Journal of Psycholinguistic Research, Vol. 29, No. 6, 2000

One Frog, Two Frog, Red Frog, Blue Frog: Factors Affecting Children's Syntactic Choices in Production and Comprehension

Felicia Hurewitz,[1,3] Sarah Brown-Schmidt,[2] Kirsten Thorpe,[1] Lila R. Gleitman,[1] and John C. Trueswell[1]

Two experiments are reported which examine children's ability to use referential context when making syntactic choices in language production and comprehension. In a recent on-line study of auditory comprehension, Trueswell, Sekerina, Hill, and Logrip (1999) examined children's and adults' abilities to resolve temporary syntactic ambiguities involving prepositional phrases (e.g., "Put the frog on the napkin into . . ."). Although adults and older children used the referential context to guide their initial analysis (pursuing a destination interpretation in a one-frog context and a modifier interpretation in a two-frog context), 4 to 5-year olds' initial and ultimate analysis was one of destination, regardless of context. The present studies examined whether these differences were attributable to the comprehension process itself or to other sources, such as possible differences in how children perceive the scene and referential situation. In both experiments, children were given a language generation task designed to elicit and test children's ability to refer to a member of a set through restrictive modification. This task was immediately followed by the "put" comprehension task. The findings showed that, in response to a question about a member of a set (e.g., "Which frog went to Mrs. Squid's house?"), 4- to 5-year-olds frequently produced a definite NP with a restrictive prepositional modifier (e.g., "The one on the napkin"). These same children, however, continued to misanalyze put instructions, showing a strong avoidance of restrictive modification during comprehension. Experiment 2 showed that an increase in the salience of the platforms that

This work was partially supported by National Science Foundation Grant SBR-96-16833, National Institute for Health Grant 1-R01-HD37507-01, and by the Institute for Research in Cognitive Science (IRCS) at the University of Pennsylvania (NSF-STC Cooperative agreement number SBR-89-20230), including an IRCS predoctoral fellowship. We thank the following schools in the Philadelphia area for the use of their facilities and for participating in these studies: Bright Start Child Care & Learning Center, Lane Montessori School, Magic Years Child Care and Learning Center, Mulberry Child Care and Preschool, and Robert J. Wilf preschool at the Kaiserman JCC.

[1] Department of Psychology, University of Pennsylvania, 3815 Walnut St., Philadelphia, Pennsylvania 19104-6196.

[2] Department of Brain and Cognitive Sciences, University of Rochester.

[3] To whom all correspondence should be addressed. email: fel@psych.upenn.edu.

0090-6905/00/1100-0597$18.00/0

distinguished the two referents increased overall performance, but still showed the strong asymmetry between production and comprehension. Eye movements were also recorded in Experiment 2, revealing on-line parsing patterns similar to Trueswell et al.: *an initial preference for a destination analysis and a failure to revise early referential commitments. These experiments indicate that child–adult differences in parsing preferences arise, in part, from developmental changes in the comprehension process itself and not from a general insensitivity to referential properties of the scene. The findings are consistent with a probabilistic model for uncovering the structure of the input during comprehension, in which more reliable linguistic and discourse-related cues are learned first, followed by a gradually developing ability to take into account other more uncertain (or more difficult to learn) cues to structure.*

KEY WORDS: referential communication; syntactic ambiguity; parsing; language learning and acquisition; eye movements.

INTRODUCTION

Developmental psycholinguistic studies in the modern era have provided rather static pictures of children's emerging knowledge of language. Studies of the rapid-fire internal processes underlying children's generation and interpretation of language are much rarer, because of primarily the paucity of laboratory techniques suitable for use with children in the preschool and early school years. Very recently, world-situated eye-tracking techniques, pioneered in the adult-sentence processing literature by Tanenhaus and his colleagues (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), have at last made it feasible to study children in the very act of comprehension (Trueswell, Sekerina, Hill, & Logrip, 1999; henceforth, TSHL). In this technique, children's eye gaze is followed as they seek out sentence interpretations while listening in the presence of a reference world. As we shall discuss further, the results of TSHL suggest that child and adult comprehension behaviors differ substantially when it comes to the process of selecting among possible analyses of syntactically ambiguous phrases: Children were unable or unwilling to use relevant facts about the referential context, which adults readily used to guide their initial syntactic choice. The experiments we present here attempt to probe these effects further so as to isolate the loci of child–adult differences.

Specifically, our experimental aim was to help adjudicate among two broad classes of explanation of these developmental effects. The first class of explanations attributes the child–adult differences to the process of comprehension itself and to its underlying mechanisms. For instance, there may exist qualitative or quantitative differences in how the child and adult comprehension systems handle structural ambiguity. The second class of explanations attributes these effects to differences that would typically be thought of as residing outside the comprehension machinery—ranging from a failure

of the child to perceive the scene in the same way as adults do to a failure to understand the pragmatic/referential conditions that support the use of certain linguistic expressions (semantic competence). For these reasons, the experiments reported here take both production and comprehension measures of children's use of the same expressions. The production task is used both to measure the child's sensitivity to contextual factors contributing to syntactic choice and to increase awareness of relevant referential and pragmatic facts immediately prior to any of the tests for comprehension. Before describing these experiments in detail, and their rather surprising outcomes, we sketch some of what is already known about the adult parsing machinery and pragmatic contributions to syntactic choice in both adults and children.

ADULT COMPREHENSION AND THE PROBLEM OF STRUCTURAL AMBIGUITY

A major source of evidence for understanding the mental processes that yield sentence comprehension in adults are ambiguities and their resolution. Particularly revealing are the ways that listeners resolve the pervasive temporary syntactic ambiguities that arise as an utterance unfolds. As an example that will be relevant to the child studies presented in this paper, consider the PP attachment ambiguity and its semantic implications in the request:

1. *Put the frog on the napkin into the box.*

At the point at which the listener hears the word *napkin,* she cannot make a warranted decision as to whether the sentence ends right there (in which case *on the napkin* is VP-attached and interpretable as the intended destination of the frog) or whether a further locative PP will follow (in which case *on the napkin* is NP-attached, a restrictive modifier of *the frog*).

The recent literature amply supports two overarching generalizations concerning the resolution of such temporary ambiguities by mature language users. The first is that listeners jump to technically unwarranted decisions, i.e., they commit to a structural–interpretive analysis although later evidence (in the sentence or in the world) may force them to abandon these decisions (e.g., Bever, 1970; Frazier & Rayner, 1982). The second is that, rather surprisingly, these apparently risky early parsing commitments do not lead to a comprehension machinery that groans and creaks under the burden of revision as countervailing information accumulates across the sentence. This is because adults integrate information "on line" (that is, during the course of hearing the sentence) from a variety of sources to inform the initial decisions, including contextual and lexical information (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; TSHL, 1999; Trueswell, Tanenhaus, & Kello, 1993).

By seizing on such probabilistic linguistic and extralinguistic information sources at the moment of their availability, adult speaker-listeners usually avoid "wandering down the garden path"—that is, following a false parsing path to its doomed end. Except in the devilishly concocted example sentences of psycholinguists, then, the riskiness of fast parsing decisions is, in practice, not great, certainly nowhere so great as it would be if the listener had to rely on syntactic information alone. Multiple statistical cues to form and meaning provide so much convergent information that, realistically, most ambiguities rapidly dissolve.

PRAGMATIC INPUTS TO SYNTACTIC CHOICE

A major focus of these experiments is understanding children's sensitivity to the surrounding context of an utterance. In particular, how does the discourse and visual context contribute to the syntactic choices made by children in both their production and understanding of particular linguistic expressions? Of central interest will be the contextual factors that contribute to restrictive noun phrase modifiers, such as *the frog on the napkin* found in (1).

