found on a napkin). For Unambiguous sentences, the inclusion of the ''that's'' forces the Modifier interpretation of ''on the napkin'', thereby removing the temporary ambiguity.

Target trials began with one of two possible configurations of objects, as illustrated in Fig. 1. The 2-Referent context contained two identical stuffed animals (e.g. two frogs), providing two possible referents for the first noun phrase of the target sentence. One animal was sitting on the table, and the other animal was sitting on a flat object (e.g. a napkin). Two other objects were present: an Incorrect Destination (e.g. an empty napkin) and a Correct Destination (e.g. an empty box). Recall that 2-Referent contexts ought to support a Modifier interpretation of ''on the napkin'', because the definite NP ''the frog'' does not uniquely specify which frog is being referred to, thereby making modification necessary. The 1-Referent contexts were identical, except that one of the animals (the one on the table) was replaced with a different animal (e.g. a horse). Recall that 1-Referent contexts ought to support a Destination interpretation of ''on the napkin'', because the definite NP ''the frog'' specifies which frog, making any modification of the NP redundant. Pairs of animals were chosen such that (a) they were not visually similar, and (b) their names were not similar sounding. For instance, we did not use both a dog and a frog in the same target trial.

Four presentation lists were constructed by randomly combining sixteen target trials with the eight distractor trials. Distractor trials contained two to three animals and one to four everyday objects. The instructions on the distractor trials and the subsequent instructions on target trials also involved the verb ''put''. These instructions used a variety of ways to indicate the destination (e.g. ''Put the X in the Y'', ''...next to the Y'', ''...back'', ''...with the Y'', etc.). If it became clear that a child was becoming fidgety or uninterested in the study, filler trials were dropped and data collection continued. This happened with only a small number of participants, and typically occurred in the second half of the experiment.

Within a presentation list, eight target trials used an Ambiguous target sentence and eight target trials used an Unambiguous target sentence. For each of these types of sentences, four trials contained 2-Referent contexts and four trials contained 1-Referent contexts. Each target item was then rotated through these four conditions, generating four different presentation lists. Each participant was assigned to one of the four lists.

4.2. Results and discussion

The presentation of the results is divided into four sections. First, an informal description of the children's eye movements is provided, including an example of an eye movement sequence. Second, a relatively coarse-grain analysis of the children's eye movements and actions is provided, which gives a general indication of the children's parsing preferences. Third, a more fine-grained temporal analysis of all eye movements is provided by examining the probability of fixating each object over time. And fourth, the eye movement patterns of correct and incorrect action trials are compared.

4.2.1. What did the children's eye movements look like?

Because little research to date has examined children's eye movements during the course of interpreting spoken sentences, we begin by providing a brief description of the children's eye movement patterns, followed by an example. Statements made in this section will be supported by quantitative analyses below. In addition, video examples of target trials can be viewed at the following web page: www.cis.upenn.edu/~ircs/Trueswellabs/video.html. The web page contains example video clips from the Five Year Old and Adult subjects reported in this paper.

An informal inspection of the eye-tracking videotapes suggests that the children in this study were capable of interrogating the visual world with respect to the perceived speech and did so in an incremental fashion. For instance, objects tended to be fixated soon after they were referred to in the speech. If the speech at a given moment in time could potentially refer to more than one object, it was common to observe alternating fixations between these competing referents. Finally, reaching toward an object was typically preceded by a fixation on that object. All of these eye movement phenomena have been observed with adults (Allopenna et al., 1998; Ballard, Hayhoe & Pelz, 1993; Tanenhaus et al., 1995), and suggest that children rapidly interrogate the scene in a manner similar to adults.

Fig. 2 provides an example of an eye movement sequence, as recorded from a particular Five Year Old subject in response to an unambiguous instruction, e.g. ''Put the frog that's on the napkin in the box''.[2] This example is not meant to be

[2] For ease of exposition, we describe this trial in terms of the example item. The actual instruction was ''Put the monkey that's on the towel in the basket'', which was uttered in a scene containing monkeys, towels and a basket.

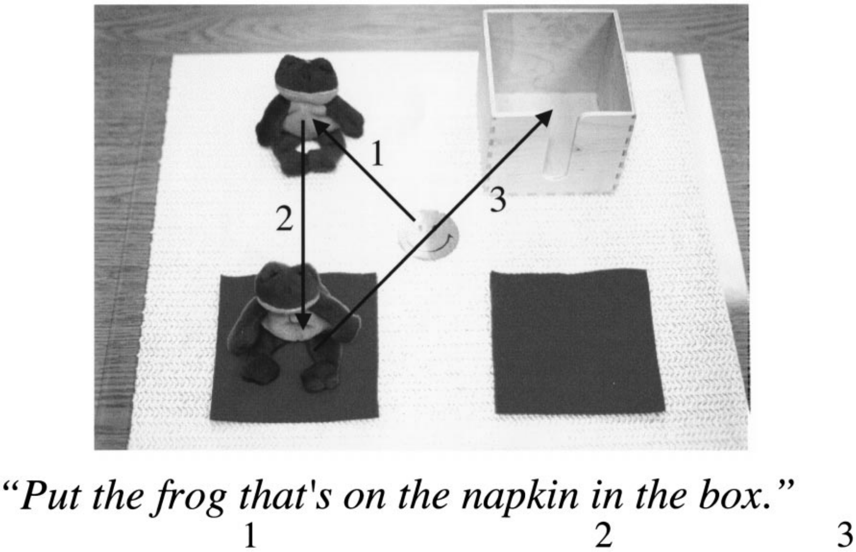

Fig. 2. Example eye movement sequence from a Five Year Old subject responding to an Unambiguous instruction.

indicative of the average eye movement pattern for this condition. Rather, it is given to help readers better evaluate the analyses reported below. In addition, the example serves to illustrate some of the eye movement phenomena mentioned above. Each arrow in the figure represents an eye movement, with the numbers indicating the sequence of these movements. These numbers also appear below the target instruction, illustrating the approximate time during the speech that each eye movement occurred. As can be seen in the figure, the trial began with the child looking at the happy face in the center of the table. Then, 300 ms after the onset of the word ''frog'' (i.e., ''Put the Frog...''), the child's direction of gaze shifted to one of the frogs (in this case, the frog that was not on a napkin). The child continued to look at this frog until 366 ms after the onset of ''napkin'', when the eyes shifted to the Target animal, i.e., the frog that was on the napkin. The final word of the sentence, ''box'', was uttered 233 ms after this eye movement. The subject then held gaze on the Target frog as he/she reached for the object. Once the object had been grasped, the direction of gaze shifted to the box, and the subject placed the frog in the box.

4.2.2. Coarse-grain analysis of eye movements and actions

Given the general phenomena described above, it seemed likely that a more formal analysis of children's eye movement patterns could yield insights into their on-going language and perceptual processes. We begin here with some relatively coarse-grain measures of the children's eye movement patterns and actions, designed to uncover children's parsing preferences for the ambiguous phrase ''on the napkin''. In particular, we examined the proportion of trials that subjects fixated the Incorrect Destination (the empty napkin) upon hearing the prepositional phrase ''on the napkin''. This measure reflects consideration of the Incorrect Destination, and indicates the extent to which subjects entertained a Destination interpretation of this phrase. In addition, this same measure was used by Tanenhaus et al. (1995) in their experiment on adults, providing a direct comparison to their research.

Eye movements to the Incorrect Destination. Digital videotapes of each participant's scene and eye-position were analyzed by hand, using slow motion and freeze frame viewing on a digital VCR. Any fixations on the Incorrect Destination (e.g. the empty napkin) were examined, starting from the onset of the preposition ''on'' in the spoken instruction, until the child's action was complete. These data were then used to compute the proportion of trials in which the Incorrect Destination was fixated during this time frame, and are plotted by condition in the upper left-hand panel of Fig. 3.[3]

If Five Year Olds were using the Referential Principle (Crain & Steedman, 1985) to interpret this phrase, we would expect them to look over to the Incorrect Destination only when the sentence is Ambiguous and in a 1-Referent context (the pattern observed for adults, see Experiment 2, Section 5, as well as Tanenhaus et al., 1995). However, as can be seen in the figure, Five Year Olds were more likely to look at the Incorrect Destination during the Ambiguous trials as compared to the Unambiguous trials, regardless of Referential Context. Approximately 70% of Ambiguous trials involved looks to the Incorrect Destination as compared to approximately 35% of the Unambiguous trials. If anything, 2-Referent contexts, which were supposed to help the listener avoid a garden-path, showed more looks to the Incorrect Destination than did the 1-Referent contexts.

Subject and item means of the Percent Looks to the Incorrect Destination were entered into separate analyses of variance (ANOVAs) with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent); and the presentation List/Item-Group factor (4 lists in the subject analysis and 4 item groups in the item analysis).[4] These ANOVAs revealed a reliable effect of Ambiguity ($F1(1, 12) = 12.62$, $P < 0.005$; $F2(1, 12) = 68.92$, $P < 0.001$), and no effect of Context ($F1(1, 12) = 2.52$; $F2(1, 12) = 3.81$, $P < 0.1$). Moreover, because a similar effect of Ambiguity was observed for both contexts, there was no significant interaction between Ambiguity and Context (both $Fs < 1$).

This eye movement pattern suggests that children's initial interpretation of the ambiguous phrase ''on the napkin'' was as a Destination rather than as a Modifier. Children were looking to the Incorrect Destination upon hearing the ambiguous phrase ''on the napkin'', indicating that they were treating this phrase as the Desti-

[3] Trials were excluded from this analysis if: (1) there was an error in the spoken instruction, (2) there was an error in how the objects were configured, or (3) the scorer deemed the track too poor to determine if a fixation had occurred on the Incorrect Destination. This made up 12% of the trials.

[4] Throughout this paper, whenever ANOVAs were conducted on proportions, identical ANOVAs were also conducted on an arcsin transformation of the data, $\arcsin((2*p) - 1)$. This was done to adjust for the fact that the proportion p is bounded at 0 and 1. Unless otherwise noted, statistically significant effects that were found for the untransformed data were also significant for the transformed data.

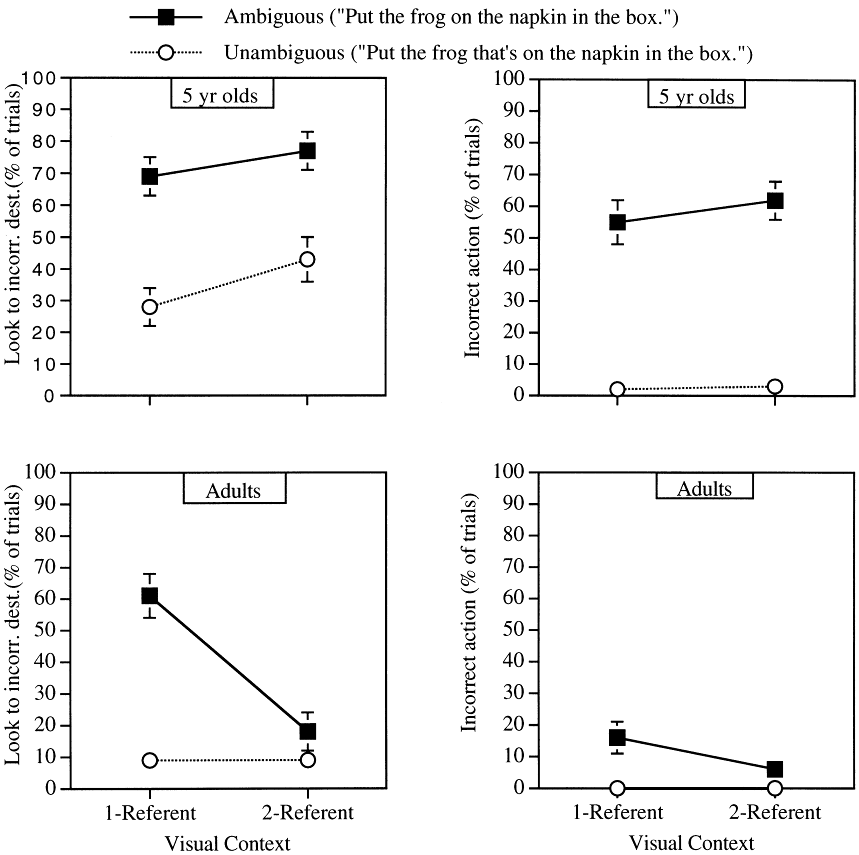

Fig. 3. Percentage of trials in which there was a look to the Incorrect Destination, e.g. the empty napkin, as measured for the onset of ambiguous phrase, e.g. ''on the napkin'' (left column). Percentage of trials in which objects were moved incorrectly (right column). Error bars indicate one standard error from the mean.

nation for ''put'' (i.e., VP attachment). The fact that Context had little effect on the eye movement measure suggests that children were not utilizing the Referential Principle to guide their interpretation of this phrase, but rather had a general bias for the Destination interpretation.