For adults, much is already understood about how context contributes to the use of restrictive modifiers. First, adult speakers generally apply a metric of informational necessity when deciding whether or not to produce restrictive modifiers (Grice, 1968). That is, adults typically utter definite NPs with restrictive modifiers, like *the green frog* or *the frog on the napkin,* rather than just saying *the frog,* when the discourse or visual context requires the use of such expressions to uniquely specify the referent, i.e., there is more than one frog and only one is green or only one is on a napkin (e.g., Clark & Wilkes-Gibbs, 1986; Horton & Keysar, 1996)—although speakers sometimes err on the side of being overly informative (Sedivy, 1999: Brennan & Clark, 1996).

Adult readers and listeners are often sensitive to these production facts when making syntactic choices about restrictive modifiers, that is, when they are resolving syntactic ambiguities whose possible alternatives include restrictive modification (e.g., Altmann & Steedman, 1988; Crain & Steedman, 1985). In particular, Crain and colleagues theorized that at points of ambiguity, readers and listeners will pursue the syntactic analysis that violates the fewest discourse presuppositions (the Referential Principle). Support

for this theory came from a number of studies. For example, Altmann and Steedman (1988) where it was found that readers of ambiguous PP attachments (*The fireman smashed down the door with the rusty lock/heavy axe*) prefer the NP-attached (*rusty lock*) reading when the discourse information supported the need for modification (in this case, two doors), but not when context did not support modification (one door).

Recent studies using the head-mounted eye-tracking methodology have examined these same issues using the visual scene as the manipulated discourse context. The eye-tracking methodology (used also in Experiment 2 of the present series) allows following the listener's eye movements across narrow time intervals (approx. 30 ms) to objects and places as he listens to a sentence and seeks out its NP referents in the scene itself. In general, these studies have again demonstrated that hearers can use contextual information afforded by the visual scene to inform parsing decisions (e.g., Spivey, Tanenhaus, Eberhard, & Sedivy, 2000; Tanenhaus *et al.,* 1995; TSHL, 1999). An example is the interpretation of sentence (1) in the presence of the beanie-baby reference worlds shown in Figure 1 (from TSHL). Adult listeners to this sentence tend to look at an empty napkin in the scene before them as they hear the phrase *on the napkin,* but only if the scene (Fig. 1a) contains a single frog: They go down the garden path taking *on the napkin* as a destination, and milliseconds later when they hear *in the box,* must revise the direction of their gaze and the parsing decision. But if there are two frogs visible in the scene (Fig. 1b), the listeners avoid the telltale glance at the empty napkin: They have interpreted *on the napkin* as a modifying phrase because only this interpretation enables selection of a unique frog.

Can children make use of the Referential Principle in the same way to resolve temporary ambiguities? A recent study using the head-mounted eye-tracking technology with children suggests that they may in fact be unable to do so. TSHL tested kindergarten-aged children on sentences like (1) in

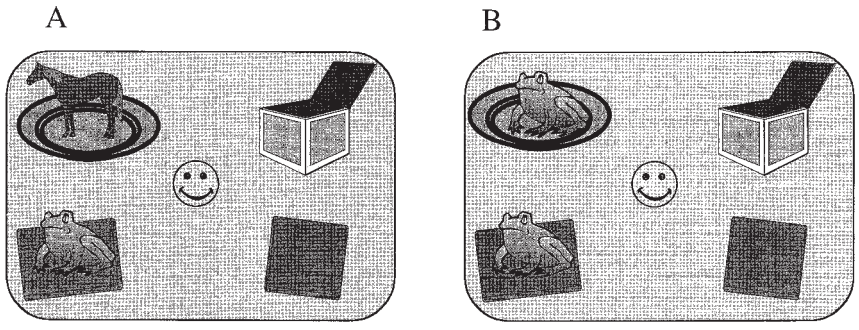

Fig. 1. Illustrations of the referential scenes used in TSHL. (a) One-referent context; (B) Two referent context.

both one-frog and two-frog and contexts (see again, Fig. 1). They found, using off- and on-line measures (their actions and eye movements, respectively) that, unlike adults, the children were not sensitive to the context manipulation. They garden pathed (glance at the empty napkin) even when they heard *on the napkin* in the two-frog context. Furthermore, these children never recovered from the false destination analysis and so, subsequently, performed incorrect actions. These actions typically involved misconstruing the intended destination, e.g., bouncing a frog onto the empty napkin, and then putting it into the box (very often with the wrong frog as well). Importantly, TSHL showed that the children's problem *was not one of sentence length or structural complexity*, as children performed nearly perfectly in response to the unambiguous form of the sentence:

2. *Put the frog that's on the napkin in the box.*

TSHL therefore suggested that the child–adult differences pertained to the developing parsing machinery and its ability to *consider and resolve temporary syntactic ambiguity:* In particular, the differences could arise from a combination of two factors. First, processing limitations may prevent young children from considering more than one syntactic analysis in parallel. Second, the analysis that *is* performed might be determined by a conspiracy of highly reliable lexical cues, which in the case of sentence (1) support a VP-attachment reading. These factors taken together might automatically lead the child parser to never consider the NP-attachment reading of *on the napkin.*

As alluded to earlier, there exist other explanations of these data, which do not appeal to developmental differences in the comprehension process *per se.* First, it is possible that the children fail to notice the referentially relevant properties of the visual scene (e.g., the presence of two identical frogs, whose only distinguishing characteristics pertain to their location relative to other objects in the scene). Second, even if the child is aware (or is made aware) of the facts about the scene, he may not view them as relevant in the current communicative situation. For instance, one can imagine that children may be less inclined than undergraduates to see the necessity of determining which of two identical frogs should be put into the box. Indeed, even adults selectively apply the need for referential specificity—upon hearing "Pass the salt," few would stop to inquire "Which one?" even if there were several saltshakers on the table.

Not all the evidence points in one direction on this question. There is some evidence that children of this age and younger may be sensitive to the contextual and pragmatic factors pertaining to restrictive modification. For instance, Crain and colleagues have argued that children are more likely to produce and understand restrictive relative clauses in experiments where the

presuppositional requirement of multiple referents (their "felicity condition") was met, as compared to experiments where it was not (Hamburger & Crain, 1982; Crain, McKee, & Emiliani, 1990, McKee, McDaniel, & Snedeker, 1998; but see also Goodluck, 1990; Goodluck & Tavakolian, 1982).

However, although these findings suggest that children can utter NP-restrictive relatives, they do not tell us whether children spontaneously notice these facts about the scene, understand the need for referential specificity in the situation, or even realize that restrictive modification is an available linguistic means to specify the intended referent. Indeed, the reports in McKee *et al.* (1998) suggest that considerable verbal prodding of the child is needed to elicit restrictive relative clauses. Consider for instance, a typical example transcript, in which a child (AZ, 3 years 8 months) was induced to use a restrictive relative clause to pick out one hotdog from other hotdogs for the benefit of a blindfolded experimenter. (The blindfolded experimenter's utterances appear in parentheses, the non-blind-folded experimenter's in brackets.) After being asked for the item to be identified, and the child replies:

3. A *hamburger* (And which hamburger? Cause there's two of 'em? Which one shall I pick up?) *this one* (Which one is "this one"?) (Can you tell her with words?) *this hamburger* (Uhm, which hamburger?) *this one the pig is eating* (p. 593).