Actions of participants. A record was also kept of how each child carried out the spoken instruction of the target sentence (i.e., where objects were put). The data indicated that Five Year Olds were often unable to recover from their initial misinterpretation, resulting in actions consistent with the Destination interpretation of the Ambiguous phrase. In particular, the upper right panel of Fig. 3 plots, by condition, the percentage of trials in which the child performed an Incorrect Action. An action was considered correct if the intended object (e.g. the frog that was on the napkin)

Table 1
Number of actions on ambiguous trials, per action type (Five Year Olds)

Type of Action	Object that was moved to the:		Type of context:		Total
	Incorrect Destination	Correct Destination	1-Referent	2-Referent	
Correct	–	Target[a]	24	25	49
Falling short	Target	–	15	3	18
Falling short	Other	–	0	12	12
Hopping	Target	Target	11	1	12
Hopping	Other	Other	0	10	10
One of each	Other	Target	1	10	11
One of each	Target	Other	4	0	4
	–	Other	0	4	4
Total					120

[a] Target = Animal which was supposed to be moved. Other = Animal which was not supposed to be moved. When same object appears in both destination columns, this means the object was ''hopped'' from one destination to the other.

was moved directly to the intended destination (e.g. the empty box).[5] As can be seen in the figure, Five Year Olds were performing nearly perfectly on unambiguous instructions (e.g. ''Put the frog that's on the...''), but were showing considerable difficulty with ambiguous trials (e.g. ''Put the frog on the...''). Only 3% of the Unambiguous trials were incorrect as compared to approximately 60% of the Ambiguous trials. Moreover, for both Ambiguous and Unambiguous trials, there was little change in performance when comparing the 1-Referent context to the 2-Referent context, suggesting that these children were not making use of the Referential Principle to help guide their actions.

Subject and item means of the Percent Incorrect Actions were entered into separate ANOVAs with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent); and the presentation List/Item-Group factor (4 lists in the subject analysis and 4 item groups in the item analysis). These ANOVAs revealed a highly reliable effect of Ambiguity ($F1(1,12) = 43.96$, $P < 0.001$; $F2(1,12) = 200.24$, $P < 0.001$) and no effect of Context ($F1(1,12) = 0.52$; $F2(1,12) = 1.13$). Again, because a similar effect of Ambiguity was observed for both Contexts, there was no significant interaction between Ambiguity and Context (both $Fs < 1$).

Table 1 presents a detailed classification of all actions on Ambiguous trials. The first two columns indicate which objects were placed on the Incorrect Destination (e.g. the empty napkin) and/or the Correct Destination (e.g. the empty box), respectively. In all cases, there were two possible objects that could have been moved: the Target animal (e.g. the frog that had been on the napkin) or the Other animal (the

[5] In all results in this paper, an action was also considered correct if both the object and the modifying object were moved to the Correct Destination (e.g. both the frog and the napkin underneath the frog were placed in the box), consisting of less than 1% of the trials overall.

animal that had been on the table, e.g. another frog in the 2-Referent context or a horse in the 1-Referent context). The first row of the table indicates correct actions (i.e., no animal was moved to the Incorrect Destination, and the Target animal was moved to the Correct Destination). As was already mentioned, there was no difference between 1-Referent and 2-Referent contexts for correct actions.

The remaining rows of Table 1 indicate different classes of incorrect actions. Despite some variation in the types of errors, all errors showed clear signs that Five Year Olds had misinterpreted the preposition ''on the napkin'' as the Destination. In particular, all but three of the errors involved moving an animal to the Incorrect Destination.

Rows 2 and 3 indicate the most common error, which we have dubbed the ''falling short'' error: placing either the Target animal (row 2) or the Other animal (row 3) on the Incorrect Destination and moving nothing to the Correct Destination. These errors reflect assigning a Destination interpretation to the first prepositional phrase (on the napkin) and ignoring the second prepositional phrase. Rows 4 and 5 indicate what we call ''hopping'' errors: the child picked up one of the animals and placed it first on the Incorrect Destination and then into the Correct Destination (always in this order). This entire action was typically done without releasing the animal. The ''hopping'' errors reflect assignment of a Destination interpretation to both prepositions. Rows 6 and 7 indicate what we call ''one of each'' errors, in which one animal was placed on the Incorrect Destination and another animal was placed in the Correct Destination (again, always in this order). These errors also reflect assignment of the Destination role to both prepositional phrases. Finally, row 8 indicates the error of moving the Other animal into the Correct Destination. Like correct responses, these actions reflect assignment of the Destination role to the second prepositional phrase.

The distribution of errors was different for 1-Referent and 2-Referent contexts, and worth evaluation. In particular, actions in the 1-Referent contexts almost always involved movement of the Target animal and not the Other animal. Recall that 1-Referent contexts involved two different animals (e.g. the Target being a frog and the Other being a horse). So, it is not surprising that the majority of errors for 1-Referent contexts involved moving the Target animal (the frog) because it was the only animal mentioned in the Instruction.

The 2-Referent contexts involved two identical animals (e.g. two frogs, one of which was on a napkin). Decisions to move these animals were essentially equal: a total of 39 trials involved movement of the Target animal, whereas 36 trials involved movement of the Other animal. (Some trials involve movement of both animals, thus the numbers do not add up to the total number of trials.) These data suggest that the phrase ''on the napkin'' was rarely taken as a Modifier, resulting in chance performance when determining a referent for the direct object NP ''the frog''. This may also mean that some of the Correct Actions resulted from a decision to ignore the first prepositional phrase, rather than treating it as a Modifier. Subjects sometimes may have stumbled upon the correct action.

However, it is also important to note that the presence of a second identical animal did have some impact on actions. In particular, movements of the Other animal were almost always to the Incorrect Destination, whereas movements of the Target animal were almost always to the Correct Destination. An inspection of Table 1 shows that this pattern appears in all three major classes of errors (''falling short'', ''hopping'', and ''one of each'' errors). Five Year Olds seem to be reasoning that because the Target frog is already on a napkin, it must be the Other frog that needs to be moved to the empty napkin. Thus, when the Incorrect Destination is interpreted as a Destination, the Other animal rushes to fill this position. We will return to this issue in Section 6.

The error data from each individual subject were also inspected to see if particular subjects favored particular errors. All subjects produced at least one error during the eight ambiguous trials. Two subjects produced only one error; one subject produced two errors; and the remaining thirteen produced three or more errors. Of the thirteen subjects who produced the majority of errors, most subjects chose to produce one class of error over others. In particular, four subjects were especially fond of the ''hopping'' error (rows 4 and 5 of Table 1). Five other subjects tended to produce the ''falling short'' error (rows 2 and 3 of Table 1). And, three other subjects tended to produce the ''one of each'' error (rows 6 and 7). The remaining subject produced a mixture of responses. Subjects' preferences were not absolute: nine of these thirteen subjects produced at least one error different from their preferred error. The general within-subject consistency in errors suggests that children may adopt a single strategy (structural or interpretive) for dealing with two prepositional phrases. Crucially, each of these strategies involved assignment of a Destination role (VP attachment) to the first prepositional phrase.

Summary of coarse-grain data. The eye movement and action data clearly indicate that Five Year Olds preferred to interpret the ambiguous phrase ''on the napkin'' as a Destination. In particular, subjects were more likely to direct their gaze to the Incorrect Destination during Ambiguous as compared to Unambiguous trials, regardless of referential context. Moreover, incorrect actions, which typically included movements to the Incorrect Destination, occurred almost exclusively in response to Ambiguous rather than Unambiguous sentences, again independent of context. Finally, subjects were at chance levels when selecting the Target over the Other animal in the 2-Referent Ambiguous condition, suggesting that the prepositional phrase ''on the napkin'' was seldom treated as a Modifier and therefore was viewed as uninformative in determining a possible referent.

Although the coarse-grain eye movement data provide valuable information pertaining to children's assignment of the Destination interpretation, the data fail to provide the temporal detail necessary to make strong conclusions about how the Destination interpretation developed over time (i.e., conclusions about on-line sentence processing). For instance, in the eye movement analysis above, fixations on the Incorrect Destination counted toward the analysis regardless of when they occurred, so long as they happened some time between the onset of the preposition ''on'' and the completion of the action. It could be the case that the majority of these fixations occurred late in the trial, and corresponded entirely to eye fixations associated with reaching (i.e., when moving the frog onto the empty napkin, subjects looked at the napkin). Thus, a more fine-grain temporal analysis of the eye movements is needed to help determine exactly when these fixations occurred, and to help determine if incremental interpretation is the norm for children in this age group.

4.2.3. Fine-grain eye movement analysis

This section provides a more fine-grained record of the fixation data over time, analyzing all fixations on all objects from the onset of the sentence to the completion of the action, in 1/30 second intervals (the sampling rate of the video record). This analysis serves at least three purposes. First, fixations on the Incorrect Destination over time should reveal exactly when the looks to the Incorrect Destination occurred, thereby telling us something about how the Destination interpretation developed during the comprehension of the sentence. Second, the more detailed analyses of looks to the Incorrect Destination can serve as a cross-check of the coarse-grain eye movement measure reported above. This is because a different scorer, who was blind to the earlier eye movement scores, performed these additional analyses. And third, the time-course of fixating other objects in the scene should provide a more complete picture of the interpretation that listeners were assigning to the input. For instance, when the phrase ''on the napkin'' is interpreted as a Modifier, we should expect it to help a listener distinguish between the two possible referents in the 2-Referent context. So, in the 2-Referent ''frog'' example, when hearing the phrase ''Put the frog…'' listeners, on average, ought to look with equal likelihood to either frog. The unambiguous phrase ''that's on the napkin…'' should increase the probability of looking at the Target frog (the one on the napkin) and eliminate looks to the Other frog. However, if children fail to treat the ambiguous phrase ''on the napkin'' as a Modifier, we expect to see few signs of referential disambiguation on 2-Referent Ambiguous trials.

Fig. 4 plots the fixation probabilities of objects over time, with separate graphs for each of the four conditions. The blue circles in these graphs indicate the probability of fixating the Target animal (e.g. the frog on the napkin). The red Xs indicate the probability of fixating the Other animal (e.g. the frog/horse). The green triangles indicate the probability of fixating the Incorrect Destination (the empty napkin). And the pink squares indicate the probability of fixating the Correct Destination (the empty box). These data were generated from the videotape records, by noting which object was being fixating during each video frame.[6]

Speech onset times. The vertical lines appearing below the x-axis indicate the mean onset of each content word in the sentence (e.g. put, frog, napkin, and box). These data are crucial for evaluating how children's eye movements relate to different points in the speech. It turns out that the speech generated by the female experimenter was extremely regular in its timing across all conditions. Mean word onset times were within 17 ms of each other in the ambiguous trials and within 33 ms of

[6] Trials were excluded from these plots if there was (a) an error in the spoken instruction, (b) an error in the configuration of objects or (c) a trackloss occurring on more than 33% of the frames. This made up 11% of the trials. The same criteria were applied when determining which trials to drop from statistical tests within a particular time slice, i.e., a trial was dropped when the relevant time slice contained a trackloss for more than 33% of its frames.

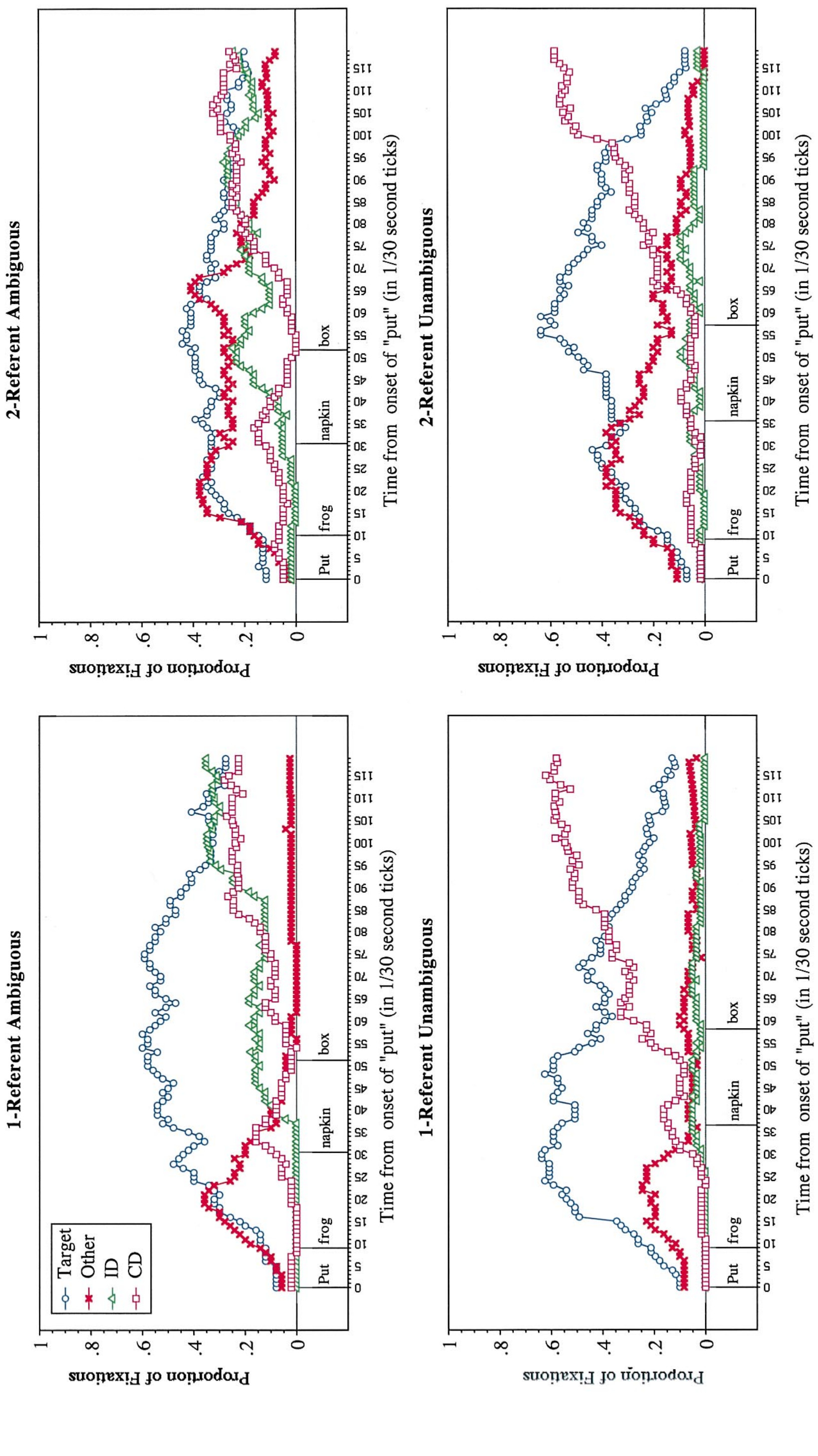

Fig. 4. Probability of fixating each object type over time for each of the four conditions (Experiment 1, Five Year Old subjects).

each other in the unambiguous trials, with all standard errors being less than 33 ms. The onsets of ''napkin'' and ''box'' were delayed by 200 ms in the Unambiguous as compared to the Ambiguous sentences because the Unambiguous versions included ''that's''.

The highly regular word-onset times should lessen concerns that prosodic cues to prepositional phrase attachment were substantially different across conditions. Indeed, some of the strongest prosodic cues to the PP-attachment ambiguity should come from timing changes like syllable lengthening and pausing (e.g. lengthening of the direct object noun and pausing after the direct object noun should signal VP attachment). Such cues, if present in different proportions across conditions, would have substantially shifted the onset times of ''napkin'' and ''box''. It is possible that intonation (tune) differences were present in the stimuli across conditions. However, we have recently conducted an eye movement study with pre-recorded stimuli that had been cross-spliced to control for prosody. Preliminary results from this study replicate the central findings reported here for Five Year Olds (see Section 6).

Given the regularity in the speech onset times for content words, we have for simplicity plotted fixation probabilities (like Fig. 4) relative to the onset of the first word in the sentence, ''put''. Individual plots relative to the onset of each content word in a sentence generate graphs that are similar to those shown in Fig. 4. All statistical tests reported below are of course performed relative to the relevant word onset time for each particular observation, so as to make sure we are comparing apples with apples.

Recognition of ''napkin'' (and syntactic ambiguity). The development of the Destination interpretation of the phrase ''on the napkin'' can be inferred by examining the proportion of fixations on the Incorrect Destination over time (the green triangles in Fig. 4). As can be seen in the figure, increased looks to the Incorrect Destination occurred approximately 300 ms after the onset of ''napkin'', with most of these fixations occurring in the Ambiguous conditions (the upper two graphs). If one takes into account that it takes about 200 ms to program an eye movement (Matin, Shao & Boff, 1993), this is the earliest possible time during the speech that we would expect to see consideration of the Incorrect Destination. These increases occur in both the 1-Referent and 2-Referent Ambiguous conditions, suggesting insensitivity to the Referential Principle.

Statistics. We quantified these early looks to the Incorrect Destination by averaging the proportion of time spent fixating the Incorrect Destination during an 800 ms time slice of the data, from 200 ms to 1000 ms after the onset of ''napkin''. Subject and item means were entered into separate ANOVAs with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent); and presentation List or Item-Group factor (4 lists in the subject analysis and 4 item groups in the item analysis). The analyses revealed a main effect of Ambiguity ($F1(1,12) = 17.63$, $P < 0.01$; $F2(1,12) = 9.58$, $P < 0.01$), no effect of Context ($F < 1$) and no interaction between these factors ($Fs < 1$).

Still finer-grain temporal analyses were performed within this time slice by conducting similar ANOVAs on each 100 ms interval. A reliable effect of Ambiguity was not observed until the 4th interval (i.e., the 300 to 400 ms window;

$F1(1,12) = 22.06$, $P < 0.01$; $F2(1,12) = 7.75$, $P < 0.05$). Thus, subjects needed to hear approximately 300 ms of the word ''napkin'' before the programming of eye movements resulted in reliably more looks to the Incorrect Destination for Ambiguous trials as compared to Unambiguous trials.

Cross-check with coarse-grain analysis. As mentioned above, it is possible to use the fine-grain analysis of fixations on the Incorrect Destination to cross-check the earlier coarse-grain eye movement analysis of this same measure, because both sets of measures were collected by different scorers. The coarse-grain eye movement analysis reflected the proportion of trials in which a fixation occurred between the onset of ''on'' and the completion of the action. We generated the same measure from the fine-grain analysis by computing the proportion of trials in which a fixation on the Incorrect Destination occurred within this same large time-window (from about frame 16 and on in Fig. 4). A total of 205 out of the 225 observations (91.1%) were in agreement between the old and new analyses. The new analysis generated values that were quite similar to those appearing in Fig. 3 (1-Ref Ambig: 65%; 1-Ref Unambig: 20%; 2-Ref Ambig: 74%; 2-Ref Unambig: 34%). An analysis of these new means revealed a very similar statistical pattern: a reliable main effect of Ambiguity ($F1(1,12) = 24.02$, $P < 0.01$; $F2(1,12) = 94.10$, $P < 0.01$), a marginal effect of Context ($F1(1,12) = 3.95$, $P < 0.1$; $F2(1,12) = 6.35$, $P < 0.05$) and no interaction between Context and Ambiguity (both $Fs < 1$). The marginal effect of Context reflects the fact that, if anything, 2-Referent contexts, rather than 1-Referent contexts, were showing more looks to the Incorrect Destination.

Both the original coarse-grain eye movement measure and the cross-check generated a relatively high proportion of looks to the Incorrect Destination for Unambiguous sentences (about 25–30%). This high rate seems at first to contradict the low rate of looks to the Incorrect Destination found in the Unambiguous probability plots found in Fig. 4. The cross-check confirms that the relatively high rate was due to an accumulation of fixations within a large time window. Subjects rarely look over to the Incorrect Destination at any given moment during the interpretation of an Unambiguous sentence, but these fixations added up over time. Indeed, this effect highlights the limitations of coarse-grain eye movement measures, which ignore the factor of time.

In sum, the more fine-grain temporal analysis of eye movements revealed that looks to the Incorrect Destination occurred more in the Ambiguous sentences, independent of context, even at the earliest stages of processing. Reliable signs of considering the Incorrect Destination on Ambiguous trials began when subjects heard approximately 300 ms of the word ''napkin''. Moreover, by accumulating the fixations over time, we were able to validate the earlier coarse-grain measure of eye movements.

Recognition of direct object noun ''frog'' (and referential ambiguity). In Fig. 4, the blue circles represent the probability of fixating the Target animal (the frog that is sitting on the napkin). The red Xs represent the probability of fixating the Other animal (the horse in the 1-Referent contexts, and the other frog in the 2-Referent contexts). As can be seen in the figure, a large proportion of fixations early in each trial were dedicated to these objects, presumably because subjects were planning to move one of them. By comparing these two probability curves, we can estimate the time-course with which listeners converged on a referent for the direct object ''the frog''.

In all four conditions, we initially observe an equal increase in the probability of fixating both the Target and the Other animal. In fact, these increases begin slightly before the onset of the noun (''frog''). These anticipatory looks to the animals are not surprising, given that the children were told during the instructions that they would be moving toy animals. Thus, hearing ''put'' led to some looks to the toy animals. In addition, we suspect that the children found these animals more visually interesting than the pots, pans, etc., leading to earlier inspection of these objects. About 250–350 ms after the onset of the noun, the 1-Referent conditions (the left graphs) show a divergence between fixations to the Target and Other animals. This divergence, with greater probability of fixating the Target, suggests that the correct referent was established soon after disambiguating phonemic material was perceived. In all 1-Referent contexts, the onset phoneme was different for the Target and the Other animals (e.g. *f*rog vs. *h*orse). Again, given that it takes about 200 ms to program an eye movement, a divergence at 250–350 ms suggests that the Five Year Olds were sometimes capable of establishing reference based upon the first few phonemes of the target noun. The timing of this divergence is consistent with recent work on spoken word recognition in adults and children, which found that children and adults were affected by the presence or absence of phoneme-initial cohorts in a scene, and that eye movements to a referent were launched soon after a disambiguating phoneme was encountered (see Allopenna et al., 1998; Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998; Swingley, Pinto & Fernald, 1999).