This is the shortest of the exchanges reported by McKee *et al.* (1998). All others required considerably more back-and-forth, almost all pertaining to the child needing verbal support or repetition to realize that a restrictive modifier is a possible means to disambiguate the referent, instead of by pointing and saying "Dat one!" Although McKee *et al.*'s subjects were younger than those in TSHL, it remains an open question whether the TSHL children (ages 4 to 5) noted the relevant facts about the scene and the conversational situation. Indeed, work from the Referential Communication literature raises similar questions about children's sensitivity to Referential Ambiguity (e.g., Robinson & Robinson, 1982; Robinson & Whittaker, 1985, but see also Ackerman, 1983).

The remaining portion of this introduction describes the means by which we tested this possibility and the means by which we coaxed children to come to an adult understanding of the situation, prior to any test of comprehension and syntactic ambiguity resolution.

EXPERIMENTAL PROSPECTUS

Under what circumstances, if ever, can young children implicitly understand the contextual conditions for restrictive modification, and build that

into their parsing activities? In two experiments, we studied the kinds of temporary PP-attachment ambiguities exemplified by sentence (1), only varying the discourse and visual environments in which the comprehension tests were done, in hopes of increasing the potency of the relevant referential and pragmatic facts.

In both experiments, a two-referent scene like that in Figure 1B was displayed for the child. The child was then told a brief story about these entities (e.g., the two frogs), so as to encourage him or her to notice the presence of the two frogs and their distinctiveness. The story was then followed by two tests: one for production and one for comprehension. The production task involved having the child answer a question about the story he just heard. In the critical conditions, the question required the child to refer linguistically to one of these frogs via restrictive modification (Q: "Which frog went to Mrs. Squid's house?" A: "The one on the napkin"). This condition measures the child's sensitivity to the contextual factors contributing to restrictive modification. Also, the very act of answering the question correctly requires the child to notice the relevant aspects of the scene (i.e., two frogs, differing by location) and to notice that this conversational situation requires specificity of referents. Immediately following the production task, the child was given the *put* instruction comprehension test as in sentence (1), and measures of comprehension were recorded (off-line measures in Experiment 1 and both off- and on-line measures in Experiment 2). Further differences between the experiments are described below.

Experiment 1

The primary manipulation in this experiment was to vary the type of question: either a specific question, which required a locative modifier referring to the visual scene (*which frog went to Mrs. Squid's house?*) or a general question, which made no such demands on the child (*What kind of cookies did the animals bake?*). If the effects of TSHL are due to children failing to notice the relevant facts about the scene and conversational situation, we might expect that correctly answering the specific question would bolster adultlike performance on the *put* comprehension test. As we shall see, although we were fairly successful at getting the child to notice the relevant scene facts and to utter phrases with restrictive modification, the child's (automatic) comprehension abilities appeared quite impervious to these experimental antics.

Experiment 2

Experiment 2 pushes our research question further, by asking whether these difficulties are again unrelated to comprehension *per se*, but rather to

some difficulty children might have contrasting objects spatially. Most simply, children might find the platforms upon which the loveable and interesting beanie babies sit to be relatively uninteresting and less-than-important to the task at hand. Indeed, in previous studies, such salience factors have been shown to exert strong effects on young children's understanding (e.g., Smith, Jones, & Landau, 1992; Hall, Waxman, & Hurwitz, 1993). We, therefore, increased the saliency of the platforms in Experiment 2: Boring plates and napkins were replaced with exotic mirrors and calico umbrellas, to see the extent to which these salient platforms would mitigate our comprehension finding. Perhaps of more central interest, however, was a subsidiary of the salience interpretation, which was also tested in Experiment 2: It has sometimes been reported (Nadig & Sedivy, 2000) that young children are more sensitive to attributes of objects, in general, than to their locations (for discussion of this "what/where" distinction, conceptually and linguistically, see Landau & Jackendoff, 1993). If so, sentences like (1), uttered in the presence of identical frogs, might be the least plausible kind of context in which to test a child's sensitivity to restrictive modification. Therefore, Experiment 2 also gave our young subjects a chance to pit their sensitivity to locations against their sensitivity to varying attributes of frogs, such as their colors and dappledness, as well as their spatial positions.

EXPERIMENT 1

Method

Participants

Sixteen children (9 males, 7 females) participated in the study. The children ranged in age from 48 to 67 months, with a mean age of 56 months. All participants were raised in English-speaking households. They received a small stuffed animal for their involvement in the study.

Procedure

Each session involved two experimenters and the participating child. Prior to the study, the child engaged in some warm-up play with the experimenters involving a puppet named Big Bird. After the warm-up, each trial in the Session began with the child hearing a story about a set of toys that were placed out on the table in front of her. The story was acted out by the experimenter, making sure the child attended to this brief theatrical presentation.

After each story, the child participated in two tests, one focusing on language generation and the other on language comprehension. The language generation test was modeled after the typical elicited production paradigm found in the acquisition literature (Thornton, 1996). Big Bird, who had been hiding under the table during the story, returns to ask the child a question about the story. (Big Bird was made absent during the story to make sure his question seemed plausible to the child.) The child, who has been instructed that Big Bird does not understand pointing, is then expected to provide a spoken response to the question.

The language comprehension test, which immediately followed the elicited production task, was identical to the procedure used in the TSHL comprehension study. The child was told to "Touch the smiley face," and was then asked to act upon three spoken instructions (e.g., "Now spin the dog. Now put it in the bowl. Now move the bowl over to the smiley face"). These stimuli had been prerecorded as digital sound files and were played over a PowerMac G3 laptop computer. Children were always given positive feedback (e.g., "Very good") regardless of their actions.

Each session began with a simplified practice trial, involving only two props so as to familiarize the child with the procedure. During this trial and when necessary throughout the experiment, the child was reminded not to point and instead to "use words" to respond to questions. The entire session was videotaped, recording both the child's spoken responses and her actions when moving objects.

Materials and Design

On each critical target trial, the visual array of objects was the same as the two-Referent condition of TSHL (see Fig. 1), with the following two exceptions. First, the competitor animal (the nontarget frog) also sat on a flat object (e.g., a frog on a towel). This was done to highlight further the contrast between the animals. Second, an additional animal was present, set apart from the other objects. For example, one target scene consisted of a frog on a book (henceforth, the Target), a frog on a towel (the Competitor), an empty book (the Incorrect Destination), an empty bowl (the Correct Destination), and a squid (the Extra Animal).

Target stories contained two separate events, each involving one of the animals in the pair, and designed to draw attention to the presence of the pair. Animals were referred to with simple noun phrases (e.g., "this frog") as the experimenter lifted up the stuffed animal. An example story is as follows. Words in parentheses were not spoken.

4. *There are two frogs. This frog* (the one on the towel) *decides to go*

swimming in the pond. See, there he goes into the water—splash. Then he goes home. This frog (the one on the book) *decides to go play with Mrs. Squid. They bake chocolate cookies together. Then they eat them all up and go home.*

During the elicited production task, the child was asked one of two questions about the story, as illustrated in 5.

5. a. *Which frog went to Mrs. Squid's house?* (specific question)
   b. *What kind of cookies did the animals bake?* (general question)

The specific question always required the child to refer to the second of the animal pair, using noun phrase modification. For instance, a correct response would be "the one on the book" or "the frog on the book."

Instructions for the comprehension task began with an initial instruction "Touch the smiley face." This was to defocus the animals in the story, so that pronominal reference or a bare noun phrase would not be a felicitous reference to the second-mentioned animal. Three additional instructions were then given, beginning with the sentence containing the temporary ambiguity (e.g., "Now put the frog on the book into the bowl"). This instruction was identical to the ambiguous condition of TSHL, except that the second prepositional phrase was headed by "into" rather than "in." This change eliminates a second potential attachment ambiguity, because *into* specifies a path and, therefore requires linkage to the verb *put.*

Four presentation lists were composed to control for order of critical trials and order of the manipulated factor. Each list contained four target trials, two involving a specific question and two involving a general question. A blocked design was used, such that presentation list 1 began with the two specific questions and presentation list 2 began with the two general questions.