A similar divergence between fixations on the Target and the Other animal was substantially delayed in the 2-Referent contexts (the two graphs on the right, Fig. 4). A divergence did not occur until the onset of the word ''napkin'' in the 2-Referent Unambiguous condition (the lower right graph), where we see increased looks to the Target animal. This delay as compared to the 1-Referent Unambiguous condition is expected because the phonemic material up to that point (e.g. ''The frog that's on the...'') is unlikely to help distinguish between the two frogs in the 2-Referent context. The exact moment of phonemic disambiguation is difficult to define, because ''on'' may help some listeners converge on the Target frog, given that the Other frog was not sitting on anything other than the table. Consistent with this, the divergence appears to be less ''crisp'' in the 2-Referent Unambiguous condition as compared to the 1-Referent Unambiguous condition.

Importantly, a divergence between the Target and the Other animal never completely occurs in the 2-Referent Ambiguous condition (the upper right graph). This lack of a clear divergence is consistent with the hypothesis that children were failing to treat the prepositional phrase ''on the napkin'' as a Modifier in the 2-Referent Ambiguous condition. If children had been treating it as a Modifier, the phrase should have helped them disambiguate between the two competing referents. Indeed, recall that the children's actions in this condition were about evenly split between moving the Target or the Other animal. The plot in Fig. 4 represents all of these trials, i.e., correct and incorrect actions. For ease of exposition, we postpone a breakdown of correct and incorrect trials until the end of the Results section (see section 4.2.4).

Statistics. We quantified the degree of referential ambiguity by comparing looking times to the Target animal with looking times to the Other animal during different time slices of the speech. For brevity, a presentation of these rather detailed analyses has been placed in the Appendix. The results can be summarized as follows. First, early in the speech, prior to having any phonemic information to discriminate between the two animals (i.e., prior to hearing the direct object ''frog''), there were no significant differences in looking times to the Target vs. the Other animal in any condition. Second, upon hearing ''frog'' and prior to hearing ''napkin'', reliably more time was spent looking at the Target animal over the Other animal in the 1-Referent contexts but not the 2-Referent contexts. Indeed, it is only in the 1-Referent contexts that the noun ''frog'' could help discriminate the Target animal (a frog) from the Other animal (a horse). Finally, upon hearing 400 ms of ''napkin'', subjects spent more time fixating the Target over the Other animal in all conditions but the 2-Referent Ambiguous condition. Thus, the phrase ''that's on the napkin'' was taken as a Modifier in the 2-Referent Unambiguous condition, helping to distinguish the Target from the Other animal. But, the ambiguous phrase ''on the napkin'' was not taken as a Modifier in the 2-Referent context, resulting in continued competition between the two possible referents.

Summary of fine-grain. In sum, the fine grain eye movement analyses indicate that children's first interpretation of the ambiguous phrase ''on the napkin'' tends to be one of Destination rather than Modifier. Reliably more looks to the Incorrect Destination were found in Ambiguous as compared to Unambiguous trials soon after hearing ''napkin'' in the speech. Crucially, the fine-grain eye movement analyses revealed that children's assignment of an interpretation (correct or otherwise) is highly incremental. Referential competition between Target and Other animals was consistently resolved at the points in speech where phonemic information could help distinguish the two referents: at ''frog'' in the 1-Referent conditions, and at ''napkin'' in the 2-Referent Unambiguous condition. The 2-Referent Ambiguous condition showed no such resolution, suggesting an inability to take the ambiguous phrase ''on the napkin'' as a Modifier.

4.2.4. Eye movements for correct and incorrect trials

Given that the Ambiguous trials contained roughly equal Correct and Incorrect Actions (40% and 60% respectively), we also separately report the eye movement patterns of correct and incorrect trials. These eye movements, especially during the early stages of a trial, may yield some insight into what caused errors in actions. The comparisons below are largely descriptive because division of the data into correct and incorrect trials resulted in a number of missing cells in subject and item means, thereby making it difficult to perform adequate statistical tests.

Fig. 5 plots the eye movement probability curves for correct actions and incorrect actions in the 1-Referent Ambiguous and 2-Referent Ambiguous conditions. The two graphs on the right show eye fixation probabilities in the 2-Referent context for correct trials (upper right) and incorrect trials (lower right). During the later part of

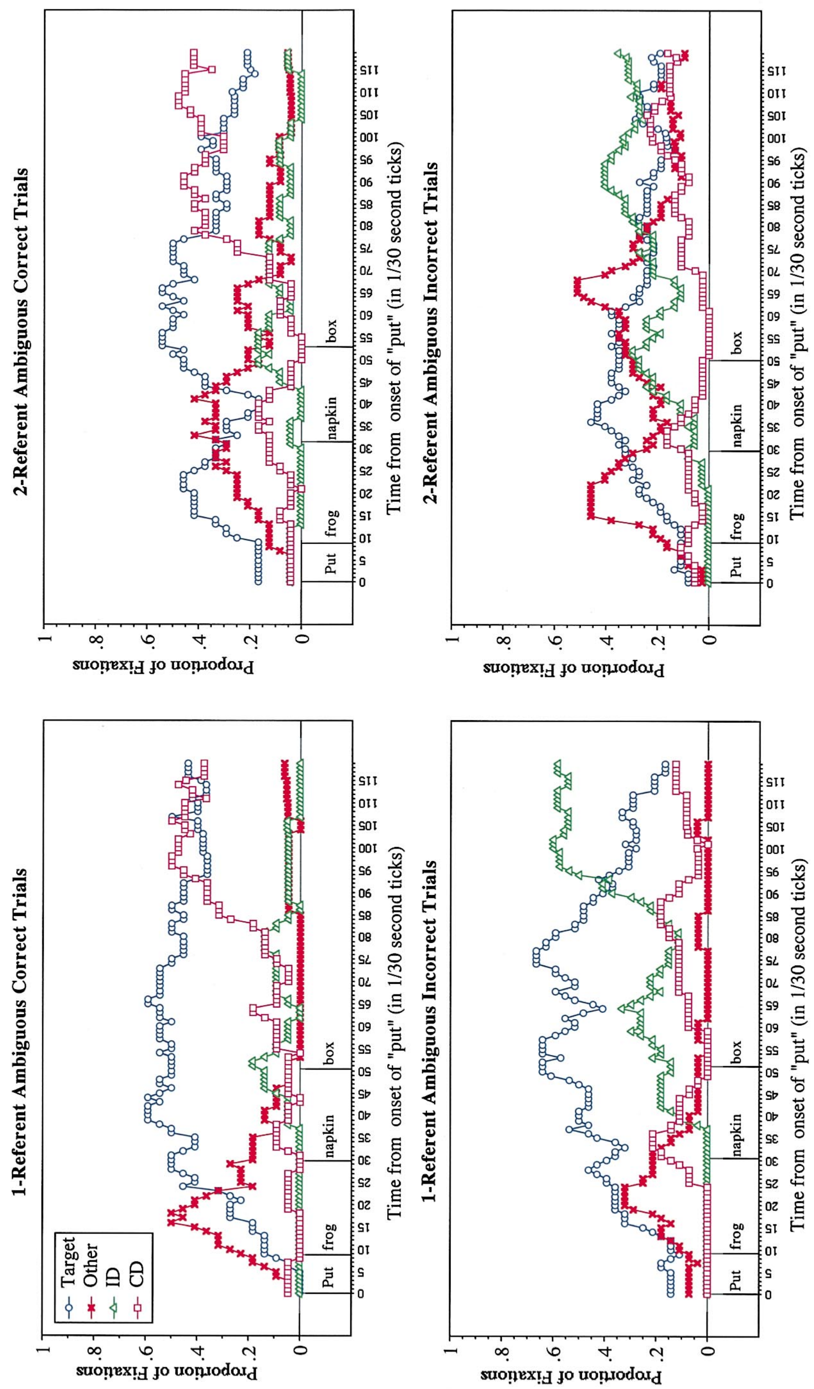

Fig. 5. Probability of fixating each object type over time for Ambiguous trials, comparing incorrect and correct trials (Experiment 1, Five Year Old subjects).

the correct trials, subjects tended to be looking at the Target animal and the Correct Destination. This is to be expected, because on correct trials children were moving the Target animal to the Correct Destination. During the later part of the incorrect trials, subjects tended to be looking at the Other animal and the Incorrect Destination. Again, this is to be expected because most incorrect actions involved these objects.

It is interesting to note, however, that the early portions of correct and incorrect 2-Referent trials are also showing a difference in fixation distributions. Correct trials appear to consist of early looks to the Target animal and/or slightly later looks to the Other animal.[7] In contrast, incorrect trials appear to consist of early looks to the Other animal and/or slightly later looks to the Target animal. Given that correct trials involve movement of the Target animal, and Incorrect trials tend to involve initial movement of the Other animal (see Table 1), one might expect that early looks to the Target or Other animal would correlate with which of the two animals was moved first in a trial. Indeed, when the Target animal was fixated first, it was moved first 63% of the time, as compared to moving the Other animal first 37% of the time. When the Other animal was fixated first, it was moved first 71% of the time as compared to moving the Target animal first 29% of the time. On the face of it, these data suggest that listeners were rarely treating the phrase ''on the napkin'' as a Modifier, and therefore found it completely uninformative with regards to selecting a referent. Subjects tended to select a referent based on which animal they happened to look at first.

Early looks to the Other animal do not correlate with the movement of the Other animal in all cases, but only in syntactically ambiguous environments. For instance, Five Year Olds showed essentially no errors on 2-Referent Unambiguous trials, even though an inspection of Fig. 4 reveals many early looks to the Other animal in this condition. It appears that when the grammar of the language requires modification, the structure guides interpretation and the ultimate decision about reference. Moreover, it is also not the case that early looks to the Other animal always led to an incorrect action. The left half of Fig. 5 compares the eye movement probabilities for Correct and Incorrect 1-Referent Ambiguous trials. If anything, these items show early looks to the Other animal leading to correct actions rather than incorrect actions. It is less clear why this pattern occurs. However, the pattern does show that accidental looks to the Other animal (the horse) simply result in a shift in attention to the Target animal (the frog) in these contexts.

We postpone further discussion of the Five Year Olds until after presenting the data of Adult subjects.

[7] Specific eye movement sequences cannot be entirely determined from graphs of this type. For instance, although some correct trials contained the sequence of early looks to the Target then the Other then the Target again, there were also sequences beginning with a delayed look to the Other animal then the Target animal. Thus, the graphs reflect the fact that the distribution of fixation probabilities was different for correct and incorrect trials.

5. Experiment 2: Adults

5.1. Method

5.1.1. Participants

12 adults (3 male, 9 female) participated in the study. The adults ranged in age from 18 to 22 years. All participants were raised in English-speaking households, where English was the dominant language. They received $6 or extra course credit for their participation.

5.1.2. Procedure

The equipment and procedure were the same as Experiment 1. An initial pilot study showed that adult subjects found the child-version of the experiment too easy, responding slowly to the instructions and often holding gaze on the central fixation throughout the course of each instruction. For these reasons, the following changes were made, making the experiment more similar to the Tanenhaus et al. (1995) study. First, participants were not told until the end of the experiment that they were being compared to children's performance on a similar task. Second, filler trials were increased to 26 trials, and frequently contained more complex instructions such as ''Put the cow in between the napkin and the towel.'' Third, 3–4 instructions were given on each trial, with all instructions being scripted in advance. Fourth, participants were reminded that they should execute their actions as quickly as possible. This was done by saying at the start of the experiment, and after trial eight, ''Remember to move as quickly as possible.'' Finally, participants were not given the ''very good'' feedback after actions.

5.1.3. Materials

The target materials and design were the same as the previous experiment, with the only exception being that the first instruction did not contain the word ''now''. In the previous experiment, the initial instruction of each target and filler trial began with a ''now'', e.g. ''Look at the happy face. Now put the cow in the box. Now put it back...etc.'' In this experiment, all trials excluded the first ''now'', e.g. ''Look at the happy face. Put the cow in the box. Now put it back...etc.'' This was done because of concerns that adults might incorrectly interpret the ''now'' as an instruction to hold gaze on the happy face while performing the first action. Tanenhaus et al. (1995) also did not use ''now'' on the first instruction, apparently for the same reason.

5.2. Results and discussion

Data were analyzed using the same measures as those reported in the previous experiment: coarse-grain eye movement and action data; and fine-grain analyses.

5.2.1. Coarse-grain analyses

The coarse-grain analyses of the eye movement and action data are presented in the lower two panels of Fig. 3. Both sets of data suggest that adults experienced a garden-path only for the Ambiguous sentences in the 1-Referent context. In particular, the eye movement pattern (left panel) showed increased looks to the Incorrect Destination only for the Ambiguous sentences in the 1-Referent context, resulting in a reliable interaction between context and Ambiguity ($F1(1,8) = 9.89$, $P < 0.05$; $F2(1,12) = 14.04$, $P < 0.005$). There were also reliable effects of Ambiguity ($F1(1,8) = 16.46$, $P < 0.005$; $F2(1,12) = 49.18$, $P < 0.001$) and Context ($F1(1,8) = 10.35$, $P < 0.05$; $F2(1,12) = 12.52$, $P < 0.005$). This statistical pattern and the values reported in the figure are essentially identical to those reported in the adult study of Tanenhaus et al. (1995).

The percentage of Incorrect Actions (right panel) showed the most errors in the 1-Referent Ambiguous condition, although the difference was quite small because there were so few errors. An ANOVA on the percent Incorrect Actions revealed that this interaction was not significant ($F1(1,8) = 1.80$; $F2(1,12) = 2.93$). The effect of Ambiguity was significant ($F1(1,8) = 6.37$, $P < 0.05$; $F2(1,12) = 5.11$, $P < 0.05$), and there was no effect of Context ($F1(1,8) = 1.80$; $F2(1,12) = 2.93$). Errors were not expected to be observed with the Adult participants. However, it is important to note that a small number of errors (less than 3 in total) were also observed with the Tanenhaus et al. (1995) adult participants, all in the 1-Referent Ambiguous Condition (Tanenhaus, personal communication). Finally, exclusion of the ten incorrect trials from the eye movement analysis revealed the same statistical pattern (1-Ref Ambiguous: 55%; 1-Ref Unambiguous: 9%; 2-Ref Ambiguous: 12%; 2-Ref Unambiguous: 8%), and included a reliable interaction between Context and Ambiguity ($F1(1,8) = 11.18$, $P < 0.05$).

5.2.2. Fine-grain eye movement analysis

Fig. 6 plots the fixation probabilities over time, in 1/30 s, for each of the four objects, broken down by condition. Trials containing track losses or other malfunctions were dropped from these plots based on the same criteria as Experiment 1 (see above), resulting in the exclusion of 7% of the trials.

Speech onset times. Vertical lines appearing below the *x*-axis indicate the mean onset of each content word in the sentence (e.g. put, frog, napkin and box). Mean onset times were within 17 ms of each other in the ambiguous trials and within 33 ms of each other in the unambiguous trials, with all standard errors less than 33 ms. The onset of ‘‘napkin’’ and ‘‘box’’ were delayed by 200 ms in the Unambiguous as compared to the Ambiguous sentences because the Unambiguous sentences included ‘‘that’s’’. Again, the data suggest a high degree of regularity in prosodic timing cues across conditions. One striking difference from the previous experiment is that the speech was considerably faster to the adult subjects. This is partially an unexpected consequence of the female speaker pressuring subjects to proceed more quickly (she spoke more quickly). It is also a natural consequence of addressing adults with a faster speech rate.

We also see in the figure that the adult subjects completed the task considerably more quickly than the children. Eye movement data are plotted only out to 3300 ms from the onset of ‘‘put’’ because most trials were complete by this point (the release

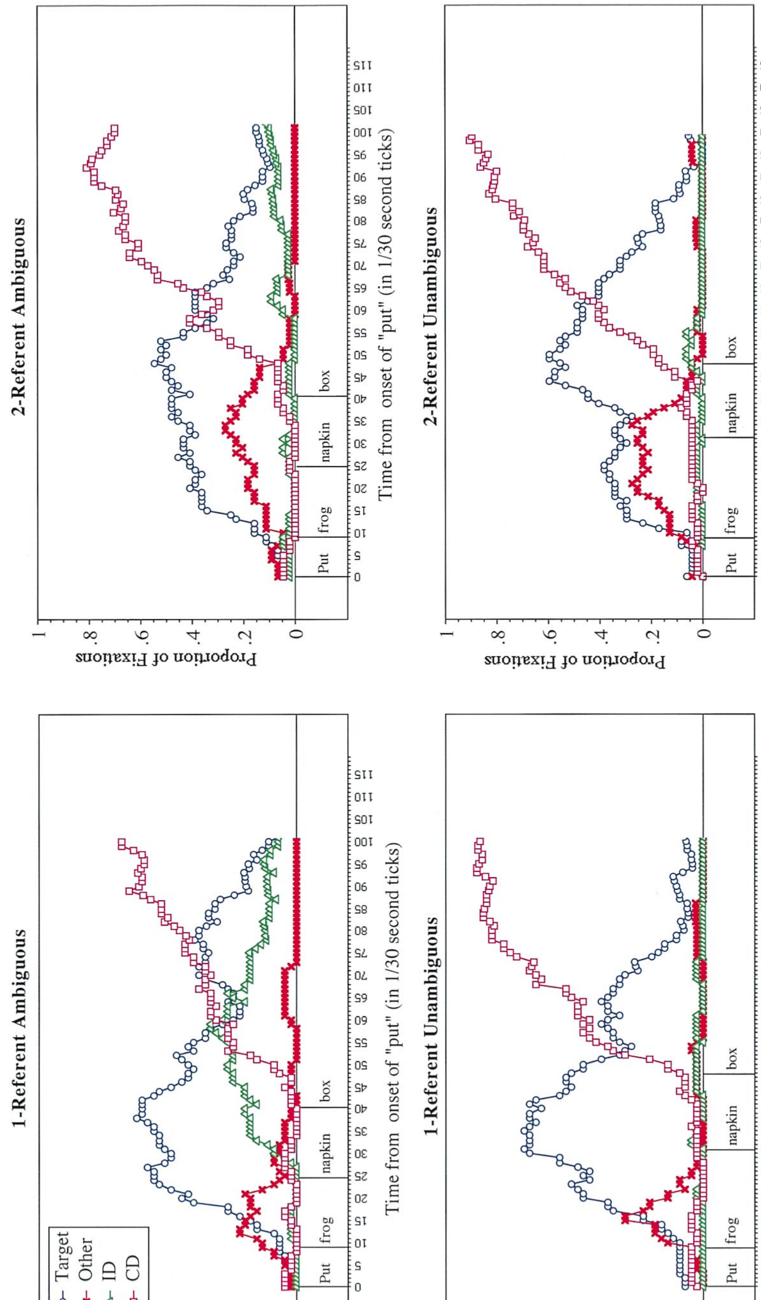

Fig. 6. Probability of fixating each object type over time for each of the four conditions (Experiment 2, Adult subjects).

of the object). This faster response in the action is likely to be due to the faster speech rate and the fact that the subjects were adults.

Recognition of ''napkin'' (and syntactic ambiguity). As can be seen in the figure, increased looks to the Incorrect Destination occur about 150–200 ms after the onset of ''napkin'', and are primarily restricted to the 1-Referent Ambiguous condition (upper left panel). This increase suggests that subjects' initial interpretation of the ambiguous phrase was as a Destination in this condition. Small increases are also observed in the 2-Referent Ambiguous condition (upper right panel), but tend to appear relatively late in the trial.

We quantified these early looks to Incorrect Destination by averaging the proportion of time spent fixating the Incorrect Destination during an 800 ms time slice of the data, from 200 ms to 1000 ms after the onset of ''napkin''. Subject and item means were entered into separate analyses of variance (ANOVAs) with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent); and presentation List or Item-Group factor (4 lists in the subject analysis and 4 item groups in the item analysis). The analysis revealed a reliable effect of Ambiguity ($F1(1,8) = 10.06$, $P < 0.05$; $F2(1,12) = 21.30$, $P < 0.01$), a reliable effect of Context ($F1(1,8) = 9.84$, $P < 0.05$; $F2(1,12) = 19.74$, $P < 0.01$) and a reliable interaction between Ambiguity and Context ($F1(1,8) = 14.97$, $P < 0.01$; $F2(1,12) = 22.15$, $P < 0.01$). Separate ANOVAs were conducted on the 1-Referent and 2-Referent context data, revealing a reliable effect of Ambiguity in the 1-Referent context ($F1(1,8) = 13.25$, $P < 0.01$; $F2(1,12) = 25.61$, $P < 0.01$) but not the 2-Referent context (both $Fs < 1$).

Even finer-grain time-course analyses were performed within this 800 ms time slice, by conducting similar ANOVAs at each 100 ms time slice. A reliable interaction between Context and Ambiguity was not observed until the 2nd time slice (100–200 ms window, $F1(1,8) = 15.12$, $P < 0.01$; $F2(1,12) = 10.76$, $P < 0.01$). This suggests that consideration of the Destination interpretation in the 1-Referent context occurred after hearing the first 100 ms of ''napkin''. Thus sensitivity to the Referential Principle in adults is essentially immediate.

Cross-check with coarse-grain analysis. In this experiment, we again performed a cross-check with the coarse grain eye movement analyses reported above. The fine-grain data generated the following proportion of trials with fixations on the Incorrect Destination (1-Referent, Ambiguous: 69%, Unambiguous: 16%; 2-Referent, Ambiguous: 25%, Unambiguous: 15%), again quite similar to earlier coarse-grain measures. A total of 165 out of the 181 observations (91%) were in agreement. Also, ANOVAs of the new data revealed the same statistical pattern, including a reliable interaction between Context and Ambiguity ($F1(1,8) = 12.74$, $P < 0.01$; $F2(1,12) = 12.58$, $P < 0.01$).

Thus, it appears that adults' consideration of the Destination interpretation in the 1-Referent context occurs at the earliest stages of processing. Reliable interactions between Ambiguity and Context were present in both the coarse- and fine-grain eye movement analyses.

Recognition of direct object noun ''frog'' (and referential ambiguity). An inspection of the referential competition between the Target and the Other animals reveals relatively strong referential disambiguation effects. Again, disambiguation occurs about 200 ms after the onset of the direct object noun (''frog'') in the 1-Referent contexts. Note that slightly later in the trial, looks to the Target animal are greater and more sharply pronounced in the 1-Referent Unambiguous as compared to the 1-Referent Ambiguous condition, because subjects are spending time inspecting the Incorrect Destination in the latter but not in the former. The 2-Referent contexts were again showing longer consideration of the Other animal over time. Divergence is quite early, especially in the 2-Referent Ambiguous condition. This suggests that subjects were using the preposition ''on'' (which occurs about 250–350 ms after the onset of ''frog'') to help determine reference. Indeed, this pattern of being sensitive to how prepositions constrain possible referents has been observed in eye movement studies using adult listeners (Chambers, Eberhard & Tanenhaus, 1998). However, as will be revealed below, this early divergence in our data is relatively noisy, and not statistically robust. Again, this should be expected because defining the exact moment of phonemic disambiguation depends upon whether ''on'' in fact tells a listener that the intended referent is the frog on the napkin as compared to the frog that is not on the napkin.

Statistics. We again quantified the degree of referential ambiguity by comparing looking times to the Target animal with looking times to the Other animal during different time slices of the speech. For brevity, a presentation of these analyses appears in the Appendix. The results can be summarized as follows. First, early in the speech, prior to having any phonemic information to discriminate between the two animals (i.e., prior to hearing the direct object ''frog''), there were no significant differences in looking times to the Target vs. the Other animal in any condition. Second, upon hearing ''frog'' and prior to hearing ''napkin'', reliably more time was spent looking at the Target animal over the Other animal, but only in the 1-Referent contexts. Thus, although adults showed some ability to use ''on'' to help determine the referent in 2-Referent contexts, statistical tests revealed this effect was unreliable. Finally, upon hearing 400 ms of ''napkin'', subjects spent more time fixating the Target over the Other animal in all conditions. Thus, the Modifier interpretation was pursued in both Ambiguous and UnAmbiguous conditions, helping to distinguish the Target from the Other animal.