For all lists, target trials were embedded in six distractor trials. Distractor trials contained a variety of different objects, questions, and instructions.

Results and Discussion

Videotapes of the sessions were analyzed for each child's spoken response to the question and each child's actions following the target instruction. There were no significant effects of list or blocked orders of presentation (all F's <1), so these conditions were collapsed in subsequent analyses.

What Children Say

In response to the specific question ("Which frog went to Mrs. Squid's house?"), children showed an ample ability to disambiguate the target item in their productions. They did so on 72% of the trials, spontaneously producing a restrictive PP modifier that described the Target (e.g., "The one on the book").

In fact, the 72% value is an underestimate of their pragmatic and syntactic knowledge associated with this structure. On an additional 15% of the trials, children gave restrictive modifier responses that failed to distinguish the referents (e.g., "The one that has the black head," or "The one with the black top whose tail is sticking up."). Thus, on 87% of trials, children realized restrictive modification was necessary and produced the structure appropriate for this linguistic need. Thirteen out of 16 children showed this ability on at least one trial.

Recall from the introductory discussion that in many experimental settings young children have to be verbally prodded and encouraged before they will respond with this kind of answer. This happened also in the present experiment, but at a much lower rate. If we look only at the subjects' first responses, before prodding, a restrictive modifier was produced on 63% of the trials. So even on the most stringent scoring, a majority of trials yielded the intended restrictive modifier response.[3]

What Children Do

Although children showed proficiency at finding an appropriate way to disambiguate the target item in production, these same children failed to use this knowledge in comprehension to select the modifier interpretation of an ambiguous PP. Instead, responses indicated a strong preference for the destination analysis to the ambiguous prepositional phrase *on the book*, regardless of condition. As we describe below, children in this study *garden pathed* in a manner similar to those in the TSHL study, who had not been treated with the techniques to increase the potency of the referent pair. Before considering how children's spoken responses related to their actions, we first examine the overall patterns of actions. We then consider whether children's comprehension performance depended upon the type of response given in the elicited production task.

Overall Action Data. As can be seen in Table I, children performed the correct action, picking up the Target frog and placing it in the bowl, on only 22% of the trials,. In fact, children tended to select a member of the animal pair at random (57% Target frog selection) and most then performed

[3] In this experiment, the frequency and type of feedback in the production task was not as carefully controlled. As we show in Experiment 2, more systematic use of feedback yields similarly high rates of restrictive modifier responses.

Table I. Action Results of Experiment 1

	Percentage of trials		
	Correct action (%)	Selection of intended referent (Target) (%)	Movement to incorrect destination (%)
Specific question	22	59	63
General question	22	56	66
Mean	22	57	64

some action that involved the incorrect destination (64%). The most common action response, observed on 52% of the trials, was to bounce either frog onto the false destination (the empty book) and then put it into the correct destination (the bowl).

The chance selection between the Target and Competitor frog suggests that children rarely if ever realized that the phrase *on the book* could be a modifier and, hence, informative about which frog to select. Movement to the incorrect destination (e.g., the empty book) indicates that a destination analysis was adopted.

The type of question used in the production component of the trial had no effect on the comprehension test. Analyses of Variance (ANOVAs) on the subject and item means of each of these three measures showed no significant effects of Type of Question (all Fs < l). Two-tailed t-tests revealed that selection of Target animal was not significantly different from 50% chance performance ($p > .25$), suggesting a failure to realize that *on the book* is a modifier.

Performance in the specific question (22% correct) is especially striking considering that 72% of these trials had been preceded by a correct utterance, i.e., one involving a restrictive modifier. Statistical tests comparing production and comprehension in this condition were significant in a two-tailed pairwise t-test, $p = .001$.

Actions by Type of Spoken Response. An important question is whether a child's performance on production task in any way corresponded with his performance on the comprehension task. Within the specific-question condition, did uttering a restrictive modifier result in more accurate understanding of the same restrictive modifier in the comprehension task? There are some hints in the data to support this. In particular, for the 28% of the trials on which children answered the question incorrectly, none was followed by a correct action. For the 72% of the trials with correct responses in production, 30% resulted in a correct action (these two distributions were marginally different, Chi-square = 3.5, $p = .06$). This difference suggests that noticing the correct contrast between the referent pairs may be a prerequi-

site for comprehension of the restrictive modifier. However, performance on the comprehension task was still poor (an abysmal 30% correct) even when the children produced the correct contrast.

Summary

We observed a strong asymmetry between production and comprehension when it came to using a restrictive modifier to disambiguate a member of a set. Children usually produced restrictive modifiers to disambiguate members of a set. Yet these same children failed to realize during language comprehension that the ambiguous *on the book* was being used in the same way to determine the referent. We will discuss some implications of these findings after considering the results of the second experiment.

EXPERIMENT 2

Experiment 1 showed that children are often able to spontaneously produce a locative modifier as a form of noun phrase modification. However, in a sizable proportion of the trials (over 25%), children did not answer the question correctly. In fact, on some trials (approximately 15%), children seemed to know that modification was required when asked the question, but were unable to verbally differentiate the objects (i.e., they did not refer to the different platforms that the objects were on.)

As alluded to in the introduction, it is possible that the platforms used in Experiment 1 (and those used in TSHL) were not salient enough for all of the children to deem worthy of their notice, given that they tended to be familiar everyday objects (napkins, plates, towels, etc.). It is also possible that contrasting items by location may be more difficult for children than by other means (e.g., an inherent attribute, such as color or size), as suggested by Nadig and Sedivy (2000).

Experiment 2 explores these issues by examining whether performance in the comprehension and production tasks will improve if the target animals rest on extremely salient and "child-friendly" platforms (toy umbrellas, beds, mirrors, etc.). In addition, all the children were asked questions that required them to distinguish between the two potential targets. The targets used in the experiment were varied by condition: In the attribute–platform condition, the set of animals had different physical attributes (e.g. a plain and a spotted pig), as well as different platforms (e.g., one on a mirror, the other on an umbrella.) In the platform-only condition, the animals were identical, and could only be distinguished by their different platforms (the pig on the mirror). This attribute–platform manipulation allowed us to test the claim that children find it easier to contrast items by attribute rather than by location.

In order to better gauge the child's spontaneous use of referential scene information, a more systematic feedback procedure was used to coax children into giving a restrictive modifier: Children who gave ambiguous responses such as "that one" when asked the production question were simply told one time to "use words because Big bird doesn't understand pointing." No further feedback was given.

Finally, a subset of the children in Experiment 2 were eye-tracked, following the methods in TSHL. Eye-gaze data provided information about the real-time assembly of interpretation during the comprehension task, allowing us to determine if initial preferences differed from the ultimate analysis of the sentence.[4]

Method

Participants

Twenty-eight children (12 males, 16 females) participated in the study. The children ranged in age from 50 months to 71 months, with a mean age of 60 months. All participants were raised in English-speaking households. They received a small stuffed animal for their participation in the study. Sixteen of the participants (8 males, 8 females) participated in the eye-tracking procedure. One subject was dropped because she failed to answer over half of the filler questions correctly.

Equipment

Eye movements were recorded using a lightweight 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 TSHL for details). 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).

Materials and Design

The materials were the same as those used in Experiment 1, with the

[4] Because of technical difficulties, it was not possible to obtain reliable eye position data during the production task. Children tended to be looking out of range of the tracker (at Big Bird, or the experimenter) when the question was asked, resulting in extended periods of track loss as the child began to answer.

following exceptions. First, animals rested on more interesting platforms, such as beds, mirrors, and umbrellas. Second, inherent properties of the animals varied depending upon condition. In the Platform-Only condition, both animals were identical, and could only be differentiated by locative utterances (e.g., the pig on the mirror). In the Attribute–Platform condition, the two animals sitting on platforms also differed by a physical attribute (e.g., spotted vs. unspotted, or green vs. brown). In this condition, participants could contrast members of the pair either by location (the different platforms) or by attribute. Platform condition was manipulated between subjects, with each child seeing either a list of Attribute–Platform items or Platform-Only items. Two presentation lists were composed of each platform type to control for item order. Each presentation list contained four target trials and six distractor trials, with distractor trials containing a variety of different objects, questions, and instructions.

Procedure

For participants who were not eye-tracked, the procedure was the same as for Experiment 1. For those who were eye-tracked, prior to the study, the eye-tracker visor was fitted on the head, until an eye signal was generated by the computer. The eye-tracking technique was the same as that reported in TSHL, with the exception that a more accurate calibration procedure was used, as follows. Prior to the study, a new "point of light" calibration procedure was done. The child was asked to look at the red light on the tip of a "magic wand" held by the experimenter. The computer simultaneously tracked the position of the red light in the scene camera image and the eye position in the eye-camera image. The correspondence between these two data sets was then used by the eye-tracking program to generate best-fit functions to predict direction of gaze in the scene image. This provided an accuracy of approximately 1/2 to 1 degree visual angle. If, at any point, the experimenter deemed the eye-tracking signal to be inaccurate, the calibration procedure was repeated.

Results

There were no significant effects of presentation list (F's <1), so these factors were collapsed in subsequent analyses.

What Children Say

Children successfully used modification to pick out the correct referent in this experiment (87% in the Attribute–Platform condition, and 87% in the

Platform-Only condition).[5] The lack of difference between these conditions suggests that contrasting the pair by location is not necessarily more difficult than contrasting them by inherent attribute. As in Experiment 1, the large majority of these restrictive modifiers were produced without prodding (79% correct in the Attribute–Platform condition, and 74% correct in the Platform-Only condition).

In most Attribute–Platform trials, children chose restrictive modification that contrasted the animals by attribute (e.g., saying "The striped one"). However, 18% of the correct responses in this condition actually involved the child disambiguating the referent spatially (e.g., "The one on the mirror"). Thus, when a choice exists, children (and presumably adults) do prefer attributes over locations, but by no means is disambiguation by location inaccessible or too complex for these children to use in this production task.

What Children Do

Although children were proficient at finding an appropriate way to disambiguate the target item in production, these same children continued to show difficulty applying this knowledge in the comprehension task. Although there was some improvement in comprehension task performance (mean of 46% correct) over that in Experiment 1 (mean of 22% correct), there continued to be a preference for the destination analysis (see Table II). In particular, children were close to chance at selecting the intended referent over the competing referent (58%) and frequently moved animals to the incorrect destination (54%).

The means in Table II suggest that multiple differences between the animals in the Attribute–Platform condition improved performance slightly on the comprehension task. However, statistical comparisons showed that these differences were significant only in item analyses (Correct Action: $F_1 = 1.12$, $F_2(1,3) = 12.79$, $p < .05$; Target Selection: $F_1 = 0.81$, $F_2(1,3) = 13.36$, $p < .05$; Incorrect Destination: $F_1 = 1.69$, $F_2(1,3) = 10.37$, $p < .05$).

Table II. Comprehension Results of Experiment 2

	Percentage of trials		
	Correct action (%)	Selection of intended referent (Target) (%)	Movement to incorrect destination (%)
Platform-only	37	52	63
Attribute–platform	54	64	45
Mean	46	58	54

[5] This analysis excludes two trials in which the child used a restrictive modifier to refer to the wrong animal.

Also, selection of the target was not above chance either in the Attribute–Platform condition (t[13] = 1.42) or in the Platform-Only condition (t[13] = .19). Thus, although there were trends suggesting an advantage for the Attribute–Platform condition, there was considerable variation across subjects regarding the direction and size of this effect.[6]

Comparing Saying and Doing

As in Experiment 1, performance on the production task was significantly better than performance on the comprehension task (platform-only condition: Wilcoxin signed rank $z = 5.45$, $p < .0001$). Thus, although the elicitation question focused children's attention on the pair of animals, and children frequently produced a correct restrictive modifier, performance was considerably worse when restrictive reference was tested in comprehension. Unlike the finding for Experiment 1 in this regard, performance on the comprehension task was not reliably better when the child had answered the question correctly (combined platform–attribute and attribute-only conditions: Chi-square = .28, $p > .5$).

Where Children Look

A subset of these children was eye-tracked to obtain a real-time measure of their comprehension process.[7] Using digital videotapes of the scene and eye position, direction of gaze was computed relative to the ongoing speech.

To summarize the temporal properties of the data, fixations were analyzed within three time regions, all of which were time locked with the portions of the linguistic input (e.g., "put the pig on the mirror into the bowl") on a trial-by-trail basis. Region 1 is used to infer initial reactions to the verb and begins at the onset of "put" and extends until 200 ms after the onset of first phoneme of "pig." Region 2 is used to infer initial reactions to the noun *pig* and begins 200 ms after the onset of "pig" and continues until 200 ms after the onset of "mirror." Region 3 is used to infer initial reactions to *mirror* (and the prepositional phrase *on the mirror*) and begins 200 ms after the onset of "mirror" and extends until 200 ms after the onset of "bowl." The

[6] In a posthoc analysis, the percentage of correct actions showed a strong interaction with the sex of the subject [males: Platform-Only: 60%; Attribute–Platform 29%; females: Platform-Only: 25%; Attribute–Platform 79%; resulting in a reliable interaction between question type and sex ($F_2(1,3) = 23.76$, $p < 0.5$]. However, a similar difference between males and females was not observed in the Specific Question condition of Experiment 1. Given also that sex was not a controlled variable, matched for age, for instance, little can be made of this interaction at this point in time. The finding does indicate that the gender of the subject should be pursued and carefully controlled in future research. It also suggests a source for the variability in the subject analyses above.

[7] This subset showed a similar pattern of performance to the overall group means (percentage correct for actions: mean = 45%, Attribute–Platform condition = 50%; Platform-Only = 41%).

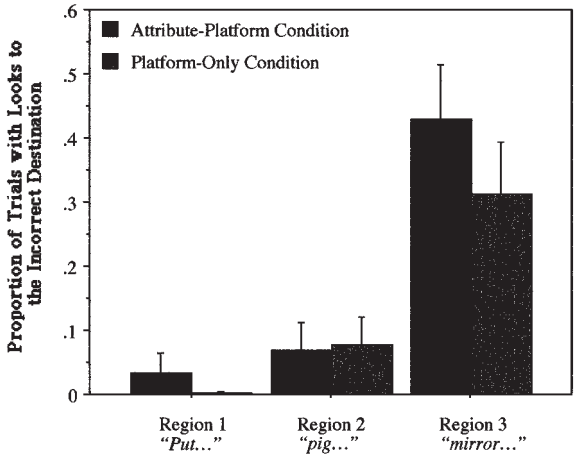

Fig. 2. Proportion of looks to the Incorrect Destination for three different portions of the utterance.

200-ms offset in these measures takes into account the 150–200 ms time to program and execute an eye movement of this type (Matin, Shao, & Boff, 1993; see also Allopenna *et al.,* 1998).