5.2.3. Comparison with 5 Year Olds

Statistical comparisons of the eye movement patterns across Experiment 1 and Experiment 2 were also made. First, both the adult and the Five Year Old coarse-grain eye movement measures (looks to the Incorrect Destination, see Fig. 3) were entered into subject and item ANOVAs having three factors: Age (Child vs. Adult); Context (1-Referent vs. 2-Referent) and Ambiguity (Ambiguous vs. Unambiguous). These ANOVAs revealed a reliable triple interaction between Context, Ambiguity and Age ($F1(1,26) = 4.57$; $P < 0.05$; $F2(1,15) = 5.41$, $P < 0.05$). Similar ANOVAs were conducted on looks to the Incorrect Destination during the 800 ms time slice occurring after the onset of ''napkin''. Again, the ANOVA revealed a reliable triple interaction between Context, Ambiguity and Age ($F1(1,26) = 5.84$; $P < 0.05$; $F2(1,15) = 20.83$, $P < 0.01$). In both cases, the interaction arose because adults' eye movements were affected by both the Context and the Ambiguity of the phrase, whereas Five Year Olds'' were affected only by the Ambiguity.

6. General discussion

These experiments demonstrate that it is now possible to study on-line sentence processing with children using the natural task of acting out spoken instructions. In addition, the results reveal several important similarities and differences between the child and adult language processing systems. With respect to similarities, both groups of subjects showed signs of rapid incremental interpretation. For instance, when a single unique referent was present in the scene for an instruction like ''Put the frog…'', both age groups launched eye movements to the intended referent within a few hundred milliseconds of perceiving the noun ''frog''. And, when two potential referents were present for a syntactically unambiguous instruction like ''Put the frog that's on the...'', both age groups showed signs of considering these two referents until the spoken message provided information that could help distinguish between them (e.g. ''napkin'').

Although all subjects came equipped with rapid incremental interpretation, the two age groups differed significantly in how they handled temporary syntactic ambiguities. Adults resolved temporary ambiguities in accord with the Referential Principle of syntactic ambiguity resolution (Crain & Steedman, 1985). Adults pursued the Modifier interpretation of the ambiguous phrase ''Put the frog on the napkin…'' when the visual context supported this interpretation (2-Referent context). They pursued the Destination interpretation of the same phrase when the visual context indicated that a Modifier interpretation would be unnecessary (1-Referent context). These parsing preferences were detected at the earliest stages of processing, with sensitivity to the Referential Principle occurring within 100 ms of perceiving the noun in the ambiguous phrase ''on the napkin''. Moreover, adults showed signs of being able to recover from the temporary consideration of an incorrect interpretation. In particular, adults performed relatively few errors in their actions, suggesting that consideration of the Destination interpretation in the 1-Referent Ambiguous condition could be revised toward the Modifier interpretation.

Five Year Olds showed a very different process for resolving a syntactic ambiguity. Children's initial parsing preferences appeared insensitive to the Referential Principle. Instead, children preferred the Destination interpretation of an ambiguous prepositional phrase (VP attachment), regardless of context. This parsing preference resulted in looks to the Incorrect Destination in both 1- and 2-Referent contexts, occurring within 300 ms of perceiving the noun of the prepositional phrase. Moreover, these children showed an inability or reluctance to revise their initial commitment to the Destination interpretation. Actions in response to Ambiguous instructions frequently involved moving an animal to the Incorrect Destination. In addition, children's choice of a referent in the 2-Referent Ambiguous condition was at chance, and was partially correlated with which object the child happened to look at first. Thus, a child's commitment to an interpretation seemed almost deterministic or ballistic, showing little or no ability to return to an earlier state.

In the remaining sections of the paper, we focus on explanations of these three aspects of the Five Year Old data, namely, strong VP-attachment preference, insensitivity to the Referential Principle, and a reluctance to revise processing commitments.

6.1. VP-attachment preference on the part of the youngest parsers

The preference for VP attachment may be the explicit outcome if, as suggested above, the children were parsing according to a general syntactic processing principle, such as ''choose the syntactically least complex structure'' (e.g. Minimal Attachment, Frazier, 1989).[8] If this were the case, we would expect an initial preference for VP attachment, which has been claimed to be syntactically less complex than NP modification (e.g. Rayner et al., 1983). Such an explanation would be most amenable to the ''rapid revision'' position of Frazier, Clifton, Mitchell and colleagues (e.g. Frazier & Clifton, 1996; Mitchell, 1989). Perhaps as the human processing system matures, revision of a Minimal Attachment parsing commitment (based on context and lexical factors) becomes increasingly rapid. In fact, Goodluck and Tavakolian (1982) explicitly argued for such a developmental situation, proposing that the underlying architecture of the child's language processing system is essentially the same as the adult system, which they characterized as the two-stage garden-path model (Frazier & Fodor, 1978).

Other explanations of the VP attachment preference focus on the lexical properties of the input provided to the child. Children may be basing their parsing commitments on their syntactic and/or semantic knowledge of verbs and possible arguments. All target items in the present study used the verb ''put'' which requires two arguments (a Theme and a Destination). In English, the Theme of the verb ''put'' is always indicated by a noun phrase, and the Destination is frequently indicated by a prepositional phrase headed by ''on'' or ''in''. We have done a preliminary examination of child-directed speech in the CHILDES Corpus (MacWhinney & Snow, 1985, 1990) portion of the Penn Treebank (Marcus, Santorini & Marcinkiewicz, 1993). A total of 179 usages of ''put'' were examined. All indicated a Destination. Of these, 76 (42%) used a full prepositional phrase for the Destination argument. The majority of other occurrences were common utterances containing particles: ''Put it back'' or ''Put it down''. A full 95% of all the PP arguments were headed by the preposition ''on'' or ''in''. Thus, a lexicalist parsing position (e.g. MacDonald et al., 1994; Trueswell & Tanenhaus, 1994) would also predict a strong VP attachment for sentences like ''Put the frog on the napkin...'', not because it is the least complex syntactic alternative, but because it is the most likely syntactic alternative given the lexical input (see also Schutze & Gibson, 1999,

[8] The adult sentence processing literature contains several other processing theories that predict a strong VP-attachment preference (see Schutze & Gibson, 1999, for a recent review). Some of these theories are based on proposed differences in the syntactic and/or semantic complexity of the two alternative interpretations (e.g. Pritchett, 1988), whereas others propose a preference for argument over adjunct attachments (e.g. Abney, 1989; Schutze & Gibson, 1999).

for a related lexicalist position that focuses on a preference for arguments over adjuncts). Also, given that ''put'' is a common verb in child-directed speech, it should be expected that by age five the syntactic and semantic preferences of the verb ''put'' would be known by the child and employed on-line.

Although the current findings do not distinguish between structural and lexical explanations of the data, existing findings in the language acquisition literature, as well as preliminary results from our own laboratory, suggest that the most plausible explanations of the current findings will need to include a lexically-specific component to children's on-line parsing preferences. In particular, research in language acquisition has found that young children exhibit strong sensitivity to lexically specific syntactic preferences, based primarily on their exposure to these lexical items in particular linguistic environments. For instance, children as young as two years of age are capable of extracting the statistical regularities of verbs pertaining to their preferred argument structures (Naigles, 1990; Naigles, Gleitman & Gleitman, 1992; see also Fisher, Hall, Rakowitz & Gleitman, 1994). In addition, a number of comprehension studies have found patterns of incorrect interpretations that are best accounted for by assuming an emerging sensitivity to lexically-specific properties such as animacy, subcategory and control information (e.g. Goodluck & Tavakolian, 1982; McDaniel and Cairns, 1990; Ratner, 1996).

Although these studies have shown lexically-specific preferences in production tasks and off-line comprehension tasks, the data do not address whether this information is used during on-line decisions concerning syntactic ambiguity. In this regard, we have begun to extend the experiments described above to include manipulations of lexical preferences, to observe if attachment preferences can be guided by this information. In the study, we have been comparing target instructions containing verbs that require a Destination argument (e.g. ''Put the frog (that's) on the napkin into the box'') with instructions containing verbs that have no expectation for a Destination argument (e.g. ''Wiggle the frog (that's) on the napkin into the box''). Both sentence types were given in visual contexts that are ''referentially supportive'' of the Modifier interpretation (2-Referent contexts).

Preliminary results (from 16 children ages 4;5 to 6;10 years) indicate that verb information plays an important role in children's commitment to an interpretation of the ambiguous prepositional phrase, but that the verb manipulation does not completely eliminate the preference for the Destination interpretation. In particular, for the verb ''put'', children incorrectly moved an animal to the Incorrect Destination (the empty napkin) on 55% of the ambiguous trials and 16% of the unambiguous trials. For the ''wiggle'' sentences, children incorrectly moved an animal to the Incorrect Destination on 23% of the ambiguous trials and 8% of the unambiguous trials, resulting in an interaction between verb type and ambiguity ($F1(1,12) = 4.51$, $P < 0.06$; $F2(1,12) = 5.27$, $P < 0.05$).[9] However, both verb types show reliable

[9] The arcsin transformation of this data yielded a marginally significant interaction in the item analysis, $F2(1,12) = 3.90$, $P < 0.08$. In addition, means of all errors in actions show a similar pattern, but with a weaker interaction of verb type (Put Ambig: 51%; Put Unambig: 19%; Wiggle Ambig: 31%; Wiggle Unambig: 13%). Interestingly, the strongest effects of verb preference occurred on trials early in the experiment, with children improving on ''put'' trials as the experiment progressed.

effects of Ambiguity. Eye movement patterns have thus far only been analyzed for eight of the sixteen children, and these data reveal a similar pattern, with looks to the Incorrect Destination occurring for ''put'' on 71% of the ambiguous trials and 44% of the unambiguous trials, and for ''wiggle'' on 52% of the ambiguous trials and 38% of the unambiguous trials. Thus, the strong preference for VP attachment is reduced for verbs that are less likely to use a Destination role.

The findings for ''put'' replicate the 2-Referent context results reported above in Experiment 1, in an experiment with two important differences. First, the stimuli were pre-recorded, being cross-spliced across conditions to control for prosody, i.e., ''wiggle'' was cross-spliced into a ''put'' utterance on half of the ambiguous item pairs, and vice versa on the other pairs. (All unambiguous trials had the ''that's'' spliced into the ambiguous utterances, hence the slightly higher error rates on unambiguous trials as compared to the study reported in this paper.) Second, the calibration of the eye-tracker was performed via a new point-of-light tracking system that permits direct calibration of even Five Year Old children.

The differences between verb types suggest that lexically-specific syntactic biases play a role in children's processing commitments. However, the data also suggest that additional constraints exist which lead children to have a general preference for the Destination role with this construction. At the moment, a wide range of explanations exist for this preference, including the possibility that the filling of argument roles is strongly preferred by young children. Indeed, verbs like ''wiggle'' belong to the semantic class of motion verbs, which typically take a Destination role.

6.2. Insensitivity to the Referential Principle in the youngest parsers

Given the discussion above, one possible explanation of the lack of referential effects in the current study is to assume that verb specific syntactic and semantic properties present in the stimuli so strongly supported the Destination interpretation that referential factors were unable to impact processing preferences. This explanation would have important implications for other developmental studies, given the inconsistent findings in the literature regarding whether children are sensitive to the Referential Principle in their off-line comprehension of restrictive relative clauses. As mentioned in the introduction, some studies have found children to be sensitive to referential contexts in comprehension tasks (e.g. Hamburger & Crain, 1982) whereas others have found smaller or no effects of referential context (e.g. Goodluck, 1990; Lee, 1992). It is possible that the apparent inconsistencies between these experiments have more to do with differences in the lexical and semantic biases associated with the particular stimuli used in the studies. For instance, verbs with more complex argument structures may result in increased working memory difficulty, which may be compounded when these verbs appear in relative clause constructions. Thus, the current debate over the presence or absence of the referential constraints in young children may be quite similar to the state of affairs that the adult sentence processing literature was in ten years ago, when some studies found effects of referential context on ambiguity resolution, and some studies did not. Studies which took into account lexico-syntactic preferences began to explain the pattern, showing that referential effects depend upon the presence of less biasing linguistic stimuli (Altmann et al., 1998; Britt, 1994; Spivey-Knowlton & Sedivy, 1995).