Following TSHL, two different measures were used to assess on-line parsing commitments. First, fixations on the Incorrect Destination (e.g., the empty mirror) were used to infer how the Destination interpretation developed during the comprehension of the sentence. Upon hearing "on the mirror," we might expect to observe increased looks to the Incorrect Destination if children think this phrase is referring to where the mirror should be put. Second, fixations on the Target animal (the one on the mirror) and the Competitor animal (the one on the umbrella) were compared to infer any development of a Modifier interpretation. Upon hearing "on the mirror," we might observe increased looks to the Target (the one on the mirror), if children are initially interpreting this phrase as referring to the platform underneath the Target and, hence, as a modifier. One caveat to this prediction is that we might expect a general preference for looking at the Target animal because it was the one referred to in the production component of the trial.[8]

Destination Analysis (Looks to the Incorrect Destination). Figure 2 graphs by Region the proportion of trials on which the Incorrect Destination was fixated. As the figure shows, participants began to look at the Incorrect

[8] In both analyses, track loss was handled in the following way. If a time region for a observation had over 50% track loss and there was no evidence of a look to the relevant object during any of the trackable portion of the time region, the observation was replaced with the grand mean for that region. This happened on fewer than 10% of the observations and caused only minimal changes in the cell means.

Destination almost immediately after hearing "on the mirror" (Region 3), suggesting an initial preference for a destination analysis of the ambiguous phrase. This pattern replicates the initial destination preference observed in TSHL. Approximately equal increases occurred for both the conditions, suggesting that initial parsing preferences were largely unaffected by potential contrasts afforded by the target set.

ANOVAs on subject and item means were conducted with the following factors: Region (Regions 1, 2, and 3) and Contrast (platform-only and attribute–platfrom). A reliable effect of Region was observed [$F_1(2,28) = 22.08$, $p < .0001$; $F_2(2,6) = 15.93$, $p < .005$), which did not interact with Contrast Type ($F_1 = 0.54$; $F_2 = 1.51$). In addition, there was no effect of contrast ($F_1 = 0.47$; $F_2 = 1.23$). Such a pattern suggests that *on the mirror* triggered a reliable and immediate consideration of the empty mirror as the possible referent in both conditions; that is, the Destination rather than the Modification interpretation.

Looks to Target and Competitor Animal. Figure 3 graphs by Region looks to the Target and Competitor animal. Upon hearing "pig," a certain amount of competition arises between looks to the Target and looks to the Competitor. This competition persists even after hearing "on the mirror," consistent with the notion that subjects were not using this phrase as a modifier to pick out the target. A slight advantage for looks to the target is discernable, as might be expected, given that this had been the referent in the production task.

Interestingly, the pattern of competition is most apparent in the Platform-

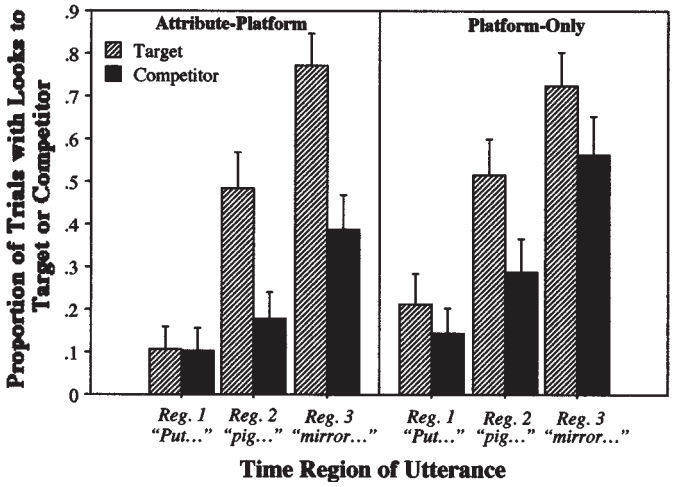

Fig. 3. Proportion of looks to the Target or Competitor object for three different portions of the utterance.

Only condition. In this condition, although there is a slight advantage for the Target (the pig on the mirror) over the Competitor animal (the pig on the umbrella), it never reaches statistical significance [Region 1: $t(5) = 1.15$, ns; Region 2: $t(5) = 1.69$, ns; Region 3: $t(5) = 2.04$), $p = .08$]. This pattern replicates the two-referent ambiguous condition of TSHL, where competition between the Target and Competitor persisted even after the potential modifier ("on the mirror") had been encountered.

The Attribute–Platform condition has a slightly greater Target advantage, which is close to significance in Region 2 and becomes significant in Region 3, after the subject has heard "on the mirror" [Region 1:$t(6) = .04$, ns; Region 2: $t(6) = 3.38$, $p = .01$; Region 3: $t(5) = 9.04$, $p < .0001$). (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, six tests were used.)

In summary, just as in the action data, participants show a trend toward converging on the correct referent in the Attribute–Platform condition. However, their actions—particularly, selection of the Target rather than the Competitor animal—indicate continued confusion about the intended referent and, hence, a failure to consider *on the mirror* as a modifier phrase.

Incorrect vs. Correct Actions

Because the eye movement data came from roughly equal numbers of correct and incorrect trials (45% and 55%, respectively), it is possible to examine whether eye movements differed depending upon the type of action. TSHL found that, in two-referent contexts like these, actions on ambiguous trials were often guided by where the child looked first. Early looks to the Target pig (upon hearing "the pig") frequently led to correct actions, whereas early looks to the Competitor pig frequently led to incorrect actions. As can be seen in Figure 4, this pattern was replicated in the current data. In reaction to *the pig* (Region 1), correct trials show a large proportion of looks to the Target, whereas incorrect trials show roughly equal looks to the Target and the Competitor.

This eye movement pattern suggests that children may be determining the referent based upon an initial guess (where they looked first), never revising this interpretation when further information is provided. Indeed, on the 27 trials in which there was an early look to the Target (during Region 1), participants went on to pick up the Target on 20 of those trials; whereas on the 12 trials in which there was an early look to the competitor animal, participants picked up the competitor animal on 8 of those trials (Chi-square = 5.83, $p < .05$).

Thus, children may sometimes stumble upon the correct action for the

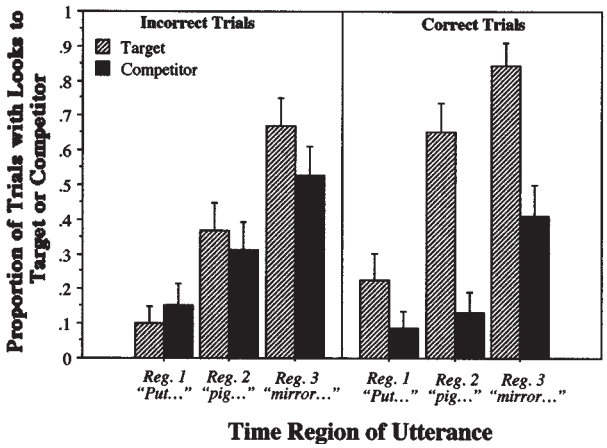

Fig. 4. Proportion of looks to the Target or Competitor object for Incorrect and Correct Trials.

wrong reasons. A child who is lucky enough to initially guess that *the pig* refers to the Target may be reluctant to move the Target animal to the Incorrect Destination (the empty mirror), because the animal is already on a platform of the same type (another mirror). The action data support this contention: Those children who chose the correct target were less likely to "bounce" the animal on the False Destination (the most common error) then those who chose the incorrect target (percentage involving False Destination: Target: 22%, Nontarget 98%, Chi-square = 60.3, $p < .0001$). One especially introspective child reported his line of reasoning. He picked up the Target animal, hesitated, and said "it's already on a brush" (the platform) and then proceeded to place the Target animal into the container (the bowl), performing the correct action.

Importantly, and as Figure 5 shows, children are looking to the incorrect destination at equal rates for both correct and incorrect trials, again suggesting that the Destination interpretation of *on the mirror* is the only analysis considered by the child.

Summary of Experiment 2

Because we made the platforms more salient and child friendly, the children showed a general increase in performance in both the production and comprehension tasks. However, the striking asymmetry between production and comprehension persisted. Children accurately disambiguated referent pairs via spatial modifiers in language production, but these same children showed great difficulty understanding these expressions during

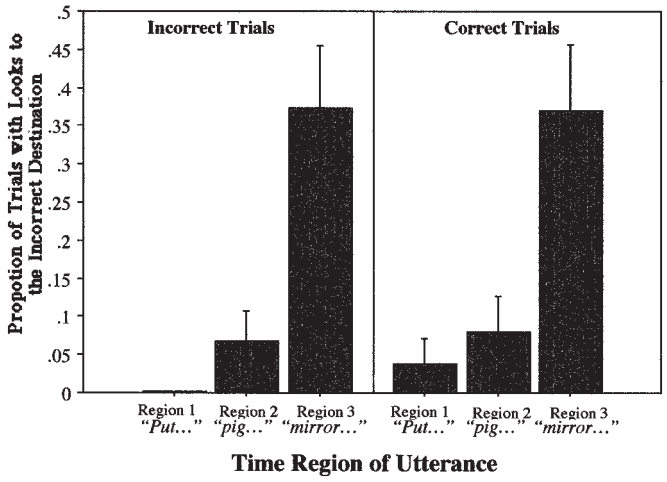

Fig. 5. Proportion of looks to the Incorrect Destination for Incorrect and Correct Trials.

comprehension.

The fact that children produced spatial contrasts at high rates (sometimes selecting them over attribute contrasts) casts some doubt on the notion that there is inherent difficulty with these phrases (e.g., Nadig & Sedivy, 2000) and, instead, suggests that the relative saliency of possible contrasts (be they spatial, attributive, or other means) better determines their rate of use in production (also see Plumert & Nicols-Whitehead, 1996; Plumert *et al.*, 1995 for further evidence of 4-year-olds' substantial abilities to refer to objects via spatial location). Another perspective is that attributes indeed do stand out over regions/locations, but that the region itself can become salient if it possesses a salient inherent attribute.

GENERAL DISCUSSION

Our goal in these experiments was to understand how children resolve temporary ambiguities, such as the one in *Put the frog on the napkin into the box.* Experimental research with adults (e.g., Altmann and Steedman, 1988; Tanenhaus *et al.*, 1995; TSHL, 1999) firmly establishes that adults are guided in their choice by aspects of the discourse and visual context in which such sentences are heard or read. Specifically, adults obey the Referential Principle and so tilt toward the Destination interpretation (VP-attachment) if the unmodified NP (*the frog*) uniquely picks out a referent in the world; symmetrically, where *the frog* is insufficient for referent selec-

tion they tilt toward the Modifier (NP-attachment) interpretation. In contrast, closely analogous experiments with children (TSHL) demonstrate that they overwhelmingly prefer the Destination interpretation of such sentences, in evidently mulish disregard of the contextual effects that are so potent for adults' on-line comprehension performance. Our present findings amply confirm this same VP-attachment bias in children. The question is why adults and children are so different in this regard.

Our experimental maneuvers were explicitly aimed toward distinguishing between two broad classes of explanation of child–adult differences in parsing strategy. First, it is possible that the child has some defect or limitation that is specific to comprehension. Our main means for probing for such an explanation was to contrast child comprehension performance with child production performance. The second possibility is that children lack a general capacity to use the visual world to infer discourse goals. Accordingly, we tried various means to make the reference world more accessible and relevant, by spicing up the interest of modifier-consistent visual aspects (pigs on mirrors), by telling a story in advance that contrasted the two animals in their actions, by asking a question so as to engage the production engine in advance of the comprehension task, and, in particular, by asking a question that requires differentiating and contrasting the animals ("*Which* . . .").

Two results stand out from these experiments. The first is that there are surprising differences between speech and comprehension behaviors in these child subjects: Although they sometimes needed prodding to suppress their preferred modes of frog selection (pointing and/or saying "That one!"), they proved perfectly capable of producing restrictive modification to choose between two frogs. However, in comprehension, no such ability or inclination surfaced in either on- or off-line measures. In particular, when responding to a question like "Which frog went to Mrs. Squid's house?," children correctly answered "the frog/one on the napkin." Yet, on hearing "Put the frog on the napkin into the box," even just after themselves producing the restrictive modifier, they persisted in the Destination interpretation, failing, arguably, to even consider the Modifier option. This was shown both off-line by their actions (such as picking up a random frog, and hopping it onto a napkin and then into a box, Tables I and II) and on-line, by their tendency to look at the empty napkin (Figs. 2 and 4).

The second effect is that pumping up the salience of relevant facts about the scene had as little specific influence as they did. There was a general improvement in both speech and comprehension, attributable to the changed platforms, between Experiments 1 and 2. However, this increased salience did not remove, or even ameliorate, the production/comprehension asymmetry.

These findings make it unlikely that the child's difficulties were due to

a global lack of understanding, either of linguistic modification or of its conditions of applicability. After all, the children could *utter* phrases and sentences containing restrictive modifiers on demand. Not only that, they did so under linguistic conditions ("Which . . .") and referential conditions (two frogs) appropriate for such modification. We do have some evidence that children were not quite so reliable or clear as adults would be as to the pragmatic conditions for making specific reference: Sometimes they responded to "Which . . ." by saying "the frog" when there were two identical frogs or by saying the "spotted frog" when both frogs were spotted. However, this occasional shakiness was rare and cannot account for the magnitude of the production/comprehension disparities in the findings. Similarly, the results have diminished the likelihood that children are so feeble in their language-use abilities as to be immobilized, unless locative properties of the referential world are made wildly salient via mirrors and other such unusual platforms.

If the differences between children and adults are not in their constructional knowledge (their "competence" with restrictive modification), not in their ability to use such structures in speech, not (or only minimally) in their understanding of when restrictive modification is situationally required, and not in their ability to notice the subtler details of the scenarios around them, then wherein does their problem reside? It seems that their differences from adults must lie inside the developing comprehension machinery itself.

Assuming now that the child's problem is one of parsing, two essential matters need to be addressed: The first is to explain the striking differences between production and comprehension. The second is to explain the developmental finding: These differences are much more extreme in young children than in adults. We believe that the answers to both these questions arise from the differing demands of talking and understanding, demands that shape the very form of the machinery that carries out these tasks.

Gross Properties of Production and Comprehension

Although much is shared about the abilities to speak and listen (the underlying linguistic requirements of the language design, for instance), there are considerable differences in the demands these tasks place on language users. Production is not simply the comprehension system run backward (or vice versa). At its heart, language comprehension is a perception process. Listeners have little control over what they are going to hear from a speaker nor at what rate they are going to hear it. No more does human memory permit unlimited buffering of incoming, unanalyzed, perceptual sources. These demands of the task require that comprehension be an autom-

atized mental guessing game (Bever, 1970; Marslen-Wilson, 1973). Indeed, this view of the comprehension mechanism motivates the vast majority of parsing proposals, which almost always include "quick-and-dirty" processing principles or heuristics as the engine-core of the comprehension device (e.g., Bever, 1970; Frazier & Fodor, 1978; and see MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994 for constraint-satisfaction theories, which recast parsing heuristics within a general probabilistic and automatic parsing device, but are still motivated by these same comprehension facts).