This explanation, however, does not address the more central developmental observation that Five Year Olds in the present study seemed unable or unwilling to use the Referential Principle to inform parsing commitments, whereas adults could do so. Adults appear to be able to override local syntactic preferences, showing a sensitivity to the Referential Principle. Indeed, Tanenhaus et al. (1995) implied that the manipulation of referential factors in a visual context (as compared to a written discourse context) is so salient that adults can use it to override the strong biases for VP attachment that have been observed in reading studies. We note first, however, that the data from our Adult participants suggest that lexical biases may not be completely overridden by contextual factors (see especially the small number of errors in actions, Fig. 3). However, the fact remains that adults are far more sensitive to the Referential Principle than children.

One possibility is that children simply lack any implicit knowledge of the Referential Principle. Children may not understand that a definite noun phrase with more than one potential referent needs modification. This explanation seems unlikely given the results of Crain and colleagues showing that children even younger than age five are more likely to understand and produce relative clause modifiers when the visual context is referentially supportive of modification (e.g. Crain et al., 1990; Hamburger & Crain, 1982; McKee, McDaniel & Snedeker, 1998; but c.f. Goodluck, 1990).

A more likely possibility is that children have the Referential Principle, but for some reason are unable to employ it under certain processing conditions. For instance, Five Year Olds may have a limited processing capacity, making it unlikely that they will entertain uncommon and/or complex syntactic alternatives. Indeed, the adult sentence processing literature on individual differences finds that adults who score low on verbal memory tests (low-span subjects) are less able to employ semantic, pragmatic and referential constraints during on-line parsing commitments (e.g. Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just & Carpenter, 1992; Pearlmutter & MacDonald, 1995 see also Gibson, 1998). There are two central explanations of these findings. One is that low-span subjects are less likely to compute multiple interpretations in parallel (e.g. MacDonald et al., 1992), and quickly abandon uncommon or complex alternatives. Another, not necessarily incompatible, explanation is that low-span subjects find it difficult to compute relevant contextual/pragmatic constraints during on-line sentence processing (e.g. Pearlmutter & MacDonald, 1995; see also MacDonald & Christiansen, 1999).

The current finding that Five Year Olds are less likely to revise parsing commitments is compatible with the first of these two processing explanations. In particular, it is possible that children of this age, for reasons of processing capacity, are more likely to abandon low probability syntactic alternatives, showing early preference for the Destination interpretation. And, this limitation is so great that they cannot reevaluate a sentence when they encounter incompatible linguistic information later in the sentence, i.e., an inability to backtrack. Thus the ultimate actions, which involve movements to the incorrect destination, reflect an accommodation of this early parsing commitment.

In addition, some aspects of the error data from Five Year Olds provide hints for the second processing capacity explanation, i.e., that Five Year Olds can compute the relevant pragmatic constraints pertaining to the Referential Principle but find it difficult to do so during on-line comprehension. These hints come from the actions in the 2-Referent context. In this context, when Five Year Olds happened to move the Target animal first (the frog on the napkin), they were very likely to perform the action correctly (86% or 25 out of 29 times). In 1-Referent contexts, children almost always chose to move the Target animal first, presumably because it was the only frog to choose. But, in this exact same action, of moving the Target animal first, children were far less likely to move the animal correctly (44% or 24 out of 54 times) and instead moved it to the Incorrect Destination. Thus, at least when moving the Target animal first, the presence of a second identical animal may have helped the child both arrive at the Modifier interpretation of the prepositional phrase and avoid the Destination interpretation. This account is somewhat tenuous because it relies exclusively on trials that the child correctly selected the Target. However, given that the eye fixation patterns showed differences during the early portions of correct and incorrect trials, it is possible that the manner in which the child inspected the visual scene may have encouraged (serendipitously) the noticing of a referential contrast, boosting the chances of arriving at the correct interpretation.

The same boosting of the correct interpretation may also be done by linguistic input, as suggested by the preliminary results contrasting ''put'' with ''wiggle''. A verb that tends not to use a Destination role (''wiggle'') increased consideration of the Modifier interpretation. It is therefore possible that Five Year Olds could show a greater sensitivity to the Referential Principle when acting upon instructions containing the ''wiggle'' verbs. This can be tested by extending the ''wiggle'' study mentioned above to include a direct manipulation of Referential context (1-Referent vs. 2-Referent). It is possible that with these verbs Five Year Olds will begin showing a sensitivity to the relevant contextual cues.

6.3. When do children start acting like adults?

The current research only provides two snapshots into the developing processing system. It would be useful to begin to map out the developmental time course of the on-line processing system, perhaps by examining older children with this same task. We have begun to do this, collecting eye movement data from children ages eight and nine years. These data were not included here because they reveal that children in this age range are a heterogeneous group when it comes to parsing preferences. Many subjects show a pattern similar to adults: very few incorrect actions and eye movements to the Incorrect Destination primarily in the 1-Referent Ambiguous condition. However, some subjects behaved much more like Five Year Olds in both their actions and eye movements, i.e., a large number of incorrect actions on Ambiguous trials, regardless of context, and frequent looks to the Incorrect Destination, regardless of context. These data suggest that more detailed estimations of linguistic age, and perhaps working memory capacity, are needed to help map out the development of on-line processing strategies.[10]

6.4. Summary and closing remarks

The current findings reveal a language processing system in Five Year Olds that relies more heavily on local linguistic factors to inform parsing preferences, along with a general inability to revise initial commitments to interpretation. Adults, however, are able to use relevant contextual factors to inform parsing commitments, and possess the ability to revise early incorrect commitments to interpretation. Despite the limitations on the Five Year Olds' ability to deal with local syntactic ambiguity, the findings indicate a highly incremental processing system at this age. Word recognition and referential resolution in syntactically unambiguous environments appears to proceed smoothly, showing patterns quite similar to adults.

Finally, these findings highlight the need to unify theories of language processing and language acquisition. Indeed, given that similar experimental techniques can now be used for studying language comprehension in children and adults, it is likely that theories of language learning will begin to emphasize the role of language processing in development. The account given here assumes a central role for these abilities, and begins to map out how the rapid and interactive adult processing system develops in young children.

Acknowledgements

This work was partially supported by National Science Foundation Grant SBR-96-16833 and by National Institute for Health Grant 1-R01-HD37507-01. We thank the Institute for Research in Cognitive Science at the University of Pennsylvania (NSF-STC Cooperative Agreement number SBR-89-20230) for use of their eye-tracking facilities. We thank Laura Carey-Anniballi for assisting us in some of the eye movement analyses. We also thank the three anonymous reviewers of the original version of this paper, who provided several useful suggestions regarding the analysis of the eye movement results.

Appendix. Referential competition between Target and Other animal

This appendix contains a statistical analysis of the referential competition that was observed between the Target and the Other animals (e.g. see Figs. 4 and 6). The data are presented below for Experiment 1 and Experiment 2.

A.1. Experiment 1: Five Year Old subjects

We quantified the degree of referential competition between the Target and the

[10] The eight year old findings were obtained using a speech rate that was identical to the Five Year Olds reported in Experiment 1. Thus, the eight year olds demonstrate that children do not always experience garden-paths at this speech rate. Most subjects (13 out of 16) exhibited little or no difficulty in the 2-Referent Ambiguous condition.

Other objects by subtracting the proportion of time spent looking at the Other animal from the proportion of time spent looking at the Target animal (see Table 2). Positive numbers reflect more time spent looking at the Target animal. Negative numbers reflect more time spent looking at the Other animal. Numbers near zero indicate approximately equal time looking at both objects. To give an indication of the time-course of competition, data were divided into three time slices. *Time slice 1*: From the onset of ''put'' to 200 ms after the onset of the direct object noun (''frog''). During this time slice, listeners should have no phonemic information to help distinguish the two animals (taking into account 200 ms for programming an eye movement). *Time slice 2*: From 200 ms after the onset of the direct object to 200 ms after the onset of the prepositional noun (''napkin''). During this time slice, listeners should have the phonemic information necessary to distinguish the two animals in the 1-Referent context but not in the 2-Referent context (taking into account 200 ms for programming an eye movement). *Time-slice 3*: From 200 ms to 1000 ms after the onset of the prepositional noun. During this time slice, listeners should have phonemic information to help distinguish the Target and the Other animals in all conditions.

Subject and item means from each time slice were entered into separate ANOVAs with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent) and presentation List or Item-Group factor (4 lists in the subject analysis and 4 item groups in the item analysis). The results of these ANOVAs appear below.

Time slice 1. When phonemic material could not help distinguish the Target from the Other animal, viewing time for these two animals was approximately equal, resulting in difference values near zero in all four conditions. The ANOVA comparing these values across conditions showed no reliable main effects or interactions. In addition, all four difference values were not significantly different from zero, as tested by two-tailed t-tests on both subject and item means. To avoid Type I errors, the significance levels of all *t*-tests took into account the number of tests by dividing the alpha by the number of tests, in this case, four tests.

Time slice 2. When phonemic material could help determine the referent in the 1-Referent context, we begin to see more looks to the Target animal (the frog) as compared to the Other animal (the horse), resulting in a reliable main effect of

Table 2
Referential competition between Target animal and Other animal. Mean difference in proportion of time fixating Target and Other animal (Five Year Olds, Experiment 1)

Context:	Time slice					
	1		2		3	
	1-Ref	2-Ref	1-Ref	2-Ref	1-Ref	2-Ref
Ambiguous	0.01	0.03	0.15	− 0.05	0.48[a]	0.18
Unambiguous	0.06	− 0.04	0.40[a]	0.00	0.42[a]	0.35[a]

[a] Value significantly greater than zero, implying a resolution toward the Target referent.

Context ($F1(1, 12) = 7.55$, $P < 0.05$; $F2(1, 12) = 21.38$, $P < 0.01$). There was no effect of Ambiguity or interaction between Context and Ambiguity. As can be seen in Table 2, only 1-Referent context conditions show differences greater than zero. Two-tailed *t*-tests revealed that this value was significantly different from zero in the 1-Referent Unambiguous condition ($t_s(15) = 4.88$, $P < 0.001$; $t_i(15) = 6.84$, $P < 0.001$), not significantly different from zero in the 1-Referent Ambiguous condition ($t_s(15) = 2.07$; $t_i(15) = 1.68$) and not significantly different from zero in either 2-Referent contexts (Unambiguous: $t_s(15) = -0.15$; $t_i(15) = 0.34$; Ambiguous: $t_s(15) = -0.65$; $t_i(15) = -0.36$).

The slightly longer delay in divergence in looks to the Target vs. the Other animal for the 1-Referent Ambiguous condition could suggest an early sensitivity to referential context. However, the failure to find an interaction between Context and Ambiguity in this position argues against this conclusion. Moreover, all other measures of eye movements and actions reported here do not support this conclusion. We suspect that the delay in the 1-Referent Ambiguous condition is actually due to listeners'' commitment to the Destination interpretation of ''on the…'', which is pragmatically odd given that the Target is already on something.