Production, on the other hand, has considerably different processing demands. Content and rate, at least at the utterance preparation stage, are largely under the control of the speaker. Its chief desideratum is accuracy in conveying the intended message. Therefore, where syntactic choice is concerned, pragmatic considerations pertaining to the goals of the speaker and perhaps her assessment of the listener directly feed this process. As a result, theories of production are ones of planning (e.g., Levelt, 1989; Bock, 1995). Obviously, to speak it is necessary to construct an explicit syntactic structure complete with its associated interpretation, rather than the rough representation at either level that may be sufficient for comprehension. Particularly relevant to the present discussion: If we leave aside punsters and poets, ambiguity is in the ear of the listener not the mouth of the speaker.

Even this very brief (we justly called it "gross") summary of the known task differences between speech and comprehension makes it a plausible facility at these tasks, for the same language materials may differ. Indeed, it is testament to the acrobatic facility of adult language users that they seem so much alike, and "perfect" across these radically different tasks. It is, therefore, not surprising, as argued below, that our results reveal a massive difference in these task performance by young children.

Evolution of the Parser in Young Children

As we remarked in the introduction to this paper, studies on the resolution of temporary ambiguity over the past 30 years have revealed that mature listeners rapidly recruit information from a variety of sources during the brief interval of time (under a second for even longish one-clause sentences) from a sentence beginning to end. This includes, among many other factors, lexical preferences of verbs (in terms of argument type, argument number, and other properties), semantic interpretation of individual phrases, and aspects of the discourse and visual world. In general, these factors are probabilistic, e.g., a certain verb may take sentence complements a large percentage of the time and NP complements much more rarely. The evidence to date strongly suggests that people rapidly extract such contingen-

cies from the database of ambient speech and make use of them in on-line parsing decisions (MacDonald *et al.,* 1994; Trueswell & Tanenhaus, 1994; and see also Saffran, Aslin & Newport, 1996, of infants' "sponge-like" abilities to extract generalizations from probabilistic input data).

If all this is so, it is easy to understand why children should exhibit special difficulties during the period when the comprehension machinery is being constructed. Insofar as the cues to structure are probabilistic, the learning child necessarily accrues information about these cues as a function of their reliability in input speech, as well as the ease with which these cues can be represented. This learning must take place incrementally, over the course of many years of listening: All other things being equal (including representational complexity), we should expect children to acquire the more reliable of these cues before the less reliable ones (for prior discussion along these lines, see TSHL). This gives special status to linguistic cues that trigger the relevant inferences about modification, including cues from the discourse. Furthermore, children may have a reduced ability to maintain multiple competing representations (e.g., Lorsbach & Reimer, 1996; Lorsbach *et al.,* 1998). This may lead them to more "greedily" pursue a single analysis, discarding lower probability, rapidly fading, syntactic analyses (e.g., THSL, 1999; Just & Carpenter, 1992; MacDonald & Christiansen, 2000; as well as Adams & Gathercole, 2000; Baddely, Gathercole, & Pagno, 1988, for some discussion of the relationship between capacity and language use in children).

The present experimental findings may be best explained in these terms. Consider the lexical preferences and discourse implications of *put*. This verb strongly suggests a destination phrase (sometimes a locative adverb such as "down" or "away", but even more frequently a locative PP such as "one the napkin" or "in your toybox"). Assuming that the child learner has acquired this highly reliable verb preference (for the very frequent item, *put*) and considers this cue first, as opposed to other sources of information, in response to sentence (1), a locative interpretation will be his first guess. Moreover, the child's first guess is likely to be his last. He never returns from the end of the garden path, because of the inability to maintain alternative analyses over an extended period of time.

Final Thoughts

Several of the finer points of our explanation for the comprehension-production disparity in young children are perforce speculative, limited as we are to evidence from these experiments and related work. In order to disentangle the issues more clearly, further lines of experimentation are clearly in order. Their overall aim should be to see if these disparities are minimized in situations where the production and comprehension tasks

become more alike. For example, one might withhold highly reliable cues from a comprehension task, thus allowing the next cue in line to take over. In the present experiments, we have made a more limited attempt to expose some of the task-boundedness of child language. From this evidence and related studies now entering the child psycholinguistic literature, we suggest that language learning should be viewed in light of this task-boundedness rather than as abstracted away from it. That is, learning itself should be viewed as a process that evolves as the direct outcome of the child's attempts to talk and understand.

REFERENCES

Adams, A., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders, 35,* 95–116.

Ackerman, B. P. (1983). Children's judgements of the functional acceptablility of referential communications in discourse contexts. *Journal of Child Language. 10,* 151–166.

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., 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, & Cognition, 20,* 209–216.

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

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

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

Bock, K. (1995) Sentence production: From mind to mouth. In P. Eimas and J. L. Miller (Eds.), *Speech, language, and communication, Handbook of perception and cognition: Language,* 2nd Ed, San Diego, California: Academic Press.

Brennan, S., & Clark H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology, Learning, Memory, & Cognition, 22,* 1482–1493.

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

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1–39.

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

Crain, S., & Steedman, M. J. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives.* Cambridge,: Cambridge University Press.

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

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehen-

sion: 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.

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

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

Grice, H. P. (1968). *Studies in the way of words.* Cambridge, Massachusetts: Harvard University Press.

Hall, D. G., Waxman, S. R., Hurwitz, W. M. (1993). How two- and four-year-old children interpret adjectives and count nouns *Child Development, 64,* 1651–1664.

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

Horton, W. S., & Keysar, B. (1996) When do speakers take into account common ground? *Cognition, 59,* 91–117.

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

Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral & Brain Sciences, 16,* 217–26.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, Massachusetts: MIT Press.

Lorsbach, T., Katz, G., & Cupak, A. (1998). Developmental differences in the ability to inihibit the initial interpretation of garden path passages. *Journal of Experimental Child Psychology, 71,* 275–296.

Lorsbach, T., & Reimer, J. F. (1996). Developmental changes in the inhibition of previously relevant information. *Journal of Experimental Child Psychology, 64,* 317–342.

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

MacDonald, M. C., & Christiansen, M. H. (2000). Reassessing working memory: A reply to Just & Carpenter and Waters & Caplan. Manuscript submitted.

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

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

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

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

Nadig, A., & Sedivy, J. (2000). Children's use of referential pragmatic constraints in production and processing. Talk presented at the *13th Annual CUNY Conference on Human Sentence Processing.* La Jolla, California.

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

Plumert, J. M., Dwert, K., & Spears, S. (1995). The early development of children's communication about nested spatial relations. *Child Development, 66,* 959–969.

Plumert, J. M., & Nicols-Whitehead, P. (1996). Parental scaffolding of young children's spatial communication. *Developmental Psychology, 32,* 523–532.

Robinson, E. J., & Robinson, W. D. (1982). Knowing when you don't know enough:

Children's judgements about ambiguous information. *Cognition, 12*(3), 267–280.

Robinson, E. J., & Whittaker, S. J. (1985). Children's responses to ambiguous messages and their understanding of ambiguity. *Developmental Psychology, 21,* 446–454.

Saffran, J. R., Aslin, R. N., Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Sedivy, J. (1999). Investigating the discourse-based properties of adjectives in on-line semantic interpretation. Talk presented at the *12th Annual CUNY Conference on Human Sentence Processing,* New York.

Smith, L. B., Jones, S., & Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology, 28,* 273–286.

Spivey, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (2000). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology,* in press.

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.

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.

Thornton, R. (1996). Elicited production. In McDaniel, D., Mckee, C. & Cairns, H. S. (Eds.), *Methods for assessing children's syntax, Language, speech, and communication,* Cambridge, MA: MIT Press. (pp. 78–102).

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

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: studying on-line sentence processing in young children. *Cognition, 73,* 89–134.

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

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,* 528–553.