Time slice 3. When phonemic material could help determine the referent in all conditions, strong signs of considering the Target animal over the Other animal were observed in all conditions except for the 2-Referent Ambiguous condition. This resulted in an interaction between Context and Ambiguity, which was marginally significant in the subject analysis and significant in the item analysis ($F1(1, 12) = 3.83$, $P < 0.08$; $F2(1, 12) = 8.02$, $P < 0.05$). There was also a main effect of Context ($F1(1, 12) = 13.21$, $P < 0.01$; $F2(1, 12) = 5.46$, $P < 0.05$). Two-tailed t-tests revealed that values were significantly different from zero in the 1-Referent Ambiguous condition ($t_s(15) = 10.85$, $P < 0.001$; $t_i(15) = 9.38$, $P < 0.001$), the 1-Referent Unambiguous condition ($t_s(15) = 9.03$, $P < 0.001$; $t_i(15) = 6.81$, $P < 0.001$), and the 2-Referent Unambiguous condition ($t_s(15) = 6.27$, $P < 0.001$; $t_i(15) = 5.18$, $P < 0.001$), but not the 2-Referent Ambiguous condition ($t_s(15) = 2.21$; $t_i(15) = 2.33$).

Finer-grain time-course analyses were performed within this 800 ms time slice by conducting a similar ANOVA on 100 ms time slices. A reliable interaction between Context and Ambiguity was not observed until the 5th time slice (400–500 ms, $F1(1, 12) = 6.78$, $P < 0.05$; $F2(1, 12) = 6.25$, $P < 0.05$). This suggests that it was not until hearing approximately 400 ms of the word ''napkin'' that subjects could discern the Target from the Other animal in the 2-Referent Unambiguous condition. This finding corresponds well with commitment to the Destination interpretation in Ambiguous conditions (see above), which occurred at approximately the same time.

A.2. Experiment 2: Adult subjects

We again quantified referential competition by subtracting the proportion of time spent looking at the Other animal from the proportion of time spent looking at the Target animal (see Table 3). As in Experiment 1, the same three time slices were used. Subject and item means from each time slice were entered into separate

Table 3
Referential competition between Target animal and Other animal. Mean difference in proportion of time fixating Target and Other animal (Adults, Experiment 2)

	Time slice					
	1		2		3	
Context:	1-Ref	2-Ref	1-Ref	2-Ref	1-Ref	2-Ref
Ambiguous	− 0.03	0.03	0.31[a]	0.19	0.50[a]	0.29[a]
Unambiguous	0.08	0.03	0.43[a]	0.10	0.50[a]	0.47[a]

[a] Value significantly greater than zero, implying a resolution toward the Target animal as referent.

ANOVAs with three factors: Ambiguity (Ambiguous, Unambiguous); Context (1-Referent, 2-Referent) and presentation List or Item-Group factor (4 lists in the subject analysis and 4 item groups in the item analysis).

Time slice 1. Prior to phonemic disambiguation, the ANOVA revealed no reliable main effects or interactions. As can be seen in Table 3, all values were near zero, implying equal viewing of Target and Other objects. Indeed, two-tailed *t*-tests on both subject and item means revealed that none of the values were significantly different from zero.

Time slice 2. When phonemic material could help determine the referent in the 1-Referent context, we see more looks to the Target animal (the frog) as compared to the Other animal (the horse). However, this effect is relatively weak, resulting in a reliable main effect of Context only in the item analysis ($F1(1,8) = 2.56$; $F2(1,12) = 5.40$, $P < 0.05$). There was no reliable effect of Ambiguity or interaction with Context. Two-tailed *t*-tests on the values in Table 3 revealed that the 1-Referent cells were reliably different from zero (1-Ref Ambig: $t_s(11) = 4.84$, $P < 0.001$, $t_i(15) = 3.76$, $P < 0.005$; 1-Ref Unambig: $t_s(11) = 4.66$, $P < 0.001$, $t_i(15) = 7.27$, $P < 0.001$) but that the 2-Referent cells were not reliably different from zero (2-Ref Ambig: $t_s(11) = 0.99$; $t_i(15) = 2.45$; 2-Ref Unambig: $t_s(11) = 0.73$; $t_i(15) = 1.85$).

Time slice 3. When phonemic material could help determine the referent in all conditions, we see strong signs of disambiguation toward the Target animal in all four conditions. Two-tailed *t*-tests on the values in Table 3 revealed that all values were significantly greater than zero (1-Ref Ambig: $t_s(11) = 11.42$, $P < 0.001$, $t_i(15) = 7.68$, $P < 0.001$; 1-Ref Unambig: $t_s(11) = 5.52$, $P < 0.001$, $t_i(15) = 7.50$, $P < 0.001$; 2-Ref Ambig: $t_s(11) = 4.08$, $P < 0.002$, $t_i(15) = 4.12$, $P < 0.001$; 2-Ref Unambig: $t_s(11) = 6.91$, $P < 0.001$, $t_i(15) = 8.47$, $P < 0.001$). The ANOVA revealed that the interaction between Context and Ambiguity was significant in the item analysis ($F1(1,8) = 4.93$, $P < 0.06$; $F2(1,12) = 5.01$, $P < 0.05$). (The item analysis conducted on the arcsin transformation of the data showed the interaction to be marginally significant, $F2(1,12) = 4.27$, $P < 0.07$). In addition, there was a reliable effect of Context in the item analysis ($F1(1,8) = 3.57$,

$P < 0.1$; $F2(1, 12) = 5.38$, $P < 0.05$) and no reliable effect of Ambiguity ($F1(1, 8) = 1.52$; $F2(1, 12) = 3.42$, $P < 0.09$).

References

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

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*, 419–439.

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

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

Altmann, G., Garnham, A., & Henstra, J. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 209–216.

Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Henstra, J. (1998). Late closure in context. *Journal of Memory and Language*, *38*, 459–484.

Ballard, D., Hayhoe, M., & Pelz, J. (1993). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80.

Bever, T. G. (1982). Regression in the service of development. In T. Bever, *Regressions in Mental Development: Basic Phenomena and Theories*, Hillsdale, NJ: Lawrence Erlbaum.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes, *Cognition and the Development of Language*, New York: Wiley.

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

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.

Chambers, C., Eberhard, K., & Tanenhaus, M. K. (1998). Words and worlds: the use of prepositions in spoken language comprehension. In M. A. Gernsbacher, & S. J. Derry, *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Erlbaum.

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

Crain, S., McKee, C., & Emiliani, M. (1990). Visiting relatives in Italy. In L. Frazier, & J. de Villiers, *Language Processing and Language Acquisition*, Dordrecht: Kluwer.

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing and modularity. In C. Clifton, K. Rayner, & L. Frazier, *Perspectives on Sentence Processing*, (pp. 443–467). Hillsdale, NJ: Erlbaum.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*, 228–231.

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

Ferreira, F., & Henderson, J. M. (1991). How is verb information used during syntactic parsing? In G. B. Simpson, *Understanding Word and Sentence*, Amsterdam: North Holland.

Fisher, C., Hall, G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give. *Lingua*, *92*, 333–375.

Frank, R. (1998). Structural complexity and the time course of grammatical development. *Cognition*, *66*, 249–301.

Frazier, L., & Clifton Jr, C. (1996). *Construal*, Cambridge, MA: MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart, *Attention and Performance XII: The Psychology of Reading*, Hillsdale, NJ: Erlbaum.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291–325.

Frazier, L. (1989). Against lexical generation of syntax. In W. D. Marslen-Wilson, *Lexical Representation and Process*, Cambridge, MA: MIT Press.

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.

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

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

Goodluck, H., & Tavakolian, S. (1982). Competence and processing in children's grammar of relative clauses. *Cognition*, *11*, 1–27.

Goodluck, H. (1990). Knowledge integration in processing and acquisition: Comments on Grimshaw and Rosen. In L. Frazier, & J. De Villiers, *Language Processing and Language Acquisition*, Dordrecht: Kluwer.

Hamburger, H., & Crain, S. (1984). The acquisition of cognitive compiling. *Cognition*, *17*, 85–136.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language Development*, II. Hillsdale, NJ: Lawrence Erlbaum.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology*, *8*, 203–241.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.

Lee, T. H. (1992). The inadequacy of processing heuristics: Evidence from relative clause acquisition in Mandarin Chinese. In T. H. Lee, *Research in Chinese Linguistics in Hong Kong*, Hong Kong: Linguistic Society of Hong Kong.

MacDonald, M.C., Christiansen, M.H., (1999). Reassessing working memory: a reply to Just and Carpenter and Waters and Caplan. Manuscript submitted for publication.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.

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

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, *32*, 692–715.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language*, *12*, 271–295.

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, *17*, 457–472.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*, 313–330.

Marslen-Wilson, W. D., & Tyler, L. K. (1987). Against modularity. In J. Garfield, *Modularity in Knowledge Representations and Natural Language Understanding*, Cambridge, MA: MIT Press.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522–523.

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

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

Mazuka, R., (1990). Japanese and English children's processing of complex sentences: An experimental comparison. Unpublished Ph.D. Dissertation, Cornell University, Ithaca, CA.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart, *Attention and Performance XII*, London: Erlbaum.

McDaniel, D., & Cairns, H. S. (1990). The processing and acquisition of control structures by young children. In L. Frazier, & J. de Villiers, *Language Processing and Language Acquisition*, Dordrecht: Kluwer.

McKee, C., Nicol, J., & McDaniel, D. (1993). Children's application of binding during sentence processing. *Language and Cognitive Processes*, *8*, 265–290.

McKee, C. (1996). On-line methods. In D. McDaniel, C. McKee, & H. Smith Cairns, *Methods for Assessing Children's Syntax: Language, Speech, and Communication*, Cambridge, MA: MIT Press.

McKee, C., McDaniel, D., & Snedeker, J. (1998). Relatives children say. *Journal of Psycholinguistic Research*, *27*, 573–596.

Mitchell, D. C. (1989). Verb guidance and other lexical effects in parsing. *Language and Cognitive Processes*, *4*, 123–154.

Naigles, L., Gleitman, L. R., & Gleitman, H. (1992). Children acquire word meaning components from syntactic evidence. In E. Dromi, *Language and Cognition: A Developmental Perspective*, Norwood, NJ: Ablex.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*, 357–374.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, *34*, 521–542.

Pritchett, B. (1988). Garden path phenomena and the grammatical basis of language processing. *Language*, *64*, 539–576.

Pynte, J., & Kennedy, A. (1993). Referential context and within-word refixations: Evidence for ''weak interaction''. In G. d'Ydewalle, & J. Van Rensbergen, *Perception and Cognition: Advances in Eye Movement Research. Studies in Visual Information Processing*, 4. Amsterdam: North-Holland.

Ratner, N. B. (1996). From ''signal to syntax'': But what is the nature of the signal? In J. L. Morgan, & K. Demuth, *Signal to syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, Mahwah, NJ: Lawrence Erlbaum.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.

Roeper, T. (1981). Introduction. In S. T. Tavakolian, *Language Acquisition and Linguistic Theory*, Cambridge, MA: MIT Press.

Schutze, C. T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*, 409–431.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, *13*, 272–281.

Spivey, M. J., & Tanenhaus, M. K. (1999). Integration of visual context and linguistic information in resolving temporary ambiguities in spoken language comprehension. Submitted for publication.

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

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

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73–108.

Swinney, D., & Prather, P. (1989). On the comprehension of lexical ambiguity by young children: Investigations into the development of modularity. In D. S. Gorfein, *Resolving Semantic Ambiguity*, Berlin: Springer–Verlag.

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., & Trueswell, J. C. (1995). Sentence comprehension. In P. Eimas, & J. L. Miller, *Handbook of Perception and Cognition: Language*, New York: Academic Press.

Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*, 1–36.

Tavakolian, S. (1981). The conjoined-clause analysis of relative clauses. In S. Tavakolian, *Language Acquisition and Linguistic Theory*, Cambridge, MA: MIT Press.

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

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19* (3), 528–553.

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

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*, 566–585.

Tyler, L. K., & Marslen-Wilson, W. (1981). Children's processing of spoken language. *Journal of Verbal Learning and Verbal Behavior*, *20*, 400–416.

Tyler, L. K. (1983). The development of discourse mapping processes: The on-line interpretation of anaphoric expressions. *Cognition*, *13*, 309–341.